Camilla A. Brooks and Barbara A. Bailar Bureau of the Census

#### I. Introduction

# A. Purpose of the Error Profile

Ideally, a user of survey data should be provided two measures of error with each statistic produced from a survey--one, a measure of the total variance, including sampling, response, and processing variability and second, a measure of the total bias, including the bias arising from the data collection procedure, the questionnaire used, and the estimator. These two measures are rarely, if ever, available. An estimate of the sampling variance, which may or may not adequately reflect all sources of variance, is all that is usually given to a user to evaluate survey data.

The interests of survey designers and some data users, however, extend beyond this to the individual components of variance and bias and how they may interact. One of the most studied surveys sponsored by the Federal Government is the Current Population Survey (CPS). Studies have been conducted to assess the impact of many of the aspects of the survey design on the estimates. Yet, even with this survey, the specific effect of each of the various sources of bias and variance on the survey statistics and their interaction are not fully quantified. This paper gives an illustration of how one can go about constructing an error profile for a survey statistic by examining each of the potential sources of nonsampling error and trying to assess the impact of each source on the survey data. These assessments would be input to a mathematical model for the total survey error. To illustrate the impact of nonsampling error, we have drawn upon a number of studies by Census Bureau staff members and others. We did not attempt to construct a mathematical model for the total error in a survey statistic. However, one of the main conclusions that can be drawn from this illustrative error profile is that much more empirical work is necessary to provide a comprehensive picture of the effects of nonsampling error on a survey statistic.

The focus of this paper is on the employment statistic. Ideally, one would list each potential source of nonsampling error and then quantify the effect of that source on the employment statistic. Because data as well as space are limited, the major steps in the survey process are mentioned, but only a few steps are described in detail. For a more thorough explanation of the Current Population Survey in its entirety the reader should refer to Hanson's draft revision of Technical Paper No. 7 (1976) which should soon be published.

# B. Specifications of the CPS Employment Statistic

The concept of "employment" used in the CPS is specified by the Bureau of Labor Statistics (BLS), and its translation into a set of questions for collecting the data is carried out jointly by BLS and the Census Bureau. Since January 1967 the definition of employment used in the CPS has been: Those at work, consisting of persons who worked one hour or more for pay or profit, or 15 hours or more without pay in a family operated enterprise, and those with a job, but not at work, such as persons temporarily absent from work because of illness, vacations, etc.

The CPS is restricted to the civilian noninstitutional population age 14 and over. A monthly housing unit sample is used which represents the universe of all households in the United States and also includes nonhousehold units, in which people live such as hotels, dormitories, flophouses, bunkhouses, and the like. In using the housing unit approach, an implicit assumption is made that each person 14 years of age and over is uniquely associated by the survey definitions and procedures with either a household or one of the nonhousehold units mentioned.

### C. The Survey Design

The CPS is redesigned after each decennial census in order to utilize the most recent census data. In the 1970 redesign, which is discussed here, 461 primary sampling units (PSU's), which are primarily counties or groups of counties, were selected from 376 strata. Of these 461 PSU's 156 were designated self-representing (SR). The other 305 PSU's were selected from 220 nonself-representing (NSR) strata and are referred to as nonself-representing PSU's.

The CPS is a multi-stage cluster sample which is essentially self-weighting. The design is such that the ultimate sampling units (USU's)that is, clusters of approximately four, usually contiguous, housing units, do not remain in sample for the entire decade. Therefore, several CPS samples must be generated for use during the decade. Each CPS sample consists of eight, approximately equal, systematically selected subsamples known as rotation groups. These rotation groups are introduced into the sample once a month for eight months using a 4-8-4 rotation scheme; i.e., each sample USU is in sample four months. Under this scheme, each month there is a 75 percent month-to-month overlap of households and a 50 percent year-to-year overlap. At the time of the 1970 redesign, data were collected from approximately 47,000 eligible households each month.

## II. The Sampling Frame

One potential source of bias in any sample survey is the lack of complete coverage of the population. The coverage is associated with the frame, and, if the frame is deficient, or information is not obtained from all persons within the sample units, there will be undercoverage.

### A. Description of the Frame

The frame for this survey is derived from a variety of sources with the main source the 1970 Decennial Census. In the CPS extensive use is made of the 229,000 enumeration districts (ED's) defined in advance of the census; these are large geographic areas, each containing about 350 housing units on the average. There were three types of ED's in the census identified by the manner of forming the address register which was a list of the housing units within an ED. These are as follows:

- 1. Tape address register (TAR) ED's in which the address register was created from a computer tape copy of a commercial mailing list and corrected by the Post Office and local agencies.
- 2. Prelist ED's in which the address register was constructed by a listing procedure conducted in advance of the census.
- 3. Conventional ED's in which the address register was prepared by the enumerators during the enumeration.

For purposes of the CPS sample, a 1970 Census ED is referred to as an address/list ED if the conditions listed below are satisfied.

- 1. The ED is a TAR ED.
- The ED is a prelist or conventional ED satisfying "a" and "b" below:
  - a. at least 90 percent of the 1970 Census addresses within the ED are recorded with complete street name and house number;
  - b. the ED is located in an area which issues building permits.

In address ED's the CPS sample is selected from the census address registers and the resulting sample referred to as an address sample. The address ED's are supplemented by the Census supplemental sample, referred to as the Cen-Sup sample. which is used to cover housing units in address ED's at addresses missed in the census or inadequately described in the address register. These units represent less than one percent of the CPS sample. All other 1970 Census ED's are referred to as area ED's. These are subdivided into area segments which are sampled and assigned to be listed by an enumerator about a month before interview. The sampling of housing units from the listings is carried out in an office operation, not by the interviewer.

Units built after April 1, 1970 are represented in address ED's and permit issuing area ED's by a sample of building permits selected from records of places issuing building permits as of January 1, 1970. This supplement to the frame is referred to as the permit universe and includes approximately 10 percent of the CPS sample. In non-permit issuing area ED's new construction is covered by interviewer listing.

# B. Potential Sources of Errors Associated with the Sampling Frame

It is known that the sampling frame used for the CPS does not fully represent the universe of housing units. - However, deficiencies discussed below are estimated to represent less than three percent of the target population of persons. Some of the frame deficiencies are discussed below.

## Permit Lag Universe

In permit issuing ED's, housing units completed after the census for which permits were issued before January 1, 1970 are not included in the CPS frame. These units are collectively referred to as the permit lag universe. There is an estimated total of 598,000 units for which permits were issued prior to January 1, 1970 that were completed after April 1, 1970 (MacKenzie, 1977).

#### Undercoverage of Special Places - Mobile Homes

Mobile homes located in address segments are a second potential source of coverage loss. Presently, in the CPS there is no general procedure for identifying or representing mobile homes in new mobile home parks, or new mobile homes at large in address ED's at addresses nonexistent in the 1970 Census; the permit universe includes regular housing units only. In addition to new mobile homes, the coverage problem of mobile homes in address ED's extends to those in existence at the time of the census but not counted in the census, and to those vacant in 1970 and therefore by design not counted in the 1970 Census but which are now occupied.

The coverage improvement program in the October 1976 AHS located approximately 300,000 mobile homes for the period April 1970-October 1976. This improvement program has not yet been included in the CPS, but these mobile homes should be represented in the sample. Though concern is greater for mobile homes, other special places including transient hotels, boarding homes, etc. could present some of the same problems as the mobile homes.

# Nonpermit Issuing TAR ED's

A small number of the TAR ED's (approximately 47-50 or about 0.3 percent of all TAR ED's) are in non-permit issuing areas. Because of irreconcilable problems in sampling them as area ED's, it was decided to treat these ED's in the same manner as permit issuing, address type ED's. Thus new construction in these ED's is not represented (Baer, 1973 and Boisen, 1971).

#### Other Structure Misses

Other problems with the frame in address ED's not addressed by the Census Supplemental sample include coverage of homes moved to a site with an address not in the 1970 Census and structures converted from nonresidential to residential use after the census.

## C. Errors Associated With the Coverage of Persons Within Sample Housing Units

Within household coverage misses and the undercoverage of individuals with no attachment to any address are believed to account for a large percentage of the frame deficiencies;

however, information on both the extent and the causes of this problem is limited. It is estimated that because of missed structures less than three percent of the target population is not included in the frame. However, Table 4 in Section V shows ratios of independent estimates of the population prepared by the Census Bureau to Current Population estimates of the population which indicate a coverage problem exceeding three percent. For white males and females the ratios are 1.049 and 1.023, respectively while for males and females of black and other races the respective ratios are 1.155 and 1.075. Further, by agreement with the data users the independent estimates of the population which are used in Table 4 do not reflect the estimated undercoverage of the census which itself did not include an estimate for the "illegal alien" population in the U.S. (Siegal, 1973). Thus, the within household coverage problem is even greater than indicated in the tables. Of blacks missed in the 1970 Census an estimated 64 percent were missed within units enumerated as occupied or occupied units enumerated as vacant; the corresponding figure for persons of white and other races was 42 percent (Jones and Blass, 1975).

This phenomenon of undercoverage compared to the decennial census, which itself is subject to undercoverage of the population, is typical of household surveys generally.

## D. The Effect of the Census Undercount on the Employment Statistics

It was noted that, although estimates of the undercoverage of the population by age-sex-race in the 1970 Census are available, the CPS coverage is adjusted to the level of the 1970 Census rather than to figures adjusted for the 1970 undercoverage. A study by Johnston and Wetzel (1969) explored the effect of the 1960 Census undercount on the average monthly labor force estimates for 1967 if this convention were not adopted. Since the labor force status of the omitted persons is unknown, the authors provided two alternative sets of "corrected" labor force estimates. In the first set, an assumption was made that the missed persons had the same labor force status as their peers (persons of the same age, sex, and race group). This is called the "comparability" assumption. In the second set, omitted persons were assumed to have labor force status comparable to people of the same age, sex, and race but living in urban poverty areas. This is called the "poverty neighborhood" assumption.

Their report shows that the most significant change was in the level of employment which would increase by 2.8 million under the comparability assumption and 2.7 million under the poverty neighborhood assumption. The report further shows that the effect of the undercount on labor force estimates is much more important for persons of black and other races than for whites because their undercoverage is greater.

The conclusion of the authors was: the population undercount leads to a substantial understatement of employment levels, and introduces further complications due to a discrepancy in level when these data are related to other time series not based on household surveys. The authors make no comment on the impact of this phenomenon for comparison of changes in level on either a current (short term) or historical (long term) basis.

Hirschberg, Yuskavage, and Scheuren (1977) have recently arrived at estimates of the effect of CPS and 1970 Census Undercoverage on the labor force estimates which are more sensitive to the coverage problem than those of Wetzel and Johnson. These are presented in a paper prepared for this year's ASA meetings.

III. The Data Collection Procedure

A. Description of the Process

The data collection procedure must be both quick and accurate. Interviews are conducted during the week (Monday through Saturday) containing the 19th day of the month, designated interview week. During interview week interviewers obtain information from respondents regarding the previous week or week containing the 12th day of the month, designated survey week.

The data collection procedure includes several potential sources of nonsampling error. For example, perhaps the way the question is worded is not clear to respondents; the use of proxy respondents for household members not at home at the time of the interview may affect the data; the use of telephone interviewing for certain households may change the kinds of data collected; the training of the interviewers may affect the way they collect or record the data. In this section of the paper, attention is focused on the actual conduct of the interview as a potential source of error.

Though any responsible adult household member 14 years of age or older is eligible to act as respondent, the interviewer is encouraged to interview the most knowledgeable household member, usually the household head or wife of head.

At the initial visit the interviewer records on a control card the name, usual residence, relationship to head, date of birth, age, race, sex, etc. for each person in the household. At each subsequent visit to the household the listing is updated. The CPS questionnaire is then completed for each household member 14 years old or older.

B. Potential Sources of Error Associated with the Conduct of the Interview

# Mode of Interview

Before the use of the telephone was instituted in CPS interviewing, a test in a limited number of PSU's was conducted to determine its effect on the data. This test, conducted in the early 1950's, showed no appreciable difference in the labor force data obtained by the two methods of interviewing, personal visit and telephone (Hanson, 1976). However, the test conducted at the time was not a completely controlled experiment and the results for today's purposes are outdated. Not only has telephone interviewing increased, but attitudes have probably changed over the years. Because of the wide use of telephone interviewing in the CPS, there is growing concern about its possible effects on the data, and because of this, studies are now being planned to learn more of its effects.

Interviewers are allowed to interview by telephone according to the following regulations:

Interviewers are instructed to conduct the first and fifth month interviews by personal visit. Second month interviews may be conducted by telephone only if no one was home at the initial personal visit. Providing the household has consented, interviews in the other months may be conducted by telephone.

The figures discussed here represent 1976 averages of telephone interviews. Though interviewers are instructed not to conduct first and fifth month interviews by telephone, 2.8 and 10.3 percent, respectively, of first and fifth month interviews were, in fact, collected by telephone, presumably to reduce the nonresponse rate that would have been experienced otherwise. Also, 44.5 percent of second month interviews and 76.0 percent of the interviews conducted in the remaining months were collected by telephone.

remaining months were collected by telephone. At the present time there is no evidence that personal interviewing and telephone interviewing yield different results on employment questions. However, it is recognized that the use of the telephone may cause a different respondent to be interviewed. Data show that there is an increase in the number of "other relatives" who are respondents in later months in sample. To the extent that "other relatives" in the household may not be as knowledgeable as the head and/or wife about the labor force status of all household members, the telephone data may not be as accurate.

### The Use of Proxy Respondents

In the CPS proxy respondents are frequently used; e.g., the wife may frequently respond for her husband, or in the case when both the head and his wife are absent, a 14 year old may respond for the entire household.

Between February 1965 and June 1966 a Methods Test was conducted outside of the regular CPS but with the purpose of testing new methods for the CPS. One of the problems evaluated in the test was the selection of best respondent for individual household members. Two different studies were conducted. In the first, three procedures were compared as follows:

- a. <u>Procedure 1</u> was similar to the present CPS procedure in that any responsible household member was accepted as a respondent for the entire household;
- b. In <u>Procedure 2</u> each adult household member was to be interviewed for himself;
- c. In <u>Procedure 3</u> an advance form containing important labor force questions was sent to each household in the test with a request that each adult household member fill the form for himself. The interviewer was then to transcribe this information to the questionnaire and ask the household respondent the remaining questions about the household members.

A comparison of the results was provided in a memorandum by Deighton (1967). The percentage of persons employed as measured by procedures 1, 2, and 3 were 55.6, 57.2, and 57.3, respectively. The self-respondent procedure resulted in more persons employed than did the household respondent procedure by approximately 1.6 percentage points. In a second experiment reported by Williams in 1969 similar results were found. However, sampling errors were not provided in either case so it is questionable whether this difference was significant.

#### Influence of Interviewers

Interviewers have many opportunities to influence the data. In this paper the focus will be on the noninterviews and misclassification of the labor force status.

The interviewer may encounter three types of noninterview situations: Type A noninterviews those households eligible for the survey for which the interviewer was unable to complete the interview; Type B noninterviews - vacant units, vacant sites, or units occupied by persons ineligible for the survey; and Type C noninterviews - units demolished, converted to permanent storage or business use, moved from site, or found to be in sample by mistake. Interviewers are strongly urged to keep the Type A noninterview households are visited each month to determine if any have become eligible for interview. Type C noninterview units are not visited again by the CPS interviewer.

Table 1 shows the noninterview misclassifications for the years 1973-76 as determined from reinterviews of subsamples of housing units in the CPS. There was an annual misclassification rate of 2.4 percent for 1976 which was significantly different from the 1974 rate only. The table shows that there are more Type A noninterviews that are misclassified as B's or C's than the reverse. This would lead to a deficit in the number of households eligible for interview, and therefore, in the coverage of the target populations and may explain part of the observed undercoverage referred to earlier.

Another aspect of the impact of interviewers is in the labor force status assigned to individuals on the basis of their responses to the survey questions. Accuracy of the data collected by the interviewer is frequently measured by using the results of the monthly CPS reinterview survey as a standard. The CPS reinterview is conducted by senior interviewers or members of the supervisory staff, and as such, are considered to be more accurate than the original interview results; however, because of limitations of the reinterview survey its results should by used with caution.

Specific to the number of persons classified as employed in the CPS, the reinterview provides information on how many persons were classified as employed in the reinterview. If one is willing to accept the reinterview as a standard, then the difference between the original interview and the reinterview can be used as a measure of bias. Table 2 shows the results of the two estimates of employment annually

Table 1. Noninterview Misclassification Rates $\frac{1}{}$  (Percentages)

Misclassification	1973 <u>-</u> /	1974 <u>-</u> /	1975 <u>-</u> 3/	1976 <u>-</u> /
Total	2.7	3.3	2.9	2.4
A's as B's	1.4	2.1	1.8	1.6
B's as A's	0.7	0.4	0.2	0.4
C's as B's	0.45	0.7	0.6	0.4
Other <u>2</u> /	0.15	0.1	0.3	0.03

1/ Moye, 1976 and Schreiner, 1977

 $\overline{2}$ / B's as C's, C's as A's, and A's as C's

 $\overline{3}$ / Base - total noninterviews

from 1956 through 1976. From 1956 to 1960 the reinterview was 0.2 percentage points lower than the original interview. However, since that time the differences have increased with a high in 1976 of 0.7 percentage points. Though these differences may seem small, they are consistent and when applied to an employment figure of 80 million, they account for between 160,000 to 560,000 persons.

Table 2. Summary of Percent of Persons			
Employed as Measured in the Original			
CPS Interview for the Reinterview			
Subsample and as Measured by the			
Reinterview after Reconciliation,			
1956,1976.			

Percent of Persons in labor force employed			Reinterview	
Year	Original	Reinterview	minus original	
1956	96.3	96.1	-0.2	
1957	95.8	95.8	0.0	
1958	93.2	93.0	-0.2	
1959	94.4	94.2	-0.2	
1960	94.6	94.4	-0.2	
1961	93.1	92.8	-0.3	
1962	94.5	94.5	0.0	
1963	94.4	94.0	-0.4	
1964	94.8	94.3	-0.5	
1965	94.9	94.7	-0.2	
1966	96.1	95.8	-0.3	
1967	96.2	95.8	-0.4	
1968	96.3	96.0	-0.3	
1969	96.3	95.9	-0.4	
1970	94.9	94.5	-0.4	
1971	94.1	93.7	-0.4	
1972	94.7	94.4	-0.3	
1973	95.0	94.7	-0.3	
1974	94.5	93.9	-0.6	
1975	91.8	91.2	-0.6	
1976	92.5	91.8	-0.7	

#### IV. Data Processing Operations

There are many activities included in the processing of the CPS questionnaires. These

operations include the coding of occupation and industry, the microfilming of the questionnaires, the conversion of the microfilm to computer tape by means of a process called FOSDIC (Film Optical Sensing Device for Input to Computers), the machine editing of the data and imputation for missing values. Though all of these activities are potential sources of error, in this paper we discuss only the FOSDIC operation.

A. Description of the FOSDIC Operation

"FOSDIC....can be described as a machine which is capable of 'reading' information from a microfilm copy of an appropriately designed schedule and transferring this intelligence to magnetic tape for processing on electronic computers" (McPherson and Volk, 1962).

Several variables play a role in the microfilming and FOSDIC procedures and their adherence to standards can determine the success or failure of the data processing. These variables include, but are not limited to, the quality of the paper used for the questionnaires; the uniformity of the index marks and marking circles on the questionnaire; and, of course, the proper operation of the FOSDIC and microfilming equipment.

## B. Potential Sources of Error in the FOSDIC Operations

Distributions of CPS questionnaires rejected by FOSDIC by cause of error are regularly prepared. In January 1976 out of 72,172 total questionnaire forms 1,074 or 1.49 percent of the questionnaires were rejected (Jablin, 1977). Of these 1,074 questionnaires, 379 or 0.5 percent were rejected because of FOSDIC/filming errors. The errors in FOSDIC/filming that are associated with filming problems, bad index marks, etc. are corrected; for example, if the questionnaire is "bad", data are transcribed to another questionnaire, questionnaires with missed pages are remicrofilmed, etc. The errors that represent a threat to data accuracy are the "invisible errors", i.e. errors that cannot be detected. An example is a FOSDIC pickup of an incomplete erasure as a mark. Also, FOSDIC itself is subject to a certain amount of measurement error; it is possible with the same tolerance levels for reading marks, that it can get different readings for the same marks read at different times.

#### CPS/FOSDIC Study

In 1974-75 a CPS/FOSDIC Study was conducted in the Operations Analysis and Quality Control Branch of the Bureau to identify specific sources of variations in the system. One major aspect of the CPS/FOSDIC Study involved the reading of two identical pages (both containing labor force data) of 300 CPS questionnaires. The two pages were filled in identically for each of the 300 documents so that there were then 600 identical pages of information. The use of different cameras for filming and different FOSDIC readers produced 32,997 attempted readings.

Out of the 32,997 attempted readings, the following errors were observed (Boisen, 1975):

Table 3. Some Results of the FOSDIC Experiment

Problem Area	Number	Percent of Total	
Drops of marks	22	.0029	
Pickups of blanks	44	.0034	
Drop-and-pickups	1	.00013	
Skipped read areas	260	.013	
Skipped pages	27	.082	

Some of the major findings resulting from this aspect of the CPS/FOSDIC experiment were that (1) basically the system as operated during the experiment was under control with system error so slight that improvement could be impossible; (2) quality control procedures should be extended to the marking of CPS questionnaires; and that (3) further investigation might pinpoint some nonrandom and significant sources of error that result in failed calibrates and missed indices.

FOSDIC error is quite small and represents a gain in accuracy when compared with the previous use of keypunching of the data. Thus processing should have less effect on the accuracy of surveys using this procedure.

#### V. Preparation of Estimates

There are several stages in the preparation of the final estimates of employment. There is a weighting procedure which attempts to adjust for noninterview and undercoverage, a series of estimates culminating in what is known as a composite estimate, the seasonal adjustment of the point estimates, the estimates of sampling variance, and the estimates of nonsampling error. In this paper only the weighting procedure is discussed.

#### A. The Weighting Procedure

The sample as selected for the CPS is essentially self-weighting, i.e. each sample household has approximately the same probability of selection. However, because of nonresponse and coverage problems, a reweighting of the records occurs before the estimates are produced. In the CPS there are five distinct steps in the reweighting process. These are as follows:

- 1. The reciprocal of the probability of selection is attached to the record for a given unit.
- During listing it was discovered that 2. certain USU's contained far more units than were expected based on census listing so subsampling took place to make the workload manageable. The interviewer weights for the subsampled units are now multiplied by a factor which inflates these units to reflect the actual number of units in the USU. This process is called duplication control. The maximum factor used in the duplication control procedure is four. When USU's are unusually large (100 or more units) and would thereby require greater subsampling, they are placed in the

rare events universe for the rotation group in which they appear. They remain members of that rotation group for eight CPS samples, thus greatly reducing the subsampling rate.

The next stage in the weighting process is 3. the adjustment for total noninterviews. For noninterview adjustment purposes, the CPS PSU's are divided into 72 clusters formed by grouping together PSU's with similar characteristics defined by the 1970 Census. The clusters are classified by geographic region, and within a region they are divided into clusters totally comprised of PSU's in SMSA's (Standard Metropolitan Statistical Areas) and those containing only non-SMSA PSU's. The clusters are further partitioned into six race-residence cells for both SMSA and non-SMSA and are applied for each of the CPS eight rotation groups.

The weighted estimates with the duplication control and the noninterview adjustment are referred to as the unbiased estimates.

4. In order to reduce the contribution to the variance arising from the sampling of PSU's, the first stage ratio adjustment procedure is applied to records in nonselfrepresenting PSU's. Separate ratios are used for race and residence categories of two groups of strata (SMSA and nonSMSA) for each of the four census regions. The adjustment factor is computed as the ratio of the 1970 Census population in the raceresidence cell for the given cluster to the estimate of this population based on the 1970 Census population for sample PSU's in the same cluster and is applied to each of the records in the given cluster. Though a few factors are higher, a maximum factor of 1.3 is used. 5.

Following the first stage ratio adjustment, the <u>second</u> <u>stage</u> <u>ratio</u> <u>adjustment</u> is applied to all sample records. The procedure attempts to bring the sample estimates into closer agreement with independent estimates of the U.S. population by age, sex, and race.

The second stage ratio adjustment has two steps. First, a separate ratio adjustment is applied to blacks and other minority race sample persons; the basic reason for this procedure is to insure that the effect of the second stage ratio adjustment to "other races" is not weakened by the adjustment to blacks. In the next stage separate ratios are computed by sex, race (white and black and other races), and 17 age groups, giving a total of 68 cells by rotation group. The age-sex-race groups are given in Table 4 as well as the second stage adjustment computed over all rotation groups in March 1975. These factors were not the actual factors used, since they were not computed by rotation group and they assume that the intermediate adjustment was not previously applied.

## Table 4. CPS Second Stage Ratio Adjustment Factors for Total Population By Age, Color and Sex $\underline{1}/$ ALL ROTATION GROUPS, MARCH 1975

Age	Whi	te	Black an Rac	d Other es <sup>2</sup> /
	Male	Female	Male	Female
Total	1.04901	1.02342	1.15468	1.07532
14-15	1.01927	.99287	1.02938	1.01576
16-17	1. <b>05079</b>	.97006	.98710	1.10733
18-19	1.08621	1.02334	1.18278	1.14973
20-21	1.04730	1.02451	1.53855	1.18612
22-24	1.12071	1.13036	1.17701	1.07878
25-29	1.07204	1.02624	1.24781	1.06004
30-34	1.03480	1.00931	1.26153	1.13815
35-39	1.07660	1.03174	1.07273	1.13769
40-44	1.05811	1.02 <b>3</b> 47	1.28741	1.10292
45-49	1.04750	1.00498	1.24003	1.03712
50-54	1.01799	1.01848	1.10833	1.07060
55-59	1.08018	1.02333	1.07344	1.02268
60-61	1.02980	.95590	.99989	1.02240
62-64	1.03221	1.07219	1.14459	.92967
65-69	1.02640	1.03032	.99706	1.05466
70-74	.96956	.98466	1.00452	.85741
75+	.98257	1.04738	1.29061	1.10894

1/ Bailar, Bailey, and Corby, 1977

2/ The factors for Black and Other Races indicate the seriousness of the undercoverage problem. These factors are not the actual adjustment factors used.

It has been shown that the ratio estimates based on age-sex-race group reduce the sampling variability of the estimates. The result of the weighting procedure is that records have weights that vary considerably. Table 5 shows the range of weights applied to records in March 1975. It is assumed that this differential weighting will reduce both biases and variances.

# B. Potential Problems with the Weighting Procedure

Some implications of the weighting procedure are as follows:

1. There is no known unbiased method of adjustment for nonresponse and undercoverage. The basic assumption is that the characteristics of the nonrespondents and missed persons are similar to those of respondents with similar demographic characteristics. An investigation of the nonrespondents was attempted in 1965 by Palmer and Jones (1967) when an intensive field followup was conducted. The results indicated that the noninterview adjustment procedures did not distort the number of employed persons. However, the results are inconclusive since less than half of the were interviewed. nonrespondents Specifically, refusals were not included in the study. It is not unreasonable to assume that the refusals have a different employment rate than the interviewed. The nature of the noninterviews has been continually changing in the last few years. In 1970 the overall Type A rate was 4.0 percent. Of these, 28 percent were not-athomes, 23 percent were temporarily absent, 39 percent were refusals, and 9 percent were "other". In 1976 the overall Type A rate was 4.4 percent. Of these, 19 percent were not-at-homes, 17 percent were temporarily absent, 59 percent were refusals, and 5 percent were "others". The percentage of refusals is growing, and we do not know the effect of the noninterview adjustment for

Table 5. Maximum, Minimum and Average Weights for Records in 13 Relationship Categories, March  $1975\frac{1}{2}$ 

Relationship Category	Maximum	Weights Minimum	Average
Male head with relatives	7448.80	33.56	1645.03
Male head without relatives	8006.21	206.62	1679.78
Wife of head	7215.72	31.46	1604.91
Female head with relatives	8549.27	33.04	1621.36
Female head without relatives	8288.90	144.76	1612.22
Male child related to head	6666.70	27.12	1617.02
Female child related to head	6597.57	30.52	1551.92
Male relative (over 18)	7060.87	67.29	1695.01
Female relative (over 18)	6296.77	39.30	1625.48
Unrelated male child	6365.42	206.62	1736.01
Unrelated female child	4496.97	1153.54	1628.61
Unrelated male (over 18)	3840.59	756.86	1695.63
Unrelated female (over 18)	4369.49	991.51	1638.46

1/ Bailar, Bailey, and Corby, 1977

those who refuse although we can place at least approximate bounds on the maximum possible effect.

2.

- To obtain the noninterview adjustment factor the household count within the raceresidence cells are tabulated by the race and residence of a designated member of the household, generally the wife of the head. However, noninterview adjustment factors are applied by the race and residence of the individual person. Thus mixed race households will not receive the same weight used in the calculation of the noninterview adjustment factor. However, mixed race households, particularly in the CPS, represent a small proportion of total households. CPS interviewers are instructed to ask race only of persons unrelated to the head; otherwise he/she records race by observation only.
- 3. There is a maximum factor of 4.0 used in the duplication control procedure to control the contribution of this procedure to variance, even when subsampling is at a rate greater than 1 in 4. However, this occurs very infrequently, since as previously mentioned, large USU's become a part of the rare events universe.
- 4. There is no evidence that the separate ratio adjustment for blacks and others which is immediately followed by the ratio adjustment for all age, sex, and race groups has a positive effect on the estimates. The factors for "other races" are highly variable.
- 5. The second stage ratio adjustment procedure is limited in that the census undercount is reflected in the independent estimates used to adjust the sample data during this stage of the weighting procedure. This need not be the case, however.

In summary, good measures are not available as to the impact of the weighting procedure on biases and on the effect of errors occurring in earlier stages of the survey.

## VIII. Conclusion

This paper has attempted to present in terms of an error profile the potential sources of nonsampling error in the Current Population Survey. Because nonsampling errors have historically been more difficult to measure than sampling error, less is known about their effect on the data. This is unfortunate since for some statistics they may dominate the mean square error. As indicated in the paper, studies have been performed on the Current Population Survey that give some indication of nonsampling error on specific survey operations, e.g. the Methods Test and the CPS/FOSDIC Study. In addition, a reinterview study is conducted each month to measure response bias. However, the reinterview survey has its limitations, some of which may be difficult to overcome, and many of the studies discussed in the paper were limited. As such, there are large gaps in our knowledge of nonsampling error. It is hoped that this error profile of a major complex survey such as the

Current Population Survey for which accuracy is a continuing concern, will stimulate the accumulation of more knowledge of this subject in all surveys.

#### Acknowledgements

This paper was written as a result of the work of the Subcommittee on Nonsampling Errors of the Federal Committee on Statistical Methodology. and the authors wish to thank the members of the Subcommittee for their suggestions, comments, and support. However, it represents the compilation of a great deal of work by Bureau staff members. The authors wish to acknowledge those contributions. They are appreciative of all the help they received from numerous persons in the Bureau who called to their attention unpublished Bureau studies and assisted in their interpretation. They wish to particularly thank persons in the Statistical Methods Division, the Engineering Division, and the Demographic Surveys Division, for making available requested information, suggesting sources of additional information, and for their many helpful comments and suggestions.

## References

- Baer, L., "Permit Status of TAR ED's in the Redesign," Bureau of the Census Memorandum to M. Boisen, January 5, 1973.
- [2] Bailar, B.A., Bailey, L., and Corby, C., "A Comparison of Some Adjustment and Weighting Procedures for Survey Data," Unpublished paper presented at Sampling Symposium, Chapel Hill, North Carolina, 1977.
- [3] Boisen, M., "Determination of Type of Segment for Tape Address Register (TAR) ED's Which Are in Non-permit Areas," Bureau of the Census Memorandum to J. Waksberg, July 26, 1971.
- [4] Boisen, M., "Final Report on CPS/FOSDIC Experiment," Bureau of the Census Memorandum to D. Levine, June 27, 1975.
- [5] Deighton, R., "Methods Test Report: Some Results of Experimentation with Self-Response Interviewing Procedures, February, 1965 - June, 1966," Bureau of the Census Memorandum, February 28, 1967.
- [6] Hanson, R.H., Statistical Methods in the Current Population Survey, Draft of the Revision of Technical Paper No. 7, 1976.
- [7] Hirschberg, D., Yuskavage, R., and Scheuren, F., "The Impact on Personal and Family Income of Adjusting the Current Population Survey for Undercoverage," Unpublished paper presented at the Annual Meetings of the American Statistical Association, Chicago, 1977.
- [8] Jablin, C., Unpublished Monthly Report of the Methods, Procedures, and Quality Control Branch of the

Demographic Surveys Division, Bureau of the Census, 1977.

- [9] Johnston, D.F. and Wetzel, J.R., Effect of the Census Undercount on Labor Force Estimates, Special Labor Force Report No. 105, Bureau of Labor Statistics, March 1969.
- [10] Jones, C. and Blass, R., "Population Undercoverage Estimates from the CPS-Census Match," Bureau of the Census Memorandum to H. Nisselson, January 17, 1975.
- [11] MacKenzie, W., Documentation of Annual Housing Survey Coverage Improvement Program, Bureau of the Census, 1977.
- [12] McPherson, J. L. and Volk, M., "FOSDIC Microfilm Problems and their Solutions," Unpublished Paper presented at the Eleventh Annual Convention of the National Microfilm Association, Washington, D.C., April 25-27, 1962.
- [13] Moye, D., "CPS Reinterview Results from the Listing Check, Check of Noninterview Classifications, and the Household Composition Check for 1975," Bureau of the Census Memorandum, May 21, 1976.

- Palmer, S., "On the Character and Influence of Nonresponse in the Current Population Survey," <u>Proceedings of the Social Statistics</u> <u>Section, American Statistical</u> <u>Association, 1967, pp 73-80.</u>
   Schreiner, I., "CPS Reinterview Results
- [15] Schreiner, I., "CPS Reinterview Results from the Listing Check, Check of Noninterview Classifications, and the Household Composition Check for 1976," Bureau of the Census Memorandum, April 27, 1977.
- [16] U.S. Bureau of the Census, Census of Population and Housing: 1970 Evaluation and Research Program PHC (E)-4, Estimates of Coverage of Population by Sex, Race, and Age: Demographic Analysis by J.S. Siegal, U.S. Government Printing Office, Washington, D.C., 1973.
  [17] Williams, L. E., "Methods Test Phase
- [17] Williams, L. E., "Methods Test Phase III: First Report on the Accuracy of Retrospective Interviewing and Effects of Nonself-Response on Labor Force Status," Bureau of the Census Memorandum to W.M. Perkins, June 24, 1969.

34