Design and Methodology

Current Population Survey

Technical Paper 66

U.S. Department Labor
U.S. BUREAU OF LABOR STATISTICS

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU

ACKNOWLEDGMENTS

Design and Methodology

Current Population Survey

U.S. Department of Labor
Elaine L. Chao,
Secretary

U.S. Bureau of Labor Statistics
Philip L. Rones
Acting Commissioner

U.S. Department of Commerce
Carlos M. Gutierrez,
Secretary

David A. Sampson,
Deputy Secretary

Economics and Statistics Administration
Cynthia A. Glassman,
Under Secretary
for Economic Affairs

U.S. CENSUS BUREAU
Charles Louis Kincannon,
Director

ECONOMICS

AND STATISTICS

ADMINISTRATION

**Economics
and Statistics
Administration**

Cynthia A. Glassman,
Under Secretary
for Economic Affairs

U.S. CENSUS BUREAU

Charles Louis Kincannon,
Director
Hermann Habermann,
Deputy Director and
Chief Operating Officer

Howard Hogan,
Associate Director
for Demographic Programs

U.S. BUREAU OF LABOR STATISTICS

Philip L. Rones,
Acting Commissioner
Philip L. Rones,
Deputy Commissioner

John Eltinge,
Associate Commissioner for Survey
Methods Research

John M. Galvin,
Associate Commissioner for Employment
and Unemployment Statistics

Thomas J. Nardone, Jr.,
Assistant Commissioner for Current
Employment Analysis

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.

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. In January 2003, the CPS adopted the 2002 census industry and occupation classification systems, derived from the 2002 North American Industry Classification System and the 2000 Standard Occupational Classification System.

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.

Charles Louis Kincannon
Director
U.S. Census Bureau

May 2006

Kathleen P. Utgoff
Commissioner
U.S. Bureau of Labor Statistics

May 2006

CONTENTS

CONTENTS | Chapter 9. Data Preparation

Chapter 10. Weighting and Seasonal Adjustment 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

CONTENTS

Figures

Illustrations

Tables

CONTENTS

Tables—Con.

U.S. Bureau of Labor Statistics and U.S. Census Bureau

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 primary source of labor force statistics for the population of the United States. The CPS is the source of numerous high-profile economic statistics, including the national unemployment rate, and provides data on a wide range of issues relating to employment and earnings. The CPS also collects extensive demographic data that complement and enhance our understanding of labor market conditions in the nation overall, among many different population groups, in the states and in substate areas.

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

The CPS is a source of information not only for economic and social science research, but also for the study of survey methodology. This report provides 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 probability selected sample of about 60,000 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 one month to the next (as well as over the year). The 4-8-4 sampling scheme has the added benefit of allowing the constant replenishment of the sample without excessive burden to respondents.

The CPS questionnaire is a completely computerized document that is administered by Census Bureau field representatives across the country through both personal and telephone interviews. Additional telephone interviewing is conducted from the Census Bureau's three centralized collection facilities in Hagerstown, Maryland; Jeffersonville, Indiana; and Tucson, Arizona.

To be eligible to participate in the CPS, individuals must be 15 years of age or over and not in the Armed Forces. People in institutions, such as prisons, long-term care hospitals, and nursing homes are ineligible to be interviewed in the CPS. In general, the BLS publishes labor force data only for people aged 16 and over, since those under 16 are 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 people who are not working.

---

[1] The sample size was increased from 50,000 occupied households in July 2001. (See Chapter 2 for details.)
[2] In the month of December, the survey is often conducted 1 week earlier to avoid conflicting with the holiday season.

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 the 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.

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. Several earlier efforts attempted 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 it 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 sample has undergone major revision on a regular basis. 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 chance of selection. 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 that are different from the rest of the population.

- **February 1952.** Document-sensing procedures were introduced into 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 individuals by race whereas only the percentage of distributions had previously been published.

- **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 during 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 estimates 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 is a continuous process, and the Census Bureau regularly updates 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.** The definition of employment status was changed. Two relatively small groups of people, both formerly classified as employed "with a job but not at work," were assigned to new classifications. The reassigned groups were (1) people on layoff with definite instructions to return to work within 30 days of the layoff date and (2) people waiting to start new wage and salary jobs within 30 days of the interview. Most of the people 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 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 (BLS) and the Census Bureau in the following 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 to be the responsibility 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 noninstitutionalized 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.** 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 people.

- **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 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 (U.S. Department of Commerce, 1976), 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[1] 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 separate ratio estimates for Negro and Other[2] race categories.

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

- **January 1971 and January 1972.** The 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.** The sample was expanded to 461 PSUs and the 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 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.** The inflation-deflation method was introduced for deriving independent estimates of the population. The derivation of independent estimates of the civilian noninstitutionalized population by age, race,

---

[1] Negro was the race terminology used at that time.
[2] Other includes American Indian, Eskimo, Aleut, Asian, and Pacific Islander.

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 people for each state. 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 data. From now on, race was 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 noninstitutionalized population 16 years and over, civilian labor force, and unemployed people. 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 going 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 that relate only to the "Black" population.

- **October 1984.** School enrollment items were added for people 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 veterans' data for females was started.

- **January 1985.** Estimation procedures were changed to use data from the 1980 census and the 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[3] 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 dramatically 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 in 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 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.

- **January 1992.** Industry and occupation codes from the 1990 census were introduced. Population estimates were converted to the 1990 census base for use in ratio estimation procedures.

- **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 to 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 data,

the redesign significantly influenced the measurement of characteristics related to employment, such as the proportion of the employed 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 the use of a very complex questionnaire without increasing respondent burden, increased consistency by reducing interviewer error, permitted editing at time of interviewing, and allowed 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.

CPS data used by the BLS were adjusted to reflect an undercount in the 1990 decennial census. Quantitative measures of this undercount are derived from a post-enumeration survey. Because of reliability issues associated with the post-enumeration survey for small areas of geography (i.e., places with populations of less than one million), the undercount adjustment was made only to state and national level estimates. While the undercount varied by geography and demographic group, the overall undercount was estimated to be slightly more than 2 percent for the total 16 and over civilian noninstitutionalized population.

- **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 TP63RV.

- **December 1994.** Starting in December 1994, a new set of response categories was phased in for the relationship to reference person question. This modification was directed at individuals not formally related to the reference person to identify whether there were 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 for 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

---

[3] Hispanics may be any race.

units. The reduced CPS national sample contained 754 PSUs. The details of the sample design changes as of January 1996 are described in Appendix H, TP63RV.

- **January 1998.** A new two-step composite estimation method for the CPS was implemented (See Appendix I). The first step involved computation of composite estimates for the main labor force categories, classified by key demographic characteristics. The second adjusted 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 provided increased operational simplicity for microdata users and improved the accuracy of labor force estimates by using different compositing coefficients for different labor force categories. The weighting adjustment method assured 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 the 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 estimates, however, until it had sufficient time to evaluate the estimates from the 60,000 household sample. See Appendix J, Changes to the Current Population Survey Sample in July 2001, for details.

- **January 2003.** The 2002 Census Bureau occupational and industrial classification systems, which are derived from the 2000 Standard Occupational Classification (SOC) and the 2002 North American Industry Classification System (NAICS), were introduced into the CPS. The composition of detailed occupational and industrial classifications in the new systems was substantially changed from the previous systems, as was the structure for aggregating them into broad groups. This created breaks in existing data series at all levels of aggregation.

Questions on race and ethnicity were modified to comply with new federal standards. Beginning in January 2003, individuals are asked whether they are of Hispanic ethnicity before being asked about their race. Individuals are now asked directly if they are Spanish, Hispanic, or Latino. With respect to race, the response category of Asian and Pacific Islanders was split into two categories: a) Asian and b) Native Hawaiian or Other Pacific Islanders. The questions on race were reworded to indicate that individuals could select more than one race and to convey more clearly that individuals should report their own perception of what their race is. These changes had little or no impact on the overall civilian noninstitutionalized population and civilian labor force but did reduce the population and labor force levels of Whites, Blacks or African Americans, and Asians beginning in January 2003. There was little or no impact on the unemployment rates of these groups. The changes did not affect the size of the Hispanic or Latino population and had no significant impact on the size of their labor force, but did cause an increase of about half a percentage point in their unemployment rate.

New population controls reflecting the results of Census 2000 substantially increased the size of the civilian noninstitutionalized population and the civilian labor force. As a result, data from January 2000 through December 2002 were revised. In addition, the Census Bureau introduced another large upward adjustment to the population controls as part of its annual update of population estimates for 2003. The entire amount of this adjustment was added to the labor force data in January 2003. The unemployment rate and other ratios were not substantially affected by either of these population control adjustments.

The CPS program began using the X-12 ARIMA software for seasonal adjustment of time series data with release of the data for January 2003. Because of the other revisions being introduced with the January data, the annual revision of 5 years of seasonally adjusted data that typically occurs with the release of data for December was delayed until the release of data for January. As part of the annual revision process, the seasonal adjustment of CPS series was reviewed to determine if additional series could be adjusted and if the series currently adjusted would pass a technical review. As a result of this review, some series that were seasonally adjusted in the past are no longer adjusted.

Improvements were introduced to both the second-stage and composite weighting procedures. These changes adapted the weighting procedures to the new race/ethnic classification system and enhanced the stability over time of national and state/substate labor force estimates for demographic groups.

More detailed information on these changes and an indication of their effect on national labor force estimates appear in "Revisions to the Current Population Survey Effective in January 2003" in the February 2003 issue of *Employment and Earnings*, available on the Internet at <http://www.bls.gov/cps/rvcps03.pdf>.

- **January 2004.** Population controls were updated to reflect revised estimates of net international migration for 2000 through 2003. The updated controls resulted in a decrease of 560,000 in the estimated size of the civilian noninstitutionalized population for December 2003. The civilian labor force and employment levels decreased by 437,000 and 409,000 respectively. The Hispanic or Latino population and labor force estimates declined by 583,000 and 446,000 respectively and Hispanic or Latino employment was lowered by 421,000. The updated controls had little or no effect on overall and subgroup unemployment rates and other measures of labor market participation.

More detailed information on the effect of the updated controls on national labor force estimates appears in "Adjustments to Household Survey Population Estimates in January 2004" in the February 2004 issue of *Employment and Earnings*, available on the Internet at <http://www.bls.gov/cps/cps04adj.pdf>.

Beginning with the publication of December 2003 estimates in Janaury 2004, the practice of concurrent seasonal adjustment was adopted. Under this practice, the current month's seasonally adjusted estimate is computed using all relevant original data up to and including those for the current month. Revisions to estimates for previous months, however, are postponed until the end of the year. Previously, seasonal factors for the CPS labor force data were projected twice a year. With the introduction of concurrent seasonal adjustment, BLS will no longer publish projected seasonal factors for CPS data. More detailed information on concurrent seasonal adjustment is available in the January 2004 issue of *Employment and Earnings* in "Revision of Seasonally Adjusted Labor Force Series," available on the Internet at <http://www.bls.gov/cps/cpsrs2004.pdf>.

In addition to introducing population controls that reflected revised estimates of net international migration for 2000 through 2003, in January 2004, the LAUS program introduced a linear wedge adjustment to CPS 16+ statewide estimates of the population, labor force, employment, unemployment, and unemployment rate.

This adjustment linked the 1990 decennial census-based CPS estimates, adjusted for the undercount (see January 1994), to the 2000 decennial census-based CPS estimates. This adjustment provided consistent estimates of statewide labor force characteristics from the 1990s to the 2000s. It also provided consistent CPS series for use in the LAUS program's econometric models that are used to produce the official labor force estimates for states and selected sub-state areas, which use CPS employment and unemployment estimates as dependent variables.

- **April 2004.** The 16-month phase-in of the redesigned sample based on the 2000 census began. This is the sample design documented in this technical paper.

- **September 2005.** Hurricane Katrina made landfall on the Gulf Coast after the August 2005 survey reference period. The data produced for the September reference period were the first from the CPS to reflect any impacts of the storm. The Census Bureau attempted to contact all sample households in the disaster areas except those areas under mandatory evacuation at the time of the survey. Starting in October, all areas were surveyed. In accordance with standard procedures, uninhabitable households, and those for which the condition was unknown, were taken out of the CPS sample universe. People in households that were successfully interviewed were given a higher weight to account for those missed. Also starting in October, BLS and the Census Bureau added several questions to identify persons who were evacuated from their homes, even temporarily, due to Hurricane Katrina. Beginning in November 2005, state population controls used for CPS estimation were adjusted to account for interstate moves by evacuees. This had a negligible effect on estimates for the total United States. The CPS will continue to identify Katrina evacuees monthly, possibly through December 2006.

REFERENCES

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.

U.S. Department of Commerce, Bureau of the Census, and U.S. Department of Labor, Bureau of Labor Statistics. "Concepts and Methods Used in Labor Force Statistics Derived from the Current Population Survey," **Current Population Reports**. Special Studies Ser. P23, No. 62. Washington, DC: Government Printing Office, 1976.

Chapter 3.
Design of the Current Population Survey Sample

INTRODUCTION

For more than six 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 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. The methods used to select the sample households for the survey are evaluated after each decennial census. Based on these evaluations, the design of the survey is modified and systems are put in place to provide sample for the following decode. The most recent decennial revision incorporated new information from Census 2000 and was complete as of July 2005.

This chapter describes the CPS sample design as of July 2005. It 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 2005:

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 noninstitutionalized population 16 years of age and older (CNP16+).

3. The CPS sample consists of independent samples in each state and the District of Columbia. Each state 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: Los Angeles County and the rest of

California; New York City and the rest of New York State.[1] 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 that 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 requirement for the monthly unemployment level for the nation, given a 6 percent unemployment rate, is 1.9 percent. The 1.9 percent CV is based on the requirement that a difference of 0.2 percentage points in the unemployment rate for two consecutive months be statistically significant at the 0.10 level.

5. The required CV on the annual average unemployment level for each state, substate area, 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 72,000 assigned housing units from 824 sample areas designed to measure demographic and labor force characteristics of the civilian noninstitutionalized population 16 years of age and older. Approximately 12,000 of the assigned housing units are sampled under the State Children's Health Insurance Program (SCHIP) expansion that has been part of the official CPS sample since July 2001. The CPS samples housing units from lists of addresses obtained from the 2000 Decennial Census of Population and Housing. The sample is updated continuously for new housing built after Census 2000. 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 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

---

[1] New York City consists of Bronx, Kings, New York, Queens, and Richmond Counties.

social and economic characteristics that are highly correlated with unemployment. One PSU is sampled in each stratum. The probability of selection for each PSU in the stratum is proportional to its population as of Census 2000.

In the second stage of sampling, a sample of housing units within the sample PSUs is drawn. Ultimate sampling units (USUs) are small groups of housing units. The bulk of the USUs sampled in the second stage consist of sets of addresses that are systematically drawn from sorted lists of blocks prepared as part of Census 2000. 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. The CPS sample is usually described as a two-stage sample, but occasionally, a third stage of sampling is necessary when actual USU size is extremely large. In addition, a sample of building permits is selected to provide coverage of construction since 2000. The sample of building permits is based on listings of new construction obtained from local jurisdictions in sample PSUs.

Each month, interviewers collect data from the sample housing units. A housing unit is interviewed for 4 consecutive months, dropped out of the sample for the next 8 months, and interviewed again 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 single 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 upholds the scientific tenets of probability sampling and each month's sample produces a true representation of the target population. The rotation system makes it possible to reduce sampling error by using 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. The 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 minimize the cost of the survey. This is done mainly by minimizing the number of field representatives needed to conduct the survey, and reducing the travel cost incurred in visiting the sample housing units. 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 or 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.

Definition of the Primary Sampling Units

PSUs are delineated so that they encompass the entire United States. The land area covered by each PSU is made reasonably compact so it 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 that arises from selecting a sample of PSUs rather than selecting 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.

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

Rules for Defining PSUs

1. Each PSU is contained within the boundary of a single state.

2. Metropolitan statistical areas (MSAs) are defined as separate PSUs using projected 2003 Core-Based Statistical Area (CBSA) definitions. CBSAs are defined as

---

[2] See Chapter 10: Estimation Procedures for Labor Force Data for more information on the composite estimation procedure.

metropolitan or micropolitan areas and include at least one county. Micropolitan areas and areas outside of CBSAs are considered nonmetropolitan areas. If any metropolitan area crosses state boundaries, each state-metropolitan area intersection is a separate PSU.[3]

3. For most states, PSUs are either one county or two or more contiguous counties. 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.

The PSU definitions are revised each time the CPS sample design is revised. Revised PSU definitions reflect changes in metropolitan area definitions and an attempt to have PSU definitions consistent with other U.S. Census Bureau demographic surveys. The following are steps for combining counties, county equivalents, and independent cities into PSUs for the 2000 design:

1. The 1990 PSUs are evaluated by incorporating into the PSU definitions those counties comprising metropolitan areas that are new or have been redefined.

2. Any single county is classified as a separate PSU, regardless of its 2000 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.

These steps created 2,025 PSUs in the United States from which to draw the sample for the CPS when it was redesigned after the 2000 decennial census.

---

[3] Final metropolitan area 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 CBSA definitions are separate PSUs. Most projected CBSA definitions are the same as final CBSA definitions (Executive Office of the President, 2003).

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 and cost considerations suggest 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 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 populations in 2000.

Sampling theory and costs dictate that highly populated PSUs should be selected for sample with certainty. The rationale is that some PSUs exceed or come close to the population 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 2000 redesign:

1. A PSU which consists of at least one county that is likely to be part of the 151 most populous metropolitan areas based on projected definitions and Census 2000 population is required to be SR.

2. The remaining PSUs are grouped into non-self-representing (NSR) strata within state boundaries. In each NSR stratum, one PSU will be selected to represent all of the PSUs in the stratum. They are formed 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 35 to 55 housing units. The number of NSR strata in a state is a function of the 2000 population, civilian labor force, state CV, and between-PSU variance[4] 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.) In Alaska, the strata are also a function of expected interview cost.
   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.

---

[4] Between-PSU variance is the component of total variance arising from selecting a sample of PSUs from all possible PSUs.

d. Stratification is performed independently of previous CPS sample designs.

   Key variables used for stratification are:

   - Number of males unemployed.

   - Number of females unemployed.

   - Number of families with female head of household.

   - Ratio of occupied housing units with three or more people, 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 7.

e. In states with SCHIP sample, the self-representing PSUs are the same for both CPS and SCHIP. In most states, the same non-self-representing sample PSUs are in sample for both surveys. However, to improve the reliability of the SCHIP estimates in Maine, Maryland, and Nevada, the SCHIP non- self-representing PSUs are selected independent of the CPS sample PSUs, with replacement. The methodology for the stratification of PSUs for SCHIP in these states is similar to the other stratifications, except that the stratification variable used is the number of people under age 18 with a household income below 200 percent of poverty.

Table 3–1 summarizes the percentage of the targeted population in SR and sampled NSR areas by state.

Several current surveys, including CPS, use the Stratification Search Program (SSP) created by the Demographic Statistical Methods Division of the Census Bureau to perform the PSU stratification. CPS strata in all states except Alaska are formed by the SSP. (A separate program performs the stratification for Alaska.) The SSP classifies certain PSUs as SR, using the criteria mentioned previously, and creates NSR strata. First, initial parameter sets for each stratification area (i.e., state or substate area) are formed by creating unique combinations of the number of NSR PSUs, the number of SR PSUs, the number of strata that must be formed, and the average monthly workload for NSR PSUs and SR PSUs. Non-self-representing PSUs are reclassified as SR if additional SR PSUs are needed to provide adequate samples and if they are among the most populous PSUs in the stratification area. Some NSR PSUs are reclassified SR if they are not similar enough to other NSR PSUs to produce a favorable stratification. Some of the created parameter sets are eliminated because of unsatisfactory PSU workloads or lack of a self-weighting design.

Next, random stratifications for each parameter set are formed. NSR PSUs are moved from one stratum to another to even out the size of the strata. Stratifications are evaluated based on the criteria in the previous section. A national stratification is then chosen by selecting one stratification from each state. The national stratification is refined through a series of moves and swaps to minimize the difference in workloads among NSR PSUs and the CV for unemployment level for each stratification area.

After the strata are defined, some state sample sizes are increased to bring the national CV for unemployment level down to 1.9 percent assuming a 6 percent unemployment rate.

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, Hawaii, 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. There are currently 446 SR PSUs. In each of the remaining 378 NSR strata, one PSU is selected for the sample following the guidelines described next. Four of the NSR strata only contain SCHIP sample.

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 hiring and training 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 Census 2000 population. The selection procedure accomplishes the following objectives:

1. Select one sample PSU from each stratum with probability proportional to the 2000 population.

2. Retain in the new sample the maximum number of sample PSUs from the 1990 design sample.

Using a procedure designed to maximize overlap, one PSU is selected per stratum with probability proportional to its 2000 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.

Table 3–1. **Estimated Population in Sample Areas for 824-PSU Design by State**

State	Self-representing (SR) areas		Non-self-representing (NSR) areas		Total sample areas	
	Population[1] in SR areas	Percentage	Population[1] in NSR sample areas	Percentage	Population[1] in sample areas	Percentage
Total.................	**161,282,405**	**76.1**	**18,957,760**	**8.9**	**180,240,165**	**85.0**
Alabama	1,556,994	46.2	746,019	22.1	2,303,013	68.3
Alaska	336,349	77.2	28,541	6.5	364,890	83.7
Arizona	3,076,477	80.5	421,667	11.0	3,498,144	91.5
Arkansas....................	876,805	43.4	388,413	19.2	1,265,218	62.6
California....................	23,690,332	94.6	587,004	2.3	24,277,336	96.9
Los Angeles	7,040,665	100.0	0	0.0	7,040,665	100.0
Remainder of California	16,649,667	92.5	587,004	3.3	17,236,671	95.8
Colorado	2,444,539	75.4	401,785	12.4	2,846,324	87.8
Connecticut..................	2,588,699	100.0	0	0.0	2,588,699	100.0
Delaware....................	595,030	100.0	0	0.0	595,030	100.0
District of Columbia...........	457,495	100.0	0	0.0	457,495	100.0
Florida	11,247,999	90.5	702,456	5.7	11,950,455	96.1
Georgia	3,921,187	64.7	730,347	12.1	4,651,534	76.8
Hawaii	902,559	100.0	0	0.0	902,559	100.0
Idaho	582,187	61.5	128,137	13.5	710,324	75.0
Illinois.......................	7,459,357	79.9	875,368	9.4	8,334,725	89.3
Indiana......................	2,504,432	54.6	705,866	15.4	3,210,298	69.9
Iowa	718,299	32.2	590,359	26.5	1,308,658	58.7
Kansas......................	1,028,154	51.5	294,193	14.7	1,322,347	66.2
Kentucky	1,414,196	45.9	525,654	17.1	1,939,850	63.0
Louisiana....................	1,779,331	54.1	570,528	17.4	2,349,859	71.5
Maine.......................	831,273	83.7	104,708	10.5	935,981	94.2
Maryland	3,635,366	91.2	249,759	6.3	3,885,125	97.5
Massachusetts................	4,915,261	100.0	0	0.0	4,915,261	100.0
Michigan	5,708,265	76.1	788,772	10.5	6,497,037	86.6
Minnesota...................	2,534,229	68.2	293,420	7.9	2,827,649	76.1
Mississippi...................	658,381	31.4	518,634	24.8	1,177,015	56.2
Missouri.....................	2,592,889	61.3	532,452	12.6	3,125,341	73.9
Montana	449,818	65.6	60,735	8.9	510,553	74.4
Nebraska....................	669,967	52.3	166,160	13.0	836,127	65.2
Nevada	1,361,183	90.3	73,785	4.9	1,434,968	95.2
New Hampshire...............	946,316	100.0	0	0.0	946,316	100.0
New Jersey..................	6,424,830	100.0	0	0.0	6,424,830	100.0
New Mexico	869,341	64.9	195,258	14.6	1,064,599	79.4
New York....................	12,976,943	89.4	604,945	4.2	13,581,888	93.6
New York City.............	6,197,673	100.0	0	0.0	6,197,673	100.0
Remainder of New York	6,779,270	81.5	604,945	7.3	7,384,215	88.8
North Carolina	3,227,560	53.0	1,112,224	18.2	4,339,784	71.2
North Dakota	258,088	53.2	97,567	20.1	355,655	73.3
Ohio	6,512,613	75.7	597,297	6.9	7,109,910	82.6
Oklahoma	1,453,949	56.4	288,633	11.2	1,742,582	67.6
Oregon......................	1,803,289	68.5	255,641	9.7	2,058,930	78.2
Pennsylvania	7,456,685	78.7	877,672	9.3	8,334,357	88.0
Rhode Island	810,041	100.0	0	0.0	810,041	100.0
South Carolina................	1,923,917	63.7	330,931	11.0	2,254,848	74.7
South Dakota	320,381	57.2	105,941	18.9	426,322	76.1
Tennessee...................	2,449,113	56.4	761,816	17.5	3,210,929	73.9
Texas.......................	11,598,894	76.6	947,146	6.3	12,546,040	82.9
Utah	1,228,567	78.1	157,145	10.0	1,385,712	88.0
Vermont.....................	472,874	100.0	0	0.0	472,874	100.0
Virginia......................	3,750,284	70.9	460.699	8.7	4,210,983	79.6
Washington..................	3,226,608	72.5	505,437	11.4	3,732,045	83.9
West Virginia	778,715	54.5	256,355	17.9	1,035,070	72.4
Wisconsin	2,021,672	49.6	877,326	21.5	2,898,998	71.1
Wyoming....................	234,672	63.3	40,965	11.0	275,637	74.3

[1] Civilian noninstitutionalized population 16 years of age and over based on Census 2000.

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 Census 2000. (See Appendix B 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.[5] 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 noninstitutionalized population, 16 years of age and older (CNP16+), for the state.

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

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.

---

[5] The variance of an estimator, u, based on a two-stage sample has the general form:
$Var(u) = Var_I E_{II}(u|$ set of sample PSUs) $+ E_I Var_{II}(u|$ 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

Substituting.

   q = 1 − p.

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

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

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 2005. Table 3-1 provides information on the proportion of the population in sample areas for each state.

The SCHIP sample is allocated among the states after the CPS sample is allocated. A sampling interval accounting for both the CPS and SCHIP samples can be computed as:

$$SI_{COMB} = (SI_{CPS}^{-1} + SI_{SCHIP}^{-1})^{-1}$$

The between-PSU variance component for the combined sample in the three states which were restratified for SCHIP can be estimated using a weighted average of the individual CPS and SCHIP between-PSU variance. The weight is the proportion of the total state sample accounted for by each individual survey:

$$\sigma_{B,COMB}^2 = (SI_{COMB}/SI_{CPS})\sigma_{B,CPS}^2 + (1 - SI_{COMB}/SI_{CPS})\sigma_{B,SCHIP}^2$$

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 civilian noninstitutionalized population.

2. Give each housing unit in the population one and only one chance of selection, with virtually all housing units in a state or substate area 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 respondent burden, cost, and other constraints.

4. Select within-PSU sample units 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 2000 Decennial Census of Population and Housing and the Building Permit Survey. Census 2000 collected information on all living quarters existing as of April 1, 2000, including characteristics of living quarters as well as the demographic composition of people 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, 2000, a sample of building permits issued in 2000 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 Census 2000 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.[6] 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 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.[7]

Development of Sampling Frames

Results from Census 2000 and the Building Permit Survey, and the relationship between these two sources, are used to develop 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 people on the property and have direct access to their living quarter 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 people who share the living quarter. About 97 percent of the population counted in Census 2000 resided in housing units.

The second type of living quarter is a group quarters. A group quarter 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 quarter is distinguished from a housing unit if it houses ten or more unrelated people. The group quarters population is classified as institutional or noninstitutionalized and as military or civilian. CPS targets only the civilian noninstitutionalized 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 noninstitutionalized housing by the time it is scheduled for interview. Less than 3 percent of the population counted in Census 2000 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

---

[6] CPS sample selection is coordinated with the following demographic surveys in the 2000 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.

[7] 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.

use of census information to reduce variance of estimates, ensure adequate coverage, and minimize 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 of the area—covered or not covered.

An address is considered complete if it describes a specific location; otherwise, the address is considered incomplete. (When Census 2000 addresses cannot be used to locate sample units, area listings must be performed 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. A USU in the unit frame consists of a geographically 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 nonresidential use, and others may be split up into several housing units. These addresses remain sample units, resulting in a small variability in cluster size.

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 people within a group quarter 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 Census 2000.

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 when USU size differs 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 Census 2000 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 Census 2000 addresses of individual group quarters or people within a group quarter are not used in the sampling. The number of housing unit equivalents is computed by dividing the Census 2000 group quarters population by the average number of people per household (calculated from Census 2000 as 2.59).

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:

$$\text{GQ block MOS} = \frac{\text{NIGQPOP}}{(4)(2.59)} + \text{MIL} + \text{IGQ} \qquad (3.4)$$

where

NIGQPOP = the noninstitutionalized group quarters population in the block from Census 2000.

MIL = the number of military barracks in the block from Census 2000.

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 Census 2000.

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 noninstitutionalized 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 Census 2000.

Special situations in old construction. During development of the old construction frames, several situations are given special treatment. 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 Census 2000.

Permit Frame

Permit frame sampling ensures coverage of housing units built since Census 2000. 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 Census 2000 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 Census 2000 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 Census 2000. 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 Census 2000 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

Table 3–2 summarizes the features of the sampling frames and CPS USU size discussed above. Roughly 80 percent of the CPS sample is from the unit frame, 12 percent is from the area frame, and less than 1 percent is from the group quarters frame. In addition, about 6 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	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 2000 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 2000 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, noninstitutionalized 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 Census 2000 and geography. Sorting minimizes within-PSU variance of estimates by grouping together units or measures with similar characteristics. The Census 2000 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.

General Sampling Procedure

The CPS sampling in the unit frame and GQ frame is a one-time operation that involves selecting enough sample for the decade. In the area and permit frames, sampling is a continuous operation. To accommodate the CPS rotation system and the phasing in of new sample designs, 21 samples are selected. A systematic sample of USUs is selected and 20 adjacent sample USUs identified. The group of 21 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.[8] Random numbers are used to calculate random starts. Random starts determine the first sampled USU in a BPC for each frame.

---

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

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 within-PSU sort criteria.

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 $RS + (n(SI_w)) \leq N$. 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 sample. The 20 USUs that follow the sequence number are selected as the next 20 samples. This method may yield hit strings with fewer than 21 samples (called incomplete hit strings) at the beginning or end of BPCs.[9] Allowing incomplete hit strings ensures that each USU has the same probability of selection.

6. A sample designation uniquely identifying 1 of the 21 samples is assigned to each USU in a hit string. The 21 samples are designated A77 through A97 for the CPS, and B77 through B97 for the SCHIP.

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 1 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 15.

Rotation group. The 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. The 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 a single county within a stratification PSU. 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 an 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.

---

[9] 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 20).

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 3–3) are selected in two BPCs.

These sample USUs are sorted and survey design codes assigned as shown in Table 3–4.

The example in Table 3–4 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 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 3–5.

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 3–3. **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 3–4. **Example of Post-Sampling Survey Design Code Assignments Within a PSU**

BPC number	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 3–5. **Examples of Post-Sampling Operational Code Assignments Within a PSU**

BPC number	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

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 for more on this procedure.

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 single 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:

   A77(1), A77(2), ..., A77(8), A78(1), A78(2), ..., A78(8), ..., A97(1), A97(2), ..., A97(8).

This 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).

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 2006. 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 2006, rotation groups 3, 4, 5, and 6 of A79; 7 and 8 of A80; and 1 and 2 of A81 are designated for interview.

2. Consecutive monthly samples have six rotation groups in common. The sample housing units in A79(4–6), A80(8), and A81(1–2), for example, are interviewed in January and February of 2006.

3. Monthly samples 1 year apart have four rotation groups in common. For example, the sample housing units in A80(7–8) and A81(1–2) are interviewed in January 2006 and January 2007.

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 2006, the sample housing units in A82(3) are interviewed for the first time and the sample housing units in A80(7) are interviewed for the fifth time after last being in sample in January.

Figure 3–1. **CPS Rotation Chart: January 2006–April 2008**

Year/month	A/B79	A/B80	A/B81	A/B82	A/B83	A/B84
2006 Jan	3456	78	12			
Feb	4567	8	123			
Mar	5678		1234			
Apr	678	1	2345			
May	78	12	3456			
June	8	123	4567			
July		1234	5678			
Aug		2345	678	1		
Sept		3456	78	12		
Oct		4567	8	123		
Nov		5678		1234		
Dec		678	1	2345		
2007 Jan		78	12	3456		
Feb		8	123	4567		
Mar			1234	5678		
Apr			2345	678	1	
May			3456	78	12	
June			4567	8	123	
July			5678		1234	
Aug			678	1	2345	
Sept			78	12	3456	
Oct			8	123	4567	
Nov				1234	5678	
Dec				2345	678	1
2008 Jan				3456	78	12
Feb				4567	8	123
Mar				5678		1234
Apr				678	1	2345

Overlap of the Sample

Table 3–6 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 3-6. **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 2004, the 2000 census-based design was phased in through a series of changes completed in July 2005 (U.S. Department of Labor, 2004).

REFERENCES

Cahoon, L. (2002), "Specifications for Creating a Stratification Search Program for the 2000 Sample Redesign (3.1-S-1)," Internal Memorandum, Demographic Statistical Methods Division, U.S. Census Bureau.

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., William N. Hurwitz, and William G. Madow, (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, David Judkins, Rajendra Singh, and Mindi Schautz, (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, E. Ann Vacca, Charles Wolters, and Rosa Hernandez, (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 Lawrence Cahoon, (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, December 2004.

Chapter 4.
Preparation of the Sample

INTRODUCTION

The Current Population Survey (CPS) 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 CPS sample more than once over the course of a decade, or in more than one of the household surveys conducted by the U. S. Census Bureau.

4. Provide cost-efficient data collection by producing most of the sampling materials needed for both the 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 a number of benefits. For example, training costs are reduced when CPS field representatives 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, sharing sampling materials helps ensure that respondents will be in only 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.

The next section is an overview, while subsequent sections provide a more in-depth description of the CPS sample preparation.

The successful completion of the CPS data collection rests on the combined efforts of headquarters and regional staff. 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. The regional offices coordinate the interview activities of the interviewing staff.

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 information (e.g., maps, lists of building permits) to a set of addresses. This conversion is described below.

Address Identification in the Unit Frame

About 80 percent of the CPS sample is selected from the unit frame. The unit frame sample is selected from a 2000 census file that contains the information necessary for within-PSU sample selection, but does not contain address information. The address information from the 2000 census is stored in a separate file. The addresses of unit segments are obtained by matching the file of 2000 census information to the file containing the associated 2000 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 2000 census information is reviewed in an attempt to complete the address before sending it to the field representative for interview. In sampling operations, this facilitates the formation of clusters of about 4 housing units, called Ultimate Sampling Units (USUs), as described in Chapter 3.

Address Identification in the Area Frame

About 12 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 individual

housing units. This is because many of the addresses in the area frame 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 an automated map of the blocks that contain the area segment, the addresses within the block(s) that are in the 2000 census file, the number of measures the block contains, and which measure is associated with the area segment. Before the individual housing units in 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 canvass the block to create a complete list of the housing units located in the block. This activity is called 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 will be designated for each month's sample.

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, 2000, but there is insufficient information about their living arrangements within the group quarters to provide a tangible sampling unit for the CPS. Measures are selected during within-PSU sampling since there is no way to associate the selected sample cases with people 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 2000 census addresses, similar to the process for the unit frame. This is a one-time operation done at headquarters. Before the individuals 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 people, rooms, or beds) or obtains a count of eligible sample units from a usable register. This is referred to as a listing operation. Then a systematic sampling pattern is applied to the listing to identify the individuals to be interviewed at the group quarters facilities.

Address Identification in the Permit Frame

The proportion of the CPS sample selected from the permit frame increases over the decade as new housing units are constructed. The CPS sample is redesigned about 4 or 5 years after each decennial census, and at this time the permit sample makes up about 6 percent of the CPS sample; this proportion has historically increased about 1 percent a year. Hypothetical measures are selected during within-PSU sampling in anticipation of the construction of new housing units. Identifying the addresses for these new units involves a listing operation at the building permit office, clustering addresses to form measures, and associating these addresses with the hypothetical measures (or USUs) in the sample.

The Census Bureau conducts the Building Permit Survey, which collects information on a monthly basis from each building permit office (BPO) nationwide 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 BPO contains the measure that is in sample. Using an automated instrument, a field representative visits the BPO to list addresses of units that were authorized to be built; this is the Permit Address Listing (PAL) operation. This list of addresses is transmitted to headquarters, where clusters are formed that correspond one-to-one with the measures. Using this link between addresses and measures, the clusters of four addresses to be interviewed in each permit segment are identified.

Forming clusters. To ensure some geographic clustering of addresses within permit measures and to make PAL listing more efficient, information collected by the Survey of Construction (SOC)[1] is used to identify many of the addresses in the permit frame. The Census Bureau collects information on the characteristics of units to be built for each permit issued by BPOs that are in the SOC. This information is used to form measures in SOC building permit offices. This data is not collected for non-SOC building permit offices.

1. SOC PALs are listings from BPOs that are in the SOC. If a BPO is in the SOC, then the actual permits issued by the BPO 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 identified 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.

---

[1] The Survey of Construction (SOC) is conducted by the Census Bureau in conjunction with the U.S. 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.

Even without geographic clustering, there is some degree of compactness inherent in the manner in which measures are formed for sampling. The units drawn from the SOC permits are assigned to measures in the same order in which they were listed in the 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 BPOS not in the SOC. At the time of sampling, the only data known for non-SOC BPOs is a cumulative count of the units authorized on all permits from a particular office for a given month (or year). Therefore, all addresses for a BPO/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.

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 without addresses 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 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 at headquarters. Performing the listing and sampling in two separate steps allows each step to be verified and allows more complete control of sampling procedures to avoid bias in designating the units to be interviewed.

In order to ensure accurate and complete coverage of the area and group quarters segments, the initial listing is updated 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 2000 census is no longer accurate for a multiunit structure and the field representative cannot adequately correct the information.

For multiunit addresses (addresses where the expected number of unit designations is two or more), the field representative receives a segment folder containing a preprinted (computer-generated) Multi-Unit Listing Aid (MULA), Form 11-12, showing unit designations for the segment as recorded in the 2000 census. The MULA displays the unit designations of all units in the structure, even if some of the units are not in sample. Other important information on the MULA includes: the address, the expected number of units at the address, the sample designations, and serial numbers for the selected sample units. If the address is incomplete (missing house number, street name), the field representative receives an Incomplete Address Locator Action Form, which provides additional information to locate the address.

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

- For addresses with 2–4 units, verifies that the 2000 census information on the MULA is accurate and corrects the listing sheet when it does not agree with what he/she finds at the address.

- For larger addresses of 5 or more units, resolves missing and duplicate unit designations only for addresses with missing or duplicate unit designations that are in any sample.

- If the changes are so extensive that a MULA cannot handle the corrections, relists the address on a blank Unit/Permit Listing Sheet.

After the field representative has an accurate MULA or listing sheet, he/she conducts an interview for each unit that has a current 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 (using the sampling pattern on the MULA) 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 necessary. The field representative may correct an in-sample listing during any visit if a change is noticed.

For single-unit addresses, a preprinted listing sheet is not provided to the field representative since only one unit is expected, based on the 2000 census information. If the address is incomplete (missing house number, no street name), the field representative receives an Incomplete Address Locator Form, which provides additional information to locate the address. If a field representative discovers other units at the address at the time of interview, he/she prepares a Unit/Permit Listing Sheet and lists the

extra units. These additional units, up to and including 15, are interviewed. If the field representative discovers more than 15 units, he/she must contact the regional office for subsampling instructions. For an example of a segment folder, MULA, Unit/Permit Listing Sheet, and Incomplete Address Locator Actions Form, see Appendix A.

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 an accurate list of living quarters.

This is a dependent operation conducted via a laptop computer using the Automated Listing and Mapping Instrument (ALMI) software. The field representative starts with a list of the addresses within the block and a digitized map of the block with some addresses mapspotted onto it. The addresses come from the Master Address File (MAF), which is a list of 2000 census addresses, updated periodically through other operations. The digitized map is downloaded from the Census Bureau Topological Integrated Geographic Encoding Reference (TIGER®) system.

The field representative updates the block information by matching the living quarters and block features found to the list of addresses and map on the laptop, and makes additions, deletions, or other corrections as necessary. The field representative collects additional data for any group quarters in the block using the Group Quarters Automated Instrument for Listing (GAIL) software.

If the area segment is within a jurisdiction where building permits are issued, housing units constructed since April 1, 2000, 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, 2000, Census Day, are already represented by segments in the permit frame. The field representative determines the year each unit was built except for: units in the 2000 census, mobile homes and trailers, group quarters, and nonstructures (buses, boats, tents). To determine "year built," the field representative inquires at each appropriate unit and enters the appropriate information on the computer.

If an area segment is not in a building-permit-issuing jurisdiction, then housing units constructed after the 2000 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 files are transmitted to headquarters where staff then apply the sampling pattern and identify the units to be interviewed.

Periodic updates of the listing are done to reflect change in the housing inventory in the listed block. 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 field representative updates the area block by verifying the existence of each unit and map feature, accounting for units or features no longer in existence, and adding any new units or features.

Listing in the Group Quarters Frame

The listing procedure is applied to group quarters addresses in the CPS sample that are in the group quarters or area frames. 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 (rooms, beds, persons, etc.) at the group quarters. The same procedures apply for group quarters found in the area frame.

Group quarters listing is an independent listing operation conducted via a laptop computer. The instrument, referred to as the Group Quarters Automated Instrument for Listing (GAIL), is 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, or to obtain a count of the number of eligible units from a register (card file, computer printout, or file) located at the group quarters. Field representatives do not list institutional and military groups quarters; however, they verify that the status has not changed from institutional or military. If it has changed, the field representative will list the addresses of noninstitutional units.

After listing group quarters units or obtaining a count of eligible units from a register, the files are transmitted to headquarters, where staff then apply the sampling pattern to identify the group quarters unit(s) (or units corresponding to sample line(s) within a register) to be interviewed.

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 BPO. This is done shortly after the permit has been issued by the BPO and is associated with a sample hypothetical measure. The second listing is required when the field representative visits the unit to conduct an interview.

PAL operation. For each BPO containing a sample measure, a field representative visits the BPO and lists the necessary permit and address information using a laptop computer. 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, a field representative visits the new construction site and draws a Permit 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 permit segments. Prior to interviewing, the field representative will receive a segment folder containing a Unit/Permit Listing Sheet (Form 11–3) and, possibly, a Permit Sketch Map to help locate the address. For an example of a segment folder, Unit/Permit Listing Sheets, and a Permit Sketch Map, see Appendix A.

At the time of the interview, the field representative verifies or corrects the basic address. Since the PAL operation does not capture unit designations at a multiunit address, the field representative will list the unit designations prior to interviewing. At both single and multiunit addresses, the field representative will also note any relevant information about the address that would affect sampling or interviewing, such as conversions, abandoned permits, construction-not-started situations, and more units found than expected for the permit address.

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 more than 15 units are in sample, the field representative must contact their regional office for subsampling instructions.

The listing of permit segments is not updated systematically; 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.

THIRD STAGE OF THE SAMPLE DESIGN (SUBSAMPLING)

Although the CPS sample is often characterized as a two-stage sample, chapter 3 describes a third stage of the sample design, referred to as subsampling. The need for subsampling depends 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 the operations that break the sample down into manageable interviewer workloads and transmit 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 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 the 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

Technological advances have changed data preparation and collection activities at the Census Bureau. Since the mid-1980s, 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. This system has two principal components: computer-assisted personal interviewing (CAPI) and centralized computer-assisted telephone interviewing (CATI). Chapter 7 explains 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 loading the monthly sample into the central database, transmission of CATI cases to the telephone centers, and database maintenance.

2. **Regional office operations in the central database.** The regional office operations include preparation of assignments, transmission of cases to field representatives, 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 that are in the sample.

Regional Office (RO) Operations

Once the monthly sample has been loaded into the central database, the ROs can begin the assignment preparation

phase. This includes making assignments to field representatives and selecting cases for the three telephone facilities. Regional offices access the central database to break down the assignment areas geographically and key in information that is used to aid the monthly field representative assignment operations. The RO 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.

The CATI assignments are also made at this time. Each RO selects at least 10 percent of the sample for centralized telephone interviewing. The selection of cases for CATI involves several steps. Cases may be assigned to CATI if several criteria are met, pertaining to the household and the time in sample. In general terms, the criteria are as follows:

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

2. The field representative may recommend that the case be sent to CATI.

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

The ROs may temporarily assign cases to CATI in order to cover their workloads in certain situations, primarily to fill in for field representatives who are ill or on vacation. When interviewing for the month is completed, these cases will automatically be reassigned for CAPI.

The majority 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 details.

The final step is the releasing of assignments. After all changes to the interview and CATI assignments have been made, and all assignments have been reviewed for geographic efficiency and the proper balance among field representatives, the ROs release the assignments. The release of assignments by all ROs signals the Master Control in the central data to create the CATI workload files.

After assignments are made and released, the ROs transmit the assignments to the central database, which places the assignments on the telecommunications server for the field representatives. 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 files include the household demographic information and labor force data collected in previous interviews. 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 ROs 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, respondent letters, and other supplies requested by the field representative.

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 important component of the Current Population Survey (CPS) is the questionnaire, also called the survey instrument. The survey instrument utilizes automated data collection methods; that is, computer-assisted personal interviewing (CAPI) and computer-assisted telephone interviewing (CATI). This chapter describes and discusses 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. Supplemental questions are added to the CPS in most months and cover a number of different topics. The order in which interviewers attempt to collect information is: (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.

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 11.)

CONCEPTS AND DEFINITIONS

Household and Demographic Information

Upon contacting a household, interviewers proceed with the interview unless 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 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 1 (presumably one of the individuals who owns or rents the unit) becomes the reference person. 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 are also 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.

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 compose the bulk of such absent household members, but people away on business or vacation are also included. (Not included are individuals in institutions or the military.) Once household/nonhousehold membership has been established for all people 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 present. Subfamilies are defined as families that live in housing units where none of the members of the family are related to the reference person. A household may contain 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, 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 is also collected. 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.

Educational attainment for each person in the household 15 or older is obtained through a question asking about the highest grade or degree completed. Additional questions are asked for several educational attainment categories to ascertain the total number of years of school or credit years completed.

Questions on race and Hispanic origin comply with federal standards. Respondents are asked a question to determine if they are Hispanic, which is considered an ethnicity rather than a race. The question asks if the individual is Spanish, Hispanic, or Latino, and is placed before the question on race. Next, all respondents, including those who identify themselves as Hispanic, are asked to choose which of the following races they consider themselves to be: White, Black or African American, American Indian or Alaska Native, Asian, or Native Hawaiian or Other Pacific Islander. Responses of "other" are accepted and allocated among the race categories. Respondents may choose more than one race.

Labor Force Information

To avoid any chance of misunderstanding, it is emphasized here that the CPS provides a measure of monthly employment—not jobs.

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, industry and related aspects of the working population. The major labor force categories are defined hierarchically and, thus, 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 during the reference week (defined below), are not asked the questions about having looked for work, and 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 people 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 not in the labor force. 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 noninstitutionalized population. In the CPS, labor force data are restricted to people 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 people. Employed people 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 people 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 people 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 people with multiple unincorporated businesses and people with multiple jobs as unpaid family workers.

CPS respondents are asked questions each month to identify multiple jobholders. First, all employed people 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?"

Hours of work. Information on both actual and usual hours of work have been collected. Published data on hours of work relate to the actual number of hours spent "at work" during the reference week. For example, people 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 people working in more than one job, the published figures relate to the number of hours worked at all jobs during the week.

Data on people "at work" exclude employed people 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 people, 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 people who usually work part-time and were at work 1 to 34 hours during the reference week for a noneconomic reason. Noneconomic reasons include illness 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, people also are classified according to their usual full- or part-time status. 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 fewer 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 fewer 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 people who usually work full-time and unemployed people who are either looking for full-time work or are on layoff from full-time jobs. The part-time labor force consists of employed people who usually work part-time and unemployed people who are seeking or are on layoff from part-time jobs.

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 according to the job at which he or she worked the largest number of hours. The unemployed are classified according to their last jobs. The occupational and industrial classification of CPS data is based on the coding systems used in Census 2000. A list of these codes can be found in the **Alphabetical Index of Industries and Occupations** at <http://www.census.gov/hhes/www/ioindex/view.html>. The class-of-worker classification assigns workers to one of the following categories: 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. The self-employed 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 individuals 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, marriage, or adoption.

Occupation, industry, and class-of-worker on second job. The occupation, industry, and class-of-worker information 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 Employment Statistics; for an explanation of this survey see BLS Handbook of Methods at <http://www.bls.gov/opub/hom/homch1_a.htm>). For the majority of multiple jobholders, occupation, industry, and class-of-worker data for their second jobs are collected only from one-fourth of the sample—those in their fourth or eighth monthly interview. 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 individual is employed or not, along with the questions an employed person typically will receive, are presented in Figure 5–1 at the end of this chapter.

Earnings. Information on what people earn at their main job is collected only for those who are receiving their fourth or eighth monthly interviews. This means that earnings questions are asked of only one-fourth of the survey respondents each month. 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" means as perceived by the respondent. If the respondent asks for a definition of usual, interviewers are instructed to define the term as more than half the weeks worked during the past 4 or 5 months. Respondents 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, and for self-employed people whose businesses are incorporated; earnings data are not collected for self-employed people whose businesses are unincorporated. (Earnings data are not edited and are not released to the public for the "self-employed 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 people. All people 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. 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 (people on layoff). Job seekers must have engaged in an 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 "help wanted" ads and taking a job training course, as opposed to actually answering "help wanted" ads or placing "employment wanted" ads. 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, people 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 composed of (a) people 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:* people who quit or otherwise terminated their employment voluntarily and began looking for work; (3)*People who completed temporary jobs:* individuals who began looking for work after their jobs ended; (4)*Reentrants:* people who previously worked but were out of the labor force prior to beginning their job search; (5)*New entrants:* individuals 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.

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 people 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 members of the civilian noninstitutionalized 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 those 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 people outside the labor force, one that includes discouraged workers as well as those who desire work but give other reasons for not searching (such as childcare problems, school, family responsibilities, or transportation problems) are also published regularly. This group is made up of people 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.

Questions about the desire for work among those who are not in the labor force are asked of the full CPS sample. Consequently, estimates of the number of discouraged workers as well as those with a marginal attachment to the labor force are published monthly rather than quarterly.

Additional questions relating to individuals' job histories and whether they intend to seek work continue to be asked only of people 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 people 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,** from <http://www.census.gov/hhes/www/ioindex/>.

U.S. Department of Labor, U.S. Bureau of Labor Statistics, **BLS Handbook of Methods,** from <http://www.bls.gov/opub/hom/>.

Chapter 6.
Design of the Current Population Survey Instrument

INTRODUCTION

Chapter 5 describes the current 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. For virtually every labor force concept, the current questionnaire wording is different from what was used previously. Data collection was redesigned so that the instrument is fully automated and is administered either on a laptop computer or from a centralized telephone facility. This chapter describes the work on the data collection instrument and changes that were made as a result of that work.

MOTIVATION FOR REDESIGNING THE QUESTIONNAIRE COLLECTING 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 remained unchanged from 1967 to 1994. The growth in the number of service-sector jobs and the decline in the number of factory jobs were 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. 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 sample became available. Spurred by all of these developments, the decision was made 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 U.S. Bureau of Labor Statistics (BLS) and U.S. Census Bureau developed 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 small. The one major definitional change dealt with the concept of discouraged workers; that is, people outside the labor force who are not looking for work because they believe that there are no jobs available for them. As 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 because 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 because 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 asked if 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 people 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 people, and people 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 people who have more than one job. Also, by separately collecting information on the number of hours multiple jobholders work on their main job and secondary jobs, estimates of the number of workers who combined two or more part-time jobs into a full-time work week, and the number of full- and part-time jobs in the economy can be made. The inclusion of the multiple job 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 the 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 of respondents who are 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 design 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 in an automated environment. For example, automated data collection allows capabilities such as (1) the use of dependent interviewing, that is carrying over 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. By automatically bringing up the next question on the interviewer's screen, computerization reduces the probability that an interview will ask the wrong set of questions. The computerized questionnaire also permits the inclusion of several built-in editing features, including automatic checks for internal consistency and 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 qualitative research, 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 using 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 suggest solutions. Analyses were based on information from item response distributions, respondent and interviewer debriefing data, and behavior coding of interviewer/ respondent interactions. For more on the evaluation methods used for redesigning the questions, see Esposito and Rothgeb (1997).

Item Response Analysis

The primary use of item response analysis was to determine whether different questionnaires produce different response patterns, which may, in turn, affect the labor force estimates. Unedited data were used for this analysis. Statistical tests were conducted to ascertain whether differences among the response patterns of different questionnaire versions were statistically significant. 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 interpretation of the response differences observed. (Response distributions were calculated on the basis of people who responded to the item, excluding those whose response was "don't know" or "refused.")

Respondent Debriefings

At the end of the test interview, respondent debriefing questions were administered to a sample of respondents to measure respondent comprehension and response formulation. 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. 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). For example, those who did not indicate in the main survey that they had done any work 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 the 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 was used to identify questions that needed revision, and to modify initial interviewer training and the interviewer manual. A secondary objective was to obtain information about the questionnaire, interviewer behavior, or respondent behavior that would 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. See Esposito and Hess (1992) for more information on interviewer debriefing.

HIGHLIGHTS OF THE QUESTIONNAIRE REVISION

A copy of the questionnaire can be obtained from the Internet at <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 new introductory statement reads, "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, "Does anyone in this household have a business or a farm?" This question determines 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 by all people 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 a business or farm is 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 holds multiple jobs. A follow-up 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 follow-up 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. The new hours series allows monthly estimates of usual hours worked to be produced for all employed people. 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 one 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 question which uses information collected during the previous month's interview in the current month's interview was implemented in the revised questionnaire for month-in-sample 2–4 households and month-in-sample 6–8 households. (Different variations of dependent interviewing were evaluated during testing. See Rothgeb et al. [1991] for more detail.)

In the revised CPS, respondents are provided 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," they 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 people 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 who do not meet the criteria for layoff status.

People 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, people reported being on layoff were never directly asked whether they expected to be recalled. In an effort to better capture the existing definition, people reported being on layoff in the revised questionnaire are asked "Has your employer given you a date to return to work?" People 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?" People 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 previous 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?" Responses that could be checked included:

- 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 were 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. 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." 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 those who report duration in whole months (between 1 and 4 months) a follow-up 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 people who are unemployed in consecutive

months. For people 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 those 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 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, 1998). 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 to "Not-in-the-Labor-Force" Questions

Response options of retired, disabled, and unable to work. 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 situation.

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 follow-up question is asked to determine whether he/she can do any gainful work during the next 6 months. Different versions of the follow-up probe are used depending on whether the person is disabled or unable to work.

Dependent interviewing for people reported to be retired, disabled, or unable to work. The revised questionnaire also is designed to use dependent interviewing for individuals 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 people outside the labor force (one that includes discouraged workers as well as those 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 people 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 a quarter of the sample, permitting estimates of the number of discouraged workers to be published monthly rather than quarterly.

Tests of the revised questionnaire showed that the quality of labor force data improved as a result of the redesign of the CPS questionnaire, and in general, measurement error diminished. 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 con-temporaneous advances in research on questionnaire design, also has helped inform the development of meth-ods for testing new or improved questions for the basic CPS and its periodic supplements (Martin [1987]; Oksen-berg; Bischoping, K.; Cannell and Kalton [1991]; Campan-elli, 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. Despite the benefits of adding new questions and improving existing ones, changes to the CPS should be approached cautiously and the effects measured and evaluated. When possible, methods to bridge differences caused by changes and 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., E. A. Martin, and J. M. Rothgeb (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., A. E. Polivka, and J. M. Rothgeb (1994), Revi-sions in the Current Population Survey Effective January 1994, **Employment and Earnings,** February 1994 vol. 41 no. 2 pp. 1337.

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

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

Esposito, J. L., and J. Hess (1992), "The Use of Inteviewer Debriefings to Identify Problematic Questions on Alternate Questionnaires," Paper presented at the Annual Meeting of the American Association for Public Opinion Research, St. Petersburg, FL.

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

Esposito, J. L. and J. M. Rothgeb (1997), "Evaluating Survey Data: Making the Transition From Pretesting to Quality Assessment," in Survey Measurement and Process Quality, New York: Wiley, pp.541–571.

Forsyth, B. H. and J. T. Lessler (1991), Cognitive Laboratory Methods: A Taxonomy, in P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman (eds.), **Measure-ment 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 Popula-tion Survey, **Proceedings of the Section on Survey Research Methods,** American Statistical Association, pp. 481–485.

Gaertner, G., D. Cantor, and N. Gay (1989), Tests of Alter-native 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 Cur-rent Population Survey, **Proceedings of the Section on Survey Research Methods,** American Statistical Associa-tion, pp. 420–424.

Martin, E. A. and A. E. Polivka (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., C. Cannell, and G. Kalton (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 S. M. Miller (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 J. M. Rothgeb (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., A. E. Polivka, K. P. Creighton, and S. R. Cohany (1992), Development of the Proposed Revised Current Population Survey Questionnaire, **Proceedings of the Section on Survey Research Methods,** American Statistical Association, pp. 56–65.

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 attempt to contact and interview a responsible person living in each sample unit selected 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 schedule, different types of interviews (due to differing MIS) are conducted by each FR within his/her weekly assignment. An introductory letter is sent to each sample household 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 Atlanta Regional Office. A personal-visit interview is required for all first month-in-sample households because the CPS sample is strictly a sample of addresses. The U.S. Census Bureau has no way of knowing who the occupants of the sample household are, or even whether the household is occupied or eligible for interview. (Note: For some MIS 1 households, telephone interviews are conducted if, 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 the CPS. There are many reasons an address may not be eligible for interview. For example, the address may have been converted to a permanent business, condemned or demolished, or it may be outside the boundaries of the area 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 the determination is made final.

Type B ineligibility includes units that 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), units occupied entirely by individuals who are not eligible for a CPS labor force interview (individuals with a usual residence elsewhere (URE), or in the Armed Forces). Such units are classified as Type B noninterviews. 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 interviews. 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), described in Chapter 11.

Additionally, one final set of households not interviewed for CPS are Type A households. These are households that the FR has determined are 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, 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 the determination is made final. Every effort is made to keep such noninterviews to a minimum. All Type A cases remain in the sample and are assigned for interview in all succeeding months. Even in cases of confirmed refusals (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 7–1. Introductory Letter

CPS-263(L) ATLANTA
(4-2004)

UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001
OFFICE OF THE DIRECTOR

FROM THE DIRECTOR
U.S. CENSUS BUREAU

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 U.S. Census Bureau obtains these figures, as well as information about people 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 people 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. We are conducting this survey under the authority of Title 13, United States Code, Section 182. Section 9 of this law requires us to keep all information about you and your household strictly confidential. We may use this information only for statistical purposes.

We have selected your address and about 58,000 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 qustion, 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,

Charles Louis Kincannon

Further information
may be obtained from:

Regional Director
U.S. Census Bureau
101 Marietta St., N.W., Suite 3200
Atlanta, GA 30303

Telephone: 404-730-3832

USCENSUSBUREAU
Helping You Make Informed Decisions

7–2   Conducting the Interviews

www.census.gov
Current Population Survey TP66
U.S. Bureau of Labor Statistics and U.S. Census Bureau

Figure 7–2 shows how the three types of noninterviews are classified and the various reasons that define each category. 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. Figure 7–3 shows the main housing unit information gathered for each noninterview category and a brief description of what each item covers.

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   Entire household in the Armed Forces
10   Entire household under age 15
11   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
7   Removed during subsampling
8   Unit already had a chance of selection
9   Unused line of listing sheet
10   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 worded exactly 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 information collected from 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, and race and ethnic 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. The FR has an opportunity 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, information about a few additional characteristics are collected after completion of the labor force portion of the interview. This information includes questions on family income and on all household members' countries of birth (and the country of birth of the member's father and mother) and, for the foreign born, 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 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) often provides 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 Items of Housing Unit Information Asked for Types A, B, and C**

Note: This list of items is not inclusive. The list covers only the main data items and does not include related items used to arrive at the final categories (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	TYPE A	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 on 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 is the type of housing unit (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	TYPE B	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 on 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	TYPE C	Which specific kind of Type C is the case.
2	PROPER	If there is any other building on 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 is the type of housing unit (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 any supplemental questions. Telephone interviewing 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. FRs must attempt to conduct a personal-visit interview for the fifth-month interview. After one attempt, a telephone interview may be conducted provided the original household still occupies the sample unit. This fifth-month interview follows 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, that is, a replacement household 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 telephone 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 prompts the FR to verify 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

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

Note: This list of items is not inclusive. The list covers only the main data items and does not include related items used to identify 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 4/1/00.
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 on 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 is the type of housing unit (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.

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 making the appropriate corrections, the instrument will take the interviewer 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 made possible by the computerization of the labor force interview. Information collected in the previous month's interview, like the household roster and demographic data, 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. Importing the previous month's job description into the current month's interview shows 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 labor force interview and reduces respondent burden. Other information collected using dependent interviewing is the duration of unemployment (either job search or layoff duration), and data on the not-in-labor-force subgroups of retired and disabled. Dependent interviewing is not used in the MIS 5 interviews or for any of the data collected solely in MIS 4 and 8 interviews.

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

Another milestone in the computerization of the CPS is the use of centralized facilities for computer- assisted telephone interviewing (CATI). The CPS has been experimenting with the use of CATI since 1983. 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 CATI facilities for interviewing, currently numbering about 7,000 cases each month. The facilities generally interview about 80 percent of the cases assigned to them. The net result is that about 12 percent of all CPS interviews are completed at a CATI facility.

Three CATI facilities are in use: Hagerstown, MD; Tucson, AZ; and Jeffersonville, IN. During the time of the initial phase-in of CATI data, use of a controlled selection criteria (see Chapter 4) allowed the analysis of the effects of the CATI collection methodology. Chapter 16 provides 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, particularly most large cities. Most of the cases sent to CATI are from major metropolitan areas. CATI is not used in most rural areas because the small sample sizes in these areas do not cause the FRs undo hardship. A concerted effort is made to hire some Spanish speaking interviewers in the Tucson Telephone Center, enabling Spanish interviews to be conducted from this facility. No MIS 1 or 5 cases are sent to the facilities, as explained above.

The facilities complete all but 20 percent of the cases sent to them. These uncompleted cases are recycled back to the field for follow-up and final determination. For this

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	HISPNON-R	What is the person's origin; the interviewer shows the respondent a flashcard from which he/she chooses the appropriate origin categories.
24	RACE*-R	What is the person's race; the interviewer shows the respondent a flashcard from which he/she chooses the appropriate race 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.

reason, the CATI facilities generally cease conducting the labor force portions of the interview on Wednesday of interview week, so the field staff has 3 to 4 days to check on the cases and complete any required interviewing or classify the cases as noninterviews. The field staff is highly successful in completing these cases as interviews, generally interviewing about 85–90 percent of them. The cases that are sent to the CATI facilities are selected by the supervisors in each of the regional offices, based on the FR's analysis of a household's probable acceptance of a CATI interview, and the need to balance workloads and meet specific goals on the number of cases sent to the facilities.

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

Note: The following list of edits is not 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 demographic edits past the Change screen.

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' Edits	
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."

Note: This list of items is not 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.

Figures 7–9 and 7–10 show the results of a typical month's (September 2004) 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 7.5 percent with an overall non-interview rate in the 21–23 percent range. In September 2004, the Type A rate was 7.56 percent. For the April 2003–March 2003 period, the CPS Type A rate was 7.25 percent. The overall noninterview rate for September 2004 was 22.98 percent, compared to the 12-month average of 23.19 percent.

Figure 7–9. **Interviewing Results (September 2004)**

Description	Result
Total HHLD	71,575
Eligible HHLD	59,641
Interviewed HHLD	55,130
Response rate	92.44%
Noninterviews	16,445
Rate	22.98%
Type A	4,511
Rate	7.56%
No one home	1,322
Temporarily absent	432
Refused	2,409
Other—specify	348
Callback needed—no progress	0
Type B	11,522
Rate	16.19%
Entire HH Armed Forces	130
Entire HH under 15	3
Temp. occupied with persons with URE	1,423
Vacant regular (REG)	7,660
Vacant HHLD furniture storage	514
Unfit, to be demolished	428
Under construction, not ready	429
Converted to temp. business or storage	167
Unoccupied tent or trailer site	354
Permit granted, construction not started	50
Other Type B	364
Type C	412
Rate	0.58%
Demolished	82
House or trailer moved	45
Outside segment	7
Converted to permanent business or storage	48
Merged	26
Condemned	9
Built after April 1, 2000	14
Unused serial no./listing Sheet Llne	54
Removed during subsampling	0
Unit already had a chance of selection	4
Other Type C	123

Figure 7–10. **Telephone Interview Rates (September 2004)**

Note: Figure 7–10 shows the rates of personal and telephone interviewing in September 2004. It is highly consistent with the usual monthly results for personal and telephone interviews.

Characteristic	Total	Telephone		Personal	
		Number	Percent	Number	Percent
Total.......	**55,130**	**35,566**	**64.5**	**19,564**	**35.5**
MIS 1&5	13,515	2,784	20.6	10,731	79.4
MIS 2–4, 6–8 ...	41,615	32,782	78.8	8,833	21.2

Chapter 8.
Transmitting the Interview Results

INTRODUCTION

With the advent of completely electronic interviewing, transmission of the interview results took on a heightened importance to the field representative (FR). The data transmissions between headquarters, the regional field office, and the FRs is all done electronically. 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 U.S. Census Bureau 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 the software and the assignment.

- Prevent unauthorized access.

- Handle mail messages.

The central database system at headquarters cannot initiate transmissions. Either the FR or the regional offices (ROs) must initiate any transmissions. Computers in the field are not continuously connected to the headquarters computers. Instead, the field computers contact the headquarters computers to exchange data using a toll free 800 number. 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 an electronic method of transmitting interview data is the need for complete security. 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 FRs' computers are compacted and

encrypted. All transmissions between headquarters, the ROs, the Centralized Telephone Facilities, and the National Processing Center (NPC) 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 an FR and the central database usually take place once per day during the survey 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 up his/her laptop computer to transmit the completed work. During the night, at a preset time, the computer modem 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 an 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 the FR has instant access to the central database as necessary. This type of procedure is used primarily around the time of closeout when an FR might have one or two straggler cases that need to be received by headquarters before the 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 require supervisory review or were completed at the RO.

Centralized Telephone Facility Transmission

Most of the cases sent to the Census Bureau's Centralized Telephone Facilities are successfully completed as computer-assisted telephone interviews (CATI). Those that cannot be completed from the telephone center are transferred to an FR prior to the end of the interview period. These cases are called "CATI recycles." Each telephone facility daily transmits both 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 has been assigned the case. Case notes that include the reason for recycle are also transmitted to the FR to assist in follow-up. Daily transmissions are performed automatically for each region every hour during the CPS interview week, sending reassigned or recycled cases to the FRs.

The RO staff also monitor the progress of the CATI recycled cases with the Recycle Report. All cases that are sent to a CATI facility are also assigned to an FR by the RO staff. The RO staff keep a close eye on recycled cases to ensure that they are completed on time, to monitor the reasons for recycling so that future recycling can be minimized, and 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 NPC coders to assign the appropriate I&O codes. This cycle repeats itself until all data are received by headquarters, usually on 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 the 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 the NPC. 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. (Week 1)

Field reps and CATI interviewers complete practice interviews and home studies. (Week 2)

Field reps pick up assignments via modem. (Week 2, Thu)

CATI interviewing / **CATI facilities transmit completed work to headquarters nightly.** (Week 3)

CATI interviewing ends*. (Week 3, Wed)

CAPI interviewing:
- FRs transmit completed work to headquarters nightly.
- ROs monitor interviewing assignments.

All assignments completed except for telephone holds. (Week 3, Sat)

All interviews completed. (Week 4, Mon)

ROs complete final field closeout. (Week 4, Tue)

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

NPC sends back completed I&O cases.

Initial process closeout. (Week 4, Wed)

NPC closeout. (Week 4, Wed)

Headquarters performs final computer processing
- Edits
- Recodes
- Weighting

(Week 4, Thu)

Census delivers data to BLS. (Week 4, Fri)

BLS analyzes data. (Week 5, Sun thru Thu)

Results released to public at 8:30 a.m. by BLS. (Week 5, Fri)

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

Chapter 9.
Data Preparation

INTRODUCTION

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

DAILY PROCESSING

For a typical month, computer-assisted telephone interviewing (CATI) starts on Sunday of the week containing the 19th of the month and continues through Wednesday of the same week. The answer files from these interviews are sent to headquarters on a daily basis from Monday through Thursday of this interview week. One file is received for all of the three CATI facilities: Hagerstown, MD; Tucson, AZ; and Jeffersonville, IN. Computer-assisted personal interviewing (CAPI) also begins on the same Sunday and continues through Monday of the following week. The CAPI answer files are again sent to headquarters daily until all the interviewers and regional offices have transmitted the workload for the month. This phase is generally completed by Wednesday of the following week.

The 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, confirming the item range checks, and rejecting invalid cases. Files containing records needing four-digit industry and occupation (I&O) codes are electronically sent to Jeffersonville for assignment of these codes. Once the Jeffersonville staff 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 Jeffersonville's I&O coding files have been returned and placed on the appropriate records on the data file, editing and imputation are performed.

INDUSTRY AND OCCUPATION (I&O) CODING

The operation to assign the I&O codes for a typical month requires ten coders for a period of just over 1 week to code data from 30,000 individuals. Sometimes the 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 new monthly CPS cases and those where a person's industry or occupation has changed since the previous month of interviewing, are sent to Jeffersonville to be coded. For those whose industry and occupation have not changed, the four-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 the four-digit numeric industry and occupation codes used in the 2000 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 processing has been completed.

After this operation, the batch of records is electronically returned to headquarters for the next stage of monthly production processing.

EDITS AND IMPUTATIONS

The CPS is subject to two sources of nonresponse. The largest is noninterview households. To compensate for this data loss, the weights of noninterviewed households are distributed among interviewed households, as explained in Chapter 10. The second source of data loss is from item nonresponse, which 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, Table 16–4).

One of three imputation methods are used to compensate for item nonresponse in the CPS. 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 (e.g., followed incorrect path of questions) and ensure that all appropriate questions have valid entries. For the most part, illogical entries or out-of-range entries have been eliminated with the use of electronic instruments; however, the edits still address these possibilities, which may arise from 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 the 3 imputation techniques described below.

The edits are run in a deliberate and logical sequence. 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 data is missing, it is assigned by looking at the age and sex of the person in conjunction with the known relationship of other household members. Missing occupation codes are sometimes assigned by analyzing the industry codes and vice versa. This technique is used as appropriate across all edits. If missing values cannot be assigned using this technique, they are assigned using one of the two following methods.

2. Longitudinal edits are used in most of the labor force edits, as appropriate. If a question is blank and the individual is in the second or later month's interview, the edit procedure looks at last month's data to determine whether there was an 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. The third imputation method is commonly referred to as "hot deck" allocation. This method assigns a missing value from a record with similar characteristics, which is the hot deck. Hot decks are defined by variables such as age, race, and sex. Other characteristics used in hot decks vary depending on the nature of the unanswered question. For instance, most labor force questions use age, race, sex, and occasionally another correlated labor force item such as full- or part-time status. This means the number of cells in labor force hot decks are relatively small, perhaps fewer 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 editing procedures, 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 and other related variables are used in this edit.

2. **Demographic edits and codes.** This processing step ensures consistency among all demographic variables for all individuals within a household. It ensures all interviewed 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/or sex and, possibly, by a related labor force characteristic.

4. **I&O edits and codes.** This processing step assigns four-digit industry and occupation codes to those I&O eligible individuals for whom 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 recode guidelines are also created. This edit uses longitudinal editing, relational allocation, and hot deck allocation. The hot decks are defined by such variables as age, sex, race, and educational attainment.

5. **Earnings edits and codes.** This processing step edits 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 and other related variables is used.

Chapter 10.
Weighting and Seasonal Adjustment 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 noninstitutionalized population and provides details by age, sex, race, and Hispanic origin. In addition, estimates for a number of other population subdomains (e.g., families, veterans, people 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 people 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 weighting steps in the estimation procedure.[1] In addition, the weighting at each step is replicated in order to derive variances for the labor force estimates. (See Chapter 14 for details.)

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

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

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

- Adjustment for nonresponse.

- First-stage ratio adjustment to reduce variances due to the sampling of primary sampling units (PSUs).

- National and state coverage adjustments to improve CPS coverage.

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

- Composite estimation using estimates from previous months to reduce the variances.

- Seasonally-adjusted estimates 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. Each of these involve additional weighting steps to produce the final estimate. The types of characteristics 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 the outgoing rotation group's 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 noninstitutionalized population. Although the processes described in this chapter have remained essentially unchanged since January 1978, and seasonal adjustment has been part of the estimation process since June 1975, modifications have been made in some of the procedures from time-to-time. For example, in January 1998, a new compositing procedure was introduced; in January 2003, new race cells for the first-stage, second-stage, national, and state coverage steps were added; in January 2005, the number of cells used in the national coverage adjustment and in the second-stage ratio adjustment was expanded to improve the estimates of children.

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's 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

---

[1] Weights are needed when the sampled elements are selected by unequal probability sampling. They are also used in poststratification and in making adjustments for nonresponse.

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, Hurwitz, and Madow, 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 between 7 and 8 percent. Other factors, such as occasional errors caused by the sample selection procedure or the omission of households or individuals missed by interviewers, can also introduce bias. These omitted households or people can be considered as having zero probability of selection. These two exceptions notwithstanding, 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 current design used in the CPS was selected with probabilities equal to the inverse of the required state sampling intervals. These sampling intervals are called the basic weights (or baseweights). Almost all sample persons within the same state have the same probability of selection. 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. New sample is continually selected and added to the CPS to provide coverage for newly constructed housing units. This results in a larger sample size with an associated increase in costs. Small maintenance sample reductions are implemented on a periodic basis to offset the increasing sample size. (See Appendix B.)

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 expected. During the estimation procedure, housing units in these USUs must receive special weighting factors (also called weighting control 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 a reduction in the bias and an increase in 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 provide 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. Item nonresponse occurs when a cooperating HU fails or refuses to provide some specific items of information. Procedures for dealing with this type of nonresponse are discussed in Chapter 9. Unit nonresponse refers to the failure to collect any survey data from an occupied sample HU. For example, data may not be obtained from an eligible household in the survey because of impassable roads, a respondent's absence or refusal to participate in the interview, or unavailability of the respondent for other reasons. This type of nonresponse in the CPS is called a Type A noninterview. Recently, the Type A rate has averaged to between 7 and 8 percent (see Chapter 16).

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 that are not interviewed is valid if 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 is present in CPS estimates when the nonresponding units differ in relevant respects from those that respond to the survey or to the particular items.

Noninterview Clusters and Noninterview Adjustment Cells

To reduce the size of the estimate's bias, the noninterview adjustment is performed based on sample PSUs that are similar in metropolitan status and population size. These PSUs are grouped together to form noninterview clusters. In general, PSUs with a metropolitan status of the same (or similar) size in the same state belong to the same noninterview cluster. PSUs classified as metropolitan are assigned to metropolitan clusters. Nonmetropolitan PSUs are assigned to nonmetropolitan clusters. Within each metropolitan cluster, there is a further breakdown into two noninterview adjustment cells (also called residence cells). Each is split into "central city" and "not central city" cells. The nonmetropolitan clusters are not divided further, making a total of 214 adjustment cells from 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} = the weighted count of interviewed households in cell j of cluster i, and

N_{ij} = 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 occurs:

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

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

- The cell contains only Type A noninterviewed households and no interviewed households.

If any one of these situations occurs, 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 people within the cluster. If after collapsing, any of the cells still meet any of the situations above, the cell is output to an "extreme cell file" that is created for review each 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 the 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 origin.[2] 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 five ratio adjustments in the CPS estimation process: the first-stage ratio adjustment, the national coverage adjustment, the state coverage adjustment, the second-stage ratio adjustment, and the composite ratio adjustment leading to the composite estimator.

In the first-stage ratio adjustment, weights are adjusted so that the distribution of the single-race Black population and the population that is not single-race Black (based on the census) in the sample PSUs in a state corresponds to the same population groups' census distribution in all PSUs in the state. In the national-coverage ratio adjustment, weights are adjusted so that the distribution of age-sex-race-ethnicity[3] groups match independent estimates of the national population. In the state-coverage ratio adjustment, weights are adjusted so that the distribution of age-sex-race groups match independent estimates of the state population. 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 the CPS match independent state population estimates by age and sex.

FIRST-STAGE RATIO ADJUSTMENT

Purpose of the First-Stage Ratio Adjustment

The purpose of the first-stage ratio adjustment is to reduce the variance of sample state-level estimates caused by the sampling of PSUs, that is, the variance that would

---

[2] Hispanics may be any race.
[3] Although ethnicity, in this chapter, only means Hispanic or non-Hispanic origin, the word ethnicity will be used.

still be associated with the state-level estimates even if the survey included 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. The relative contribution of the between-PSU variance at the national level is generally quite small.

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 accuracy gained 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 the population by race (Black alone[4]/non-Black alone[5]) by age groups 0–15 and 16+ satisfies all three criteria.

By using the Black alone/non-Black alone categories, the first-stage ratio adjustment compensates for the fact that the racial composition of an NSR (non-self-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 Census 2000 data and are applied only to sample data for the NSR PSUs. Factors are computed for the two race categories (Black alone/non-Black alone) for each state containing NSR PSUs. The following formula is used to compute the first-stage adjustment factors for each state:

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

where

FS_{sj} = the first-stage factor for state s and age/race cell j (j=1, 2, 3, 4)

C_{sij} = the Census 2000 civilian noninstitutional population for NSR PSU i (sample or nonsample) in state s, age/race cell j

C_{skj} = the Census 2000 civilian noninstitutional population for NSR sample PSU k in state s, age/race cell j

_____
[4] Alone is defined as single race.
[5] Non-Black alone can be defined as everyone in the population who is not of the single race Black.

π_{sk} = 2000 probability of selection for sample PSU k in state s

n = total number of NSR PSUs (sample and non-sample) 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 Census 2000 civilian noninstitutionalized population in the appropriate age/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 alone and non-Black alone 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 an age/race cell in the state.

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 factor)
x (noninterview adjustment factor)
x (first-stage ratio adjustment factor)

The weight after the first-stage adjustment is called the first-stage weight.

NATIONAL COVERAGE ADJUSTMENT

Purpose of the National Coverage Adjustment

The purpose of the national coverage adjustment is to correct for interactions between race and ethnicity that are not addressed in the second-stage weighting. For example, research has shown that the undercoverage of certain combinations (e.g., non-Black Hispanic) cannot be corrected with the second-stage adjustment alone. The national coverage adjustment also helps to speed the convergence of the second-stage adjustment (Robison, Duff, Schneider, and Shoemaker, 2002).

Computing the National Coverage Adjustment

The national coverage adjustment factors are based on independently derived estimates of the population. (See Appendix C.) Person records are grouped into four pairs based on months-in-sample (MIS) (MIS 1 and 5, MIS 2 and 6, MIS 3 and 7, and MIS 4 and 8). This increases cell size

and preserves the structure needed for composite weighting. Each MIS pair is then adjusted to age/sex/race/ethnicity population controls (see Table 10–1) using the following formula:

$$F_{jk} = \frac{C_j}{E_{jk}}$$

where

F_{jk} = national coverage adjustment factor for cell j and month-in-sample pair k

C_j = national coverage adjustment control for cell j; national current population estimate for cell j

E_{jk} = weighted tally for the cell j and month-in-sample pair k (using weights after the first-stage adjustment)

For example, in October 2005, 80 percent of the respondents who were adjusted by a national coverage adjustment factor, received a F_{jk} value between 0.9868 and 1.2780. The factor F_{jk} is computed for each of the cells listed in Table 10–1.

The independent population controls used for the national coverage adjustment are from the same source as those for the second-stage ratio adjustment.

Extreme cells are identified for any of the following criteria:

- Cell contains less than 20 persons.
- National coverage adjustment factor is greater than or equal to 2.0.
- National coverage adjustment factor is less than or equal to 0.6.

No collapsing is performed because the cells were created in order to minimize the number of extreme cells.

Table 10–1: **National Coverage Adjustment Cell Definitions**

Black alone non-Hispanic			White alone non-Hispanic			White alone Hispanic			Non-White alone Hispanic			Asian alone non-Hispanic			Residual race non-Hispanic		
Age	Male	Female	Age	Male	Female	Age	Male	Female	Age	Male	Female	Age	Male	Female	Age	Male	Female
0–1			0			0			0–15			0–4			0–4		
2–4			1			1											
5–7			2			2						5–9			5–9		
8–9			3			3											
10–11			4			4						10–15			10–15		
12–13			5			5											
14			6			6											
15			7			7											
16–19			8			8			16+			16–24			16–24		
20–24			9			9											
25–29			10–11			10–11						25–34			25–34		
30–34			12–13			12–13											
35–39			14			14						35–44			35–44		
40–44			15			15											
45–49			16–19			16–19						45–54			45–54		
50–54			20–24			20–24											
55–64			25–29			25–29						55–64			55–64		
65+			30–34			30–34						65+			65+		
			35–39			35–39											
			40–44			40–44											
			45–49			45–49											
			50–54			50–54											
			55–59			55–64											
			60–62			65+											
			63–64														
			65–69														
			70–74														
			75+														

Weights After National Coverage Adjustment

After the completion of the national coverage adjustment, the weight for each person is the product of:

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

This weight will usually vary within households due to different household members having different demographic characteristics.

STATE COVERAGE ADJUSTMENT

Purpose of the State Coverage Adjustment

The purpose of the state coverage adjustment is to adjust for state-level differences in sex, age, and race coverage. Research has shown that estimates of characteristics of certain racial groups (e.g., Blacks) can be far from the population controls if a state coverage step is not used.

Computing the State Coverage Adjustment

The state coverage adjustment factors are based on independently derived estimates of the population. Except for the District of Columbia (DC), person records for the non-Black-alone population are grouped into four pairs based on MIS (MIS 1 and 5, MIS 2 and 6, MIS 3 and 7, and MIS 4 and 8). This increases cell size and preserves the structure needed for composite weighting. Person records for the Black-alone population for all states and the non-Black-alone population for DC are formed at the state level with all months-in-sample combined. For the Black-alone component of the adjustment, states were assigned to different tables (Tables 10–2A, 10–2B and 10–2C) based on the expected number of sample records in each age/sex cell. For the non-Black-alone component, all states except DC were assigned to Table 10–2D. (DC was assigned to Table 10–2C.) Each cell is then adjusted to age/sex/race population controls in each state[6] using the following formula:

$$F_{jk} = \frac{C_j}{E_{jk}}$$

where

F_{jk} = state coverage adjustment factor for cell j and MIS pair k.

C_j = state coverage adjustment control for cell j; state current population estimate for cell j.

E_{jk} = weighted tally for the cell j and MIS pair k.

---

[6] In the state coverage adjustment, California is split into two parts and each part is treated like a state—Los Angeles County and the balance of California. Similarly, New York is split into two parts with each part being treated as a separate state—New York City (New York, Queens, Bronx, Kings and Richmond Counties) and the balance of New York.

For example, in October 2005, 80 percent of the respondents who were adjusted by a state coverage adjustment factor, received a F_{jk} value between 0.8451 and 1.1517. The independent population controls used for the state coverage adjustment are from the same source as those for the second-stage ratio adjustment. (See next section for details on the second-stage adjustment.)

Extreme cells are identified for any of the following criteria:

- Cell contains less than 20 persons.

- State coverage adjustment factor is greater than or equal to 2.0.

- State coverage adjustment factor is less than or equal to 0.6.

No collapsing is performed because the cells were created in order to minimize the number of extreme cells.

Table 10–2: **State Coverage Adjustment Cell Definitions**

Table 10–2A — Black Alone

Age	Male and female combined
0+	

States assigned to Table 10–2A: HI, ID, IA, ME, MT, ND, NH, NM, OR, SD, UT, VT, WY
(all rotation groups combined)

Table 10–2B — Black Alone

Age	Male	Female
0+		

States assigned to Table 10–2B: AK, AZ, CO, IN, KS, KY, MN, NE, NV, OK, RI, WA, WI, WV
(all rotation groups combined)

Table 10–2C — Black Alone and Non-Black Alone for DC

Age	Male	Female
0–15		
16–44		
45+		

States assigned to Table 10–2C: Black Alone: AL, AR, CT, DC, DE, FL, GA, IL, LA, MA, MD, MI, MO, MS, NC, NJ, OH, PA, SC, TN, TX, VA, LA county, bal CA, NYC, bal NY; Non-Black Alone: DC
(all rotation groups combined)

Table 10–2D — Non-Black Alone

Age	Male	Female
0–15		
16–44		
45+		

States assigned to Table 10–2D: Non-Black Alone: All states + LA county + NYC + bal CA + bal NY (DC is excluded)
(by rotation group pair)

Weights After State Coverage Adjustment

After the completion of the state coverage adjustment, the weight for each person is the product of:

(baseweight) x (special weighting factor)
x (noninterview adjustment factor)
x (first-stage ratio adjustment factor)
x (national coverage adjustment factor)
x (state coverage adjustment factor)

This weight will usually vary within households due to different household members having different demographic characteristics.

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 month-in-sample pair 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 noninstitutionalized population for the 50 states and the District of Columbia by sex and age (0–15, 16–44, 45 and older).

- Total national civilian noninstitutionalized population for 26 Hispanic and 26 non-Hispanic age-sex categories (see Table 10–3).

- Total national civilian noninstitutionalized population for 56 White, 36 Black, and 34 "Residual Race" age-sex categories (see Table 10–4).

The adjustment is done separately for each MIS pair (MIS 1 and 5, MIS 2 and 6, MIS 3 and 7, and MIS 4 and 8). 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 ten 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 is also known as iterative proportional fitting or raking ratio estimation.

In addition to reducing the error in many CPS estimates and converging to the population controls within ten 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 weight for the ith sample record after the second-stage adjustment, and

W_{1i} = the weight for the ith record after the first-stage adjustment.

Thus, the raking adjusts the weights of the records so that the sample estimates converge to the population controls while minimally affecting the weights after the state coverage adjustment. The article by Ireland and Kullback (1968) provides more details on the properties of raking ratio estimation.

Table 10–3: Second-Stage Adjustment Cell by Ethnicity, Age, and Sex

Hispanic			Non-Hispanic		
Age	Male	Female	Age	Male	Female
0–4			0–4		
5–9			5–9		
10–15			10–15		
16–19			16–19		
20–24			20–24		
25–29			25–29		
30–34			30–34		
35–39			35–39		
40–44			40–44		
45–49			45–49		
50–54			50–54		
55–64			55–64		
65+			65+		

Table 10–4: **Second-Stage Adjustment Cell by Race, Age, and Sex**

Black alone			White alone			Residual race		
Age	Male	Female	Age	Male	Female	Age	Male	Female
0–1			0			0–1		
2–4			1			2–4		
5–7			2			5–7		
8–9			3			8–9		
10–11			4			10–11		
12–13			5			12–13		
14			6			14–15		
15			7			16–19		
16–19			8			20–24		
20–24			9			25–29		
25–29			10–11			30–34		
30–34			12–13			35–39		
35–39			14			40–44		
40–44			15			45–49		
45–49			16–19			50–54		
50–54			20–24			55–64		
55–64			25–29			65+		
65+			30–34					
			35–39					
			40–44					
			45–49					
			50–54					
			55–59					
			60–62					
			63–64					
			65–69					
			70–74					
			75+					

Sources of Independent Controls

The independent population controls used in the second-stage ratio adjustment and in the coverage adjustment steps are prepared by projecting forward the population figures derived from Census 2000 using information from a variety of other sources that account for births, deaths, and net migration. Subtracting estimated numbers of resident Armed Forces personnel and institutionalized people from the resident population gives the civilian noninstitutionalized population. 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 C 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/sex/age
- Hispanic/sex/age
- Race/sex/age

There is no collapsing done for the second-stage adjustment. The cells are designed to avoid having small cells. Instead, a small or extreme cell is identified as follows:

- It contains fewer than 20 people.
- It has an adjustment factor greater than or equal to 2.0.
- It has an adjustment factor less than or equal to 0.6.

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. These three steps are repeated through ten 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	$E_e = 822$ $F_r = \dfrac{1000}{822+221} = 0.959$ $E_r = E_eF_r = 788$	$E_e = 221$ $F_r = \dfrac{1000}{822+221} = 0.959$ $E_r = E_eF_r = 212$	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 10:

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 ten 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 30 adjustment factors (ten iterations of three rakes). The product of these 30 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 factor)
x (noninterview adjustment factor)
x (first-stage ratio adjustment factor)
x (national coverage adjustment factor)
x (state coverage adjustment factor)
x (second-stage ratio adjustment factor)

COMPOSITE ESTIMATOR

Once each record has a second-stage weight, an estimate of level for any given set of characteristics identifiable in the CPS can be computed by summing the second-stage 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.

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 estimate at the completion of the second-stage ratio adjustment which is 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 the part of the sample that 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.

Effective with the release of January 1998 data, the Bureau of Labor Statistics (BLS) implemented a new composite estimation method for the CPS. The new technique provides increased operational simplicity for microdata users and allows optimization of compositing coefficients for different labor force categories.

Under the procedure, weights are derived for each record that, when aggregated, produce estimates consistent with those produced by the composite estimator. Under the previous 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 CPS

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 (about 75 percent of the sample) continue in the 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 one 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 people 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 Mansur and Shoemaker, 1999, Breau and Ernst, 1983, Bailar, 1975.) Also, see section "Time in Sample" of Chapter 16. "Time in Sample" is concerned with the effects on labor force estimates from the CPS as the result of interviewing the same CPS respondents several times.

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.(1999).

Computing Composite Weights

Composite weights are produced only for sample people aged 16 or older. As described in previous sections, 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 nonresponse, and four successive stages of ratio adjustments to population controls are applied. The second-stage raking procedure ensures that sample weights add to independent population controls for states by sex and age, 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 the 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[7] and the District of Columbia (53 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 20 national age/sex/ethnicity cells and 46 national age/sex/race cells. These computations use cell definitions specified in Tables 10–5 and 10–6. Coefficients K = 0.4 and A = 0.3 are used for all UE estimates in all categories.

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 nine more times for a total of 10 iterations.

An analogous procedure is done for estimating the number of people employed using coefficients K = 0.7 and A = 0.4.

For the not-in-labor force category (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).

Extreme cells are identified if a cell size is less than 10, or an adjustment factor is greater than 1.3 or less than 0.7.

Final Weights

For people 16 years and older, the composite weights are the final weights. Since the procedure does not apply to persons under age 16, their final weights are the second-stage weights.

---

[7] California is split into two parts and each part is treated like a state—Los Angeles County and the balance of California. Similarly, New York is split into two parts with each part being treated as a separate state—New York City (New York, Queens, Bronx, Kings and Richmond Counties) and the balance of New York.

Table 10–5: **Composite National Ethnicity Cell Definition**

Hispanic		
Age	Male	Female
16–19		
20–24		
25–34		
35–44		
45+		

Non-Hispanic		
Age	Male	Female
16–19		
20–24		
25–34		
35–44		
45+		

Table 10–6: **Composite National Race Cell Definition**

Black alone		
Age	Male	Female
16–19		
20–24		
25–29		
30–34		
35–39		
40–44		
45+		

White alone		
Age	Male	Female
16–19		
20–24		
25–29		
30–34		
35–39		
40–44		
45–49		
50–54		
55–59		
60–64		
65+		

Residual race		
Age	Male	Female
16–19		
20–24		
25–34		
35–44		
45+		

Since data for all eight rotation groups are combined for the purpose of computing composite weights, summations of final weights within rotation group will not match independent population controls for people 16 years and older. Summations of final weights for the entire sample will be consistent with these second-stage controls, but will only match a selected number of those controls. If the

composite ethnicity x gender x age detail or race x gender x age detail is the same as the second-stage detail, then there will be a match. The composite-age detail is coarser than the second-stage age detail. For example, the composite procedure controls for White alone, male, ages 60–64; the second stage has more age detail (60–62 and 63–64). Summations of final (composite) weights for all White alone males by age will not match the corresponding second-stage population control for either the 60–62 or the 63–64 age group. However, the summed final weights for White alone, males, ages 60–64 will match the sum of the two second-stage population controls.

PRODUCING OTHER LABOR FORCE ESTIMATES

In addition to basic weighting to produce estimates for individuals, 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 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 estimates related to families and family composition. They also provide the basis for household weights. The family weight is derived from the second-stage 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 second-stage 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. This is because there is better coverage of females than for males. 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 can be used to produce estimates at the household level, such as the number of households headed by a woman.

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 they are 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 second-stage 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.

The outgoing rotation weights are obtained by multiplying the outgoing ratio adjustment factors by the second-stage weights. For consistency, an outgoing rotation group weight equal to four times the basic CPS family weight is assigned to all people in the two outgoing rotation groups who were not eligible for this special weighting (mainly military personnel and those 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 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 composite weights. Each respondent is classified as a veteran or a nonveteran. Veterans' records are classified into cells based on veteran status (Vietnam-era, Not Vietnam-era or Peactime), age and sex.

The composite weights for CPS veterans are tallied into type-of-veteran/sex/age 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 composite 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 alone/non-White alone), labor force status (unemployed, employed, and not-in-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 nonveteran control totals. The composite 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 nonveteran controls derived above. The factor is the ratio of the nonveteran control total to the

sample estimate. The composite weight for each nonveteran is multiplied by the appropriate factor to produce the nonveteran weight.

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 beyond those available in the stock data.

The estimation of monthly gross flows, and any other longitudinal use of the CPS data, 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 the sample according to the 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 people 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 from 1994 forward, a longitudinal weight allows users to estimate gross labor force flows by 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 inflates all estimates or final weights by the ratio of the current month's population controls to the sum of the second-stage weights for the current month in the

matched cases by sex. Although the technique provides 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 has 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. For information about more recent work on gross flows estimation methodology, refer to two articles in the *Monthly Labor Review*, accessible online through <http://www.bls.gov/opub/mlr/mlrhome.htm>. (See Frazis, Robison, Evans and Duff, 2005; Ilg, 2005.) Test tables produced by the BLS using the methodology described in the two articles cannot be reproduced using the longitudinal weights specified in this section.

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 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.

SEASONAL ADJUSTMENT

Short-run movements in labor-force time series are strongly influenced by seasonality, which refers to periodic fluctuations that are associated with recurring calendar-related events such as weather, holidays, and the opening and closing of schools. Seasonal adjustment is the process of estimating and removing these fluctuations to yield a seasonally-adjusted series. The reason for doing so is to make it easier for data users to observe fundamental changes in the level of the series, particularly those associated with general economic expansions and contractions.

For example, the unadjusted CPS levels of employment and unemployment in June are consistently higher than those for May because of the influx of students into the labor force. If the only change that occurs in the unadjusted estimates between May and June approximates the normal seasonal change, then the seasonally-adjusted estimates for the two 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 often will be in a direction opposite to the movement in the unadjusted series.

Refinements of the methods used for seasonal adjustment have been under development for decades. The current procedure used for the seasonal adjustment of national CPS series is the X-12-ARIMA program from the U.S. Census Bureau. This program is an enhanced version of earlier programs utilizing the widely used X-11 method, first developed by the Census Bureau and later modified by Statistics Canada. 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 outliers. In most uses of the X-11 method, including that

for national CPS labor force series, the seasonality is estimated as evolving rather than fixed over time. A detailed description of the X-12-ARIMA program is given in the U. S. Census Bureau (2002) reference.

The current official practice for the seasonal adjustment of CPS national labor force data is to run the X-12-ARIMA program monthly as new data become available, which is referred to as concurrent seasonal adjustment.[8] The season factors for the most recent month are produced by applying a set of moving averages to the entire data set, including data for the current month. While all previous-month seasonally-adjusted data are revised in this process, no revisions are made during the year. Revisions are applied at the end of each year for the most recent five years of data.

Seasonally-adjusted estimates of many national labor force series, including the levels of the civilian labor force, total employment and total unemployment, and all unemployment rates, are derived indirectly by arithmetically combining the series directly adjusted with X-12-ARIMA. For example, the overall national unemployment rate is computed using eight directly-adjusted series: females aged 16–19, males aged 16–19, females aged 20+, and males aged 20+ for both employment and unemployment. 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 are directly adjusted along with the components, such consistency would generally not occur, since X-11 is not a sum- or ratio-preserving procedure. It is not generally appropriate to apply factors computed for an aggregate series to its components because various components tend to have statistically significant different patterns of seasonal variation.

For up-to-date information and a more thorough discussion on seasonal adjustment of national labor force series, see any monthly issue of **Employment and Earnings** (U.S. Department of Labor).

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 S. C. Hillmer (1990), "The Time Series Approach to Estimation for Repeated Surveys," **Survey Methodology**, 16, pp. 195–215.

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.

Copeland, K. R., F. K. Peitzmeier, and C. E. Hoy (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., R. B. Tiller, and T. S. Zimmerman (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 R. McGuinness (1993), "Correlations and Adjustment Factors for CPS (VAR80 1)," Internal Memorandum for Documentation, January 6th, Demographic Statistical Methods Division, U.S. Census Bureau.

Frazis, H.J., E.L. Robison, T.D. Evans, and M.A. Duff (2005), "Estimating Gross Flows Consistent With Stocks in the CPS," **Monthly Labor Review**, September, Vol. 128, No. 9, pp. 3–9.

Hansen, M. H., W. N. Hurwitz, and W. G. Madow (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.

Ilg, R. (2005), "Analyzing CPS Data Using Gross Flows," **Monthly Labor Review**, September, Vol. 128, No. 9, pp. 10–18.

Ireland, C. T., and S. Kullback (1968), "Contingency Tables With Given Marginals," **Biometrika**, 55, pp. 179–187.

Kostanich, D. and P. Bettin (1986), "Choosing a Composite Estimator for CPS," presented at the International Symposium on Panel Surveys, Washington, DC.

Lent, J., S. Miller, and P. Cantwell (1994), "Composite Weights for the Current Population Survey," **Proceedings of the Survey Research Methods Section**, American Statistical Association, pp. 867–872.

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**, Vol. 15, No. 3, pp. 431–438.

Mansur, K.A. and H. Shoemaker (1999), "The Impact of Changes in the Current Population Survey on Time-in-Sample Bias and Correlations Between Rotation Groups," **Proceedings of the Survey Research Methods Section**, American Statistical Association, pp. 180–183.

---

[8] Seasonal adjustment of the data is performed by the Bureau of Labor Statistics, U.S. Department of Labor.

Oh, H. L. and F. Scheuren (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.

Robison, E., M. Duff, B. Schneider, and H. Shoemaker (2002), "Redesign of Current Population Survey Raking to Control Totals," **Proceedings of the Survey Research Methods Section**, American Statistical Assocation.

Scott, J. J., and T. M. F. Smith (1974), "Analysis of Repeated Surveys Using Time Series Methods," **Journal of the American Statistical Association**, 69, pp. 674–678.

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. 182–185.

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. Census Bureau (2002), *X-12-ARIMA Reference Manual* (Version 0.2.10), Washington, DC.

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., T. D. Evans, and R. B. Tiller (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. These studies keep the nation informed of the economic and social well-being of its people and are used 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. In addition, any supplemental inquiry must 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. An additional weighting method is often necessary because the supplement targets a different universe from that of the basic CPS. A supplement can also engender a different level of response or cooperation from respondents.

CRITERIA FOR SUPPLEMENTAL INQUIRIES

A number of criteria to determine the acceptability of undertaking supplements for federal agencies or other sponsors have been developed and refined over the years by the U.S. Census Bureau, in consultation with the U.S. 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 and not introduce 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 allowed.

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. Competing requirements for the use of Census Bureau staff or facilities that arise in dealing with a supplemental inquiry are resolved by giving the basic CPS first priority. The Census Bureau will not jeopardize the schedule for completing the CPS or other Census Bureau work to favor completing a supplemental inquiry within a specified time frame.

5. The subject matter must not be 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 supplemental survey's 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 a person's 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.

The Census Bureau does not make the final decision regarding the appropriateness or utility of the supplemental survey. The Office of Management and Budget (OMB), through its Statistical Policy Division, reviews the proposal to make certain it meets government-wide standards regarding the need for the data and the 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 a supplement is asked the additional questions that are included in 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 2004.

The Housing Vacancy Supplement (HVS) and American Time Use Surveys (ATUS) are unusual in that they are separate survey operations that base their sample on the results of the basic CPS interview. The HVS supplement collects additional information (e.g., number of rooms, plumbing, and rental/sales price) on housing units identified as vacant in the basic CPS and the ATUS collects information about how people spend their time. The HVS is collected at the time of the basic CPS interview, while the ATUS is collected after the household's last CPS interview. Probably the most widely used supplement is the Annual Social and Economic (ASEC) Supplement, 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 nonresponse rates. In addition, supplement universes may be different from the basic CPS universe. Thus, some supplements require weighting procedures that are different from those of the basic CPS. These variations are described for three of the major supplements—the Housing Vacancy Survey (HVS), the American Time Use Survey (ATUS), and the Annual Social and Economic (ASEC) supplement—in the following sections.

Housing Vacancy Survey (HVS) Supplement

Description of supplement

The Housing Vacancy Survey (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 people 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 the 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 that 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 statistical 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 2004**

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, December 1999	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 Social and Economic Supplement (formally known as the AnnualDemographic Supplement)	March 1995–2004	Provide data concerning family characteristics, household composition, marital status, education attainment, health insurance coverage, foreign-born population, previous year's income from all sources, work experience, receipt of noncash benefit, poverty, program participation, and geographic mobility. Periodicity: Annual	Census/BLS
Food Security	April 1995, September 1996, April 1997, August 1998, April 1999, September 2000, April 2001, December 2001, 2002, 2003, 2004	Provide data that will measure hunger and food security. It will provide data on food expenditure, access to food, and food quality and safety.	FNS
Race and Ethnicity	May 1995, July 2000, May 2002	Test alternative measurement methods to evaluate how best to collect these types of data.	BLS/Census
Marital History	June 1995	Collect information from "ever-married" persons on marital history.	Census/BLS
Fertility	June 1998, 2000, 2002, 2004	Provide data on the number of children that women aged 15-44 have ever had and the children's characteristics. Periodicity: Biennial.	Census/BLS
Educational Attainment	July 1995	Test several methods of collecting these data. Test 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, September 1997, 1999, August 2001, 2003	Provide data for veterans of the United States on Vietnam-theater and Persian Gulf-theater 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 1994–2004	Provide information on population 3 years old and older 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, February 2002, February 2003, June 2003, November 2003	Provide data for population 15 years and older on current and former use of tobacco products; restrictions of smoking in workplace for employed persons; and personal attitudes toward smoking. Periodicity: As requested.	NCI
Displaced Workers	February 1996, 1998, 2000, January 2002, 2004	Provide data on workers who lost a job in the last 5 years due to plant closing, shift elimination, or other work-related reason. Periodicity: Biennial.	BLS
Job Tenure/ Occupational Mobility	February 1996, 1998, 2000, January 2002, 2004	Provide 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, 2002, 2004	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 or awarded. April data will be matched to March data.	OCSE
Voting and Registration	November 1994, 1996, 1998, 2000, 2002, 2004	Provide demographic information on persons who did 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, 2004	Provide information about multiple job holdings and work schedules and telecommuters who work at a specific remote site.	BLS

Table 11–1. **Current Population Survey Supplements September 1994–December 2004**—Con.

Title	Month	Purpose	Sponsor
Computer Use/Internet Use	November 1994, October 1997, December 1998, August 2000, September 2001, October 2003	Provide information about household access to computers and the use of the Internet or World Wide Web.	NTIA
Participation in the Arts	August 2002	Provide data on the type and frequency of adult participation in the arts; training and exposure (particularly while young); and their musical artistic activity preferences.	NEA
Volunteers	September 2002, 2003, 2004	Provide a measurement of participation in volunteer service, specifically about frequency of volunteer activity, the kinds of organizations volunteered with, and types of activities chosen. Among nonvolunteers, questions identify what barriers were experienced in volunteering, or what encouragement is needed to increase participation.	USA Freedom Corps
Cell Phone Use	February 2004	Provide data about household use of regular landline telephones and household use of cell phones. If both were used in the household, it asked about the amount of cell phone usage.	BLS/Census

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, two HVS adjustments and a regional housing unit (HU) adjustment. (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 2000 census data. There are separate first-stage factors for year-round and seasonal housing units. For each state, they are calculated as the ratio of the state-level census count of vacant year-round or seasonal housing units in all NSR PSUs to the corresponding state-level estimate of vacant year-round or seasonal housing units from the NSR PSUs in sample.

The appropriate first-stage adjustment factor is applied to every vacant year-round and seasonal housing units in the 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 regional HU adjustment is the final stage in the HVS weighting procedure. The factor is calculated as the ratio of the HU control estimates by the four major geographic regions of the United States (Northeast, South, Midwest, and West) supplied by the Population Division, to the sum of estimated occupied (from the CPS) plus vacant HUs, through the HVS second stage adjustment.[1] This factor is applied to both occupied and vacant housing units.

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. 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.)

American Time Use Survey (ATUS)

Description

The American Time Use Survey (ATUS) collects data each month on how people spend their time. Data collection for the ATUS began in January 2003. There are seventeen major categories for activities such as work, sleep, eating and drinking, leisure and sports. Data are tabulated by variables such as sex, race/ethnicity, age, and education.

Sample

Approximately 2,200 households are selected for ATUS each month, and each household is interviewed just once. The ATUS sample comes from CPS sample households that

[1] The estimate of occupied housing units from the CPS is obtained by aggregating the family weight of each interviewed CPS household.

have completed their eighth and final CPS interview (i.e., CPS MIS 8). There is a two-month lag from the last CPS interview, to initial contact with the household for an ATUS interview. For example, households selected from the January 2005 CPS sample would be part of the March 2005 ATUS sample.

The CPS households from which the ATUS sample is selected are stratified by race/ethnicity and presence of children. Non-White households are sampled at a higher rate to insure that valid comparisons can be made across major race/ethnicity groups. Half of the households are assigned to report on Saturdays and Sundays, and half are assigned to report on weekdays. One person 15 years or older is selected from each ATUS sample household for participation in the ATUS.

Weighting procedure

The basic weight for each ATUS record is the product of the CPS first-stage weight, an adjustment for the CPS state-based design, the within-stratum sampling interval and a household size factor. The ATUS basic weight is then adjusted by an ATUS noninterview factor, plus a set of population control adjustments; the population control adjustments are based on sex, age, race/ethnicity, education, labor force status, and presence of children in the household. A day adjustment factor is computed separately for weekdays, Saturdays, and Sundays. This factor accounts for increased numbers of weekend interviews, and for varying frequencies of each day of the week for a particular month.

Annual Social and Economic Supplement (ASEC)

Description of supplement

The ASEC is sponsored by the Census Bureau and the BLS. The Census Bureau has collected data in the ASEC since 1947. From 1947 to 1955, the ASEC took place in April, and from 1956 to 2001 the ASEC took place in March. Prior to 2003, the ASEC was known as the Annual Demographic Supplement or the March Supplement. In 2001[2], a sample increase was implemented that required more time for data collection. Thus, additional ASEC interviews are now taking place in February and April. 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, education attainment, health insurance coverage, the foreign-born population, work experience, income from all sources, receipt of

_____
[2] The expanded sample was first used in 2001 for testing and was not included in the official ADS statistics for 2001. The statistics from 2002 are the first official set of statistics published using the expanded sample. The 2001 expanded sample statistics were released and are used for comparing the 2001 data to the official 2002 statistics.

noncash benefit, poverty, program participation, and geographic mobility. A major reason for conducting the ASEC in 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 universe for the ASEC is slightly different from that for the basic CPS. It includes certain members of the armed forces in the estimates. This requires some minor changes to the sampling procedures and to the weighting methodology.

The ASEC sample consists of the March CPS sample, plus additional CPS households identified in prior CPS samples and the following April CPS sample. Table 11–2 shows the months when the eligible sample is identified for years 2001 through 2004. Starting in 2004, the eligible ASEC sample households are:

1. The entire March CPS sample.

2. Hispanic households—identified in November (from all month-in-sample (MIS) groups) and in April (from MIS 1 and 5 groups).

3. Non-Hispanic non-White households—identified in August (MIS 8), September (MIS 8), October (MIS 8), November (MIS 1 and 5), and April (MIS 1 and 5).

4. Non-Hispanic White households with children 18 years or younger—identified in August (MIS 8), September (MIS 8), October (MIS 8), November (MIS 1 and 5), and April (MIS 1 and 5).

Prior to 1976, no additional sample households were added. From 1976 to 2001, only the November CPS households containing at least one person of Hispanic origin were added to the ASEC. The households added in 2001, along with a general sample increase in selected states, are collectively known as the State Children's Health Insurance Program (SCHIP) sample expansion. The added households improve the reliability of the ASEC 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 cases from the August, September, October, November and April CPS are completely different from those in the March CPS. The additional sample cases increase the effective sample size of the ASEC compared with the March CPS sample alone. The ASEC sample includes 18 MIS groups for Hispanic households, 15 MIS groups for non-Hispanic non-White households, 15 MIS groups for non-Hispanic White households with children 18 years or younger, and 8 MIS groups for all other households.

Table 11–2. MIS Groups Included in the ASEC Sample for Years 2001, 2002, 2003, and 2004

CPS month/Hispanic status		Month in sample							
		1	2	3	4	5	6	7	8
August	Hispanic[1]	NI[2]							
	Non-Hispanic[3]	NI[2]							2004
September	Hispanic[1]	NI[2]							
	Non-Hispanic[3]	NI[2]							2004
October	Hispanic[1]	NI[2]							
	Non-Hispanic[3]	NI[2]							2003 2004
November	Hispanic[1]				2001 2002 2003 2004				
	Non-Hispanic[3]	2001 2002 2003 2004	NI[2]			2001 2002 2003 2004	2001 2002	2001 2002 2003	2001 2002 2003
March	Hispanic[1]				2001 2002 2003 2004				
	Non-Hispanic[3]								
April	Hispanic[1]	2001 2002 2003 2004	NI[2]			2001 2002 2003 2004	NI[2]		
	Non-Hispanic[1]								

[1] Hispanics may be any race.
[2] NI - Not interviewed for the ASEC.
[3] The non-Hispanic group includes both non-Hispanic non-Whites and non-Hispanic Whites with children 18 years old or younger.

The March and April ASEC eligible cases are administered the ASEC questionnaire in those respective months (see Table 11–3). The April cases are classified as "split path" cases because some households receive the ASEC supplement questionnaire, while other households receive the supplement scheduled for April. The November eligible Hispanic households are administered the ASEC questionnaire in February for MIS groups 1 and 5, during their regular CPS interviewing time, and the remaining MIS groups (MIS 2–4 and 6–8) receive the ASEC interview in March. (November MIS 6–8 households have already completed all 8 months of interviewing for the CPS before March, and the November MIS 2–4 households have an extra contact scheduled for the ASEC before the 5th interview of the CPS later in the year.)

The August, September, October, and November eligible non-Hispanic households are administered the ASEC questionnaire in either February or April. November ASEC eligible cases in MIS 1 and 5 are interviewed for the CPS in February (in MIS 4 and 8, respectively), so the ASEC questionnaire is administered in February. (These are also split path cases, since households in other rotation groups get the regular supplement scheduled for February). The August, September, and October MIS 8 eligible cases are split between the February and April CPS interviewing months.

Mover households are defined at the time of the ASEC interview as households with a different reference person when compared to the previous CPS interview, or the person causing the household to be eligible has moved out (i.e., the Hispanic person or other race minority moved out, or a single child aged the household out of eligibility.) Mover households identified from the August, September, October, and November eligible sample are removed from the ASEC sample. Mover households identified in the March and April eligible samples receive the ASEC questionnaire.

The ASEC sample universe is slightly different from that in the CPS. The CPS completely excludes military personnel while the ASEC includes military personnel who live in households with at least one other civilian adult. These differences require the ASEC 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 ASEC 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. (Chapter 10 contains a

Table 11–3. Summary of 2004 ASEC Interview Months

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
August	Hispanic[1]	NI[2]								NI[2]							
	Non-Hispanic[3]	NI[2]								NI[2]							Feb.
September	Hispanic[1]	NI[2]								NI[2]							
	Non-Hispanic[3]	NI[2]								NI[2]							Feb.
October	Hispanic[1]	NI[2]								NI[2]							
	Non-Hispanic[3]	NI[2]								NI[2]							Apr.
November	Hispanic[1]	NI[2]								Feb	March			Feb	March		
	Non-Hispanic[3]	NI[2]								Feb.	NI[2]			Feb.	NI[2]		
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]																

[1] Hispanics may be any race.
[2] NI - Not interviewed for the ASEC.
[3] The non-Hispanic group includes both non-Hispanic non-Whites and non-Hispanic Whites with children 18 years old or younger.

description of these and the following adjustments.) The ASEC also receives an additional noninterview adjustment for the August, September, October, and November ASEC sample, a SCHIP Adjustment Factor, a family equalization adjustment, and weights applied to Armed Forces members.

The August, September, October, and November eligible samples are weighted individually through the CPS noninterview adjustment and then combined. A noninterview adjustment for the combined samples and the CPS first-stage ratio adjustments are applied before the SCHIP adjustment is applied.

The March eligible sample, and the April eligible sample are also weighted separately before the second-stage weighting adjustment. All 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 ASEC sample.

Households from August, September, October, and November eligible samples: The households from the August, September, October, and November eligible samples start with their basic CPS weight as calculated in the appropriate month, modified by the appropriate CPS special weighting factor and appropriate CPS noninterview adjustment. At this point, a second noninterview adjustment is made for eligible households that are still occupied, but for which an interview could not be obtained in the February, March, or April CPS. Then, the ASEC sample weights for the prior sample are adjusted by the CPS first-stage adjustment ratio and the SCHIP Adjustment Factor.

The ASEC noninterview adjustment for the August, September, October, and November eligible sample. The second noninterview adjustment is applied to the August, September, October, and 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 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 mover households receive no further ASEC weighting. 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 August, September, October, and November eligible sample households interviewed in the February, March, or April CPS in cell j of cluster i.

N_{ij} = the weighted number of August, September, October, and 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 August, September, October, and 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 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 August, September, October, and 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 11–4.) After this adjustment is applied, the August, September, October, and November eligible sample households are ready to be combined with the March and April eligible samples for the application of the second-stage ratio adjustment.

Eligible households from the April sample: The households in 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 size 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 11–4 summarizes these weight adjustments.

Eligible households from the March 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 size 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 11-4 summarizes these weight adjustments.

Combined sample of eligible households from the August, September, October, November, March, and April CPS: At this point, the eligible samples from August, September, October, November, March, and April are combined. The remaining adjustments are applied to this combined sample file.

ASEC second-stage ratio adjustment: The second-stage ratio adjustment adjusts the ASEC estimates so that they agree with independent age, sex, race, and Hispanic-origin population controls as described in Chapter 10. The same procedure used for CPS is used for the ASEC.

Additional ASEC weighting: After the ASEC weight through the second-stage procedure is determined, the next step is to determine the final ASEC weight. There are two more weighting adjustments applied to the ASEC sample cases. The first is applied to the Armed Forces members. The Armed Forces adjustment assigns weights to the eligible Armed Forces members so they are included in the ASEC estimates. The second adjustment is for family equalization. Without this adjustment, there would be more married men than married women. Weights, mostly of males, are adjusted to give a husband and wife the same weight, while maintaining the overall age/race/sex/Hispanic control totals.

Armed Forces. Male and female members of the Armed Forces living off post or living with their families on post are included in the ASEC as long as at least one civilian adult lives in the same household, whereas the

Figure 11-1. Diagram of the ASEC Weighting Scheme

```
August¹          September¹       October²         November³                    March⁴                    April⁵
Sample           Sample           Sample           Sample                       Sample                    Sample
  │                │                │                │                             │                         │
  ▼                ▼                ▼                ▼                             ▼                         ▼
Base             Base             Base             Base                         Base                      Base
Weight           Weight           Weight           Weight                       Weight                    Weight
  │                │                │                │                             │                         │
  ▼                ▼                ▼                ▼                             ▼                         ▼
Special          Special          Special          Special                      Special                   Special
Weighting        Weighting        Weighting        Weighting                    Weighting                 Weighting
Factor           Factor           Factor           Factor                       Factor                    Factor
  │                │                │                │                             │                         │
  ▼                ▼                ▼                ▼                             ▼                         ▼
August NI        September NI     October NI       November NI                  March NI                  April NI
Adjustment       Adjustment       Adjustment       Adjustment                   Adjustment                Adjustment
```

ASEC NI Adjustment

Mover Households Removed ◄── ASEC NI Adjustment

First-Stage Ratio Adjustment

SCHIP Adjustment

First-Stage Ratio Adjustment (March)

Non-Mover Hispanic, Non-Hispanic non-White, Non-Hispanic Whites with Children Households

Mover Hispanic, Non-Hispanic non-White, Non-Hispanic Whites with Children Households and All Others

SCHIP Adjustment

First-Stage Ratio Adjustment (April)

SCHIP Adjustment

Second-Stage (SS) Ratio Adjustment

Armed Forces Weighting

Family Equalization

ASEC Weight

Eligible Cases for:
[1] From MIS 8 rotations that are Non-Hispanic Non-Whites and Non-Hispanic Whites with children; interviewed in February.
[2] From MIS 8 rotations that are Non-Hispanic Non-Whites and Non-Hispanic Whites with children; interviewed in April.
[3] Hispanics from MIS 2-4 and 6-8 groups; interviewed in March. From MIS 1 and 5, Hispanic, Non-Hispanic Non-Whites, and Non-Hispanic Whites with children; interviewed in February as part of their MIS 4 and 8 interviews.
[4] All cases interviewed from March.
[5] From MIS 1 and 5, all Hispanics, Non-Hispanic Non-Whites, and Non-Hispanic Whites with children; all interviewed in April.

Table 11–4. **Summary of ASEC SCHIP Adjustment Factor for 2004**

CPS month/Hispanic status		Mover — Month in sample								Nonmover — Month in sample							
		1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
August	Hispanic[1]	0[2]								0[2]							
	Non-Hispanic[3]	0[2]								0[2]							8/15
September	Hispanic[1]	0[2]								0[2]							
	Non-Hispanic[3]	0[2]								0[2]							8/15
October	Hispanic[1]	0[2]								0[2]							
	Non-Hispanic[3]	0[2]								0[2]							8/15
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				8/15				8/15				8/15			

[1] Hispanics may be any race.
[2] Zero weight indicates the cases are ineligible for the ASEC.
[3] The non-Hispanic group includes both non-Hispanic non-Whites and non-Hispanic Whites with children 18 years old or younger.

CPS excludes all Armed Forces members. Households with no civilian adults in the household, i.e., households with only Armed Forces members, are excluded from the ASEC. The weights assigned to the Armed Forces members included in the ASEC are the same weights civilians receive through the SCHIP adjustment. Control totals, used in the second-stage factor, do not include Armed Forces members, so Armed Forces members do not go through the second-stage ratio adjustment. During family equalization, the weight of a male Armed Forces member with a spouse or partner is reassigned to the weight of his spouse/partner.

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 of households
6. Male partners in male/male unmarried partner households
7. All other civilian males

Three different methods, depending on the household composition, are used to assign the ASEC 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 ASEC weight.

SUMMARY

Although this discussion focuses on only three CPS supplements, the HVS, the ATUS, and the ASEC Supplement, every supplement has its own unique objectives. The particular 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, ATUS, and ASEC indicate only 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 U.S. Bureau of Labor Statistics and the U.S. Census Bureau through broad publication programs that 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, more CPS-based products will be electronically accessible in the future.

BUREAU OF LABOR STATISTICS PRODUCTS

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 also is available electronically on the Internet and can be accessed at <http://www.bls.gov/news.release/empsit.toc.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 Asians and Hispanics,[1] are published every quarter. Data also are 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 10,000 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://www.bls.gov/data/>. 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 also is used to collect detailed information on particular segments or 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 extent of work experience of the population during the calendar year; the marital and family characteristics of workers; and the employment of school-age youth, high school graduates and dropouts, and recent college graduates. Surveys are also made periodically on subjects such as contingent workers, job tenure, displaced workers, disabled veterans, and volunteers. 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 and can be used to create additional cross-sectional detail. The actual identities of the individuals are protected on all versions of the files made available to noncensus staff. 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.

Annual averages from the CPS for the four census regions and nine census 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. The publication is available electronically on the Internet and can be accessed at <http://www.bls.gov/opub/gp/laugp.htm>.

Table 12–1 provides a summary of the CPS data products available from BLS.

CENSUS BUREAU PRODUCTS

The 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,

---

[1] Hispanics may be any race.

demographic, and economic topics. In addition, special reports on many subjects also have 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 news 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 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, locations inside and outside metropolitan areas, the 50 states and the District of Columbia, and the 75 largest metropolitan areas. Information is also made available on national home ownership rates by age of householder, family type, race, and Hispanic ethnicity. A quarterly news release and quarterly and annual data tables are released on the Internet.

Supplemental 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 non-census 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 <http://www.bls.census.gov/cps/>. The Census Bureau plans 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 and Unemployment Among Youth in the Summer	An analysis of the labor force, employment, and unemployment characteristics of 16- to 24-year-olds between April and July with the focus on July	Annual	Monthly CPS	Free ([1])
Employment Characteristics of Families	An analysis of employment and unemployment by family relationship and the presence and age of children	Annual	Monthly CPS	Free ([1])
Employment Situation of Veterans	An analysis of the work activity and disability status of persons who served in the Armed Forces	Biennial	August CPS supplement	Free ([1])
Job Tenure of American Workers	An analysis of employee tenure by industry and a variety of demographic characteristics	Biennial	January CPS supplement	Free ([1])
Labor Force Characteristics of Foreign-Born Workers	An analysis of the labor force characteristics of foreign-born workers and a comparison with the labor force characteristics of their native-born counterparts	Annual	Monthly CPS	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])
Volunteering in the United States	An analysis of the incidence of volunteering and the characteristics of volunteers in the United States	Annual	September CPS supplement	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	Annual Social and Economic Supplement to the CPS (February–April)	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	$53.00 domestic; $74.20 foreign; per year
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	$49.00 domestic; $68.60 foreign; per year
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	Annual Social and Economic Supplement to the CPS (February–April)	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	$25.00 domestic; $35.00 foreign

Table 12–1. **Bureau of Labor Statistics Data Products From the Current Population Survey**—Con.

Product	Description	Periodicity	Source	Cost
Highlights of Women's Earnings	An analysis of the hourly and weekly earnings of women by a variety of characteristics	Annual	Monthly CPS; outgoing rotation groups	Free
Issues in Labor Statistics	Brief analysis of important and timely labor market issues	Occasional	CPS; other surveys and programs	Free
Microdata Files				
Annual Demographic Survey		Annual	Annual Social and Economic Supplement to the CPS (February–April)	(4)
Contingent Work		Biennial	January CPS supplement	(4)
Displaced Workers		Biennial	February CPS supplement	(4)
Job Tenure and Occupational Mobility		Biennial	January CPS supplement	(4)
School Enrollment		Annual	October CPS supplement	(4)
Usual Weekly Earnings (outgoing rotation groups)		Annual	Monthly CPS; outgoing rotation groups	(4)
Veterans		Biennial	August CPS supplement	(4)
Work Schedules/ Home-Based Work		Occasional	May CPS supplement	(4)
Time Series (Macro) Files				
National Labor Force Data		Monthly	Labstat(5)	(1)
Unpublished Tabulations				
National Labor Force Data		Monthly	Electronic files	Free

[1] Accessible from the Internet <http://www.bls.gov/bls/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://www.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 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 for error from both sources has been examined relative to the particular use at hand.

This chapter provides an overview of how these two sources of potential error can affect data quality, discusses their relationship to each other from a conceptual viewpoint, and defines 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 the 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 would result from taking all possible samples that could be selected using the specified 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 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. The mean squared error is 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 that 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 also 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 from 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 two sets of components, one resulting from the fact that estimates are based on a sample of units rather than the entire population (sampling error) and the other 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, and can include issues such as individual behavior.) 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.

It is frequently possible to construct an unbiased estimator of variance. In the case of complex surveys like the CPS, estimators have been developed that 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 the CPS, 160 subsamples or replicates are used in variance estimation for the 2000 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 statistically 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 people in the sample unit the questions, coding responses into predefined categories, and creating estimates that 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 deviating during implementation from the predefined specifications.

For example, we now recognize that the initial labor force question used in the 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 that 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 to help managers and staff working on a survey understand the process is quality. They can also aid users of the various products of the survey process (both individual responses and their aggregations into statistics) 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 the CPS, is a function of 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. In the case of the 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, assuming that 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 concentrating on doing the best job possible. Process monitoring through quality indicators, such as coverage and response rates, can determine when additional training or revisions in process specification are needed. Continuing process improvement is a vital component for achieving the survey's quality goals.

REFERENCES

Gonzalez, M. E., J. L. Ogus, G. Shapiro, and B. J. Tepping (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., W. N. Hurwitz, and M. A. Bershad (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., B. Swensson, and J. Wretman (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. Analyze the survey design by evaluating the effect of each of the stages of sampling and estimation on the overall precision of the survey estimates.

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. Chapter 13 provides additional information on these topics. Certain aspects of the CPS sample design, such as the use of one sample PSU per non-self-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, and captures the effects of sample selection and estimation on the total variance. Variance estimates of selected characteristics and tables, which 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 out (replicated) several times. The dispersion of the resulting estimates can be used to measure the variance of the full sample.

Method

One would not likely repeat the entire CPS 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 those used for the full sample, and to apply the regular CPS estimation procedures to these subsamples, which are called replicates. Determining the number of replicates to use involves balancing the cost and the reliability of the variance estimator; because 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 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 the calculation of variance estimates for many characteristics with replication of all stages of the weighting through compositing. 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 non-self-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 the 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 the matrix determined differing weights for the half samples. 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

$$\text{Var}(\hat{Y}_0) = \frac{4}{48} \sum_{r=1}^{48} (\hat{Y}_r - \hat{Y}_0)^2.$$

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 below) 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 AND 2000 DESIGNS

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 and 2000 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

---

[1] Usually balanced half-sample replication uses replicate factors of 2 and 0 with the formula,

$$\text{Var}(\hat{Y}_0) = \frac{1}{k} \sum_{r=1}^{k} (\hat{Y}_r - \hat{Y}_0)^2$$

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.

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 the 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 samples. Between-PSU variance cannot be estimated directly using this methodology; it is the difference between the estimates of total variance and within-PSU variance. NSR strata are combined into pseudo- strata 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, for the 1990 design, 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 assuming roughly equal sizes of the original strata. However, for the 2000 design, these factors were further adjusted to account for the unequal sizes of the original strata within the pseudostratum. All USUs in a pseudostratum are assigned the same row number(s).

For an SR sample, two rows of the Hadamard matrix are assigned to each pair of USUs creating replicate factors, f_r for r = 1,...,160

---

[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.

$$f_{ir} = 1 + (2)^{\frac{-3}{2}} a_{i+1,r} - (2)^{\frac{-3}{2}} a_{i+2,r}$$

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.

As in the 1980 methodology, the unbiased weights (base-weight 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, national and state coverage adjustments, the second-stage ratio adjustments, 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 ($\hat{Y}_0$). The formula is

$$\text{Var}(\hat{Y}_0) = \frac{4}{160} \sum_{r=1}^{160} (\hat{Y}_r - \hat{Y}_0)^2.$$

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 remain 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 an SR sample. For an 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. For more detailed information regarding the formation of replicates, see the internal Census Bureau memoranda (Gunlicks, 1996, and Adeshiyan, 2005).

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. However, estimates for states and areas that have substantial contributions from NSR sample areas have variance estimation problems that are more difficult to resolve.

Complicating Factors for State Variances

Most states contain a small number of NSR sample PSUs, so variances at the state level are based on fairly small sample sizes. Pairing these PSUs into pseudostrata further reduces the number of NSR SECUs 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.

Some research has been done to produce improved state and local variances obtained directly from successive difference replication and the modified balanced half-sample methods. Griffiths and Mansur (2000a, 2000b, 2001a and 2001b) and Mansur and Griffiths (2001) examine methods based on times series and generalized linear modeling techniques to address the small sample size problem and the bias induced by collapsing NSR strata to estimate between-PSU variances.

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 provides more stable variance estimates by taking advantage of these similarities.

Generalization Method

The GVF that is used to estimate the variance of an estimated population total, $\hat{X}$, is of the form

$$Var(\hat{X}) = a\hat{X}^2 + b\hat{X} \qquad (14.1)$$

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

$$Var(\hat{X}) = -(deff)\frac{N}{n}\left(\frac{\hat{X}^2}{N}\right) + (deff)\frac{N}{n}\hat{X}.$$

Letting

$$a = -\frac{b}{N}$$

and

$$b = \frac{(deff)N}{n}$$

gives the functional form

$$Var(\hat{X}) = a\hat{X}^2 + b\hat{X}.$$

We choose

$$a = -\frac{b}{N}$$

where N is a control total so that the variance will be zero when $\hat{X} = N$.

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 individuals within housing units. For example, estimates of total people classified by a characteristic of the housing unit or of the household, such as the total urban population, number of recent migrants, or people of Hispanic[4] origin, would tend to have fairly large design effects. The reason is that these characteristics usually appear among all people 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.

For many subpopulations of interest, N is a control total used in the second-stage ratio adjustment. In these subpopulations, as $\hat{X}$ approaches N, the variance of $\hat{X}$ approaches zero, since the second-stage ratio adjustment guarantees that these sample population estimates match independent population controls (Chapter 10).[5] 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

$$V_x^2 = a + \frac{b}{X},$$

where the relative variance (V_x^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 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 is 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. Using a model to estimate the relative variance of an estimate in this way introduces 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 estimates 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.

Table 14–1. **Parameters for Computation of Standard Errors for Estimates of Monthly Levels**

Characteristic	a	b
Unemployed:		
Total or White	−0.000016350	3095.55
Black	0.000151396	3454.72
Hispanic origin[1]	−0.000141225	3454.72

[1] Hispanics may be any race.

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 9 million unemployed men in the civilian labor force

($\hat{X}$ = 9,000,000). From Table 14–1

a = -0.000016350 and b = 3095.55 , so

$$s_x = \sqrt{(-0.000016350)(9,000,000)^2 + (3095.55)(9,000,000)} \approx 163,000$$

An approximate 90-percent confidence interval for the monthly estimate of unemployed men is between 8,739,200 and 9,260,800 [or 9,000,000 ± 1.6(163,000)].

VARIANCE ESTIMATES TO DETERMINE OPTIMUM SURVEY DESIGN

The following tables show variance estimates computed using replication methods by type (total and within-PSU), and by stage of estimation. The estimates presented are based on the 1990 sample design and new weighting procedures introduced in January 2003. Updated estimates based on the 2000 design will be provided when they are available.

Averages over 13 months have been used to improve the reliability of the estimated monthly variances. The 13-month period, March 2003–March 2004, was used for estimation because the sample design was essentially unchanged throughout this period (there was a maintenance reduction for CPS in April 2003). Data from January and February 2003 were not used in order to allow new weighting procedures to be fully reflected in the composite estimation step.

Variance Components Due to Stages of Sampling

Table 14–2 indicates, for the estimate after the second-stage (SS) adjustment, how the several stages of sampling affect the total variance of each of the given characteristics. The SS estimate is the estimate after the first-stage, national coverage, state coverage 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 estimate from the total variance estimate. 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 population and White-alone population employed in agriculture, the within-PSU component still accounts for 70 to 80 percent of the total variance, while the between-PSU component accounts for the remaining 20 to 30 percent.

Table 14–2. **Components of Variance for SS Monthly Estimates**

[Monthly averages: March 2003–March 2004]

Civilian noninstitutionalized population 16 years old and over	SS[1] estimate (x 10^6)	Standard error (x 10^5)	Coefficient of variation (percent)	Percent of total variance	
				Within	Between
Unemployed, total..........	**8.81**	**1.70**	**1.93**	**98.7**	**1.3**
White alone	6.34	1.43	2.25	99.0	1.0
Black alone...................	1.80	0.76	4.20	100.6	−0.6
Hispanic origin[2]	1.46	0.75	5.16	101.0	−1.0
Teenage, 16–19	1.24	0.56	4.49	99.5	0.5
Employed—agriculture, total.......................	**2.26**	**1.12**	**4.98**	**75.4**	**24.6**
White alone	2.12	1.10	5.18	74.7	25.3
Black alone...................	0.06	0.15	23.04	94.8	5.2
Hispanic origin[2]	0.45	0.59	13.06	86.2	13.8
Teenage, 16–19	0.12	0.19	16.67	88.3	11.7
Employed—nonagriculture, total.......................	**135.76**	**3.84**	**0.28**	**97.4**	**2.6**
White alone	112.34	3.29	0.29	102.2	−2.2
Black alone...................	14.66	1.41	0.96	94.9	5.1
Hispanic origin[2]	17.01	1.53	0.90	92.7	7.3
Teenage, 16–19	5.84	1.04	1.77	97.6	2.4
Civilian labor force, total ...	**146.83**	**3.56**	**0.24**	**92.3**	**7.7**
White alone	120.80	3.07	0.25	95.9	4.1
Black alone...................	16.52	1.31	0.79	93.7	6.3
Hispanic origin[2]	18.92	1.32	0.70	91.3	8.7
Teenage, 16–19	7.19	1.09	1.52	93.7	6.3
Not-in-labor force, total	**74.79**	**3.56**	**0.48**	**92.3**	**7.7**
White alone	60.77	3.07	0.50	95.9	4.1
Black alone...................	9.24	1.31	1.42	93.7	6.3
Hispanic origin[2]	8.74	1.32	1.51	91.3	8.7
Teenage, 16–19	8.93	1.09	1.22	93.7	6.3

[1] Estimates after the second-stage ratio adjustments (SS).
[2] Hispanics may be any race.

Because the SS estimate of the total civilian labor force and total not-in-labor-force populations 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.

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 weighting control factors applied. The noninterview estimate includes the baseweights, the weighting control adjustment, and the noninterview adjustment. The SS estimate includes baseweights, the weighting control adjustment, the noninterview adjustment, the first-stage, national and state coverage adjustments, and second-stage ratio adjustments.

In Table 14–3, the figures for unemployed show, for example, that the relative variance of the SS estimate of level is 3.741×10^{-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 1.11 times as large. If the noninterview stage of estimation is also included, the relative variance is 1.10 times the size of the relative variance for the SS estimate of level. Including the first stage of estimation maintains the relative variance factor at 1.11. The relative variance for total unemployed after applying the second-stage adjustment without the first-stage adjustment 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. However, as illustrated in Rigby (2000), removing the first-stage adjustment increased the variances of some state-level estimates.

The second-stage adjustment, however, appears to greatly reduce the total variance, as intended. This is especially true for characteristics that belong to high proportions of age, sex, or race/ethnicity subclasses, such as White alone, Black alone, or Hispanic people in the civilian labor force or employed in nonagricultural industries. Without the second-stage adjustment, the relative variances of these characteristics would be 5–10 times as large. For

Table 14–3. **Effects of Weighting Stages on Monthly Relative Variance Factors**

[Monthly averages: March 2003–March 2004]

Civilian noninstitutionalized population 16 years old and over	Relative variance of SS estimate of level (x 10^{-4})	Relative variance factor[1]			
		Unbiased estimator	Unbiased estimator with—		
			NI[2]	NI & FS[3]	NI & SS[4]
Unemployed, total	**3.741**	**1.11**	**1.10**	**1.10**	**1.00**
White alone	5.065	1.12	1.11	1.11	1.00
Black alone	17.654	1.27	1.26	1.25	1.00
Hispanic origin[5]	26.600	1.28	1.27	1.27	1.00
Teenage, 16–19	20.186	1.11	1.11	1.11	1.00
Employed—agriculture, total	**24.847**	**0.99**	**0.99**	**1.00**	**0.99**
White alone	26.829	0.99	0.99	0.99	1.00
Black alone	530.935	1.07	1.07	1.04	1.00
Hispanic origin[5]	170.521	1.08	1.09	1.10	1.00
Teenage, 16–19	277.985	1.03	1.05	1.05	1.00
Employed—nonagriculture, total	**0.080**	**6.53**	**6.34**	**6.14**	**1.00**
White alone	0.086	8.16	8.00	7.70	1.00
Black alone	0.920	6.67	6.57	5.89	1.00
Hispanic origin[5]	0.813	6.48	6.43	6.42	1.00
Teenage, 16–19	3.146	1.79	1.78	1.77	1.00
Civilian labor force, total	**0.059**	**8.28**	**8.01**	**7.75**	**1.00**
White alone	0.064	10.38	10.15	9.76	1.00
Black alone	0.631	9.02	8.87	7.86	1.00
Hispanic origin[5]	0.487	10.45	10.35	10.34	1.00
Teenage, 16–19	2.310	2.04	2.03	2.01	1.00
Not-in-labor force, total	**0.226**	**2.67**	**2.55**	**2.50**	**1.00**
White alone	0.254	3.01	2.91	2.80	1.00
Black alone	2.015	3.84	3.73	3.10	1.00
Hispanic origin[5]	2.282	3.58	3.54	3.54	1.00
Teenage, 16–19	1.499	2.64	2.60	2.60	1.00

[1] Relative variance factor is the ratio of the relative variance of the specified level to the relative variance of the SS level.
[2] NI = Noninterview.
[3] FS = First-stage.
[4] SS = Estimate after Second-stage when the First-stage adjustment is skipped.
[5] Hispanics may be any race.

smaller groups, such as the unemployed and those employed in agriculture, the effect of second-stage adjustment is not as dramatic.

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 SS estimates. For example, the estimated variance of the composited estimate of unemployed Hispanics is 5.445×10^9. The variance factor for this characteristic is 0.96, implying that the variance of the composited estimate is 96 percent of the variance of the estimate after the second-stage adjustments. The relative variance of the composite estimate of unemployed Hispanics, which takes into account the estimate of the number of people with this characteristic, is about the same as the size of the relative variance of the SS estimate (hence, the relative variance factor of 0.99). 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 and within-PSU variances 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 (SRS) design. The design effects in this table were computed by solving the 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 13 months of data.

For the unemployed, the design effect for total variance is 1.377 for the uncomposited (SS) estimate and 1.328 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 total variance by about 32.8 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 total variance by about 20.6 percentage points. The design effects for composited estimates are generally lower than those for the SS estimates, indicating again the tendency

Table 14–4. Effect of Compositing on Monthly Variance and Relative Variance Factors

[Monthly averages: March 2003–March 2004]

Civilian noninstitutionalized population, 16 years old and over	Variance of composited estimate of level (x 10^9)	Variance factor[1]	Relative variance factor[2]
Unemployed, total	**27.729**	**0.96**	**0.97**
White alone	19.462	0.96	0.97
Black alone	5.163	0.90	0.93
Hispanic origin[3]	5.445	0.96	0.99
Teenage, 16–19	3.067	0.98	1.01
Employed—agriculture, Total ..	**12.446**	**0.98**	**0.99**
White alone	11.876	0.98	1.00
Black alone	0.213	1.01	1.00
Hispanic origin[3]	3.425	0.99	1.00
Teenage, 16–19	0.354	0.95	0.99
Employed—nonagriculture, total	**115.860**	**0.79**	**0.79**
White alone	87.363	0.81	0.81
Black alone	15.624	0.79	0.79
Hispanic origin[3]	20.019	0.85	0.86
Teenage, 16–19	9.604	0.90	0.92
Civilian labor force, total	**98.354**	**0.78**	**0.78**
White alone	74.009	0.79	0.79
Black alone	14.172	0.82	0.82
Hispanic origin[3]	14.277	0.82	0.82
Teenage, 16–19	10.830	0.91	0.93
Not-in-labor force, total	**98.354**	**0.78**	**0.77**
White alone	74.009	0.79	0.78
Black alone	14.172	0.82	0.83
Hispanic origin[3]	14.277	0.82	0.80
Teenage, 16–19	10.830	0.91	0.88

[1] Variance factor is the ratio of the variance of a composited estimate to the variance of an SS estimate.
[2] Relative variance factor is the ratio of the relative variance of a composited estimate to the relative variance of an SS estimate.
[3] Hispanics may be any race.

Table 14–5. Design Effects for Total and Within-PSU Monthly Variances

[Monthly averages: March 2003–March 2004]

Civilian noninstitutionalized population, 16 years old and over	Design effects for total variance		Design effects for within-PSU variance	
	After second stage	After compositing	After second stage	After compositing
Unemployed, total	**1.377**	**1.328**	**1.359**	**1.259**
White alone	1.325	1.277	1.312	1.220
Black alone	1.283	1.180	1.291	1.195
Hispanic origin[1]	1.567	1.530	1.583	1.533
Teenage, 16–19	1.012	1.008	1.007	1.000
Employed—agriculture, Total....	**2.271**	**2.248**	**1.712**	**1.703**
White alone	2.304	2.281	1.722	1.715
Black alone	1.344	1.351	1.274	1.279
Hispanic origin[1]	3.089	3.071	2.662	2.658
Teenage, 16–19	1.292	1.255	1.141	1.112
Employed—nonagriculture, total	**1.124**	**0.883**	**1.094**	**0.804**
White alone	0.782	0.632	0.799	0.590
Black alone	0.579	0.456	0.550	0.403
Hispanic origin[1]	0.601	0.513	0.557	0.478
Teenage, 16–19	0.756	0.688	0.738	0.655
Civilian labor force, total	**1.024**	**0.794**	**0.946**	**0.693**
White alone	0.685	0.540	0.657	0.471
Black alone	0.452	0.371	0.423	0.323
Hispanic origin[1]	0.404	0.332	0.369	0.319
Teenage, 16–19	0.689	0.633	0.645	0.589
Not-in-labor force, total	**1.025**	**0.795**	**0.946**	**0.693**
White alone	0.854	0.671	0.819	0.586
Black alone	0.780	0.643	0.730	0.559
Hispanic origin[1]	0.833	0.676	0.761	0.650
Teenage, 16–19	0.560	0.501	0.524	0.466

[1] Hispanics may be any race.

of the compositing to reduce the variance of most esti-mates. Since the same denominator (SRS variance) is used in computing deff for both the total and within-PSU vari-ances, deff is directly proportional to the respective vari-ances. As a result, the total variance design effects tend to be higher than the within-PSU variances. This is consistent with results from Table 14–2 where total variance esti-mates are expected to be higher than within-PSU variance estimates.

REFERENCES

Adeshiyan, S. (2005), "2000 Replicate Variance System (VAR2000–1)," Internal Memorandum for Documentation, Demographic Statistical Methods Division, U.S. Census Bureau.

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 Sec-tion 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 Statis-tical Association, pp. 212–217.

Fay, R., C. Dippo, and D. Morganstein, (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 G. Train, (1995), "Aspects of Survey and Model-Based Postcensal Estimation of Income and Poverty Char-acteristics for States and Counties," **Proceedings of the Section on Government Statistics,** American Statistical Association, pp. 154–159.

Griffiths, R. and K. Mansur, (2000a), "Preliminary Analysis of State Variance Data: Sampling Error Autocorrelations (VAR90–36)," Internal Memorandum for Documentation, May 31st, Demographic Statistical Methods Division, U.S. Census Bureau.

Griffiths, R. and K. Mansur, (2000b), "Preliminary Analysis of State Variance Data: Autoregressive Integrated Moving Average Model Fitting and Grouping States (VAR90–37)," Internal Memorandum for Documentation, June 15th, Demographic Statistical Methods Division, U.S. Census Bureau.

Griffiths, R. and K. Mansur, (2001a), "Current Population Survey State-Level Variance Estimation," paper prepared for presentation at the Federal Committee on Statistical Methodology Conference, Arlington, VA, November 14–16, 2001.

Griffiths, R. and K. Mansur, (2001b), "The Current Popula-tion Survey State Variance Estimation Story (VAR90–38)," Internal Memorandum for Documentation, October 29th, Demographic Statistical Methods Division, U.S. Census Bureau.

Gunlicks, C. (1993), "Overview of 1990 CPS PSU Stratifica-tion (S–S90–DC–11)," Internal Memorandum for Documen-tation, February 24th, 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,** Wash-ington, D.C.: 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 Subcom-mittee, June 17th, Demographic Statistical Methods Divi-sion, 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 Sec-tion on Survey Research Methods,** American Statistical Association, pp. 11–20.

Mansur, K. and R. Griffiths, (2001), "Analysis of the Cur-rent Population Survey State Variance Estimates," paper presented at the 2001 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Meth-ods.

Rigby, B. (2000), "The Effect of the First Stage Ratio Adjust-ment in the Current Population Survey," paper presented at the 2000 Joint Statistical Meetings, American Statistical Association, **Proceedings of the Section on Survey Research Methods**.

Train, G., L. Cahoon, and P. Makens, (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, **Employ-ment and Earnings, 42,** p.152, Washington, D.C.: Gov-ernment 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

For a given estimator, the difference between the estimate that would result if the sample were to include the entire population and the true population value being estimated is known as nonsampling error. 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 to 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).

Sources of nonsampling error can include the following:

1. Inability to obtain information about all sample cases (unit nonreponse).

2. Definitional difficulties.

3. Differences in the interpretation of questions.

4. Respondent inability or unwillingness to provide correct information.

5. Respondent inability to recall information.

6. Errors made in data collection, such as recording and coding data.

7. Errors made in processing the data.

8. Errors made in estimating values for missing data.

9. Failure to represent all units with the sample (i.e., undercoverage).

It is clear that there are two main types of nonsampling error in the CPS. The first type is error imported from other frames or sources of information, such as decennial census omissions, errors from the Master Address File or its extracts, and errors in other sources of information used to keep the sample current, such as building permits. All nonsampling errors that are not of the first type are 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, proxy response, 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 focuses on the sources and operational efforts used to control the occurrence of error in the survey processes. Each section discusses a procedure aimed at reducing 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 broad-reaching effects on the total survey.

This chapter deals with many sources of and controls on nonsampling error, but it is not exhaustive. It includes 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 the full effects of nonsampling error on the survey estimates are unknown, research in this area is being conducted (e.g., latent class analysis; Tran and Mansur, 2004). Ultimately, the CPS attempts to prevent such errors from entering the survey process and tries to keep these that occur 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 (e.g., 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. On the other hand, 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 an undercount of population (resulting from undercoverage).

The primary sources of CPS coverage error are:

1. **Frame omission.** Frame omission occurs when the list of addresses used to select the sample is incomplete. This can occur in any of the four sampling frames (i.e., unit, permit, area, and group quarters; see Chapter 3). Since these erroneously omitted units cannot be sampled, undercoverage of the target population results. Reasons for frame omissions are:

- Master Address File (MAF): The MAF is a list of every living quarters nationwide and their geographic locations. The MAF is updated throughout the decade to provide addresses for delivery of Census 2000 questionnaires, to serve as the sampling frame for the Census Bureau's demographic surveys, and to support other Census Bureau statistical programs. The MAF may be incomplete or contain some units that are not locatable. This occurs when the decennial census lister fails to canvass an area thoroughly or misses units in multiunit structures.

- New construction: This is a sampling problem. Some areas of the country are not covered by building permit offices and are not surveyed (unit frame) to pick up new housing. New housing units in these areas have zero probability of selection. (Based on initial results, the level of errors from this source of frame omission is expected to be quite small in the 2000 CPS Sample Redesign.)

- Mobile homes: This is also a sampling problem. Mobile homes that move into areas that are not surveyed (unit frame) also have a zero probability of selection.

- Group quarters: The information used to identify group quarters may be incomplete, or new group quarters may not be identified.

2. **Erroneous frame inclusion.** Erroneous frame inclusion of housing units occurs when any of the four frames or the MAF (extracts) contains units that do not exist on the ground; for example, a housing unit is demolished, but still exists in the frame. 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, overcoverage of the target population results if the errors are not detected and corrected.

3. **Misclassification errors.** Misclassification errors occur when out-of-scope units are misclassified as in-scope or vice versa. For example, a commercial unit is misclassified as a residential unit, an institutional group quarter is misclassified as a noninstitutional group quarter, or an occupied housing unit with no one at home is misclassified as vacant. Errors can also occur when units outside an appropriate area boundary are misclassified as inside the boundary or vice versa. These errors can cause undercoverage or overcoverage of the target population.

4. **Within-housing unit omissions.** Undercoverage of individuals can arise from failure to list all usual residents of a housing unit on the household roster or from misclassifying a household member as a nonmember.

5. **Within-housing unit inclusions.** Overcoverage can occur because of the erroneous inclusion of people on the roster for a household, for instance, when people with a usual residence elsewhere are incorrectly treated as members of the sample housing unit.

Other sources of coverage error are omission of homeless people from the frame, unlocatable addresses from building permit lists, and missed housing units due to the start dates for the sampling of permits.[1] For example, housing units whose permits were issued before the start month and not built by the time of the census may be missed. Newly constructed group quarters are another possible source of undercoverage. If they are noticed on the permit lists, the general rule is purposely to remove these group quarters; if they are not noticed and the field representative (FR) discovers a newly constructed group quarters, the case is stricken from the roster of sample addresses and never visited again for the CPS until possibly the next sample redesign.

Through various studies, it has been determined that the number of housing units missed is small 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 that attempts to quantify the overall coverage of the survey, despite its coverage errors. For example, the coverage ratio for the total population that is at least 16 years old has been approximately 88 percent since the CPS started phasing in the new 2000-based sample in April 2004. See Chapter 16 for further discussion of coverage error.

CONTROLLING COVERAGE ERROR

This section focuses on those processes during the sample selection and sample preparation stages that control housing unit omissions. Other sources of undercoverage can be viewed as response error and are described in a later section of this chapter.

Sample Selection

The CPS sample selection is coordinated with and does not duplicate the samples of at least five other demographic surveys. These surveys basically undergo the same verification processes, so the discussion in the rest of this section is not unique to the CPS.

Sampling verification consists of testing and production. The intent of the testing phase (unit testing and system testing) is to design the processes better in order to

---

[1] For the 2000 Sample Redesign, sampling of permits began with those issued in 1999; the month varied depending on the size of the structure and region. Housing units whose permits were issued before the start month in 1999 and not built by the time of the census may be missed.

improve quality, and the intent of the production phase is to verify or inspect the quality of the final product in order to improve the processes. Both phases are coordinated across the various surveys. Sampling intervals, measures of expected sample sizes, and the premise that a housing unit is selected for only one survey for 10 years are all part of the sample selection process. Verification in each of the testing and production phases is done independently by two verifiers.

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 are used to test the performance of the system in extreme situations.

The production phase of sampling verification consists of verification of output using the actual sample, focusing on whether the system ran successfully. The following are illustrations and do not represent any particular priority or chronology:

1. PSU probabilities of selection are verified; for example, their sum within a stratum should be one.

2. Edits are done on file contents, checking for blank fields and out-of-range data.

3. When applicable, the files are coded to check for logic and consistency.

4. Information is centralized; for example, the sampling rates for all states and substate areas for 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[2]

Sample preparation activities, in contrast to those of initial sample selection, are ongoing. The listing review[3] and check-in of sample are monthly production processes, and the listing check is an ongoing field process.

Listing review. The listing 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. Automation of the review is a major advancement in improving the timing of the process and, thus, a more accurate CPS frame.

Various listing reviews are performed by both the regional offices (ROs) and the National Processing Center (NPC) in Jeffersonville, IN. Performing these reviews in different parts of the Census Bureau is, in itself, a means of controlling nonsampling error.

After an FR makes an initial visit either to a multiunit address from the unit frame or to a sample address from the permit frame, the RO staff reviews the Multiunit Listing Aid or the Permit Listing Sheet, respectively. (Refer to Appendix A for a description of these listing forms.) This review occurs the first time the address is in the sample for any survey. However, if major changes appear at an address on a subsequent visit, the revised listing is reviewed. If there is evidence that the FR encountered a special or unusual situation, the materials are compared to the instructions in the manual *"Listing and Coverage: A Survival Guide for the Field Representative"* (see Chapter 4) to ensure that the situation was handled correctly. Depending on whether it is a permit-frame sample address or a multiunit address from the unit frame, the following are verified:

1. Were correct entries made on the listing when either fewer or more units were listed than expected?

2. Did the FR relist the address if major structural changes were found?

3. Were no units listed (for the permit frame)?

4. Was an extra unit discovered (for the permit frame)?

5. Were additional units interviewed if listed on a line with the current CPS sample designation?

6. Did the unit designation change?

7. Was a sample unit demolished or condemned?

Errors and omissions are corrected and brought to the attention of the FR. This may involve contacting the FR for more information.

The Automated Listing and Mapping Instrument (ALMI) features an edit that prevents duplication of units in area frame blocks[4]. The RO review of area-frame block updates is performed for new FRs or those with limited listing experience. This review focuses on deletions, moved units, added units, and house number changes. There is

---

[2] See Chapter 4, *Preparation of the Sample*, specifically its section *Listing Activities*, for more details. Also see Appendix A, *Sample Preparation Materials*.
[3] Because of the automated instrument, it is appropriate to use the generic term "listing review" for the area, unit, and group quarters frames. "Listing sheet" and "listing sheet review" are appropriate for the permit frame. Regardless, these sections will use the generic terms.

---

[4] Other features of the ALMI are: pick lists that minimize the amount of data entry and, therefore, keying errors; edit checks that check the logic of the data entry and identify missing/critical data.

no RO review of group quarters listings in the area and group quarters frames; however, basic edits are included with the Group Quarters Automated Instrument for Listing (GAIL).

As stated previously, listing reviews are also performed by the NPC. The following are activities involved in the reviews for the unit-frame sample:

1. Verify that the FR resolved all unit designations that were missing or duplicated in Census 2000; i.e., if any sample units contained missing or duplicate unit designations.

2. Review for accuracy any large multiunit addresses that were relisted.

3. Check whether additional units (found during listing) already had a chance of selection. If so, the RO is contacted for instructions for correcting the Multiunit Listing Aid.

In terms of the review of listing sheets for permit frame samples, the NPC 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, the NPC determines whether the additional units already had a chance of selection. If so, the RO is contacted with instructions for updating the listing sheet.

The majority of listings are found to be correct during the RO and NPC reviews. When errors or omissions are detected, they usually occur in batches, signifying a misinterpretation of instructions by an RO or a newly hired FR.

Check-in of sample. Depending on the sampling frame and mode of interview, the monthly process of check-in of the sample occurs as the sample cases progress through the ROs, the NPC, and headquarters. Check-in of sample describes the processes by which the ROs verify that the FRs receive all their assigned sample cases and that all are completed and returned. Since the CPS is now conducted entirely by computer-assisted personal or telephone interview, it is easier than ever before to track a sample case through the various systems. Control systems are in place to control and verify the sample count. The ROs monitor 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. A check-in also occurs when the cases are transmitted to headquarters.

Listing check. Listing check is a quality assurance program for the Demographic Area Address Listing (DAAL). It uses the same instrument as the area-frame listing, i.e., the ALMI. Each month, a random sample of FRs who have completed sufficient listing work is selected for a listing check. The goal of this sampling process is to check each FR's work at least once a year. If the FR is selected for a listing check, then a sample of the work performed by this

FR is selected and is checked by a senior staff member. Various types of errors, including coverage errors, content errors, and mapping errors, are recorded during the listing check. Classification of errors is completely automated within the ALMI instrument. The results are then compared against the 3-month aggregate average error rate for each RO to determine a "pass" or "fail" status for the FR. This allows ROs to monitor the performance of FRs and provides information about the overall quality of the DAAL that allows continual improvements of the listing process. Depending on the severity and type of errors, the ROs must give feedback to FRs who failed the listing check and retrain them as necessary. The automated system will warn the supervisors if they try to assign work to an FR who failed a listing check but has not been retrained.

SOURCES OF NONRESPONSE ERROR

There are three main sources of nonresponse error in the CPS: unit nonresponse, person nonresponse, and item nonresponse. Unit nonresponse error occurs 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. The weights of eligible households who do respond to the survey are increased to account for those who do not, but nonresponse error can be introduced if the characteristics of the interviewed households differ from those that are not interviewed.

Individuals 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 others in the household as a proxy reporter. Panel nonresponse exists when those who live in the same household during the entire time they are in the CPS sample do not agree to be interviewed in any 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 (explained in Chapter 9) 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.

One example of the magnitude of the error due to nonresponse is the national Type A household noninterview

rate, which was 7.70 percent in July 2004[5]. For item nonresponse, a few of the average allocation rates in July 2004, by topical module were: 0.52 percent for household, 1.99 percent for demographic, 2.35 percent for labor force, 9.98 percent for industry and occupation, and 18.46 percent for earnings. (See Chapter 16 for discussion of various quality indicators of nonresponse error.)

CONTROLLING NONRESPONSE ERROR

Field Representative Guidelines[6]

Response/nonresponse rate guidelines have been developed for FRs 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 FRs needing performance improvement. An FR whose response rate, household noninterview rate (Type A), or minutes-per-case falls below the fully acceptable range based on one quarter's work is considered in need of additional training and development. The CPS supervisor then takes appropriate remedial action. National and regional response performance data are also provided to permit the RO staff to judge whether their activities are in need of additional attention.

Summary Reports

Another way to monitor and control nonresponse error is the production and review of summary reports. Produced by headquarters after the release of the monthly data products, they are used to detect changes in historical response patterns. Since they are distributed throughout headquarters and the ROs, other indications of data quality and consistency can be focused upon. The contents of some of the summary report tables are: noninterview rates by RO for both the basic CPS and its supplements, 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, percent of personal-visit cases actually conducted in person, allocation rates by topical module, 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 to help the ROs understand CPS field operations, to solicit and share the ROs' views on the causes of the increasing nonresponse rates, and to evaluate methods to decrease these rates. All of the answers provide insight into the CPS operations that may affect nonresponse and follow-up procedures for household noninterviews. A few are:

1. The majority of ROs responded that there is written documentation of the follow-up process for CPS household noninterviews.

2. The standard process is that an FR must let the RO know about a possible household noninterview as soon as possible.

3. Most regions attempt to convert confirmed refusals to interviews under certain circumstances.

4. All regions provide monthly feedback to their FRs on their household noninterview rates.

5. About half of the regions responded that they provide specific region-based training/activities for FRs on converting or avoiding household noninterviews. Much research has been distributed regarding whether and how to convince someone who refused to be interviewed to change their mind.

Most offices use letters in a consistent manner to follow-up with noninterviews. Most ROs also include informational brochures about the survey with the letters, and they are tailored to the respondent.

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 can include:

1. The respondent misinterprets the question, does not know the true answer and guesses (e.g., recall effects), exaggerates, has a tendency to give an answer that appears more 'socially desireable,' or chooses a response randomly.

2. The interviewer reads the question incorrectly, does not follow the appropriate skip pattern, misunderstands or misapplies the questionnaire, or records the wrong answer.

3. A proxy responder (i.e., a person who answers on someone else's behalf) provides an incorrect response.

4. The data collection modes (e.g., personal visit and telephone) elicit different responses.

---

[5] In April 2004, CPS started phasing in the new 2000-based sample.
[6] See Appendix D for a detailed discussion, especially in terms of the performance evaluation system.

5. The questionnaire does not elicit correct responses due to a format that is not easy to understand, or has complicated or incorrect skip patterns or 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,[7] while their interactions are discussed in the section *The Reinterview Program—Quality Control and Response Error*.[8]

In terms of magnitude, measures of response error are obtainable through the reinterview program, specifically, the index of inconsistency. This index 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 interview to interview, 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 2003 is considered moderate at 37.9. Other statistical techniques being used to measure response error include latent class analysis (Tran and Mansur, 2004), which is being used to look at how responses vary across time-in-sample.

CONTROLLING RESPONSE ERROR

Survey Instrument[9]

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 FRs and telephone calls made by interviewers at centralized telephone centers. Regardless of the interview mode, the questionnaire and the software are basically the same (see Chapter 7).

Software. Computer-assisted interviewing technology in the CPS allows very complex skip patterns and other procedures that 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 balances nonsampling error existing for the item with the possibility of a total noninterview. The instrument provides opportunities for the FR 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 to the FR to be verified and corrected. In addition to reducing response and interviewer burden, the instrument avoids erratic variations in industry and occupation codes among pairs of months for people who have not changed jobs but who describe their industry and occupation differently in the 2 months.

Questionnaire. Two objectives of the design of the CPS questionnaire are 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) are: short and clear question wording, splitting complex questions into two or more questions, building concept definitions into question wording, reducing reliance on volunteered information, using explicit and implicit strategies for the respondent to provide numeric data, and using precoded response categories for open-ended questions. Interviewer notes recorded at the end of the interview are critical to obtaining reliable and accurate responses.

Modes 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 FR or by an interviewer from a centralized telephone facility. (In July 2004, nationally, 81 percent and 62.8 percent of personal visit cases were actually conducted in person in months-in-sample 1 and 5, respectively.) Although each mode has its own set of performance guidelines that must be adhered to, similarities do exist. The controls detailed in *The Interviewer* and *Error Due to Nonresponse* sections are mainly directed at personal visits but are also basically valid for the calls made from the centralized facility via the supervisor's listening in.

Continuous testing and improvements. Insights gained from past research on questionnaire design have 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 is tested to gauge whether respondents are

---

[7] Most discussion in this section is applicable whether the interview is conducted via computer-assisted personal interview or computer-assisted telephone interview by the FR or in a centralized telephone facility.

[8] Appendix E provides an overview of the design and methodology of the entire reinterview program.

[9] Many of the topics in this section are presented in more detail in Chapter 6.

correctly interpreting the questions. Chapter 6 provides an extensive list of the various methods of testing. The Census Bureau has also developed a set of protocols for pretesting demographic surveys.

To improve 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 the CPS interviewers and supervisors are conducted periodically.

Despite the benefits from adding new questions and improving existing ones, changes to the CPS are approached cautiously until the effects are measured and evaluated. In order to avoid the distruption of historical series, methods to bridge differences caused by changes or techniques are included in the testing whenever possible. In the past, for example, parallel surveys have been conducted using the revised and unrevised procedures. Results from the parallel survey have been used to anticipate the effect the changes would have on the survey's estimates and nonresponse rates (Kostanich and Cahoon, 1994; Polivka, 1994; and Thompson, 1994).

Interviewer

Interviewer training, observation, monitoring, and evaluation are all methods used to control nonsampling error, arising from inaccuracies in both the frame and data collection activities. For further discussion, see this chapter's section *Field Representative Guidelines* and Appendix D. Group training and home study are continuing efforts in each RO to control various nonsampling errors, and they are tailored to the types of duties and length of service of the interviewer.

Field observation is an extension of classroom training and provides on-the-job training and on-the-job evaluation. It is one of the methods used by the supervisor to check and improve the performance of the FR. It provides a uniform method for assessing the FR's attitudes toward the job and use of the computer, and evaluates the FR'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 such as the following:

1. Asking questions as worded and in the order presented on the questionnaire.

2. Adhering to instructions on the instrument and in the manuals.

3. Knowing how to probe.

4. Recording answers in the correct manner and in adequate detail.

5. Developing and maintaining good rapport with the respondent conducive to an exchange of information.

6. Avoiding questions or probes that suggest a desired answer to the respondent.

7. Determining the most appropriate time and place for the interview.

The emphasis is on correcting habits that interfere with the collection of reliable statistics.

Respondent: Self Versus Proxy

The *CPS Interviewing Manual* states that any household member 15 years of age or older is technically eligible to act as a respondent for the household. The FR attempts to collect the labor force data from each eligible individual; however, in the interests of timeliness and efficiency, any knowledgeable adult household member is allowed to provide the information. Also, the survey instrument is structured so that every effort is made to interview the same respondent every month. The majority of the CPS labor force data is collected by self-response, and most of the remainder is collected by proxy from a household respondent. The use of a nonhousehold 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.

There has been a substantial amount of research into self-versus-proxy reporting, including research involving CPS respondents (Kojetin and Mullin, 1995; Tanur, 1994). Much of the research indicates that self-reporting is more reliable than proxy reporting, particularly when there are motivational reasons for proxy and self-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 1 and 5 whenever possible. By showing a sincere understanding of 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.

Reinterview Program—Quality Control and Response Error[10]

The reinterview program has two components: the Quality Control Reinterview Program and the Response Error Reinterview Program. One of the objectives of the Quality Control Reinterview Program is to evaluate individual FR's performance. It checks a sample of the work of an FR and identifies and measures aspects of the field procedures that may need improvement. It is also critical in the detection and prevention of data falsification.

The Response Error Reinterview Program 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 survey's original results; as a by-product, 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 potential 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

Industry and Occupation Coding Verification

To be accepted into the CPS processing system, files containing records needing three-digit industry and occupation codes are electronically sent to the NPC 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 online industry and occupation reference materials, the NPC 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 and 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 editing is one of the checks used to control nonsampling 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 with that 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 made for zero weighting 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 four-page tabulation of monthly summary statistics that highlights the six edit modules and the two main weighting programs for the current month and the preceding 12 months. This is called the "CPS Monthly Check-in Sheet."

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 agencies compute numerous tables in composited and uncomposited modes. These tables are then checked cell-by-cell to ensure that the cell entries across the two agencies are identical. If so, the data are deemed to have been computed correctly.

Extensive verification is done every time any change is made to any part of the CPS estimation process and its impact is evaluated. Examples are weighting changes,

---

[10] Appendix E provides an overview of the design and methodology of the entire reinterview program.

changes in replication methodology, and the introduction of new noninterview cluster codes and transitional base-weights during the phasing-out of one sample design and the phasing-in of another design. In fact, there are several guidelines and specifications that deal solely with CPS verification roles and responsibilities.

CPS Population Controls

National and state-level CPS population controls are developed by the Census Bureau independently from the collection and processing of the CPS data. These monthly and independent projections of the population are used for the iterative, second-stage and composite weighting of the CPS data. All of the estimates start with data from the last census (currently Census 2000), and use administrative records and projection techniques to provide updates. (See Appendix C 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 details 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 are files that contain overlapping/redundant data. Third, a two-person clerical review of all data with comments/notes is performed. An important means of assuring that quality data are used as input into the CPS population controls is continuous research into improvements in methods of making population estimates and projections.

Modeling Errors

This section focuses on a few of the methods to reduce nonsampling error that are applied to the seasonal adjustment programs and monthly model-based state labor force estimates. (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

Kojetin, B.A. and P. Mullin (1995), "The Quality of Proxy Reports on the Current Population Survey (CPS)," **Proceedings of the Section on Survey Research Methods**, American Statistical Association, pp. 1110–1115.

Kostanich, D.L. and L.S. Cahoon, (1994), **Effect of Design Differences Between the Parallel Survey and the New CPS,** CPS Bridge Team Technical Report 3, dated March 4.

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.

Tanur, J.M. (1994), "Conceptualizations of Job Search: Further Evidence From Verbatim Responses," **Proceedings of the Section on Survey Research Methods**, American Statistical Association, pp. 512–516.

Thompson, J. (1994), **Mode Effects Analysis of Major Labor Force Estimates,** CPS Overlap Analysis Team Technical Report 3, April 14.

Tran, B. and K. Mansur, (2004), "Analysis of the Unemployment Rate in the Current Population Survey—A Latent Class Approach," **Journal of the American Statistical Association**, August 12.

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 an 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 a larger nonresponse bias than a survey that has a higher nonresponse rate (if the difference between respondents and nonrespondents is larger in the survey with the lower nonresponse rate than it is in the survey with the higher nonresponse rate), one would generally expect that larger nonresponse indicates a greater potential for bias. 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 being selected for 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 people erroneously included or excluded from the survey differ from those of individuals correctly included in the survey. Historically in the CPS, the net effect of coverage errors has been an underestimate of the size of the total population for most major demographic population subgroups before the population controls are applied (known as undercoverage).

Coverage Ratios

One way to estimate the coverage error present in a survey is to compute a coverage ratio. A coverage ratio is the outcome of dividing the estimated number of people in a specific demographic group from the survey by 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 national and state coverage adjustments and the second-stage ratio adjustment. See Chapter 10 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 C 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 overcoverage exists when the ratio is greater than 1.0. Figure 16–1 shows the average monthly coverage ratios for September 2001 through September 2004.

In terms of race, Whites have the highest coverage ratio (90.7 percent), while Blacks have the lowest (82.2 percent). Females across all races have higher coverage ratios than males. Hispanics[1] 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 people who are missed by the CPS are on the average likely to be different from those who are included. People who are missed are accounted for in the CPS, but they are given the same labor force characteristics as those of the people who are included. This produces bias in the CPS estimates.

This graph, as well as two other graphs of coverage ratios by race and gender, can be found at <http://www.bls.census.gov/cps/basic/perfmeas/coverage.htm>. (Their updates, with more current data, will be posted on this site as they are made available.) The three graphs provide

---

[1]Hispanics may be any race.

Figure 16–1.
CPS Total Coverage Ratios: September 2001–September 2004[1], National Estimates
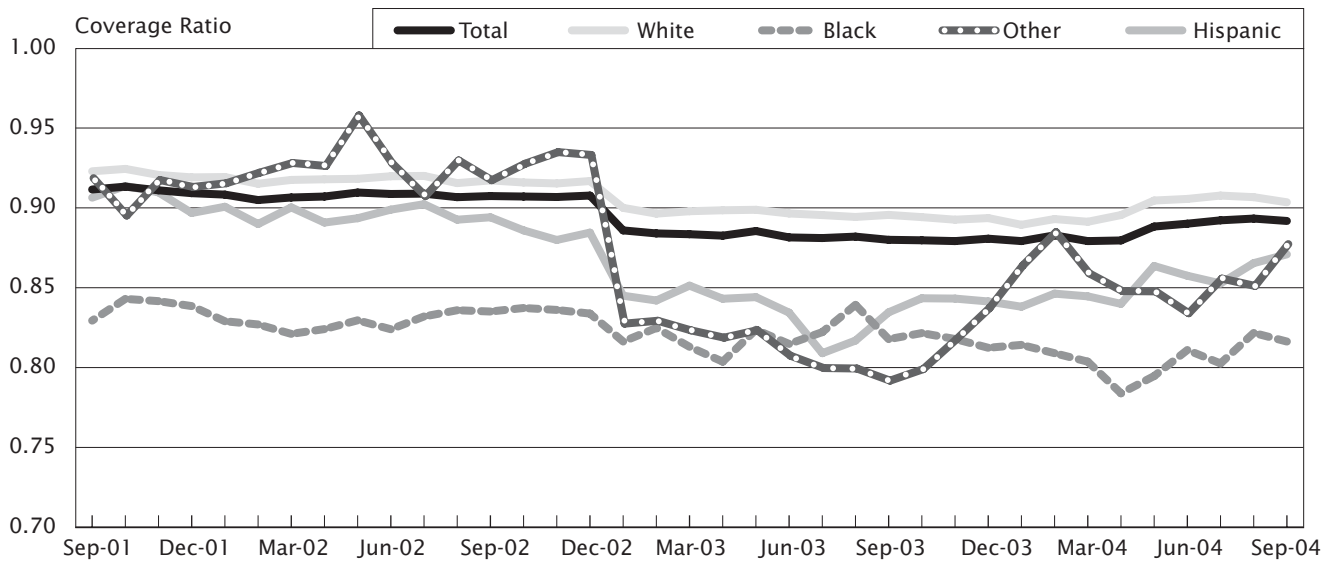
[1] There is a drop in January 2003. This is when the new definitions for race and ethnicity were introduced, as well as some adjusted population controls based on Census 2000.

a picture of coverage for the population at least 16 years old in the CPS from September 2001 through September 2004. The first one gives the coverage ratios for racial groups. Traditionally, Blacks and Hispanics have been the most underrepresented groups in the CPS. The other two graphs show the coverage ratios by race and gender. Coverage ratios are lowest for Black and Hispanic males.

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, or 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 that respond in one 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 seen in Figure 16–2, the noninterview rate for the CPS remained relatively stable at around 4 to 5 percent for most of 1964 through 1993; however, there have been some changes since 1993. Figure 16–2 shows 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 was chiefly because of disruptions in data collection due to shutdowns of the federal government (see Butani, Kojetin, and Cahoon, 1996). After 1994, refusals, noninterviews, and noncontacts increased. The relative stability of the overall noninterview rate from 1960 to 1994 masked some underlying changes that occurred. Specifically, the refusal portion of the noninterview rate increased over this 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. A corresponding decrease in the rate of noncontacts and other noninterviews compensated for this increase in refusals.

Seasonal variation also appears in both the overall noninterview rates and the refusal rates (see Figure 16–3). During the year, the noninterview and refusal rates have

Figure 16–2.
Average Yearly Type A Noninterview and Refusal Rates for the CPS 1964–2003, National Estimates
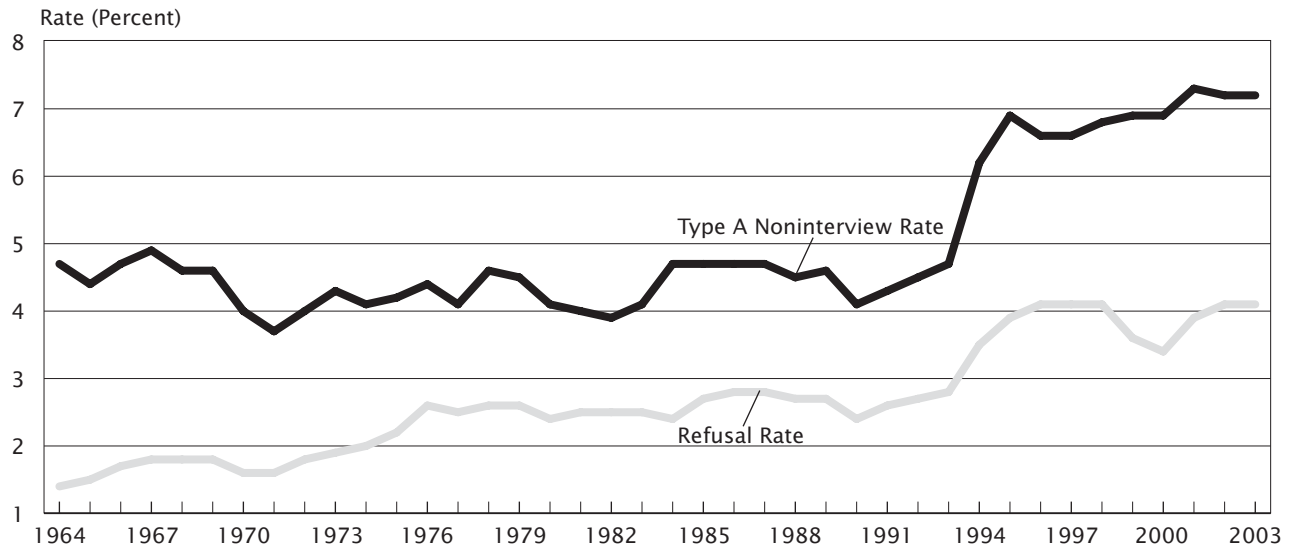
Figure 16–3.
CPS Nonresponse Rates: September 2003–September 2004, National Estimates
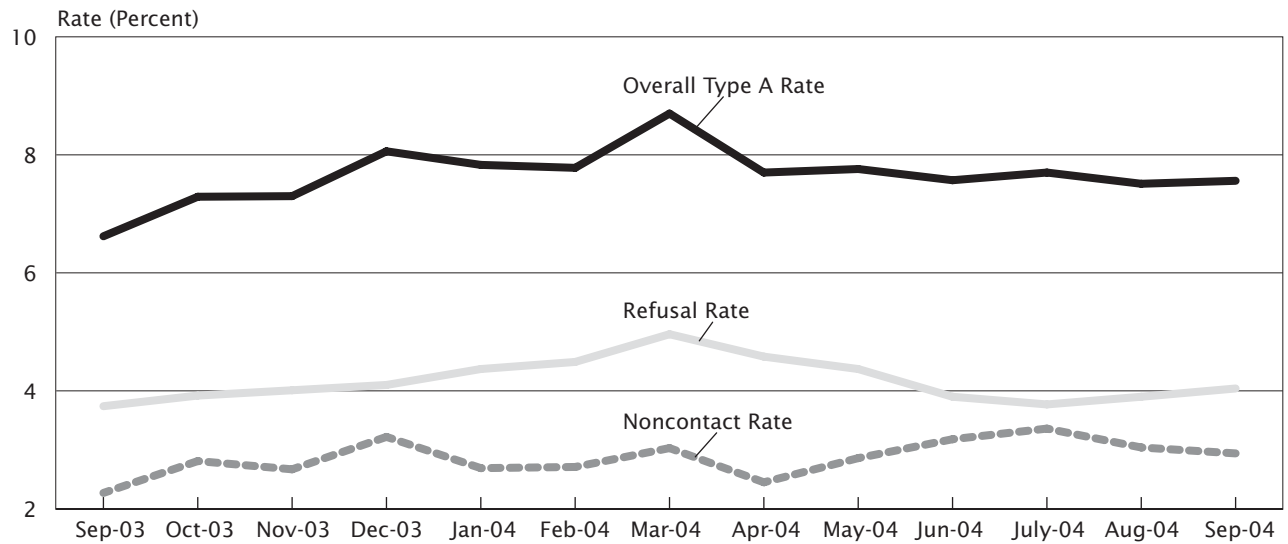
tended to increase after January until they reach a peak in March or April, at the time of the Annual Social and Economic Supplement. At this point, there is a drop in noninterview and refusal rates that extends below the starting point in January until they bottom out in July or August. The rates then increase and approach the initial level. This pattern has been evident most years in the recent past and appears to be similar for 2003–2004. The noncontact rates are higher during the winter holidays and the summer, when some household members are either away from home or difficult to contact.

Effect on noninterview rates of the transition to a redesigned survey with computer-assisted data collection. With the transition to the redesigned CPS questionnaire using computerized data collection in January 1994, Type A nonresponse rates increased as seen in Figure 16–2. This transition included several procedural changes in the collection of the data, and the adjustment to these new procedures may account for this increase. For example, the computer-assisted personal interviewing (CAPI) instrument requires 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. Households not interviewed by the computer-assisted telephone interviewing (CATI) centers are recycled back to the field representatives continuously during the survey week. However, cases recycled late in the survey week (some reach the field as late as Friday morning) 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. The annual overall Type A rate, the refusal rate, and noncontact rate (which includes temporarily absent households and other noncontacts) are shown in Table 16–1 for the period 1993–1996 and 2003.

Table 16–1. **Components of Type A Nonresponse Rates, Annual Averages for 1993–1996 and 2003, National Estimates**

[Percent distribution]

Nonresponse rate	1993	1994	1995	1996	2003
Overall Type A	4.69	6.19	6.86	6.63	7.25
Noncontact	1.77	2.30	2.41	2.28	2.58
Refusal	2.85	3.54	3.89	4.09	4.10
Other	.13	.32	.34	.25	.57

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 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 16–2 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, about 82 percent, completed interviews each month, and only 2 percent never participated (for further information, see Harris-Kojetin and Tucker, 1997). Dixon (2000) compared those who moved out to those who moved in. Out-movers were more likely to be unemployed but more likely to respond compared with in-movers. Unemployment may be slightly underestimated due to the combination of these two effects.

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

Table 16–2. **Percentage of Households by Number of Completed Interviews During the 8 Months in the Sample, National Estimates**[1]

[January 1994–October 1995]

Number of completed interviews	Percent 1994–1995
0	2.0
1	0.5
2	0.5
3	0.6
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 they were in the sample were not included in these tabulations. Movers were not included in this tabulation.

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 people 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 and households that responded one month but failed to respond in the other month. However, the labor force status of people in households that were nonrespondents for both months is unknown.

Monthly labor force data were used for each consecutive pair of months for January through June 1997, for households whose members responded for each consecutive pair of months and separately for households whose respondents responded only one month and were nonrespondents the other month (see Tucker and Harris-Kojetin, 1997). The top half of Table 16–3 shows the labor force classification in the first month for people in households who were respondents the second month compared with people who were in households that were noninterviews the second month. People from households that became nonrespondents had higher rates of participation in the labor force, employment, and unemployment than those from households that responded in both months. The bottom half of Table 16–3 shows the labor force classification for the second month for people in households that were respondents in the previous month compared with people 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

Table 16–3. **Labor Force Status by Interview/Noninterview Status in Previous and Current Month, National Estimates**

[Average January–June 1997[1] percent distribution]

First month labor force status	Interview in second month	Nonresponse in second 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

Second month labor force status	Interview in first month	Nonresponse in first 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).

Type A noninterview rate is quite small, the effect of the nonresponding households on the overall unemployment rate is also relatively small. However, the labor force characteristics of the people 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).

Dixon (2001) found similar effects in data from 1996–1999, and further found that the bias came from noncontacts more than refusals. These results were supported by an analysis of data from the CPS matched to the long form of Census 2000 (Dixon, 2004), where nonresponders (and their census employment status) in the CPS could be identified. This addressed the problem of those who never responded to the CPS. The estimates were in the same direction as found by Tucker and Kojetin, although the bias was higher for the 2000 data.

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, allocation rates of all interviewed cases were examined from both January 1997 and January 2004. The average levels of item-missing data are represented by the allocation rates in Table 16–4. It is likely that the bias due to imputation or allocation for the demographic and labor force items is quite small, but there may be a concern for industry/occupation and earnings data. Allocation rates have increased over the years, as indicated by differences between 1997 and 2004.

Dixon (2002) found that item refusals predicted unit refusals and were negatively related to unemployment. Respondents who refused to answer particular items (especially items related to income data) were more likely to refuse to participate in subsequent surveys and were less likely to be unemployed.

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 unknown biases in the estimates (e.g., a high noninterview rate.)

Table 16–4. **CPS Items With Missing Data (Allocation Rates, %), National Estimates**

Month/year	Household		Demographic		Labor force		I&O		Earnings	
	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean
Jan. 1997......	0.16–5.84	0.41	0.08–9.2	1.35	0.15–12.0	1.57	2.12–5.0	4.09	0.99–32.0	11.02
Jan. 2004......	0.01–3.42	0.56	0.12–19.0	2.41	0.21–15.0	2.18	5.43–10.0	8.74	0.01–36.0	17.68

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 people reported as unemployed (see Table 16–5). In this schematic representation, the number of identical people giving a different response

Table 16–5. **Comparison of Responses to the Original Interview and the Reinterview, National Estimates**

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

in the original interview and the reinterview for the characteristic unemployed is given by (b + c). If these responses are two independent measurements for identical people using identical measurement procedures, then an estimate of the simple response variance component $\left(\sigma_R^2\right)$ 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)[2]. 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 to determine 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 the 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 will trigger a review of field procedures to determine and remedy the cause.

Table 16–6 provides estimates of the index of inconsistency shown as percentages for selected labor force characteristics for 2003.

Table 16–6. **Index of Inconsistency for Selected Labor Force Characteristics in 2003, National Estimates**

Labor force characteristic	Estimate of index of inconsistency	90-percent confidence limits
Employed at work.........	14.0	19.0–20.8
Employed absent	50.4	45.7–55.0
Unemployed layoff........	60.8	49.8–71.8
Unemployed looking	38.5	34.8–42.1
Not-in-labor force retired ..	7.8	6.8–8.6
Not-in-labor force disabled.	21.1	18.4–24.0
Not-in-labor force other ...	31.1	29.4–32.8

MODE OF INTERVIEW

Incidence of Telephone Interviewing

As described in Chapter 7, 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

---

[2] 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.

supposed to be done in person, the entire interview may 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 is shown in Table 16–7. Overall, about 66 percent of the cases are done by telephone, with almost 80 percent of the cases in month-in-samples (MIS) 2–4 and 6–8 done by telephone. Furthermore, a number of cases in months 1 and 5 are completed by telephone, despite standing instructions to the field interviewers to conduct personal visit interviews.

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, causing the interview to be recorded as a telephone interview. In June 1996, an additional item was added to the CPS instrument 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 16–7. It was expected that in MIS

Table 16–7. **Percentage of Households With Completed Interviews With Data Collected by Telephone (CAPI Cases Only), National Estimates**

[Percent]

Month in sample	Last contact with household was telephone (average, Jan.–Dec. 2004)	Majority of data collected by telephone (average, Jan.–Dec. 2004)
1	13.5	20.5
2	73.0	75.6
3	76.0	78.5
4	76.3	79.0
5	28.3	37.9
6	76.2	78.8
7	77.8	80.0
8	77.7	80.5

1 and 5, interviewers would have reported that the majority of the data was collected though personal visit more often than was revealed by data on 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 to obtain missing information not collected at the initial personal visit. However, the percentage of cases where the majority of the data was reported to have been collected by telephone was larger than the percentage of cases where the last contact with the household was by telephone for MIS 1 and 5. The percentage increased in

2004 compared with 1995. The explanation for this pattern of results is not clear at the present time. Some reasons hypothesized are that some respondents may be more comfortable over the telephone, or using CATI may solve a difficult scheduling problem.

TIME IN SAMPLE

The rotation pattern of the CPS sample was described in detail in Chapter 3, and the use of composite estimation in the 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, 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 that shows the relationships of all the month-in-sample groups. This index is the ratio of the estimate based on 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 16–8 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 March 2003 to February 2004. For the percentage of the total population that is unemployed, the index of 106.26 for the first-month households indicates that the estimate from households in the sample in the first month is about 1.0626 times the average overall eight month-in-sample groups; the index of 99.29 for unemployed for the fourth-month households indicates that it is about the same as the average overall month-in-sample groups.

Estimates from one of the month-in-sample groups should not be taken as the standard, in the sense that they 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 that are in the sample for the first time, are likely to be less successful at calling when a responsible household member is available. Thus, noninterview rates generally start above average with first-month households, decrease with more time in the sample, go up a little for households in the fifth month of the sample (after 8 months, the household

Table 16–8. Month-in-Sample Bias Indexes (and Standard Errors) in the CPS for Selected Labor Force Characteristics

[Average March 2003–February 2004]

Employment status and sex	Month-in-sample							
	1	2	3	4	5	6	7	8
TOTAL POPULATION, 16 YEARS AND OLDER								
Civilian labor force..............	101.00	100.97	100.64	100.50	99.74	98.95	99.15	99.06
Standard error	0.34	0.33	0.36	0.36	0.37	0.40	0.38	0.36
Unemployment level	106.26	103.07	100.56	99.29	101.37	99.14	95.64	94.67
Standard error	1.33	1.41	1.32	1.58	1.04	1.40	1.29	1.40
MALES, 16 YEARS AND OLDER								
Civilian labor force..............	100.73	101.04	100.66	100.37	99.86	98.92	99.36	99.07
Standard error	0.39	0.40	0.40	0.41	0.43	0.43	0.40	0.38
Unemployment level	105.69	102.44	102.43	101.90	99.66	97.42	96.07	94.39
Standard error	1.65	1.63	1.64	1.92	1.49	1.74	1.89	1.95
FEMALES, 16 YEARS AND OLDER								
Civilian labor force..............	101.32	100.89	100.61	100.64	99.60	98.99	98.90	99.04
Standard error	0.39	0.35	0.43	0.44	0.40	0.45	0.44	0.43
Unemployment level	106.98	103.87	98.22	96.03	103.49	101.29	95.10	95.02
Standard error	1.77	1.98	2.09	2.30	1.43	1.72	1.31	2.06
BLACK POPULATION, 16 YEARS AND OLDER								
Civilian labor force..............	101.87	100.75	101.37	100.76	99.17	98.62	98.12	99.35
Standard error	1.08	1.15	1.18	1.19	1.18	1.29	1.28	1.34
Unemployment level	112.74	105.44	103.25	98.69	96.63	99.99	93.23	90.03
Standard error	3.41	3.44	3.04	3.21	3.14	3.39	2.94	3.23
BLACK MALES, 16 YEARS AND OLDER								
Civilian labor force..............	101.69	100.70	101.14	100.05	98.76	98.55	99.42	99.68
Standard error	1.23	1.42	1.42	1.50	1.51	1.60	1.63	1.61
Unemployment level	109.55	105.16	103.97	99.15	97.13	97.03	98.56	89.46
Standard error	4.45	4.78	4.50	4.56	4.30	4.55	4.20	4.44
BLACK FEMALES, 16 YEARS AND OLDER								
Civilian labor force..............	102.03	100.79	101.57	101.38	99.52	98.68	96.99	99.05
Standard error	1.29	1.30	1.34	1.33	1.25	1.39	1.40	1.52
Unemployment level	115.86	105.72	102.55	98.25	96.13	102.89	88.01	90.59
Standard error	4.10	4.09	4.43	4.53	3.55	4.32	3.21	4.46

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.) Figure 16-4, showing the Type A noninterview rate by month-in-sample for July–September 2004, illustrates this pattern. An index of the noninterview rate constructed in a similar manner to the month-in-sample group index can be seen in Table 16–9. This noninterview rate index 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 noninterview

changes are unlikely to be distributed proportionately among the various labor force categories (see Table 16–3). 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.

Figure 16–4.
Basic CPS Household Nonresponse by Month in Sample, July 2004–September 2004, National Estimates
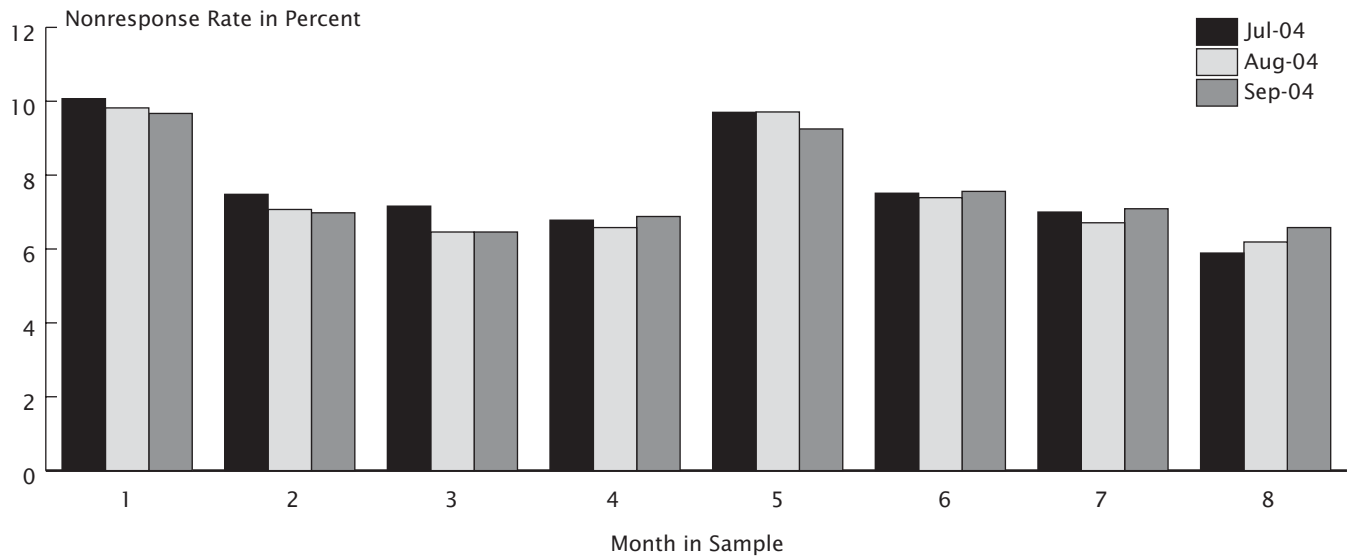
Table 16–9. **Month-In-Sample Indexes in the CPS for Type A Noninterview Rates January–December 2004**

Month in sample	Noninterview rate index
1	129.1
2	93.7
3	89.1
4	88.6
5	119.8
6	98.2
7	94.2
8	87.1

PROXY REPORTING

Like many household surveys, the CPS seeks information about all people in the household, whether they are available for an 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 other household members, nonsampling error may occur because of the use of proxy reporters.

The level of proxy reporting in the CPS was generally around 50 percent in the past and continues to be so in the revised CPS. As can be seen in Table 16–10, the

month-in-sample has very little effect on the level of proxy reporting and the levels have remained the same over time. 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 on proxy use.

Table 16–10. **Percentage of CPS Labor Force Reports Provided by Proxy Reporters**

Percent reporting	1995	2004
All Interviews		
Proxy reports	49.68	48.84
Both self and proxy	0.28	1.14
MIS 1 and 5		
Proxy reports	50.13	48.98
Both self and proxy	0.11	0.78
MIS 2–4 and 6–8		
Proxy reports	49.53	48.79
Both self and proxy	0.35	1.25

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 the process. This chapter also includes information on the potential effects on the CPS estimates due to nonresponse, centralized telephone interviewing, and the duration that groups are in the CPS sample. This research comes close to identifying the presence and effects of nonsampling error in the CPS, but the results are not conclusive.

The full extent of nonsampling error in the CPS cannot be known without an external source for comparison. Recent work cited here indicates undercoverage has increased since 2000, and nonresponse continues to be a concern. However, the effect of nonresponse bias is considered to be small. The month-in-sample bias persists, and item nonresponse has increased over the last ten years. On the other hand, there is evidence that, while some inconsistency in reporting exists, the measurement of employment status is on a sounder conceptual footing than ever before. (See Biemer (2004), Miller and Polivka (2004), Tucker (2004), and Vermunt (2004).)

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.

Biemer, P. P. (2004), "An Analysis of Classification Error for the Revised Current Population Survey Employment Questions," **Survey Methodology**, Vol. 30, 2, 127–140.

Brooks, C. A. and B. A. Bailar (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, 16–10.

Butani, S., B. A. Kojetin, and L. Cahoon (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.

Dixon, J. (2000), "The Relationship Between Household Moving, Nonresponse, and Unemployment Rate in the Current Population Survey." Paper presented at the Joint Statistical Meetings of the American Statistical Association, Indianapolis, Indiana.

Dixon, J. (2001), "Relationship Between Household Nonresponse, Demographics, and Unemployment Rate in the Current Population Survey." Paper presented at the Joint Statistical Meetings of the American Statistical Association, Atlanta, Georgia.

Dixon, J. (2002), "The Effects of Item and Unit Nonresponse on Estimates of Labor Force Participation." Paper presented at the Joint Statistical Meetings of the American Statistical Association, New York City, New York.

Dixon, J. (2004), "Using Census Match Data to Evaluate Models of Survey Nonresponse." Paper presented at the Joint Statistical Meetings of the American Statistical Association, Toronto, Canada.

Hansen, M. H., W. N. Hurwitz, H. Nisselson, and J. Steinberg (1955), "The Redesign of the Current Population Survey," **Journal of the American Statistical Association,** 50, 701–719.

Hansen, M. H., W. N. Hurwitz, and L. Pritzker (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, U.S. Census Bureau, Washington, DC.

Harris-Kojetin, B. A. and C. Tucker (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.

Miller, S. H. and A. E. Polivka (2004), "Comment," **Survey Methodology**, Vol. 30, 2, 145–150.

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 B. A. Harris-Kojetin (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.

Tucker, C. (2004), "Comment," **Survey Methodology**. Vol. 30, 2, 151–153.

Vermunt, J. K. (2004), "Comment," **Survey Methodology**, Vol. 30, 2, 141–144.

Williams, W. H. and C. L. Mallows (1970), "Systematic Biases in Panel Surveys Due to Differential Nonresponse," **Journal of the American Statistical Association,** 65, 1338–1349, 16–11.

Appendix A
Sample Preparation Materials

INTRODUCTION

Despite the conversion of CPS to an automated listing and interviewing environment, some paper materials are still in use. These materials include segment folders, listing sheets, 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 Folder, BC–1669 (CPS)
Illustration 1

A field representative receives a segment folder for each unit segment with incomplete addresses or multi-unit addresses.

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 designation for the segment, and basic geography.

2. 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: Multi-UnitListing Aids (MULA) (Form 11-12), Unit/Permit Listing Sheets (Form 11-3), and Incomplete Address Locator materials.

Multi-Unit Listing Aids (Form 11–12)
Illustration 2

MULAs are provided for all unit segments with multi-unit addresses. For each multi-unit,the field representative receives a preprinted listing that shows addresses and unit information as recorded from Census 2000.

Unit/Permit Listing Sheets (Form 11–3)
Illustration 3

Unit/Permit Listing Sheets are used only when the entire address needs to be relisted and the field representative prefers not to relist directly on the MULA. The field representative has a supply of blank Unit/Permit Listing Sheets for this use.

Incomplete Address Locator Action Form (Form NPC–1138)
Illustration 4

The Incomplete Address Locator Action Form is used when the information obtained from Census 2000 is in some way incomplete (i.e. missing a house number, unit designation, etc.). The form provides the field representative the additional information that can be used to locate the incomplete address. The information on the form includes:

1. The address as it appeared in Census 2000 and possibly a suggested 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 that the field representative should take to locate the address.

After locating the address, the field representative completes or corrects the basic address in the laptop case management (for single and multi-units)and on the MULA (for multi-units).

AREA FRAME MATERIALS

There are no longer paper sample preparation materials for the area frame. Sample preparation has been fully automated.

GROUP QUARTERS FRAME MATERIALS

There are no longer paper sample preparation materials for the group quarters frame. Sample preparation has been fully automated.

PERMIT FRAME MATERIALS

Segment Folder, BC–1669 (CPS)
Illustration 1

See the description and illustration in the Unit Frame Materials above. A field representative receives a segment folder for every permit segment. The folder will contain a Unit/Permit Listing Sheet (Form 11-3) and possibly a Permit Sketch Map.

Unit/Permit Listing Sheets (Form 11–3)
Illustration 3, 5, and 6

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 sheets(s) to list all units at a multi-unitaddress.

Permit Sketch Map (Form 11–187)
Illustration 7

When completing the permit address list (PAL) operation, 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 and streets in the vicinity of the site.

Illustration 1. Segment Folder, BC-1669 (CPS)

FORM **BC-1669** (CPS)

SEGMENT FOLDER (**SURVEY CPS**)

(BAR CODE INFORMATION)

DATE

	MO/YR	MO/YR	MO/YR	MO/YR
REDUCED				
REINSTATED				
REASSIGNED				

RO   PSU   SEG#

FRAME (Unit or Permit)

ZIP

PLACE NAME or MCD NAME

COUNTY

SPECIAL INSTRUCTIONS

PSU

REMARKS

SEG #

(fold is here) . . . .

Illustration 2. Multi-Unit Listing Aid, Form 11-12

FORM 11-12	U.S. DEPARTMENT OF COMMERCE U.S. CENSUS BUREAU	RO 32	PSU 06037	Segment 1999	Survey: CPS
MULTI-UNIT LISTING AID (MULA)		Tract: 004893	Block 3005	Expected Number of Units 16	

Address * 4827 Labor Force Ave. North Hollywood, CA 91601	State CA	County Los Angeles

Line No.	Unit Designation (or apartment number)	Sample Desig.	Serial Number	Line No.	Unit Designation (or apartment number)	Sample Desig.	Serial Number
1		M		21		A78	
2		M	A79	03	22			
3	# 101	A78	04	23		A77	
4	# 102	A77	01	24		A79	
5	# 103	D	A78	02	25			
6	Apt. 103	D		26		A78	
7	# 104	A77	04	27		A79	
8	# 105	A79	02	28		A77	
9	# 201			29			
10	# 202	A78	03	30		A79	
11	# 204	A79	01	31		A78	
12	# 205	A77	03	32		A77	
13	#301			33			
14	#302	A79	04	34		A79	
15	#303	A78	01	35		A78	
16	#304	A77	02	36		A77	
17				37		A78	
18		A79		38			
19		A78		39		A77	
20		A77		40		A79	

Footnotes:	Contact Person:
	Title:
	Telephone:

Resolve all Missing Unit Designations. Resolve all Duplicate Unit Designations.	FR Code							
	FR Initials							
	Month/Year							

BSAID: 0002
ESD within segment: 200408

Date Printed: 05/09/04
Page: 1 of 1

A-4 Sample Preparation Materials

Current Population Survey TP66

U.S. Bureau of Labor Statistics and U.S. Census Bureau

Illustration 3. Unit/Permit Listing Sheet, 11-3 (Blank)

FORM 11-3   U.S. DEPARTMENT OF COMMERCE
 (6-11-2001) Economics and Statistics Administration
               U.S. CENSUS BUREAU
      UNIT/PERMIT LISTING SHEET
	RO	PSU	Segment		Survey name	

Type of segment   **Expected no. of units**

Address   **Permit office name**

Post office	State	ZIP Code	Urban/Rural	Permit date of issue ----------------------- Year	Month	Day	PAL sequence/line no.

County   **Permit number** (or **BSAID** or **SPID/GQID**)   **Combined address**

PAL keyed remarks

Line No. (1)	Unit Designation (or apartment number)	Sample Desig.	Serial Number	Remarks (Reason and date for changes)
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				

19. Multi-Units	20. Listed and Updated					
Name of Complex:	FR Code					
Contact Person:	FR Initials					
Title:	Month/Year					
Phone No.	Total No. of Units					
Footnotes						

ESD: 200411   Date Printed: 08/13/04   Sheet __1__ of __1__ Sheets

Illustration 4. Incomplete Address Locator Actions Form, NPC 1138

NPC-1138 (March 19, 2004)	U.S. Department of Commerce Economics and Statistics Administration U.S. Census Bureau

<div align="center">

2000 SAMPLE REDESIGN

INCOMPLETE ADDRESS LOCATOR ACTIONS

</div>

SECTION I – IDENTIFICATION

Survey	Sample	County	BSAID	Block	MAFID
Sample Date	Tract/BNA	RO	PSU	Segment	Serial#

SECTION II - NPC LOCATOR REVIEW/RESULTS

1. Address AFTER NPC Review: Expected
    Name: # units:
    Address:
    Post Office: ST: ZIP:

2. Research

[ ] Future File:

[ ] MAF Browser:

[ ] Power Finder 2000

[ ] Internet

[ ] Zip +4:

[ ] Directory Assistance

[ ] Post Office:

[ ] Outcome Code:
    (NPC use only)

SECTION III - ACTIONS THE RO/FR MUST PERFORM

To locate the sample unit:
    [ ] Use Locator Materials Enclosed
    [ ] Refer to attached list..
    [ ] Use Multi-Unit Listing Sheet Enclosed.
    [ ] Use Multi-Unit Listing Sheet, previously sent to you for:
            Sample: Sample Date:
    [ ] Other – Specify:

REMARKS/COMMENTS

Illustration 5. Unit/Permit Listing Sheet, 11-3 (Single unit in Permit Frame)

FORM 11-3 U.S. DEPARTMENT OF COMMERCE (6-11-2001) Economics and Statistics Administration U.S. CENSUS BUREAU **UNIT/PERMIT LISTING SHEET**	**RO** 30	**PSU** 48155	**Segment** 2001	**Survey name** CPS

	Type of segment PERMIT	**Expected no. of units** 1

Address 350 POPULATION PLACE	**Permit office name** WALKER

Post office WALKER	**State** TX	**ZIP Code** 77484	**Urban/Rural** RURAL	**Permit date of issue** ----------------------- **Year** 2000 | **Month** 10 | **Day** 01	**PAL sequence/line no.** 70035215/11

County FOARD COUNTY	**Permit number** (or **BSAID** or **SPID/GQID**) 0256	**Combined address** NO

PAL keyed remarks

Line No. (1)	Unit Designation (or apartment number)	Sample Desig.	Serial Number	Remarks (Reason and date for changes)
1		A79		
2		A79		
3		A79		
4		A79		
5		A79		
6		A79		
7		A79		
8		A79		
9		A79		
10		A79		
11		A79		
12		A79		
13		A79		
14		A79		
15		A79		

19. Multi-Units	20. Listed and Updated					
Name of Complex:	FR Code					
Contact Person:	FR Initials					
Title:	Month/Year					
Phone No.	Total No. of Units					

Footnotes

ESD: 200409   Date Printed: 06/10/04   Sheet __1__ of __1__ Sheets

Illustration 6. Unit/Permit Listing Sheet, 11-3 (Multi-unit in Permit Frame)

FORM 11-3 (6-11-2001)	U.S. DEPARTMENT OF COMMERCE Economics and Statistics Administration U.S. CENSUS BUREAU	RO 23	PSU 42101	Segment 5001	Survey name CPS
	UNIT/PERMIT LISTING SHEET		Type of segment PERMIT		Expected no. of units 71

Address 2100 SURVEY ST				Permit office name PHILADELPHIA	

Post office PHILADELPHIA	State PA	ZIP Code 19001	Urban/Rural URBAN	Permit date of issue ----------------------- Year 2002	Month 07	Day 21	PAL sequence/line no. 45639851/0001

County PHILADELPHIA COUNTY	Permit number (or BSAID orSPID/GQID) 78126	Combined address NO

PAL keyed remarks

Line No. (1)	Unit Designation (or apartment number)	Sample Desig.	Serial Number	Remarks (Reason and date for changes)
1		A78	01	
2		A78	02	
3		A78	03	
4		A78	04	
5		A79	01	
6		A79	02	
7		A79	03	
8		A78		
9		A78		
10		A78		
11		A78		
12		A79		
13		A79		
14		A79		
15		A78		

19. Multi-Units		20. Listed and Updated						
Name of Complex:		FR Code						
Contact Person:		FR Initials						
Title:		Month/Year						
Phone No.		Total No. of Units						

Footnotes

ESD: 200404   Date Printed: 01/09/04   Sheet __1__ of __1__ Sheets

Illustration 7. Permit Sketch Map, Form
11-187

FORM 11-187 U.S. DEPARTMENT OF COMMERCE (10-15-02) Economics and Statistics Administration U.S. CENSUS BUREAU	1. RO **27**	2. PSU **06703**	3. Permit month/year **07/2005**	4. Sequence number **69003100**

PERMIT SKETCH MAP

PERMIT ADDRESS LIST OPERATION

5. Permit Office name **Alameda County**			
6. Locality (Post Office) **Oakland City**		7. County **Alameda**	
8. State **CA**	9. ZIP Code **20709**	10. Permit number **47325**	11. PAL line **32**

ATTENTION REGIONAL OFFICE ▶ Keep both copies in the Regional Office files

12. U/R **Rural**	13. Field Representative name **Sue Hood**	Code **M5**

Quimby Street

X ←—— 1 mile ——→

Hill Rd.

Hwy 64

N ↑

Notes

Red brick house, garage on left side.

U.S. GOVERNMENT PRINTING OFFICE 1992-750-111/40529

Appendix B.
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. The most recent sampling maintenance reductions were implemented in 1999 and 2003.

These maintenance reductions are different from changes to the base CPS sample size resulting from modifications to 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.

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 so that dropped cases will not be interviewed.

Housing 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 maintenance reduction. The post-reduction 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-level 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 that 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 the sample to reduce. However, a large number of reduction groups can lead to imbalances in the distribution of sample cuts across PSUs, since small PSUs may not have enough sample to have all reduction groups represented.

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/non-self-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 about which PSUs will actually have permit samples over the life of the design. The sort sequence is:

1. Stratification PSU.

2. Basic PSU Component (BPC)

3. Original 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 the 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 2000 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 further 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 it is gradually phased in with each incoming rotation group to minimize the effect on survey estimates and reliability and to prevent sudden changes to the 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.

Appendix C.
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. In addition to the CPS, 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 civilian noninstitutionalized population of the United States and the 50 states (including the District of Columbia). The projections for the country are distributed by demographic characteristics in two ways: (1) age, sex, and race, and (2) age, sex, and Hispanic origin.[1] The projections for the states are distributed by race (Black only and all other race groups combined), age (0–15, 16–44, and 45 and over), and sex. 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 the 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 measurements of 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.

This appendix seeks first to present this framework systematically, 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 the total population from a census date via the demographic components of change and explains how each of the requisite inputs to the mathematical framework is measured. This analysis is presented separately for the total population of the United States and its distribution by demographic characteristics and by state of residence by demographic characteristics. The final two sections are, 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. The final section, "Procedural Revisions," contains 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, which are essential to understanding the derivation of census-based population controls for surveys but not necessarily prevalent in the literature on survey methodology.

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.

A *census-level* population estimate or projection is an estimate or projection of population change that can be linked to a census count.

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 people defined to be on active duty by one of the five branches of the Armed Forces, as well as people 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

---

[1] Hispanics may be any race.

demographic components often cited in the literature consist of births, deaths, and net migration. This appendix also refers to changes in the institutional and active-duty Armed Forces populations, which affect the CPS control universe.

The *CPS control universe* is characterized by two attributes: (1) restriction to the civilian noninstitutionalized population, and (2) modification of census data by race. Each of these defining concepts appears separately in this glossary.

Emigration is the departure of a person from a country of residence, in this case the United States. In the present context, it refers to people legally resident in the United States but is not confined to legally permanent residents (immigrants) who make their usual place of residence outside of the United States.

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 data). Population estimates discussed in this appendix stipulate an enumerated base population, coupled with estimates of population change from the enumeration date of the base population to the date of the estimate.

The *institutional population* refers to a population universe consisting of inmates or residents 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 population 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 people intending to live or make their usual residence abroad, in Puerto Rico, or the outlying areas.

Legal permanent residents are people 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 race describes the census population after the definition of race has been aligned with definitions from other administrative sources.

The *natural increase* of a population over a time interval is the number of births minus the number of deaths during the interval.

Population *projections* are population figures relying on modeled or assumed values for some or all of their components, and 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 2001 include some count resolution corrections to Census 2000 not included in original census publications.

A population *universe* is the population represented in a data collection, in this case, the CPS. It is determined by a set of rules or criteria defining inclusion and exclusion of people 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 as the civilian noninstitutional population residing in the United States.

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 close to the 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 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 could be 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.

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 useful.

POPULATION UNIVERSE FOR CPS CONTROLS

In the concept of "population universe," as used in this appendix, are found not only the rules specifying who is 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. This section considers three population universes: the resident population universe defined by Census 2000, 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, the extensive detail that it provides on the geographic distribution and characteristics of the U.S. population and the legitimacy accorded it by the U.S. Constitution.

The universe defined by Census 2000 is the U.S. resident population, including people 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, the Virgin Islands of the United States, and the 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. People 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 counted as 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 for 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 latest census (2000) if this census had been held on the reference date of the estimates. However, subgroups within this universe are distinct from those in the census universe due to race recoding, which does not affect the total population.

The estimates differ from the census in their definition of race, with categories redefined to be consistent with other administrative-data sources. Specifically, people enumerated in Census 2000 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 people (18.5 million nationally) were Hispanics who gave a Hispanic-origin response to the race question. Publications from Census 2000 do not incorporate this modification, while all official population estimates do incorporate it.

The universe underlying the CPS sample is confined to the civilian noninstitutionalized population. Thus, independent population controls are sought for this population to define the CPS control universe. The previously mentioned alternative definitions of race, associated with the universe for official estimates, is carried over to the CPS control universe. Three additional changes are incorporated, which distinguish the CPS control universe (civilian noninstitutional population) from the universe for official Census Bureau estimates (resident population).

1. The CPS control universe excludes active-duty Armed Forces personnel, including those stationed within the United States, while these individuals 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 people 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 institutionalized 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 difference affects estimation procedures at the state level, but not at the national level.

CALCULATION OF POPULATION PROJECTIONS FOR THE CPS UNIVERSE

This section explains 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, age, sex, and race.

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 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 that the population at some reference date is 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

$$PRES_{t1} = PRES_{t0} + B - D + NM \qquad (1)$$

where:

$PRES_{t1}$ = Resident population at time t_1 (the reference date),
$PRES_{t0}$ = Resident population at time t_0 (the base date),
B = births to U.S. resident women, from time t_0 to time t_1,
D = deaths of U.S. residents, from time t_0 to time t_1,

NM = net migration to the United States (migration into the United States minus migration out of the United States) from time t_0 to time t_1.

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, deaths in the civilian noninstitutional population 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:

$$PCN_{t1} = PCN_{t0} + B - CND + (IM - NRAF - NEI) \qquad (2)$$

where:

PCN_{t1} = civilian noninstitutionalized population, time t_1,
PCN_{t0} = civilian noninstitutionalized population, time t_0,
B = births to the U.S. resident population, time t_0 to time t_1,
CND = deaths of the civilian noninstitutionalized population, time t_0 to time t_1,
IM = net civilian migration, time t_0 to time t_1,
$NRAF$ = net recruits (inductions minus discharges) to the U.S. Armed Forces from the domestic civilian population, from time t_0 to time t_1,
NEI = net admissions (enrollments minus discharges) to institutions, from time t_0 to time t_1.

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 institutionalized 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 t_1 and t_0, 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.

$$NEI = PCI_{t1} - PCI_{t0} + DI \qquad (5)$$

where PCI_{t1} and PCI_{t0} represent the civilian institutionalized population at times t_1 and time t_0, 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 measuring institutionalized inmate deaths or deaths of the Armed Forces members residing in the United States.

where IM = net international migration in the time interval (t_0, t_1).

At this point, we restate equation (2) to incorporate equations (3), (4), and (5), and simplify. The resulting equation, which follows, is the operational equation for population change in the civilian noninstitutionalized population universe.

$$PCN_{t1} = PCN_{t0} + B - D + IM - (WWAF_{t1} - WWAF_{t0} + DAFO - NRO) - (PCI_{t1} - PCI_{t0}) \quad (6)$$

Equation (6) identifies all the components that must be separately measured or projected to update the total civilian noninstitutionalized 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 (6) 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 (6) 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 and residential definition, and this requirement is a major aspect of the way they are measured. A description of the individual inputs to equation (6), which applies to the CPS control universe, and their data sources follows.

The census base population (PCN_{t0}). While any base population (PCN_{t0}, in equation [6]), whether a count or an estimate, can be the basis of 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. For the survey controls in this decade, Census 2000 results were not adjusted for undercount. The census base population includes minor revisions arising from count resolution corrections. These corrections arise from the tabulation of data from the enumerated population. As of May 21, 2004, count resolution corrections incorporated in the estimates base population

amounted to a gain of 2,697 people from the originally published count. In order to apply equation (6), which is derived from the balancing equation for the civilian noninstitutionalized 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:

$$PCN_0 = R_0 - RAF_0 - PCI_0 \quad (7)$$

where R_0 = the enumerated resident population, RAF_0 = the Armed Forces resident in the United States on the census date, and PCI_0 = the civilian institutionalized 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 (6). In practice, this is what occurs when the monthly CPS population controls are produced.

Births and deaths (B, D). In estimating total births and deaths of the resident population (B and D, in equation [6]), we assume the population universe for vital statistics matches 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 this 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 2004 CPS control series were for calendar year 2001. Final birth and death data are summarized in NCHS publications (see Martin et al., 2002, and Arias et al., 2003). Monthly births and deaths for calendar year(s) up to 2 years before the CPS reference dates (e.g., 2002 for 2004 controls) are based on provisional estimates by NCHS (see Hamilton et al. and Kochanek et al., 2004). 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.

A concern exists relative to the consistency of the NCHS vital statistics universe with the census resident universe and the CPS control universe derived from it.

Birth and death statistics obtained from NCHS exclude 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.

This source of bias, which is assumed to be minor, affects the accounting of total population. Other 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.

Net International Movement. The net international component combines three parts: (1) net migration of the foreign-born, (2) emigration of natives, and (3) net movement from Puerto Rico to the United States. Data from the American Community Survey (ACS) is the basis for estimating the level of net migration of the foreign-born between 2000 to 2001 and 2001 to 2002 because it provided annually updated data based on a larger sample than the CPS. After determining the level of migration for the foreign-born (the net difference between two time periods for the foreign-born population), we accounted for deaths to the entire foreign-born population during the periods of interest to arrive at the final estimate of net migration of the foreign-born. We applied the Census 2000 age-sex-race-Hispanic-origin county-level distribution of the non-citizen foreign-born population who entered in 1995 or later to the national-level estimate of net migration of the foreign-born. The remaining two parts of the net international migration component, the net movement from Puerto Rico to the United States and the emigration of natives, were produced in similar ways. For both parts, we do not have current annually updated information. Therefore, we used the levels of movement produced from change in the decennial censuses.[2] For both parts, we applied the age-sex-race-Hispanic origin-county distributions from Census 2000 that were most similar to the population of interest. For the net movement from Puerto Rico, the underlying distribution was based on those who indicated that their place of birth was Puerto Rico and who had entered the United States in 1995 or later. We assumed that natives who emigrated were likely to have the same distributions as natives who currently reside in the United States. Therefore, the characteristics of natives who emigrated were assumed to be the same age-sex-race-Hispanic-origin county-level distribution as natives residing in the 50 states and the District of Columbia in Census 2000.

_____

[2] For more information on the estimate of 11,133 for the net movement from Puerto Rico, see Christenson. For information on estimates of native emigration, see Gibbs, et al.

Once the net migration of the foreign-born, net movement from Puerto Rico, and emigration of natives were estimated, all three parts were combined to estimate a final net international migration component. Adding up the parts of net international migration, the net migration of the foreign-born (1,300,000) plus the net migration from Puerto Rico (+11,133) minus the net emigration of natives (−18,012) yielded 1,293,121 as the annual estimate for net international migration for the 2000–2001 year and the 2001–2002 year.

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

$$(\text{WWAF}_{t1} - \text{WWAF}_{t0} + \text{DRAF} + \text{DAFO} - \text{NRO})$$

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 one 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 on active duty.

The last three components of net recruits to the Armed Forces from the U.S. civilian population, deaths in the resident Armed Forces (DRAF), deaths in 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 populations, 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 individuals 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 people with home of record outside the U.S. yield the series for net recruits to the Armed Forces from overseas.

Change in the institutionalized population (CIP$_{t1}$ - CIP$_{t0}$). The change in the civilian population residing in institutions is currently held constant, equal to the civilian institutionalized group quarters population obtained from Census 2000.

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. Thirty-one categories of race and two classes of ethnic origin (Hispanic, non-Hispanic) are required, with no cross-classifications of race and Hispanic origin. The population projection program now uses the full cross-classification of age by sex and by race and by Hispanic origin in its monthly series, with 101 single-year age categories (single years from 0 to 99, and 100 and over), two sex categories (male, female), 31 race categories (White; Black; American Indian, Eskimo, and Aleut; Asian; Native Hawaiian and Pacific Islander; and all multiple-race groupings), and two Hispanic-origin categories (not Hispanic, Hispanic).[3] The resulting matrix has 12,524 cells (101 x 2 x 31 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 124 cells, defined by variables not naturally changing over time (2 values of sex by 31 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 (6). Similarly, the derivation of the civilian noninstitutionalized base population from the census resident population follows equation (7).

Distribution of Census Base and Components of Change by Age, Sex, Race, and Hispanic Origin

The 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.

Census Base Population by Age, Sex, Race and Hispanic Origin: Modification of the Census Race Distribution

The race modification conforms to the Office of Management and Budget's (OMB) 1997 revised standards for collecting and presenting data on race and ethnicity. The revised OMB standards identified five race categories: White, Black or African American, American Indian and Alaska Native, Asian, and Native Hawaiian and Other Pacific Islander. Additionally, the OMB recommended that respondents be given the option of marking or selecting one or more races to indicate their racial identity. For respondents unable to identify with any of the five race categories, the OMB approved including a sixth category, "Some other race," on the Census 2000 questionnaire.

No modification was necessary for responses indicating only an OMB race alone or in combination with another race. However, about 18.5 million people checked "Some other race" alone or in combination with another race. These people were primarily Hispanic and many wrote in their place of origin or Hispanic-origin type (such as Mexican or Puerto Rican) as their race. For purposes of estimates production, responses of "Some other race" alone were modified by blanking the "Some other race" response and imputing an OMB race alone or in combination with another race response. The responses were imputed from a donor who matched on response to the question on Hispanic origin. Responses of both "Some other race" and an OMB race were modified by blanking the "Some other race" response and keeping the OMB race response.

---

[3] Throughout this appendix, "American Indian" refers to the aggregate of American Indian, Eskimo, and Aleut; "API" refers to Asian and Pacific Islander.

The resulting race categories (White, Black, American Indian and Alaska Native, Asian, and Native Hawaiian and Other Pacific Islander) conform with OMB's 1997 revised standards for the collection of data on race and ethnicity and are more consistent with the race categories in other administrative sources, such as vital statistics.

Distribution of births by sex, race, and Hispanic origin. Data on registered births to U.S. residents by birth month, sex of child, and race and Hispanic origin of mother and father were supplied by the National Center for Health Statistics (NCHS). For the 2004 CPS controls, final data were available through 2001 and preliminary data were available for 2002.

Registered births to U.S. resident women were estimated from data supplied by NCHS. At present, NCHS continues to collect birth certificate data using the 1977 race standards of White, Black, American Indian, Eskimo or Aleut, and Asian or Pacific Islander, under the "mark one race" scenario. To produce post-2000 population estimates, it was necessary to develop birth data that coincided with the new race categories: White, Black, American Indian and Alaska Native, Asian, and Native Hawaiian and other Pacific Islander, under the "mark one or more races" scenario. Because of this inconsistency in data on race, it was necessry to model the full 31 possible single- and multiple-race combinations. For a detailed description of this modeling process, see Smith and Jones, 2003.

Data on births by birth month, sex, and race and Hispanic origin of the mother and father are based on final micro-data files for calendar year 2001 from the NCHS registration system. The model was based on information from Census 2000 on race and Hispanic origin reporting within households for the age zero (under 1 year of age) population and their parent(s). First, the NCHS births were tabulated for each of the combinations of parents' race and Hispanic origin. These births by parents' race and Hispanic origin were then distributed according to the Census 2000 race and Hispanic origin distribution for the age zero population for the matching combination of parents' race and Hispanic origin. Race and Hispanic origin modeling was done separately for mother-only and two-parent households.

To estimate the distribution of births for calendar year 2002, data on preliminary 2002 births received from NCHS from their 90-percent sample of final births were distributed according to the 2001 births by birth month, sex and modeled race and Hispanic origin.

To estimate the distribution of births by race and Hispanic origin of the mother for the first half of 2003, age-specific birth rates centered on July 1, 2002, for women in each race group and for women of Hispanic origin were applied to preliminary estimates of the number of resident women in the specified age groups.

Births through the remaining period through the end of 2004 were then projected from the births estimated above.

Deaths by age, sex, race, and Hispanic origin. Data on registered deaths of U.S. residents by death month, age, sex, race, and Hispanic origin were supplied by NCHS. Final data were available through 2001 and preliminary data were available for 2002.

It was again necessary to model the race distribution because death certificates ask for the race of the deceased using only four race categories (1977 race categories for White, Black, American Indian, Eskimo or Aleut, Asian or Pacific Islander). Separate death rates were calculated for the 1977 race categories by age, sex, and Hispanic origin. Rates were constructed using the 1998 mortality[4] and 1998 population estimates.[5] These rates were then applied to the July 2001 population in the 31 modified race categories from the Vintage 2002 estimates. Death rates for the White, the Black, the American Indian, Eskimo or Aleut, and the Asian and Pacific Islander groups were applied to the corresponding White alone, Black alone, American Indian and Alaska Native alone, Asian alone, and Native Hawaiian and Other Pacific Islander alone populations. The Asian and Pacific Islander death rate was applied to both the Asian alone population and the Native Hawaiian and Other Pacific Islander alone population. Multiple-race deaths were estimated as the difference between total 2001 deaths as reported by NCHS and the sum of deaths estimated for the single-race groups. Consequently, a constant death rate was applied to each of the 26 multiple- race groups.

The 2002 deaths were distributed using 2001 deaths by modeled race, death month, age, and sex and were controlled to the 2002 preliminary deaths from NCHS by Hispanic origin.

To estimate the distribution of deaths by race and Hispanic origin for the first half of 2003, projected age-specific mortality rates for July 1, 2002, were applied to preliminary 2003 estimates of the population by single year of age, sex, race, and Hispanic origin.

Deaths by race and Hispanic origin for the period beyond the first half of 2003 through the end of 2004 are projected from the deaths produced above.

---

[4] The race distribution for the 1998 deaths as set in the processing of the national estimate is used here because it adjusts for NCHS/Census race inconsistencies. In the production of national population estimates in the 1990s, preliminary deaths to the American Indian and Asian and Pacific Islander populations by sex were projected using life tables, with proportional adjustment to sum to the other races total. Hispanic origin deaths by sex and race were estimated for all years using life tables applied to a distribution of the Hispanic population by age, sex, and race.

[5] The 1998 population estimate from the vintage 2000 population estimates.

International migration by age, sex, race, and Hispanic origin. For the purposes of the estimates program, the net international migration component consists of three parts: (1) net migration of the foreign-born, (2) emigration of natives, and (3) net movement from Puerto Rico to the United States.

Net migration of the foreign-born component.
Although a component basis was used throughout most of the 1990s, the findings from an extensive evaluation showed that the use of administrative sources for developing migration components by migrant status for the foreign-born was problematic.[6] Therefore, we assumed that the net difference between two time periods for the foreign-born population using the same data source would be attributable to net migration. To maximize the use of available data, the American Community Survey (ACS) data for 2000, 2001, and 2002 was the basis for the net migration between 2000 to 2001, and 2001 to 2002.[7] The calculation of the change in the size of the foreign-born population, must be adjusted for deaths in the population during the period of interest.

Characteristics for net migrants were imputed by taking the estimated net migration of the foreign-born and then redistributing that total to the Census 2000 distribution by single year of age (0–115+), sex-race (31 groups), Hispanic origin and county distribution of the noncitizen, foreign-born who reported that they entered in 1995 or later. The rationale was that the most recent immigrants in Census 2000 would most closely parallel the distribution of the foreign-born in the ACS, by age-sex-race-Hispanic origin and county.

Net migration from Puerto Rico component. Since no data set would easily provide updated information on movement from Puerto Rico, there was no way to update the estimate that was based on the average movement of the 1990s.[8]

Characteristics were imputed for the net migration from Puerto Rico using the age-sex-race-Hispanic origin-county distribution of those who indicated that their place of birth was Puerto Rico and who had entered the United States in 1995 or later. The underlying assumption was that the characteristics of those who indicated that they were born in Puerto Rico and had migrated since 1995 would parallel the distribution of recent migrants from Puerto Rico.

Net native emigration component. Since no current data set would easily provide updated information on native emigration, there was no systematic way to update the estimate.[9] We assumed that natives who emigrated were likely to have the same distributions as natives who currently reside in the United States. Therefore, the level was distributed to the same age-sex-race-Hispanic origin-county proportions as those of the native population (born in the 50 states and DC) living in the 50 states and the District of Columbia in Census 2000.

Combined net international migration component.
The estimates of net migration of the foreign-born, net movement from Puerto Rico, and net emigration of natives were combined to create the estimate of the total net international migration component.

The net international migration component was kept constant for years 2000 to 2003. The change in the foreign-born population from the 2000 to 2001 period and from the 2001 to 2002 period was not statistically significant. This finding, in conjunction with limited time for additional research, led to the decision to keep the estimate constant for the projected years 2004 and 2005. The additional parts of the net international migration component are kept constant because they are currently measured only once per decade.

Migration of Armed Forces by age, sex, race, and Hispanic origin. Estimation of demographic details for the remaining components of change requires estimation of the details for the Armed Forces residing overseas and in the United States. These data are 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/Hispanic 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 people residing on board ships to conform to census definitions, which is accomplished through a special tabulation (also provided by Defense Manpower Data Center) of people assigned to sea duty. These data 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-Hispanic-origin distribution must also be adapted. The Armed Forces "race-ethnic" categories supplied by the Defense Manpower Data Center

---

[6] An overview of the Demographic Analysis-Population Estimates (DAPE) project along with a summary of findings is in Deardorff, K. E. and L. Blumerman. Mulder, T., et.al., gives an overview of the methods used throughout the 1990s' estimates series.

[7] Throughout, "ACS" will refer to the ACS combined with the Supplementary Surveys for the corresponding years.

[8] See Christenson, M., for an explanation of how the annual level of 11,133 was derived.

[9] See Gibbs, et.al., for an explanation of how the estimated annual level of 18.000 native emigrants was derived. Because of rounding, the sum of the detailed age, sex, race, and Hispanic origin estimates used in the actual estimates process is 18,012.

treat race and Hispanic origin as a single variable, with the Hispanic component of Blacks and Whites included under "Hispanic," and the Hispanic components of the American Indian and the API populations assumed to be nonexistent. A residual "other race" category is also caculated. The method of converting this distribution to be consistent with Census-modified race categories employs, for each age-sex category, the Census 2000 race distribution of the total population (military and civilian) to supply all race information missing in the Armed Forces race-ethnic categories. Hispanic origin is thus prorated to the White and Black populations according to the total race-modified population of each age-sex category in 2000; American Indian and API are similarly prorated to Hispanic and non-Hispanic categories. 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. This adaptation is robust; it requires imputing the distributions of only small race and Hispanic-origin categories.

The two 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 Hispanic origin based on data from an aggregate of overseas censuses, in which Puerto Rico dominates numerically. The age and sex distribution follows the worldwide Armed Forces.

The civilian institutionalized population by age, sex, race, and Hispanic origin. The fundamental source for the distribution of the civilian institutionalized population by age, sex, race, and Hispanic origin is a Census 2000 tabulation of institutional population by age, sex, race (modified), Hispanic origin, and type of institution. The last variable has four categories: nursing home, correctional facility, juvenile facility, and a small residual. Data for the resident armed forces are also obtained from Census 2000 and updated with data from the Department of Defense. The size of the civilian population is obtained by subtracting the resident armed forces from the resident population. The civilian noninstitutionalized population is obtained by subtracting the institutionalized population from the civilian population.

Population Controls for States

State coverage adjustment for the CPS requires a distribution of the national civilian noninstitutionalized population by age, sex, and race to the states. The second-stage weighting procedure requires a disaggregation only by age and sex by state. This distribution is determined by a linear extrapolation through two population estimates for each characteristic group 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 for each age, sex, race and Hispanic-origin group by proportional adjustment. For example, the state distribution for June 1, 2003, was determined by extrapolation along a straight line determined by two data points (July 1, 2001, and July 1, 2002) for each state. The resulting distribution for June 1, 2003, was then proportionately adjusted to the national total for each age, sex and race group, 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 forms 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 for 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 noninstitutionalized population but to the household population under 65 years of age and 65 years of age and over. Estimates procedures for the population under 65 and the population 65 and older are similar, but differ in that they employ a different data source for developing mea- sures of internal migration. Furthermore, while national-level population estimates are produced for the first of each month, state estimates are produced at one-year intervals, with July 1 reference dates.

The household population under 65 years of age. The procedure for estimating the nongroup quarters population under 65 is analogous to the national-level procedures for the total population, with the following major differences.

1. Because the population is restricted to individuals under 65 years of age, the number of people advancing to age 65 must be subtracted, along with deaths of those 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 people under 65 and people 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. The immigration total and the emigration of native-born residents are distributed on the basis of the various segments of the population enumerated in Census 2000.

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 net 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 ZIP-CRS program (Sater, 1994) 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.

A detailed explanation of the use of IRS data and other facets of the subnational estimates can be found at <http://www.census.gov/popest/estimates.php>.

The population 65 years of age and over. For the population ages 65 and over, the approach is similar to that used for the population under age 65. The population ages 65 and over in Census 2000, is added to the population estimated to be turning age 65. This quantity, as discussed in the previous section, is subtracted from the base under-age-65 population. The Census Bureau subtracts deaths and adds net immigration, occurring for this same age group. Migration is estimated using data on Medicare enrollees, which are available by county of residence from the U.S. Centers for Medicare and Medicaid Services. Because there is a direct incentive for eligible individuals to enroll in the Medicare program, the coverage rate for this source is high. To adapt this estimate to the Census 2000 level, the Medicare enrollment is adjusted to reflect the difference in coverage of the population 65 years of age and over.

Exclusion of the Armed Forces population and inmates of civilian institutions. Once the resident population by the required age, sex, and race groups has been estimated for each state, the population must be restricted to the civilian noninstitutionalized 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). 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.

PROCEDURAL REVISIONS

The process of producing estimates and projections, like most 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, reflecting the introduction of data from a new decennial census, which can be associated with new policies regarding population universe or estimating methods. However, revisions may occur at any time, depending on new information obtained regarding the components of population change, or the availability of resources to complete methodological research.

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 Census 2000 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

REFERENCES

Ahmed, B. and J. G. Robinson, (1994), **Estimates of Emigration of the Foreign-Born Population: 1980–1990**, U.S. Census Bureau, Population Division Technical Working Paper No. 9, December, 1994, <http://www.census.gov /population/www/documentation/twps0009 /twps0009.html>.

Arias, Elizabeth, Robert Anderson, Hsiang-Ching Kung, Sherry Murphy, and Kenneth Kochaneck,"Deaths: Final Data for 2001," **National Vital Statistics Reports**, Volume 52, Number 3, Division of Vital Statistics, National Center for Health Statistics, <http://www.cdc.gov/nchs /data/nvsr/nvsr52/nvsr52_03.pdf>.

Christenson, M., "Evaluating Components of International Migration: Migration Between Puerto Rico and the United States," Population Division Technical Working Paper No. 64, January 2002, <http://www.census.gov/population /www/documentation/twps0064.html>.

Deardoff, K.E. and L. Blumerman, "Evaluating Components of International Migration: Estimates of the Foreign- Born Population by Migrant Status in 2000," Population Division Technical Working Paper No. 58, December, 2001, <http:// www.census.gov/population/www/documentation /twps0058.html>.

Gibbs, J. G. Harper, M. Rubin, and H. Shin, "Evaluating Components of International Migration: Native-Born Emigrants," Population Division Technical Working Paper No. 63, January, 2003, <http://www.census.gov/population /http://www/documentation/twps0063.html>.

Hamilton, Brady E., Joyce A. Martin, and Paul D. Sutton (2003), "Births: Preliminary Data for 2002," **National Vital Statistics Reports**, Volume 51, Number 11, Division of Vital Statistics, National Center for Health Statistics, <http://www.cdc.gov/nchs/data/nvsr/nvsr51 /nvsr51_11.pdf>.

Kochaneck, Kenneth; and Betty Smith (2004), "Deaths: Preliminary Data for 2002," **National Vital Statistics Reports**, Volume 52, Number 13, Division of Vital Statistics, National Center for Health Statistics, <http://www.cdc .gov/nchs/data/nvsr/nvsr52/nvsr52_13.pdf>.

Martin, Joyce A., Brady E. Hamilton, Stephanie Ventura, Fay Menacker, Melissa M. Park, and Paul D. Sutton (2002), "Births: Final Data for 2001," **National Vital Statistics Reports**, Volume 51, Number 2, Division of Vital Statistics, National Center for Health Statistics, <http://www .cdc.gov/nchs/data/nvsr/nvsr51/nvsr51_02.pdf>.

Mulder, T., Frederick Hollmann, Lisa Lollock, Rachel Cassidy, Joseph Costanzo, and Josephine Baker, "U.S. Census Bureau Measurement of Net International Migration to the United States: 1990 to 2000," Population Division Technical Working Paper No. 51, February, 2002, <http:// www.census.gov/population/www/documentation /twps0051.html>.

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, <http://www.census.gov/population/www/documentation /twps0007/twps0007.html>.

Smith, Amy Symens, and Nicholas A. Jones (2003), **Dealing With the Changing U.S. Racial Definitions: Producing Population Estimates Using Data With Limited Race Detail**. Paper presented at the 2003 meetings of the Population Association of America, Minneapolis, Minnesota, May, 2003.

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. Immigration and Naturalization Service (1997), **Statistical Yearbook of the Immigration and Naturalization Service,** 1996, Washington, DC: U.S. Government Printing Office.

Appendix D.
Organization and Training of the Data Collection Staff

INTRODUCTION

The data collection staff for all U.S. Census Bureau programs is directed through 12 regional offices (ROs) and 3 telephone centers. The ROs collect data using two modes: CAPI (computer-assisted personal interviewing) and CATI (computer-assisted telephone 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), located in Jeffersonville, Indiana.

ORGANIZATION OF REGIONAL OFFICES/CATI FACILITIES

The staffs of the ROs and CATI facilities carry out the Census Bureau's field data collection programs for both sample surveys and censuses. Currently, the ROs supervise over 6,500 part-time and intermittent field representatives (FRs) who work on continuing current programs and one-time surveys. Approximately 1,900 of these FRs work on the Current Population Survey (CPS). When a census is being taken, the field staff increases dramatically.

The location of the ROs and the boundaries of their responsibilities are displayed in Figure D–1. RO areas were originally defined to evenly distribute the office workloads for all programs. Table D–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 two or three CPS program supervisors on staff. The program supervisors have a staff of two to four office clerks working essentially full-time. Most of the clerks are full-time federal employees 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 20–35 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 nonresponse and quality assurance reinterview 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, 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 leadership responsibilities, 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) is 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 or near 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 or fewer hours 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

2000 Decennial and Survey Boundaries
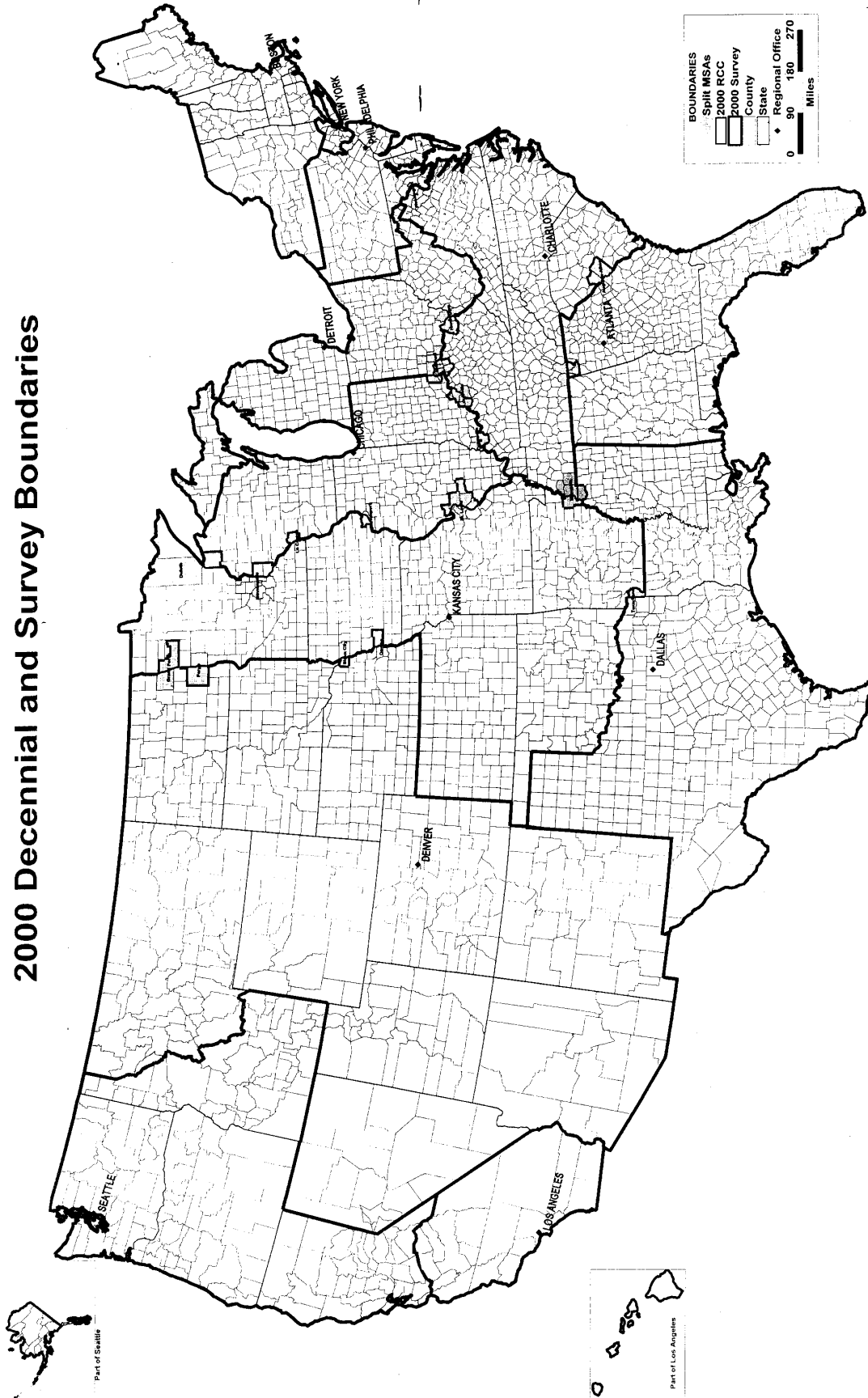
Table D–1. **Average Monthly Workload by Regional Office: 2004**

Regional office	Base workload	CATI workload	Total workload	Percent
Total	**67,383**	**6,839**	**74,222**	**100.00**
Boston	8,813	844	9,657	13.01
New York	2,934	275	3,209	4.32
Philadelphia	6,274	620	6,894	9.29
Detroit	4,848	495	5,343	7.20
Chicago	4,486	507	4,993	6.73
Kansas City	6,519	657	7,176	9.67
Seattle.........................	5,251	676	5,927	7.99
Charlotte	5,401	541	5,942	8.01
Atlanta	4,927	455	5,382	7.25
Dallas	4,418	391	4,809	6.48
Denver	9,763	1,029	10,792	14.54
Los Angeles	3,749	349	4,098	5.52

to starting his/her assignment. The initial training program consists of up to 24 hours of pre-classroom home study, 3.5 to 4.5 days of classroom training (dependent upon the trainee's interviewing 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 FR's 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.

Trainees receive extensive training on interviewing skills, such as how to handle noninterview situations, how to probe for information, ask questions as worded, and implement face-to-face and telephone techniques. 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, compared with 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 their performance.

Each CPS supervisor is responsible for developing each employee to his/her fullest potential. Employee development requires 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, accurate payrolls submitted on time, deadlines met, reports to team leaders, training sessions, and administrative responsibilities.

The most useful tool to help supervisors evaluate the FRs performance is the CPS 11–39 form. This report is generated monthly and is produced using data from the Regional Office Survey Control System (ROSCO) and Cost and Response Management Network (CARMN) systems.

This report provides the supervisor 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 that NPC 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 ROs establish their own standards, they 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 is actually interviewed. It is calculated by dividing the total number of interviewed households by the sum of the number of interviewed households, the number of refusals, noncontacts, and noninterviewed households for other reasons, including temporary refusals. (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 include interview time and travel time. The measure 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 FR's 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 assignments that cover multiple PSUs.

- Large number of inherited or confirmed refusals.

- Working part of another FR's assignment.

- Inordinate number of temporarily absent cases.

- High percentage of Type B/C noninterviews that decrease the base for nonresponse rate.

- Other substantial changes in normal assignment conditions.

- Personal visits/phones.

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 FR's attitude toward the job, use of the computer, and the extent to which FRs apply CPS concepts and procedures during actual work situations. There are three types of observations:

1. Initial observations (N1, N2, N3)

2. General performance review (GPR)

3. Special needs (SN)

Initial observations are an extension of the initial classroom training for new hires and provide 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 are an integral part of the initial training given to all FRs. A 2-day initial observation (N1) is scheduled during the FR's first CPS CAPI assignment. A second 1-day initial observation (N2) is scheduled during the FR's second CPS CAPI assignment. A third 1-day initial observation (N3) is scheduled during the FR's 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 FRs. Each FR is regularly observed at least once a year.

Special-needs observations are made when an FR has 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 an FR has a high Type A noninterview rate, a high minutes-per-case rate, a failure on reinterview, an unsatisfactory evaluation on a previous observation, made 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 notes 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 also stresses vehicular and personal safety practices.

The observer reviews the FR's 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 an 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, and with guidance from the Human Resources Division, 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, provide specific suggestions on ways to improve performance, alert the FR to actions that will be taken by the survey supervisor to assist the FR to improve his/her performance, and notify the FR that he/she is subject to separation if the work does not show improvement in the allotted time. Once placed on a PIP, the regional office provides performance feedback during a specified time period (usually 30 days, 60 days, and 90 days).

Appendix E.
Reinterview: Design and Methodology

INTRODUCTION

A continuing program of reinterviews on subsamples of Current Population Survey (CPS) households is carried out every month. The reinterview is a second, independent interview of a household conducted by a different interviewer. The CPS reinterview program has been in place since 1954. Reinterviewing for CPS serves two main quality control (QC) purposes: to monitor the work of the field representatives (FRs) and to evaluate data quality via the measurement of response error (RE) where all the labor force questions are repeated. The reinterview program is a major tool for limiting the occurrence of nonsampling error and is a critical part of the CPS program.

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 samples, 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 the 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. It provides a means of limiting nonsampling error, as described in Chapter 15. The RE reinterview currently estimates simple response variance, a measure of 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 previous 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, 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 the sample. This rule applies to both the RE and QC samples. Any household contacted for reinterview is ineligible for reinterview selection during its remaining months in the 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 (ROs). Only households that 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 uses 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. A falsification study determined that a relationship exists between tenure and both frequency of falsification and the percentage of assignments 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. 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. 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.[1] 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 the reional 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.[2] 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 were asked in the original interview.[3] Currently, no reconciliation is conducted.[4]

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 resign or are terminated.

Results from the RE reinterview 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., J. A. Dewey, and G. Weller, (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] Prior to October 2001, QC reinterviews asked the full set of labor force questions for all eligible household members.

[2] Beginning in February 1998, the RE reinterview respondent rule was made the same as the QC reinterview respondent rule.
[3] Prior to October 2001, RE reinterviews asked the industry and occupation questions independently.
[4] Reconciled reinterview, which was used to estimate reponse bias and provide feedback for FR performance, was discontinued in January 1994.

Acronyms

ADS	Annual Demographic Supplement	NCES	National Center for Education Statistics
API	Asian and Pacific Islander	NCEUS	National Commission on Employment and Unemployment Statistics
ARMA	Autoregressing Moving Average		
ASEC	Annual Social and Economic Supplement	NCHS	National Center for Health Statistics
BLS	Bureau of Labor Statistics	NCI	National Cancer Institute
BPC	Basic PSU Components	NEA	National Endowment for the Arts
CAPE	Committee on Adjustment of Postcensal Estimates	NHIS	National Health Interview Survey
		NPC	National Processing Center
CAPI	Computer-Assisted Personal Interviewing	NRAF	Net Recruits to the Armed Forces
CATI	Computer-Assisted Telephone Interviewing	NRO	Net Recruits to the Armed Forces From Overseas
CCM	Civilian U.S. Citizens Migration		
CCO	CATI and CAPI overlap	NSR	Non-Self-Representing
CES	Current Employment Statistics	NTIA	National Telecommunications and Information Administration
CESEM	Current Employment Statistics Survey Employment		
		OB	Number of Births
CNP	Civilian Noninstitutional Population	OCSE	Office of Child Support Enforcement
CPS	Current Population Survey	OD	Number of Deaths
CPSEP	CPS Employment-to-Population Ratio	OMB	Office of Management and Budget
CPSRT	CPS Unemployment Rate	ORR	Office of Refugee Resettlement
CV	Coefficient of Variation	PAL	Permit Address List
DA	Demographic Analysis	PES	Post-Enumeration Survey
DAFO	Deaths of the Armed Forces Overseas	PIP	Performance Improvement Period
deff	Design Effect	POP	Performance Opportunity Period
DoD	Department of Defense	PSU	Primary Sampling Unit
DRAF	Deaths of the Resident Armed Forces	PWBA	Pension and Welfare Benefits Administration
FNS	Food and Nutrition Service		
FOSDIC	Film Optical Sensing Device for Input to the Computer	QC	Quality Control
		RDD	Random Digit Dialing
FR	Field Representative	RE	Response Error
FSC	First- and Second-Stage Combined	RO	Regional Office
FSCPE	Federal-State Cooperative Program for Population Estimates	SCHIP	State Children's Health Insurance Program
		SECU	Standard Error Computation Unit
GQ	Group Quarters	SFR	Senior Field Representative
GVF	Generalized Variance Function	SMSA	Standard Metropolitan Statistical Area
HUD	Housing and Urban Development	SOC	Survey of Construction
HVS	Housing Vacancy Survey	SR	Self-Representing
I&O	Industry and Occupation	SS	Second-Stage
IM	International Migration	STARS	State Time Series Analysis and Review System
INS	Immigration and Naturalization Service		
IRCA	Immigration Reform and Control Act	SW	Start-With
LAUS	Local Area Unemployment Statistics	TE	Take-Every
MARS	Modified Age, Race, and Sex	UE	Number of People Unemployed
MIS	Month-in-Sample	UI	Unemployment Insurance
MLR	Major Labor Force Recode	URE	Usual Residence Elsewhere
MOS	Measure of Size	USU	Ultimate Sampling Unit
MSA	Metropolitan Statistical Area	WPA	Work Projects Administration

Index

V

Vacancy rates   11-2, 11-4
Variance estimation
  definitions   13-1–13-2
  design effects   14-7–14-9
  determining optimum survey design   14-5–14-6
  generalizing   14-3–14-5
  objectives of   14-1
  replication methods   14-1–14-2
  state and local area estimates   14-3
  successive difference replication   14-2–14-3
  total variances as affected by estimation   14-6–14-7
Veterans, data for females   2-4
Veterans' weights   10-14

W

Web sites
  CPS reports   12-2
  employment statistics   12-1
  questionnaire revision   6-3
Weighting Procedure
  American Time Use Survey (ATUS)   11-5
  Annual Social and Economic Supplement (SEC)
   11-6–11-10
  Housing Vacancy Survey (HVS)   11-4
Weighting samples   15-8–15-9
Work Projects Administration (WPA)   2-1

X

X-11 and X-12 ARIMA program   2-6, 10-15–10-16