OVERVIEW OF THE CURRENT POPULATION SURVEY OPERATIONS AND ESTIMATION
Margaret Schooley, Lawrence Cahoon, Chester Ponikowski and Kenneth Riccini
U.S. Bureau of the Census

I. Introduction
The Current Population Survey (CPS) is the nation's primary vehicle for obtaining labor force data by demographic characteristics of the civilian noninstitutional population. In addition to labor force data, the CPS also provides a wealth of other socioeconomic data.

Since the Current Population Survey's (CPS) inception almost 50 years ago and its uninterrupted operation since then, many changes and refinements have taken place in every component of the survey. Most of these changes in sample design, questionnaire design, field enumeration and processing procedures are described in various publications of the Bureau of the Census and the Bureau of Labor Statistics. See [1]. An overview of the Current Population Survey Sample Design by Moore, et al., which was also presented at these meetings, describes the overall sample design and the three supplementary samples that have been or will be added to the CPS in the 1970's. The purpose of this paper is to present new developments and changes to both the operational and estimation procedures used in the CPS and to touch upon some of the aspects of the survey not covered by prior publications.

II. Overview of Interviewing Procedure
Each month during the week containing the 19th day of the month known as interview week; CPS interviewers begin their job of contacting the households in their monthly assignment. Households are in sample for 4 months, drop out of sample for 4 more months, and return in sample for 4 more months. Households that are in sample for the first and second months and those returning in sample after the 8-month dormant period (5th contact) must be visited by the interviewer in person. Other sample households may be contacted by telephone. During 1978 approximately 57 percent of the CPS interviews were completed by use of the telephone.

The first form to be completed in the household is the control card. Household identification and personal characteristics data for each household member are recorded on the control card which serves as the basic record for the household and persons living or staying in the household. Control card household items include address, tenure, type of housing unit, etc. while personal characteristic items include age, race, sex, marital status, etc. Labor force data are then obtained for each civilian household member 14 years of age or older via the CPS-I questionnaire (see Section III). For the remaining months the unit is in sample the control card information is updated, as necessary, and the labor force questions on the CPS-I document are again asked of each eligible person. Since the CPS-I document is the sole source for data capture, certain household and personal characteristic items are transcribed from the control card to the CPS-I each month. In some months, supplemental questions are included in the CPS-I and are asked after the labor force information is obtained for each eligible person.

III. The Survey Instruments
A. The Control Card

1. Basic Changes
Currently there are two versions of the CPS Control Card (form CPS-260) in use. The old version, which will be entirely phased out of the sample by January 1980, differs from the revised version in that:

a. The Census Bureau's long standing practice of designating the husband, if present, as the head of household and therefore as the central person to whom all other household members' relationship is obtained is being phased out. This concept has been replaced in the revised control card with an owner/renter approach to determining a central reference person or householder.

b. Race was previously marked by observation and consisted of three categories--White, Black, or other. Currently, race is asked once for each person in the household. The revised racial categories are White, Black; American Indian; Aleut; Eskimo; and Asian or Pacific Islander. For respondents who cannot classify themselves or other household members into one of the four categories, the interviewer marks 'other' in the control card. Currently, less than .1 percent of the persons eligible are recorded as 'other.'

c. Origin or descent, previously asked only for household members 14 years or older, is asked once of the household respondent who reports for all household members. Since demands for children's data are increasing, the universe for this item was expanded to include persons in the 0-13 years old age bracket.

The revisions to the race and origin or descent items were made in compliance with the Office of Fed. Statistical Policy and Standards, Directive 5, which specifies the Race and Ethnic Standards for Federal Statistics and Administrative Reporting.

2. Method of Introduction—Revised Control Card
Because of the rotation scheme for the CPS sample, 16 months elapse between the time an address enters the sample for the first time and completes its last month in sample. Since the new control cards are being introduced one rotation group at a time beginning with the October 1978 survey. The entire sample will not be converted until January 1980. An interim procedure was developed, however, that enabled the full sample to utilize the reference person concept by January 1979.

B. The CPS-I Document
Up until January 1979 the basic CPS document was virtually unchanged since the late 1960's and was comprised of four types of pages: a household page, a page for information on children (used periodically), the basic labor force page, and the supplement page for adults (persons 14 years of age or older). In January 1979, the CPS-I was reformatted. The same four types of pages are still used in the new CPS-I. The major changes to the CPS-I are (1) personal characteristics data for children 0-13 years of age were permanently added to the CPS document and will be collected on a monthly basis, (2) the basic
labor force page was modified to include a series of items concerning usual hourly and weekly earnings of wage and salary workers, and (3) the personal characteristics items were modified and moved from the basic labor force page to the page facing it which previously was used only for supplemental material.

The last modification includes: (a) use of the reference person concept and expansion of the "other relative" category to include separate categories for 'own child,' 'parent,' and 'brother/sister,' (b) expansion of marital status categories to include separate categories for 'separated,' 'widowed,' and 'divorced,' and (c) expansion of the race codes to reflect the racial categories contained on the revised version of the control card.

IV. Field Operations
A. Field Procedures - Hiring/Training/Supervision/Control
The Census Bureau's field organization is comprised of 12 regional offices, each staffed by a regional director and a staff of program assistants. One or more of these program assistants is designated as the CPS Program Supervisor and is responsible for the monthly field collection, control and regional office processing procedures. Other activities carried out in the regions include the hiring, training, supervision of the interviewing staff and control of the CPS materials. See [1] for more details. Currently, there are approximately 1,500 CPS interviewers each of whom must undergo a rigorous training program before their first "solo" interview. The interviewer's performance is closely monitored and evaluated on the basis of error rates, Type A noninterview rates and interviewer efficiency.

Type A noninterviews are occupied units eligible for the survey for which the interviewer was unable to complete the interview. If an interviewer's type A noninterview rate is in excess of 5 percent of the occupied eligible households in sample, the interviewer "needs improvement." Type A nonresponse rates, especially refusals which account for over half the type A noninterviews, have been increasing over the years. Concerns about privacy and confidentiality, general mistrust of government, and fear of crime being suggested as possible reasons for the increase [4].

Other functions of the regional office include the control of and preparation of assignments for reinterview.

B. Washington, D.C.
1. Pre-processing
Records for documents are deleted from the computer depending upon various reject conditions (e.g., incomplete identification, incomplete number of pages on document (torn), CPS-I document for the wrong month processed). CPS-I's so identified are then located, corrected as necessary and resubmitted through the system for acceptance by the Data Acceptance Check (DAC).

A simulation of the computer edit used in the final processing is run on the documents. Errors (blanks, inconsistencies, etc.) are identified, tallied, and very closely monitored. No actual transformation of the data takes place at this point as would occur during the actual edit/allocation. A record leaves the system exactly as it entered. A count of individuals by their labor force status is also done. From these counts a rough "advance estimate" of the level of unemployment is computed and sent to the Bureau of the Census.

2. Final Processing
After all the documents have been processed through Jeffersonville and the data are accepted as specified above, CPS preprocessing close-out has been achieved. Production processing consists of:

a. Editing
The edit guarantees that the final data for each person or household are complete and internally consistent, (i) Household or Personal Characteristics Edit—This phase of the edit identifies interview and noninterview households and edits the personal characteristics of persons within the interviewed households. Errors are first identified. If the problem cannot be resolved by referring to existing information already entered in the record, allocation is made from a record for a person with similar characteristics, (ii) Labor Force Edit—The first step in the labor force edit is to determine each person's labor force status. At least two positive indications of labor force participation are required in order for a person to be classified as in the labor force. Once the labor force status is determined, any errors or missing information are identified and corrected or imputed.

At this point, computer listings are produced that identify each interviewer and the number of noninterviews and/or errors that he/she may have been responsible for. These listings are provided to each regional office and form the basis of the evaluation of the interviewer's performance as explained in Section IV above.

b. Estimation Procedure
The preparation of the current estimates from the CPS includes several distinct steps: preparation of base estimates; adjustment for noninterview; two stages of ratio estimation; preparation of a composite estimate; and application of a seasonal adjustment to major labor force data. Most of these steps have been modified with the additions of supplementary samples in the 70's; when part of the CPS sample was restratified and new sets of PSU's were added and with new research findings, the procedures for applying the noninterview adjustment, the first-stage adjustment, and the second-stage adjustment have been changed. The modified noninterview and second-stage adjust-
ments were implemented in January of 1978. The new first-stage adjustment was implemented in January of 1979. A description of the estimation procedure currently in use follows. The future supplementation may require additional changes in the estimation procedure.

i. Preparation of Base Estimates
The raw survey data is first weighted by multiplying each sample unit--person or household--by the inverse of the probability with which that unit was selected.

ii. Noninterview Adjustment
The purpose of noninterview adjustment is to reduce the bias caused by nonresponse. Beginning in January 1978, the adjustment is performed by the noninterview clusters which have been formed by grouping sample primary sampling units (PSU's) with similar demographic characteristics within each State and the District of Columbia. Prior to January 1978, the adjustment was performed by noninterview clusters of sample PSU's formed within each census region. The procedure used prior to 1978 is described in detail in [1]. The clusters are totally comprised of PSU's either from SMSA or non-SMSA areas. Within each noninterview cluster, 24 adjustment cells are defined by: (i) three residence categories--central city of SMSA, balance urban of SMSA, balance rural of non-SMSA, (ii) two race categories--White and Black and Other, and (iii) four pairs of rotation groups. The noninterview adjustment factor is then computed for each cell. The factor is defined as the ratio of the total number of interviewed plus noninterviewed households to the number of interviewed households. The factors are then multiplied by the basic weights of the units in the appropriate cells.

iii. First-Stage Ratio Estimation
The purpose of applying the first-stage ratio estimation procedure is to reduce the bias caused by undercoverage. Prior to January 1978 the procedure was done in two steps. First, Blacks and Other races only were adjusted separately to an independent estimate of 48 age-sex-race categories. The second step in the procedure was applied to all sample persons and the adjustment was made to an independent national population estimate within 68 age-sex-race categories. This procedure is described in detail in [2]. Beginning in January 1978 the procedure is done in three steps. First, the sample estimates for each State and the District of Columbia are adjusted to an independent control for the population 16 years and over for the State. The second step involves a separate ratio adjustment for persons of races other than White. Blacks and Other races are adjusted separately to an independent estimates of 48 age-sex-race categories across the whole Nation. The third and a final step in the second-stage ratio estimation procedure is applied to all sample persons including those involved in the second step of the procedure. The final adjustment is made to an independent national population estimates within 68 age-sex-race categories. The entire second-stage ratio estimation is iterated six times in order to insure that the sample estimates both of State population and of national age-sex-race categories will be virtually equal to the independent population controls.

iv. Composite Estimate
The composite estimate is a weighted average of two estimates for the current month for any particular characteristic. The first of these two estimates is the result of all estimation steps described above. The second estimate consists of the composite estimate for the preceding month to which has been added an estimate of the change from the preceding month to the present month, based on approximately 75 percent of the sample which is common to both months. In CPS, the weights used for combining these two estimates are each one-half.

v. Seasonal Adjustment
Seasonal adjustment is applied to estimates for the major labor force data. This is done to adjust for normal seasonal fluctuations in the data so that meaningful comparisons of month-to-month changes can be made. Some past studies indicate that the variances are about the same for seasonally adjusted data as for nonadjusted data [1]. Seasonal adjustment of the data is done by the Bureau of Labor Statistics.

VI. Processing Modification
Rather than incorporate all the questionnaire design changes (as explained in Section III) into the processing system at the same time as they were made, recoding and reformatting of the new CPS data into the old CPS data configurations were done as much as possible so that the existing computer programs could be used. Thus, except for the new earnings items and
children's items, all final CPS data are in the same format as prior to the changes. Also, because of the imminence of the 1980 census and resulting revisions to the CPS anticipated subsequent to the census and the analysis of the CPS program by the National Commission on Employment and Unemployment (NCEUS), major edit and processing changes to the CPS are being held to a minimum. Plans for incorporating the changes into the CPS processing procedures, however, are currently underway.

With respect to the earnings items computer program modifications to the data acceptance check (DAC) to scan for legal/illegal entries in the earnings items, the inclusion of an earnings pre-edit, expansion of the record size for CPS up to and including the estimation procedure, and program modifications to separate the portion of the sample eligible for the earnings items (two rotations) were made to the basic CPS processing system. After this point an independent system for processing the earnings items has been developed. This system includes separate edit, allocation and estimation procedures, identifies interviewer errors, and will produce three detailed tabulation packages.

At the current time, two of three tabulation packages have been produced for each month since October 1978 and submitted to the Bureau of Labor Statistics for their review. Although data will be tabulated monthly, publication will be on a quarterly basis beginning at the end of this year.

VII. Impact of Operational Changes on National Estimates

Most of the impact of the changes have surfaced as changes in coverage ratios and increases in the number of errors. The change in the race procedures had an impact in the sample coverage for races other than White. The coverage for the non-White population other than Black increased from .882 to 1.025 between October and December 1978. Much of this increase was anticipated and is attributed to the differences in race data collection procedures (observation vs. asking). The coverage ratios for this population group for January (1.127) and February (1.127) were unexpectedly large, and are believed to be the result of a combination of factors including incorrect classification on the control card and errors resulting from transcribing the race code from two versions of the control card to the reformatted CPS-I document. The large increase in coverage in January and February is primarily a "false" increase.

After the transcription problems were pointed out to the interviewers and they became accustomed to the new control card and CPS-I format, the coverage ratio's dropped (1.067, 1.047, 1.021, 1.034 for March through June, respectively). Beginning in June 1979, revised procedures for recoding race for problem cases were implemented in the field. These procedures are in conformity with the procedures to be used for the 1980 census. These procedures were developed to address the problem of improper recording of the race data on the control card (as opposed to improper transcription from the control card to the CPS-I).

A detailed analysis of the effects of the new procedure on the published estimates is currently underway. More monthly data are necessary, however, before the impact can be accurately assessed.

Preliminary reports indicate that there is some difference in average estimates based on the first quarter 1979 data but the difference is not large enough to be considered statistically significant.

The increase in the coverage of the non-White population other than Black has also raised some questions concerning the reliability of the independent estimates of this population used in the second-stage CPS ratio estimation. Research in this direction is underway.

To date, no impact of the inclusion of earnings items on basic CPS data has been discovered. Detailed analyses will continue until a sufficient number of months of data have been processed and statistical inferences can be made.

VIII. Supplements

In addition to providing labor force data, the Current Population Survey is frequently used by the Bureau of the Census and other Government agencies as a means of obtaining a variety of demographic data. The most well known, and perhaps the most important is the Annual Demographic Survey conducted in March of each year. This is a comprehensive supplemental inquiry which provides the Nation's main source of data concerning individual and family income, migration, household and family composition and work experience during the preceding calendar year. Other annual supplements include those on dual jobholding, school enrollment, fertility and birth expectations, and immunization. Other one-time surveys are also conducted.

IX. The Accuracy of the Data

There are two types of errors which occur in sample surveys - sampling and nonsampling errors. Sampling error is the error incurred because only a portion of the total population is interviewed. The Census Bureau has had the goal of estimating variances, and thus standard errors, for every type of characteristic for which estimates are published. The standard errors as estimated for CPS are primarily measures of the sampling error; they also partially measure the effect of some nonsampling errors in response and enumeration, but do not measure any systematic biases in the data.

Nonsampling errors in surveys can be attributed to many sources, e.g., noninterviews, definitional difficulties, inability or unwillingness of the respondents to provide correct information, recall errors, processing errors, and undercoverage of the target population. The full extent of nonsampling error in the CPS is unknown. However, several special studies have been conducted to quantify some sources of nonsampling error in the CPS. There are cases where it is difficult to identify and quantify the magnitude of the errors.
The major ongoing study of nonsampling error in the CPS is the reinterview program. This program is used to estimate various sources of error as well as to evaluate and control the work of the interviewers. A subsample of the CPS is reinterviewed in the eighth month. The ratio estimation provides an indication of the magnitude of response variance and response bias. The primary purpose of the program is control and evaluation of interviewer performance. This use reduces errors caused by the interviewers, and thus helps to control nonsampling error in the CPS. The results of this program indicate that the data published from the CPS are subject to moderate systematic biases. It is impossible in this limited space to adequately discuss the other results of the reinterview program. As a more complete description of the results can be found in [1].

Some sources of nonsampling errors are evident in the monthly CPS tabulations. One of the more prominent of these is the existence of rotation group bias. This bias is best illustrated by the observation that first-month-in-sample households produce, in general, a higher estimated unemployment rate than does the remainder of the sample despite the fact that both parts of the sample are representative of the entire population. The presence of this type of bias in the CPS is not surprising as such biases are frequently observed in panel surveys. While it is possible to quantify the magnitude of this bias in a relative sense among the eight rotation groups used in the CPS, it has not been possible to determine which, if any, of the eight rotation groups gives the best estimate. It may be that the first-month-in-sample households provide the best estimates as there has been no conditioning of the respondents. Then again these households may not produce better estimates because higher than average noninterview rates occur for first month households. It is not known if this effects the magnitude of the bias. It is known that some of the questions asked of fourth and eighth month in sample households to obtain more data concerning persons not in the labor force (i.e., identify discouraged workers) result in a higher reporting of unemployment, thus affecting the rotation group bias indices [4].

Undercoverage in the CPS and in the census is another source of nonsampling error in the survey. This undercoverage results from missed housing units, and from missed persons within sample housing units. The census is estimated to have missed 2.5 percent of the population while the CPS misses about 7.5 percent of the population. Undercoverage varies by age, sex, and race and is generally larger for males than females and Blacks than Whites. A ratio estimation procedure as described earlier adjusts for CPS undercoverage relative to the census. No adjustment is made for census undercoverage. Thus, biases exist in that data exist due to the lack of an adjustment for census undercoverage. The ratio estimation partially corrects for survey undercoverage relative to the census. However, biases exist in the estimates to the extent that missed persons have different characteristics than interviewed persons.

A subtle means by which nonsampling error can be introduced into the CPS results from the ratio estimation used in the survey. The adjustments are based on race as identified by the 1970 census with an appropriate adjustment for the aging of the population since then. As described earlier, the CPS is in the process of switching to a system where the respondent identifies his or her race. This new system is essentially consistent with methods used in the census. The use of the ratio estimation assumes that race concepts have remained unchanged since the time of the census. That is, that a person who identified himself as Black at the time of the census will do the same today. This may not be the case; to the extent that it is not the ratio estimation procedure introduces nonsampling bias into the data. This may be the cause of the coverage problem discussed in Section VIII.

This description has not illustrated many of the possible sources of nonsampling error, but rather has focused on only a few of the more obvious sources. A more complete discussion of the sources of such errors in the CPS is given in [4].

Variances have been directly estimated for the CPS 461 PSU and earlier designs. As a result, much is known about the reliability of the estimates for data from the national sample in use prior to the supplementations in the 1970's. There has been as yet no direct estimation of variances for national and subnational data from the supplemented designs described in the companion paper (Moore, et al), variances for some of these estimates have been obtained by indirect means. Thus the variance estimation procedure discussed in this section has been used only for CPS designs prior to the supplementations.

There are two steps in the process through which the standard errors presented with the data published monthly are obtained. First, variances are directly estimated for a small subset of the characteristics published each month. A modified random group technique and Taylor linearization technique are used. Variances for most levels of estimation and within-PSU, between-PSU, within-stratum and between-stratum components are estimated for each characteristic. The variance estimation for CPS is discussed in detail in an appendix to [2]. Also, the techniques used and a brief, undetailed description of the CPS variance estimator are presented in [5].

After the variances are directly estimated for the subset of characteristics, they are used to estimate parameters of generalized variance functions for groups of characteristics. These groups are distinguished by such characteristics as age, sex, race and labor force status. The combination of the two steps of variance estimation allows us to estimate variances for all the published data at a much smaller cost than would be incurred if the variances were directly estimated for every published characteristic.

The standard errors for CPS are presented in generalized tables in the monthly publication. From these tables a standard error can be obtained for almost every characteristic shown in the data tables. Descriptions of how to use the standard error tables are also provided. The Bureau of Labor Statistics will make statements of comparison with 90 percent confidence.

The form of the generalized variance function for most characteristics is

\[ \text{Var}(x) = ax^2 + bx \]
where \( x \) is the estimate for characteristic of interest and \( a \) and \( b \) are the estimated parameters. Use of the function allows flexibility in estimating standard errors for any characteristic.

In addition to using the standard errors for analysis, the variances are used to evaluate the effectiveness of the sample design and estimation. As mentioned above, variances for different levels of estimation and components are estimated. From this we know, for example, that the second-stage ratio estimate dramatically reduces the variance on the estimate of the total civilian labor force and that the between-PSU variance is relatively small for most characteristics. An analysis of the CPS variances is presented in [4]. These variances can be monitored and used as indicators of a need to change the design and/or estimation.

X. Cost

The total cost for CPS interviewing including field collection, data preparation and processing is approximately $25,000 per interview. The total CPS budget for Fiscal Year 1979 is $20,000,000.

XI. Availability of CPS Data

Data produced from the CPS are primarily in microfiche and data tape format. Data are published on a monthly basis by the Bureau of Labor Statistics in Employment and Earnings, and the Monthly Labor Review. Additionally, annual average estimates for States and large metropolitan areas are published.

Computer tapes containing monthly data from the basic CPS and any supplemental inquiries that may have been asked for that particular month are also available. Children's data and earnings data are expected to become part of the monthly tapes in the near future.

In order to meet the Bureau's confidentiality requirements, an area with a population of less than 250,000 cannot be uniquely identified. (Of course no information concerning identification of individuals is released, nor is this type of information even included on the Census Bureau's internal use CPS data files.) These computer tapes are produced from standard public use tape procedures and records for each State and many SMSA's are identifiable on the tape. Because of the standardization of the production of the public use tapes, the cost of the CPS tapes have been considerably reduced from the "tailor-made" versions previously produced. Further information concerning the CPS computer tapes may be obtained by contacting the Data Users Services Division of the Bureau of the Census.

XII. Future Developments

The increased interest in State and area data will most likely result in a complete redesign of the CPS sample. The review of the CPS by the National Commission on Employment and Unemployment Statistics (NCEUS) may also result in various questionnaire/concept changes to the Current Population Survey program. Implementation of any conceptual or questionnaire changes into the CPS program will first require thorough investigation as to the validity of the data being obtained and its impact on the labor force responses. The Methods Development Survey currently underway in four geographic areas, and expanding to eight areas beginning in December 1979, may be utilized as "proving grounds" for any NCEUS recommendations. The main purpose of this survey, however, is to test new procedures, questions, and concepts that may be used for the redesign of the CPS.

XIII. Conclusion

The growth and refinement of the CPS program as well as its leading role in providing statistical data for National and subnational areas will undoubtedly continue well into the future. The constant "fine tuning" of the program is one of the main reasons for its continuing success.

REFERENCES


