Accuracy of Data for Employment Status as Measured by the CPS-Census 2000 Match

FINAL REPORT

This evaluation reports the results of research and analysis undertaken by the U.S. Census Bureau. It is part of a broad program, the Census 2000 Testing, Experimentation, and Evaluation(TXE) Program, designed to assess Census 2000 and to inform 2010 Census planning. Findings from the Census 2000 TXE Program reports are integrated into topic reports that provide context and background for broader interpretation of results.

Thomas Palumbo
Paul Siegel

Assisted By Mai Weismantle

Housing and Household Economic
Statistics Division

USCENSUSBUREAU

Helping You Make Informed Decisions

CONTENTS

i

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

Introduction

This report presents the results of an exact-match study that used the Current Population Survey-Census 2000 Match to evaluate the labor force data in Census 2000 by making estimates of their content error, which refers to the accuracy of the data, as opposed to *coverage* error, which refers to how completely people and housing units are counted . The report describes the methods used to create the file for the Current Population Survey-Census 2000 Match and how the Match data were used to measure levels of content error.

For people in Census 2000 who were also in the Current Population Survey sample in February through May 2000, the Current Population Survey-Census 2000 Match brought together each person's census report with the same person's Current Population Survey report. Ideally, this linkage provided the opportunity to compare two independent observations (one from Census 2000, the other from the Current Population Survey) of the same event (the person's relationship to the work force at a particular time), and to use the outcome of the one observation (the person's labor force classification in the Current Population Survey) to ascertain the validity of the outcome of the other (the same person's labor force classification in Census 2000). The verdicts from these individual comparisons were combined to form a mosaic that, when viewed, so to speak, from various angles or through special lenses, revealed much about the accuracy of the Census 2000 employment-status estimates.

The Current Population Survey was used because it is considered to be the standard of comparison for census labor force data. The Current Population Survey is a large, well-designed sample survey that focuses on labor-force measurements, is conducted by trained and experienced enumerators, and is continuously fielded. Other things being equal, these attributes should make it more accurate than the multi-purpose, highly self-enumerated, and intermittent census.

Methods

Although there is considerable emphasis on small-area geography in Census 2000, for practical reasons, the analysis in this report was restricted to the national level. The study centered around a detailed cross-tabulation of the employment status in Census 2000 of the people in the civilian noninstitutional population 16 years and over, by their employment status in the Current Population Survey in the first month between February and May 2000 that they were represented in the Current Population Survey (this tabulation is the "mosaic" mentioned above). This primary cross-tabulation is weighted to national totals, and displayed for combinations of sex, age, race, and Hispanic origin groupings.

The cross-tabulation presents estimates of the quantities of *response error* in published census figures. A response error is said to occur when a person's labor force classification in Census 2000 as either *employed*, *unemployed*, or *not in labor force* differs from that same person's classification in the Current Population Survey. To make these quantities meaningful, two relative measures of response errors (percentage distributions) and two summary measures of response errors were derived from them; these derived measures are the focus of the report (they

represent, respectively, the metaphorical "angles" and "special lenses" mentioned above). The percentage distributions reveal the success rates of Census 2000 in classifying people to their *correct* (same as Current Population Survey) labor force categories and away from *incorrect* (different from Current Population Survey) categories. The summary measures are the *net difference rate*, an estimator of statistical bias (to the extent that the Current Population Survey accurately reflects reality) that can be used to adjust published census estimates; and the *index of inconsistency*, a measure to detect response variance that is especially useful for evaluating the adequacy of the data-collection instrument for providing valid measures of a characteristic. The derived measures are valid, of course, only to the extent that their underlying assumptions are met. Known and presumed departures of the methods and data of this study from these assumptions do not invalidate the results, but they do impose the need for caution in interpreting and applying them.

Response-error measures in exact-match studies are most valid and useful only when the classifications for all people in the scope of the study actually do represent separate and accurate observations of the same event for the same person. In the Current Population Survey-Census 2000 Match, this condition unfortunately is not true for the labor force classifications of many people, either because the timing of their Census 2000 observation differs from that of their Current Population Survey observation (different reference weeks), or one or the other of the corresponding observations is faulty (in which case the labor force classification was either *assigned* on the basis of incomplete information, or *imputed* when useable information was unavailable) . To address this concern – historically the bane of exact-match evaluations of census labor force data – the authors, after computing the response-error measures for all people in the Match, recomputed them for various subsets of people whose corresponding observations were judged to have a high likelihood of being accurate representations of the same event. The report describes the methods and criteria used to select these subsets, and compares the response-error measurements for them among themselves and with those of the Match population in general.

This report also presents the results of efforts to use the Current Population Survey-Census 2000 Match to gain insights into why the aggregate labor force estimates in Census 2000 differed substantially from the official estimates of the Bureau of Labor Statistics for the Census 2000 time period based on the Current Population Survey. Appendices report on research into the roles of the Census 2000 edit and imputation procedures, of the differences between the Current Population Survey and Census 2000 in their reference periods for employment status, and of several facets of the Census 2000 employment-status questions.

Findings

- Census 2000 and the Current Population Survey are reasonably consistent in classifying people to the employed and not in labor force categories, but they exhibit considerable variability in classifying people to the unemployed category.
- Previous studies of Current Population Survey-census employment classifications, which were done for the 1960 and 1970 censuses, *but not the 1980 and 1990 censuses*, revealed patterns similar to those described in the above statement. However, for Census 2000, the consistency for all three categories slipped somewhat from the 1970 levels, in spite of

efforts, particularly after the 1990 census, to make the census employment questions conform more closely with the Current Population Survey questions.

- As was true in the 1970 and 1960 studies, the values of the index of inconsistency for the unemployed category were in the high range (above 50), which suggests that improvements are needed in the method used to collect these data (if, indeed, the unemployed concept is measurable at all in a census context, or, more generally, outside a context like that of the Current Population Survey). The shortcomings of the Match methodology, especially as applied to a generally short-lived phenomenon like unemployment, probably exaggerated these values, however. Hence, considerable caution must be exercised in interpreting them.

- The analysis suggested that the failure of the census questionnaire to distinguish between active and passive methods of searching for a job, and between active job-seekers and discouraged workers, is an important, but likely not a decisive, factor in creating the census overcount of unemployed people compared with the count of the Current Population Survey.

- The results for the employed and not in labor force categories indicated that, although the census is able to measure these concepts reasonably well, improvements are needed. The study suggested, for example, that it may have been a mistake to use the Current Population Survey wording for the "work last week" question in Census 2000.

- The underestimate of employment and the overestimate of people not in the labor force in Census 2000 relative to the Current Population Survey is likely related to the failure of the census classification system to filter more employed people out of the not in labor force category and into the employed category. This failure may be related to the change in wording between the 1990 and 2000 census in the "work last week"question, which is the key question in the census decision to classify a person to the employed category.

- The difference between the reference periods for the labor force estimates of Census 2000 and the Current Population Survey is probably not a major contributor to the gaps between the two surveys' estimates.

- Census 2000 may have had problems correctly classifying the employment status of people who had a job or business in the census reference week, but who did not work during that week for various reasons. When the census successfully identified that such "absent" people had jobs, it often failed to determine that they were *not at work* in the reference week. This problem does not affect census estimates of employed people, but it has the potential to bias census data on the counts and characteristics of people at work; accurate data on the "at-work" population are critical for the census journey-to-work data that are used in transportation-planning studies. A worse problem for employment status is that Census 2000 sometimes failed to determine that absent people had jobs at all. This latter problem may be related to a failure of the census to clarify for people who were on maternity or paternity leave from jobs just how they should answer the question about temporary absences from work. The problem, however, can likely explain only a small part of the Current Population Survey-Census 2000 gap in corresponding estimates of employment.

- A tendency for people classified as employed in the Current Population Survey to be classified as *not* employed in Census 2000 appeared to be associated with particular age categories (16 to 19; 20 to 24 years; 65 years and over), class of worker categories (self-employed, unincorporated; without-pay worker), and educational attainment categories

(high school or less, no diploma). The finding suggests that some groups of workers may have had difficulty in understanding or correctly responding to the work-last-week question in the census. On the surface, it is consistent with the hypothesis that the increasing difficulty of the census to accurately measure employment status may be related to a growing presence in the workforce of people with nontraditional work arrangements, such as so-called contingent workers, for whom traditional census terms such as "work,"and "temporary absence" may be ambiguous, and, even more foreboding, for whom the official concept of employment status may be too rigid to describe their fluid relationships to the labor market.

Recommendations

- The results of this study should be useful in improving the quality of employment status data collected in future demographic surveys and censuses, particularly in the new American Community Survey (ACS), which uses the same employment questions as those used in Census 2000. Preliminary comparisons of aggregate-level American Community Survey labor force estimates with CPS estimates reveal that the American Community Survey has many of the same shortcomings relative to the CPS as Census 2000 does. The results of this Census 2000 evaluation should have considerable applicability to the American Community Survey. In particular, it is likely that the suggested problems with the Census 2000 questions discussed above will also be detrimental to the collection of accurate labor force data in the American Community Survey. Substantial research should be devoted to revising the American Community Survey questions by addressing these issues, though it should not be limited to them.
- Research aimed at improving the accuracy of the American Community Survey employment data through questionnaire improvements must include a large component of cognitive/behavioral research to develop new questions or approaches prior to pre-testing them. This evaluation suggests that the effects of shortcomings in the employment-status questions may be too subtle to detect in pre-tests alone.
- The American Community Survey will have the opportunity to collect labor force data through respondent-enumerator interactions, primarily via computer-assisted instruments, to a much greater extent than was true in Census 2000. The kinds of flaws in the Census 2000 employment-status questions, and by implication in those same questions in the American Community Survey, suggested by this evaluation, may be especially amenable to amelioration or even elimination through the use of such methods. Hence, special attention should be devoted to the development of the enumerator versions of the employment-status questions in the American Community Survey. In this effort, however, consideration must be given to how differences in the effectiveness of various collection modes may differentially impact the quality of the data for various segments of the population.
- Attempts to revise the American Community Survey employment status questions should proceed by evolutionary or incremental means. The evaluation results suggests that the existing questions, in spite of their likely flaws, likely have many virtues as well.
- Efforts should be made to measure the amount of bias and response variability in the American Community Survey employment status data. It is especially important to make users aware of the potentially serious consequences of response variability on the

accuracy of cross-tabulations of employment status data by other characteristics.

1. BACKGROUND

This report presents information on estimates of the content error associated with the employment status characteristic as measured in Census 2000. These estimates are based on comparisons of data for the same people from two independent sources (referred to in this study as *dual-observational* data): the Census 2000 long-form sample and Current Population Survey (CPS) in the months of February 2000 through May 2000.[1] The universe for this study was restricted to persons in the civilian non-institutional population, as identified in the CPS.[2]

The CPS has been conducted since the 1940s as an ongoing national monthly survey with a sample, in the year 2000, of about 50,000 eligible households per month.[3] Its purpose is to provide monthly and annual data on the economic and social characteristics of the population; it is specifically designed to produce the official household estimates of employment and unemployment for the United States each month .

The CPS is considered to be the standard for comparison for census employment data because the CPS data, although not likely to be error-free[4], are believed to be more accurate than the census data . Employment and unemployment estimates from Census 2000 generally differ from the official labor force data collected in the CPS and released by the Bureau of Labor Statistics, if, for no other reason, than that the design and collection methodology of the census and the CPS meet different purposes.[5] Census 2000 was primarily a mail-out/ mail-back data collection

---

[1] Corresponding studies were produced after the 1950, 1960, and 1970 censuses, but not after the 1980 and 1990 censuses. The report for the1960 study is: U.S. Bureau of the Census, *Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Accuracy of Data on Population Characteristics as Measured by the CPS-Census Match*, Series WER60, No.5., U.S. Government Printing Office, Washington, D.C., 1964. The report for the 1970 study is: U.S. Bureau of the Census, 1970 Census of Population and Housing, Evaluation and Research Program*, Accuracy of Data for Selected Population Characteristics as Measured by the 1970 CPS-Census Match*, Series PHC(E)-11, U.S. Government Printing Office, Washington, D.C., 1975.

[2] That is, the study excludes people on active duty in the U.S. Armed Forces, and people living in institutional group quarters such as prisons, hospitals, and nursing homes.

[3] The survey was initiated by the Works Project Administration (WPA) in 1940 and transferred to the Bureau of the Census in 1942. In 1959, the responsibility for planning, analysis, and publication of the labor force data was assigned to the Bureau of Labor Statistics. The CPS sample was expanded to approximately 60,000 eligible households in 2001.

[4] The report for the 1970 CPS-Census Match (U.S. Census Bureau, 1975) states (page 20): "Even though the CPS response is usually assumed to be the standard of accuracy, the CPS is obviously subject to some degree of error. In fact , for some characteristics, the CPS may be as error prone as the census."

[5] Specifically, at the national level, Census 2000 estimates of employment were considerably below, and estimates of unemployment above, the corresponding CPS estimates. Sub-national estimates from the two sources may exhibit even wider relative differences. See Table B and the Census 2000 Auxiliary Evaluation, titled "Comparing Employment, Income, and Poverty: Census 2000 and the Current Population Survey" available at

designed to collect general information about the labor force for very small geographic areas on a one-time basis.[6] CPS data collection consists of personal interviews of respondents by field representatives who ask a more extensive and detailed set of probing questions about labor force activities than it is possible to ask in the general-purpose census (see Box 1). The CPS utilizes a staff of full-time, experienced interviewers, and is conducted under more extensive controls and training procedures than the census. Appendix A describes other differences between the census and the CPS that support the presumption that the CPS employment estimates are more accurate than corresponding census estimates; the appendix also compares the questions and approaches of the two surveys, and elucidates the reasons for their major differences.

Box 1: Census and CPS Batteries of Employment Questions

Census Battery of Employment Questions (Form D-2, mail-out long form)
In Census 2000, individuals in the civilian non-institutional population were classified as employed if they responded "yes" to either questions 1 or 3. Otherwise, such individuals who were available to work (" yes" in 6) were classified as unemployed if they responded "yes" in 5, or "yes" in both 2 and 4. All remaining individuals (16 years and over) were classified as not in labor force.
1. LAST WEEK, did you do ANY work for either pay or profit? *If 1 is "no" , ask 2.*
2. LAST WEEK, were you on layoff from a job? *If 2 is "yes" , ask 4; otherwise, ask 3.*
3. LAST WEEK, were you TEMPORARILY absent from a job or business?
4. (*For people on layoff*) Have you been informed that you will be recalled to work within the next 6 months OR been given a date to return to work?
5. Have you been looking for work during the last four weeks?
6. Could you have started a job last week if offered one, or returned to work if recalled ?

---

http://www.census.gov/hhes/www/laborfor.html. A known problem in Census 2000 increased the number of unemployed people for some places with relatively large numbers of people living in civilian non-institutional group quarters, such as college dormitories, worker dormitories, and group homes, and may have affected comparisons of labor force data for higher levels of geography. For more information on this specific problem, see Data Note 4 in Chapter 9 of the technical documentation for Census 2000 Summary File 3 available at http://www.census.gov/prod/cen2000/doc/sf3.pdf.

[6] Roughly 70 percent of the population in the employment-status universe (people 16 years old and over) was enumerated on mail-out/mail-back forms (based on calculations of the authors; excludes people in group quarters).

CPS Battery of Employment Questions (Extracted from Figure 5-1, page 5-6, of Current Population Survey: Design and Methodology, Technical Paper 63RV (TP63RV), available at http://www.bls.census.gov/cps/tp/tp63.htm:

In the CPS, individuals are classified as employed if they say "yes" to question 2, or 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 ("yes" in 10 or 11) are classified as unemployed if they say "yes" to 5 and either 6 or 7, or if they say "yes" to 8 and provide in 9 a job search method that could have brought them into contact with a potential employer .

1. Does anyone in the 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?

10.Could you have started a job LAST WEEK if one had been offered?

11.(*For persons who answered "yes" in 6 or 7.*) Could you have returned to work LAST WEEK if you had been recalled?

The method of evaluating census results by using dual-observational data is only one of many possible evaluation procedures. Reinterview of a sample of cases, which are then matched with the census returns, and record checks, which consist of the matching of data collected in the census with independent records of establishments, are two other methods utilizing exact-match methods[7]. In addition, there are analytic methods of evaluation, such as modeling and comparisons of statistical aggregates from the census with aggregated data for the same population groups from other sources. For example, the Census 2000 Auxiliary Evaluation B.8, *Comparing Employment, Income, and Poverty: Census 2000 and the Current Population Survey*, compares aggregated (macro-level) employment estimates from the Census 2000 with corresponding estimates from the CPS.[8]

Response errors in the census employment-status statistics could have resulted from:
• erroneous or inconsistent reporting of characteristics;
• failure to obtain responses for all of the information requested from all of the people in the sample;
• errors in the clerical or computer processing of the data; or
• errors or imprecision in the editing and imputation procedures for unacceptable or missing data.

In this study, *unless otherwise noted*, the comparison of CPS and census figures reflects data in final form after all editing and imputation procedures have been completed. Therefore, the data presented here reflect the quality of published Census 2000 statistics. The data have been weighted to national totals, but, owing to the nature of the weighting procedures (see section 2.2.a, "Weighting", and section 3, "Limitations of the Data"), the resulting weighted estimates are only approximately equal to published CPS or Census 2000 figures, and cannot be substituted for published figures. Primarily for this reason, the main body of this report presents only percentage distributions and index measures of CPS-census classification comparisons; the numbers used to calculate these data are provided in Appendix C.

Although there is a considerable emphasis on geographic detail in the census, the analysis in this report is restricted to the national level. The cost of producing separate evaluations of each area for which census data are shown would be prohibitive. The measures of error presented here do not, therefore, necessarily apply to individual States, cities, or other local areas.

---

[7] A reinterview study was conducted as part of the evaluation program for Census 2000, but it did not include observations of employment status because the reference period of the original observations could not be replicated. A description of the study is presented in Census 2000 Evaluation B.5, *Census 2000 Content Reinterview Survey: Accuracy of Data for Selected Population and Housing Characteristics as Measured by Reinterview*, available at: http://www.census.gov/pred/www/rpts .

[8] Available at http://www.census.gov/hhes/www/laborfor.html .

2. METHODS

2.1 The CPS-Census 2000 Match dataset

The CPS-Census 2000 Match attempted to link the record for each address in the CPS sample in February, March, April, or May of 2000 (hereafter called the "combined-month" CPS sample) with its record in Census 2000. It also attempted to link the record for each person associated with an address in the CPS sample with his or her record in the Census. All interviewed and not-interviewed survey addresses were eligible for matching, except those identified by the CPS Field Representative as "outside the survey segment," "built after April 1, 1990,[9]" or "unused serial number or listing sheet line". All people associated with these addresses — including those described as "household members," "non-household members," and "proxy respondents" — were eligible for matching. The 53,000 person (and address) records which do not figure in official estimates from the CPS [10] were included in the match to make it possible to pursue research interests beyond those undertaken here, for instance to study the census characteristics of survey non-respondents, differences in the construction of household membership in Census and CPS, and so on. Two "match" datasets were created: one using the entire combined-month sample; the other using records from the March CPS sample only (hereafter referred to as the "March CPS sample" or simply as the "March sample"). The combined-month sample consists of all March addresses and those addresses from the February and April surveys which were not in sample in March and those addresses from the May survey which were not in sample in March or April. It includes the March special Hispanic supplementary sample[11]. The matching had four distinct stages: computer matching, computer geocoding, clerical review, and post-clerical manual processing.

2.1.a Computer Matching

The CPS files were matched to the Census unedited files containing names and addresses using the commercial software, *Automatch* . People were matched on name, sex, and birth date (reported or computed), and addresses were matched on address characteristics, in independent operations. In each, the search for matches was limited to the state in which the survey address

---

[9] The CPS building permit sample is designed to represent housing units constructed since the previous census. In order to maintain the "correct" probability of selection, units constructed since the previous census (1990 in the present case) are ineligible for inclusion in the area, group quarters, or (1990 Census) samples.

[10] Person records associated with Type A, B, or C non-interviewed housing units and records for non-members or proxies in interviewed housing units. (See Table B.)

[11] Thus the sample consists of addresses in all rotation groups in March, rotation groups in their first or fifth month in sample in April or May, and in their fourth or eighth month in sample in February. Only the March records were retained for addresses which were both in sample in February and included in the March Hispanic supplementary sample.

was located. Both Census and CPS addresses were standardized before matching, using the 2002 version of *CodeOne*, a commercial software package which attempts to resolve certain address ambiguities. Individual names were subject to much less refined standardization, beyond the removal of place holders like "child 1" or "Mr." (Full specifications for the computer linking are contained in Judson, et al., 2002.) It should be noted that in neither the Census nor the Current Population Survey are names *required*; they are merely conveniences used to distinguish the individuals in a household who are the subjects of the inquiry. Even when first and last name are present, many of the person records in the CPS do not contain sex or birth date. This is particularly true for people in noninterviewed housing units and non-members of the interviewed household. The computer match was restricted to eligible addresses, and to household members in addresses which were interviewed or which refused to be interviewed (type A noninterviews); non-members and proxies were excluded. All person records associated with addresses which contained no eligible people (type B noninterviews), or addresses which contained no eligible units (type C noninterviews) were withheld from computer matching. Table A summarizes the results of the operation.

Table A. – Match results for records eligible for computer match.

Disposition of survey records	Addresses		People	
	Records	**Percent of records**	**Records**	**Percent of records**
Total	109,654	100.0%	230,774	100.0%
Linked by computer	87,534	79.8%	197,059	85.4%
Linked after computer	15,136	13.8%	16,850	7.3%
Linked in final dataset (computer/clerical/post-clerical)	102,670	93.6%	213,909	92.7%

Source: Unpublished tabulations of the CPS-Census 2000 match dataset.
Note: Only members of occupied addresses (interviews and refusals) in the sampling frame were submitted for computer linking. The total numbers of CPS person records in the file, and the number linked in the final dataset, are 275,883 and 219,710, respectively.

The computer match linked about 80 percent of the eligible survey housing unit records to Census housing unit records and about 85 percent of eligible survey person records to Census person records. Among the records which were submitted to computer matching, an additional 14 percent of addresses and 7 percent of people were matched by subsequent stages of the process (which will be explained below).

2.1.b Computer Geocoding

All survey addresses were submitted to the Census Bureau's Geography Division for geocoding. This was the only way to associate the CPS addresses with areas recognized in the administration

of Census 2000 and the organization of its data records. These areas formed the basis for grouping CPS addresses and candidate census records into about 20,000 "work-units" – small "local" batches for clerical review. This grouping reduced the number of Census 2000 address records that had to be handled from 116 million – the entire census – to about 15 million, and the number of person records from 280 million to about 35 million.[12] The geocoding also identified blocks adjacent to each block to which an address was geocoded, and permitted the identification of the set of Census 2000 maps that had to be made available (as viewable digital files) to the clerical analysts.

2.1.c Clerical Matching

Clerical analysts at the Census Bureau's National Processing Center reviewed the computer match links between Census and CPS records and attempted to find links for people and addresses not linked by computer. Neither the Census returns nor the CPS interview records were available on paper. The CPS is conducted as computer-assisted telephone or personal interviews (CATI/CAPI) and, for the most part, never exists on paper. There was no system in place to permit access to the paper enumerator-filled and mail-return Census 2000 questionnaires after they were converted to digital images and then to electronic data files. Special software was created to assemble a database of survey records, census records, and census maps. The heart of the database is the "work unit" – the restricted set of Census 2000 records to which a small set of survey records could be linked. Because the only Census 2000 records to which a CPS address could be linked were those in its work unit, the logic of their construction is worth examining briefly. (Full details are available in Gunnison, 2002a.)

Survey addresses in the "area" and "unit" sampling frames were geographically clustered, and therefore treated similarly, and differently from survey addresses in the "permit" and "group quarters" sampling frames. Each sampling segment (identified by distinct primary sampling unit (PSU) and segment numbers) in the area and unit frames is a distinct work unit. Every Census 2000 block to which a survey address in the work unit is linked by geocoding is part of that work unit.[13]

Survey addresses in the "permit" and "group quarters" sampling frames which were geocoded to blocks in work units formed from addresses in the area or unit frame were assigned to those work units. The remaining geocoded addresses in the permit or group quarters frames were assigned to distinct work units for each Census 2000 block to which they were geocoded. Permit and group quarters addresses that were not geocoded, but were linked to Census 2000 addresses by the computer matching, were assigned to the block to which that address was

---

[12] As a byproduct, this geocoding provides census housing unit identification numbers in most cases, because the first stage of geocoding addresses consists of standardizing them (using the Census Bureau's address standardizing software) and matching them to the Census Bureau's Master Address File.

[13] A Census 2000 block can be part of more than one work unit, as will become apparent below.

geocoded [14]. If this block was not already in a work unit, it became a work unit. Permit and group quarters addresses that were not geocoded, and were not linked to Census 2000 addresses, but which contained one or more people who were linked to a Census 2000 person by the computer matching, were assigned to the block containing the first linked person [15]. If this block was not already in a work unit, it became a work unit. Finally, each remaining survey address – that is, a permit or group quarters frame housing unit which was not geocoded and not computer matched and contained no linked people – was assigned a distinct work unit number. There were 423 work units with no census addresses – 416 with one survey address and seven with more than one. Of the 454 CPS addresses in these work units, three quarters were either vacant or ineligible units (type B or type C noninterviews) in the survey; 79 were survey interviews; and 26 were refusals (type A noninterviews).

It is not difficult to see the crucial influence of geocoding on the success of the clerical matching. If the analysts were presented with census addresses from the wrong area, finding an address to which to link a survey household was impossible. Such was the case for about 500 survey addresses (about 0.5 percent) where the clerical analysts indicated they could tell the address they were attempting to match was likely in a block for which they had not been provided the Census records. Providing census addresses for a one-block ring around blocks to which survey addresses were geocoded offered some margin for error, and the links formed in the computer matching phase offered further cushioning . Nevertheless, geocoding to blocks is much less precise in some areas than in others, and this doubtless played a role in the differential success of matching addresses and people from the four sampling frames of the CPS.

Working with the output of the computer match, the clerical analysts were required to attempt to find links for survey people and addresses without them, review "weak" links – links regarded as suggestive, but not strong enough to go unexamined – formed in the computer matching, and review multiple links for the same address or person. They could search for matches within the records in a work group. Special computer software allowed them to "view" selected survey and census records simultaneously, and to form links between them. (Gunnison, 2002a, 2002b, 2002c, 2003.) The software made the appropriate digital Census 2000 maps for a particular work unit accessible to an analyst. Analysts also had access to paper copies of the CPS field maps on which area and permit frame sample addresses were annotated, and to copies of the original CPS field listing sheets. They had access to all the links formed in the computer match, and many were changed or edited as the review proceeded.

Review of address matches was conducted separately from review of person matches, in the sense that only address characteristics were employed to link housing unit records. The last name of the first household member in a survey housing unit could be used as an aid in searching for

---

[14] Where the computer matching linked a survey address to more than one census address, the first was used here.

[15] Recall that only household members were submitted to computer matching.

census addresses, but could not form the sole basis for linking addresses. Where a survey address was linked to more than one census address, this set of census addresses was regarded as a set of duplicates, and one was chosen to represent the set and be the "primary" link to the survey address. The primary was chosen on the basis of the person links within the household, and is treated as the sole match. There were a small number of survey households which were duplicates. The software made no provision for them, and they had to be dealt with manually, in a post-clerical operation (see below). People were matched on the basis only of person characteristics available in the two datasets (race and Hispanic origin were excluded). Where primaries had to be chosen among sets of person records regarded as duplicates, links between the addresses containing them were taken into consideration.[16]

2.1.d Post-clerical processing

Several minor final manipulations were required to complete the match data:
- The unedited census file to which the CPS was matched contained some addresses recognized as represented by duplicate returns. One or more of the duplicate returns for a given address could have been deleted in subsequent stages of census processing. Some survey records were linked to the subsequently deleted members of such a group. Such links were manually reviewed to make sure that the survey record was linked to the retained census record.
- In the CPS, the rooms or apartments in group quarters are treated as separate addresses, while in the Census the entire group quarters is treated as a single unit. For example, in CPS each room in a college dormitory is a distinct addresses. In the Census the entire dormitory is a single group quarters unit. The match data had to be reviewed manually to provide links from the several units in CPS to the single unit in Census that would not make the CPS records appear to be duplicate records for the same Census record.
- The Census mail-back forms gathered detailed information for five or six people (depending on form type) and provided spaces on a roster on which to list the names of any other residents. The clerical match identified some possible links between CPS people and people on these rosters. Where "data-defined person" records were created in the Census to represent these people, the "possible" links were manually associated with them and treated as "matches". In most cases, these survey people are represented by imputed data in the Census.
- Samples of 5,000 computer matches of addresses and people were unlinked and sent to clerical review as not matched, in order to assess the quality of the computer links. About two dozen people and two dozen addresses were not linked by the analysts, and the analysts made different matches for 119 addresses and 110 people. Where the two operations produced different links, the clerical link

---

[16] Further information on the clerical matching operation can be found in Adams, 2003a .

was retained. Where the analysts failed to link an address or person that the computer matching had linked, the computer link was accepted.

- A handful of addresses (about 30 of the 109,654 address in the combined-month sample) were discovered to be in the CPS more than once with distinct identification numbers. They were retained as survey duplicates because they form part of the survey estimates. Not all are linked to Census people or addresses.
- After clerical review, there were 1,207 survey person records in groups of duplicates. There were 582 such groups, and in each group one survey record was designated the primary. These were reviewed, and, where necessary, the primary and duplicates were exchanged, in order to make sure that the primary was a household member in the survey and if possible, interviewed in March. A total of 86 people were switched from primary to duplicate and therefore 86 from duplicate to primary.

The success of matching the CPS and Census 2000 must be judged separately for each inquiry because the CPS file contains records for household members, some former household members, and non-member informants, for interviewed households and for some addresses in which interviews were not conducted in the reference month (see Table B). Different sets of cases will be relevant for different analyses. In many cases, the relevant figure is 93 percent of the members of interviewed survey households (people with positive weights in the survey) are linked to census records [17]. For other analyses, the relevant figure may be the 98 percent of interviewed survey addresses that were matched to census records, or the 83 percent of the members of housing units where a CPS interview was refused. It cannot be determined whether the unmatched survey addresses or people could not be linked to Census records because they were not included in the Census or because the information available for them was insufficient to form a link by any of the means employed. The fact, however, that the match rate for interviewed survey addresses exceeds that for members of interviewed survey units suggests that coverage of people within housing units (HU) plays a large role in the match rates.

---

[17] True for both the combined-month sample and the March sample.

Table B. – Matching results for people and addresses in the combined-month (February-May) sample and the March sample, CPS-Census 2000 Match.

CPS interview outcome	Addresses		People			
			All		Housing Unit (HU) Members[1]	
	Total	Matched	Total	Matched	Total	Matched
Combined February-May CPS Sample						
All Records	109,654	93.6%	275,883	79.6%	234,639	91.5%
Interviewed Housing units (HU)	85,943	98.3%	242,035	86.5%	222,453	93.0%
HU Type A – refusals[2]	6,503	96.6%	11,604	61.5%	7,590	82.7%
HU Types B and C[3]	17,208	69.4%	22,244	14.3%	4,596	32.1%
March CPS Sample						
All Records	64,739	94.0%	166,235	79.6%	141,710	91.4%
Interviewed Housing units	51,016	98.3%	145,775	86.1%	133,710	93.1%
HU Type A – refusals[2]	3,747	97.2%	7,160	64.4%	5,019	82.8%
HU Types B and C[3]	9,976	71.0%	13,300	15.7%	2,981	31.6%

Source: Unpublished tabulations of the CPS-Census 2000 match dataset.

[1] Armed Forces members are not treated as members of interviewed addresses in the combined-month sample. There are 701 of them, of whom 648 are matched. In March, Armed Forces members are treated as members of interviewed housing units.

[2] Occupied addresses which refused interview. Information on people may be available from prior month interviews or from proxy respondents.

[3] Addresses eligible for the sample which contain no eligible people (Type B) and addresses which are not eligible for the sample – not residential (Type C).

2.2 Inference

The CPS-Census 2000 match dataset can be regarded as a universe for analysis in its own right, but it is designed to offer a basis for inference to larger universes. Because the linking is

conducted *from* CPS records *to* Census records, the match dataset for the combined-month sample represents the CPS universe consisting of the civilian non-institutional population; the match dataset for the March sample represents the CPS universe consisting of the civilian non-institutional population, plus members of the Armed Forces living off base or with family on base. The match datasets include records for both interviewed and noninterviewed addresses, and for people who are not household members, in both interviewed and noninterviewed households, though the analyses presented here are limited to members of interviewed households – the cases which form the basis of the official estimates from the survey. This section briefly discusses two of the tools that permit inference from the match datasets to the larger universes – adjustments that have been made to the CPS survey weights to permit estimates of the CPS universe, and replicate variances to represent the uncertainty in those estimates arising from sampling. For a more detailed discussion of these matters, see Zbikowski, 2003.

2.2.a Weighting

The weights in the CPS survey files permit the construction of estimates of the civilian non-institutional population, or the civilian non-institutional population plus members of the Armed Forces living off base or with family on base. These weights must be modified to reflect its construction in order to permit similar estimates from the CPS-Census 2000 match datasets.

Weights for the dataset for the combined-month sample must reflect the use of cases from samples over several months and the inclusion of the March Hispanic supplement cases. Weights for individual cases must be <u>reduced</u> because there are cases from extra months in the sample. The inclusion of the March Hispanic supplementary sample cases complicates matters further, because these cases do not ordinarily receive weights for estimating the civilian non-institutional population. The weights used are based on the two-stage weights, not the composite weights designed for labor force estimates from CPS. This basis provides weights for many analytic foci, but it means that estimates from this sample will differ from the official published estimates of labor force categories. Weights for the match dataset for the March CPS sample are also constructed from the two-stage weights, but do not require the modifications outlined above.

Records for some members of interviewed survey households could not be linked to records from Census 2000. This necessitates a nonmatch adjustment to the weights. For the combined-month sample, the nonmatch adjustment makes the sum of the adjusted weights for matched people equal the sum of the unadjusted weights for matched and unmatched people within cells formed by state, MSA (metropolitan statistical area) status (MSA Central City, MSA non-central city, non-MSA), and interview month (March and April vs. February and May). Then a second-stage ratio adjustment (see below) is applied to the results of this adjustment. Subsequent analysis suggests that these cells might not be the most useful in reducing nonmatch bias, but the second-stage adjustment appears to have compensated for it, so that overall, the nonmatch adjustment for the combined-month sample seems adequate. A brief digression into nonmatch rates may clarify the issue.

Of the 15,000 unmatched civilian members of interviewed survey households in the combined-month sample, 82.6 percent (12,783) are in addresses that *are* linked to a census address. Some survey characteristics of these people are shown in Table C. The Census and CPS often differ in the address they attribute to college students – for the most part, CPS treats them as members of their parents' household, while Census treats them as residents of college group quarters – and, sure enough, 20.9 percent of survey records for people age 16 to 24 who are enrolled in college full-time are not linked to a record for a person in Census 2000, though their survey address is [18] . This is the only sub-population of the CPS which would be expected to have high nonmatch rates because of differences between the Census and the CPS. But full-time college students are a relatively small population, and these nonmatched students form only 8.7 percent of the unmatched survey people in matched addresses. Only 9.5 percent of the survey records of people age 16 to 24 who are not full-time college students are in addresses that are matched to a census address but not linked to a census person, but this is a somewhat larger population, so they comprise 15.7 percent of the unmatched survey people in matched addresses.

The population ages 45 years and over has a relatively low fraction of people who are not matched but live in matched addresses – 3.1 percent – and it is *under*-represented among the nonmatched survey people in matched addresses – 18.8% of nonmatched people in matched housing units are in this age group, while almost twice the fraction (34.5%) of all members of interviewed housing units are. All other age groups except those under age 9 have nonmatch rates that are close to the average for all people – 5.8 percent – and form fractions of the unmatched survey people in matched addresses that are similar to their proportion of the total population. By way of contrast, the percentages of survey people in Central Cities of MSAs, outside the Central Cities but in MSAs, or in non-MSAs who are in survey addresses matched to census addresses but not matched to a census person record, are 7.6 percent, 5.0 percent, and 4.9 percent, respectively.

Table C. – Unmatched people in matched housing units – civilian members of interviewed housing units, combined-month sample. (Unweighted.)

Sub-population	Unmatched members of matched interviewed housing units†			Members of interviewed housing units†	
	Rate*	Number	Percent	Number	Percent
Total	5.8%	12,783	100.0%	222,453	100.0%
Under 9 years	7.3%	2,146	16.8%	29,362	13.2%
9-15 years	5.4%	1,302	10.2%	24,146	10.9%
16-24 years	11.8%	3,122	24.4%	26,471	11.9%

---

[18] For similar results in matching CPS and Census, see Bancroft, 1958, p 161; Fay, 1989.

Sub-population	Unmatched members of matched interviewed housing units†			Members of interviewed housing units†	
	Rate*	Number	Percent	Number	Percent
enrolled college full time	20.9%	1,117	8.7%	5,343	2.4%
not enrolled college full time	9.5%	2,005	15.7%	21,128	9.5%
25-44 years	5.8%	3,805	29.8%	65,620	29.5%
45 years and over	3.1%	2,408	18.8%	76,854	34.5%
Central City of MSA	7.6%	4,951	38.7%	64,792	29.1%
MSA, not Central City	5.0%	5,337	41.8%	106,593	47.9%
Not MSA	4.9%	2,495	19.5%	51,068	23.0%
Black	10.9%	2,587	20.2%	23,639	10.6%
Non-Black Hispanic	8.3%	2,437	19.1%	29,499	13.3%
All other	4.6%	7,759	60.7%	169,315	76.1%
Reference person w/relatives present	3.5%	2,070	16.2%	59,041	26.5%
Reference person with no relatives present	6.1%	1,602	12.5%	26,483	11.9%

Sub-population	Unmatched members of matched interviewed housing units†			Members of interviewed housing units†	
	Rate*	Number	Percent	Number	Percent
Spouse of reference person	2.9%	1,344	10.5%	45,688	20.5%
Child of reference person	2.2%	4,927	38.5%	69,551	31.3%
Other relative of reference person	12.1%	1,419	11.1%	11,733	5.3%
Non-relative of reference person	14.3%	1,421	11.1%	9,957	4.5%
"Owner"	4.0%	6,294	49.2%	156,781	70.5%
"Renter"	9.9%	6,489	50.8%	65,672	29.5%

$$* \text{ "rate"} = 100 * \frac{\textit{unmatched members of matched interviewed housing units}}{\textit{all members of interviewed housing units}}$$

†Excludes members of the Armed Forces.

 The percentage of records for survey people in matched addresses that are not matched to a census person record are 10.9 percent for Blacks, 8.3 percent for non-Black Hispanics, and 4.6 percent for all other survey people. For people in housing units that are owned with or without a mortgage this peculiar nonmatch rate for people is 4 percent, while for people in housing units that are not owned the rate is 9.9 percent, and "renters" are greatly over-represented among the unmatched people in matched addresses. Finally, household reference persons with other relatives present, their spouses, and their children are under-represented among the unmatched people in matched addresses, while "other relatives" and "non-relatives" of the reference person are over-represented.

In sum, the picture of nonmatch of survey to census people that emerges is largely one of within-household nonmatch, strongly responsive to age, race and Hispanic origin, tenure, and relationship to the householder. With the exception of full-time college students, some of these are precisely the dimensions on which the second-stage controls adjust the survey weights (see

U.S. Census Bureau, Bureau of Labor Statistics, 2000, Chapter 10.) Others represent dimensions subject to more "behavioral" explications of census undercoverage (see Martin and de la Puente, 1993). Thus, it is not surprising that the weights for the combined-month sample appear to be much more successfully adjusted for nonmatch than is true for those for the March sample, which did not undergo the second stage adjustment after nonmatch adjustment. This result is consistent with the " belief that the second-stage controls reduce the bias due to coverage errors" (U.S. Census Bureau, Bureau of Labor Statistics, 2000., chapter 15.) A new nonmatch adjustment using household and person characteristics has been applied to the March sample.

Most official estimates from the Current Population Surveys in 2000 use weights controlled to independent population estimates based on the 1990 Census. In the match dataset, this source of difference from Census 2000 is eliminated by using Census 2000 population controls. The cells in which this second-stage ratio adjustment is carried out are those of the 1990 Census-based sample design, but the control totals are taken from Census 2000 [19] .

When analysis focuses on characteristics measured only in the Census 2000 long form, the weights must be adjusted to represent the structure of the Census sample. This is accomplished by multiplying the adjusted CPS weight by the Census 2000 weight. This procedure was applied to the data in the tables in this report based on the match dataset for the combined sample.

2.2.b Variances

Variances for estimates from the match datasets are formed by using replicate weights representing 160 independent samples from the dataset. (See U.S. Bureau of the Census and U.S. Bureau of Labor Statistics, 2000, Chapter 14.) The replicates for the combined-month dataset are adjusted for nonmatch and then have the second-stage ratio adjustment applied. In the March dataset, the nonmatch adjustment is applied separately to each replicate after it has had the second-stage ratio adjustment. When analyzing data from the census long-form, the nonmatch adjusted CPS sample weights are multiplied by the Census 2000 sample weights to represent the effect of Census 2000 sampling. The estimated sampling variance of an estimate is obtained by using the adjusted replicate weights to make 160 separate estimates, and estimating their variance as $4 * \dfrac{\sum_{i=1}^{160} (X_i - X_0)^2}{160}$, where X_0 is the statistic of interest estimated on the full sample, X_i is the estimate formed using the ith set of replicate weights, and the fraction 4/160 represents the treatment of self-representing and non-self-representing primary sampling units. (See U.S. Bureau of the Census and U.S. Bureau of Labor Statistics, 2002, chapter 14.)

---

[19]This set of weights was made available for this study by the Small Area Income and Poverty Estimates staff of the Housing and Household Economics Statistics Division (HHES) of the Census Bureau.

2.3 Data Presentation

This report presents estimates of response variability and net error or bias [20] associated with census statistics on employment status. The modern concept of employment status, developed in essentially its current form in the 1930s , is intended to measure the success of the labor market in gainfully employing all people actively interested in such employment. The concept is defined operationally in the same way in both the CPS and the census. It classifies people 16 years and over – the working-age population — in the civilian noninstitutional population into five categories:

- employed, at work;
- employed, with a job, but not at work;
- unemployed, on layoff;
- unemployed, looking for work;
- not in labor force [21].

These categories are collapsed into three major categories: employed; unemployed; not in labor force. Box 2 presents the definitions of the categories of the employment status concept.

---

[20] The estimates of bias presented in this report are measures of the discrepancy of census estimates from corresponding CPS estimates, and not necessarily from the truth. They measure departures from the truth only to the extent that the CPS faithfully represents the truth.

[21] This category represents a collapsing of three categories in the Current Population Survey: not in labor force - retired; not in labor force - disabled; not in labor force - other.

Box 2: Definitions of Categories of the Employment Status Concept Used in the Census and the CPS

Beginning in 1970, the census has used the following definitions of employment status concepts, which are the same official concepts used in the Current Population Survey. In the census, these concepts are applied through a series of questions (see Box 1) to identify, in this sequence: (1) people who worked at any time during the reference week; (2) people who did not work during the reference week, but who had jobs or businesses from which they were temporarily absent (excluding people on layoff); (3) people on temporary layoff who expected to be recalled to work within the next six months or who had been given a date to return to work, and who were available for work during the reference week; and (4) people who did not work during the reference week, who had looked for work during the reference week or the three previous weeks, and who were available for work during the reference week.

Employed. All civilians 16 years old and over who were either (1) "at work" — those who did any work at all during the reference week as paid employees, worked in their own business or profession, worked on their own farm, or worked 15 hours or more as unpaid workers on a family farm or in a family business; or (2) were "with a job but not at work" — those who did not work during the reference week, but who had jobs or businesses from which they were temporarily absent because of illness, bad weather, industrial dispute, vacation, or other personal reasons. Excluded from the employed are people whose only activity consisted of work around their own house (painting, repairing, or own home housework) or unpaid volunteer work for religious, charitable, and similar organizations. Also excluded are all institutionalized people and people on active duty in the United States Armed Forces.

Unemployed. All civilians 16 years old and over were classified as unemployed if they were neither "at work" nor "with a job but not at work" during the reference week, were actively looking for work during the last four weeks, and were available to start a job. Also included as unemployed were civilians 16 years old and over who: did not work at all during the reference week, were on temporary layoff from a job, had been informed that they would be recalled to work within the next six months or had been given a date to return to work, and were available to return to work during the reference week, except for temporary illness. Examples of active job seeking methods are:
• Registering at a public or private employment office
• Meeting with prospective employers
• Investigating possibilities for starting a professional practice or opening a business
• Placing or answering advertisements
• Writing letters of application
• Being on a union or professional register

Civilian labor force. Consists of people classified as employed or unemployed in accordance with the criteria described above.

Not in labor force. All people 16 years old and over who are not classified as members of the labor force. This category consists mainly of students, individuals taking care of home or family, retired workers, seasonal workers enumerated in an off-season who were not looking for work, institutionalized people (all institutionalized people are placed in this category regardless of any work activities they may have done in the reference week), and people doing only incidental unpaid family work (fewer than 15 hours during the reference week).

Reference week. In the census, the data on employment status related to a one-week time period, known as the reference week. For each person, this week is the full calendar week, Sunday through Saturday, preceding the date the questionnaire was completed. This calendar week is not the same for all people since the enumeration was not completed in one week, nor is the week necessarily interpreted the same way by respondents to the mail form. The occurrence of holidays during the enumeration period probably had no effect on the overall measurement of employment status. The CPS data always relate to the calendar week during the month that contains the 12th day of the month.

The tables in this study focus on estimates of the differences between the employment-status classifications of people in the census and of these *same* people in the CPS. The basic data unit represents the union or match of two observations of the same individual: one observation of the employment-status classification of the person in the census, and the other, the employment-status classification of the identical person in the CPS. For this reason, the data are referred to in this study as *dual-observational* data. *Unless otherwise noted*, all tables are based on the records in the CPS-Census 2000 Match dataset created from the entire combined-month sample (see Section 2.1); many tables have additional restrictions that are indicated in the table headings. Note that, hereafter, the terms *CPS-Census 2000 Match* and *Combined-month CPS-Census 2000 Match* are used interchangeably in this report, and relate exclusively to the dataset created from the entire combined-month sample.

Each person in the scope of the study has one and only one census employment-status classification [22]; this census value is matched with the person's CPS employment-status classification for the first month in the February 2000 to May 2000 period that the person was in the CPS sample. Some tables display both dual-observational comparisons of employment status, and provide aggregate-level groupings of people based on their age, race, and Hispanic origin

---

[22] This classification is represented by the value of the employment status recode (ESR) for the person on the Census 2000 Sample Edited Detail File (SEDF).

characteristics as observed in the CPS [23]. For instance, Table 1A (see section 4.5) presents a percentage distribution of people in each of the employment-status categories in the CPS by their employment-status classification in the census, by their age, race, and Hispanic origin characteristics in the CPS.

The data in these tables are based on sample statistics that have been weighted to population totals. Data are presented as percentages and indices. The nature of the weighting procedure invalidates any direct use of the figures in this report for making absolute estimates of the people in the employment-status categories in Census 2000 (the reader should use the published figures instead) and for comparing differences in absolute figures between the census and CPS. Descriptions of the index measures are presented in the following section; the computational forms of these measures are provided in Appendix H (also see the publication, U.S. Bureau of the Census, 1970 Census of Population and Housing, Evaluation and Research Program, *Accuracy of Data for Selected Population Characteristics as Measured by the 1970 CPS-Census Match*, Series PHC(E)-11).

2.4 The Concept of Response Error

For categorical (qualitative) measures, such as employment status, a response error results, in simple terms, from the assignment of a person to an incorrect category in a classification system. For example, if a person actually belongs in the employed category, a response error will result from the assignment of that person to one of the other categories. Such errors affect census categorical data in at least two ways: (1) the errors may introduce bias into the estimates of the population characteristic; and (2) the errors distort the relationships among variables. If only a single observation is available for each person, it is not possible to directly estimate the bias and variability associated with the classification process, although the bias may be estimated when aggregated data from an independent source are available. For this evaluation, estimates of response error for the employment-status characteristic were obtained by comparing the classification made in the census with the corresponding classification made in the CPS across all people for whom both a census observation and a CPS observation were available. CPS classifications are not error free, so it is *not* appropriate to say that a difference between the census and the CPS classification for a person always reflects error in the census [24]. Furthermore, for employment status, the difference may reflect a true change in category because of the close connection between an individual's employment status and the timing of its observation, a subject discussed below. Indeed, because of timing, differences between the CPS and census classifications may reflect valid changes in employment status to a greater extent than response errors. Even so, such comparisons do, among other things, provide an estimate of the variability

---

[23] The age variable shown in the boxheads of the tables under the "Census classification" heading is based on age data collected in Census 2000.

[24] See the discussion in Appendix A concerning bias in CPS estimates.

in the classification of an individual over repeated trials and, therefore, provide meaningful insights into the quality of the census data (see section 2.5.b).

2.5 Measures of Response Error and Variability

This report presents three measures of response error and one measure of response variability based on exact-match comparisons of CPS and census classifications. The percentage distributions exemplified in Tables 1A and 1B are *descriptive measures* of response error; the *net difference rate* and the *index of inconsistency* exemplified in Table 1C are *summary measures* of response error and response variability, respectively [25] [26].

2.5.a Descriptive Measures of Response Error

2.5.a.i Census-Based Percentage Distributions

The first descriptive measure is the percentage distribution of the people in each census employment category by their CPS categories, which is shown in Table 1A (see section 4.5) . The percentage for a category lying on the diagonal (shaded cells) of the table represents the proportion of the people in the category whose census classification matched their CPS classification. (For example, the "All Races, Both Sexes" rows of the "16 years and over, Unemployed" column for Table 1A indicate that of all those people classified as unemployed in the census, 33.2 percent were also designated as unemployed in the CPS.) The off-diagonal percentages represent various kinds of mismatches. The data provide insights into the capacity of the census classification system to divert, or screen-out, from a category those people who belong in another category. Hence, the census-based percentage distributions are indicators of both the compositional integrity of the census categories [27] and the filtering-out capability of the census classification system (to return to the above example, Table 1A indicates that 32.0 percent of the people in the unemployed category in the census were designated as employed, and 34.8 percent as not in labor force, in the CPS; hence, the census failed about two-thirds of the time (32.0

---

[25] The net difference rate, as applied in this report, measures response bias of the census in relation to the CPS, and not necessarily in relation to the truth. The index of inconsistency measures the impact of response errors on the total variance of a variable, and is not a direct measure of response error. See the appendix in U.S. Bureau of the Census, *Evaluating Censuses of Population and Housing*, Statistical Training Document, ISP-TR-5, Washington, D.C., 1985.

[26] The estimates in this report are based on responses from a sample of the population. As with all surveys, estimates may vary from the actual values because of sampling variation or other factors. All comparisons made in this report have undergone statistical testing and are significant at the 90-percent confidence level unless otherwise noted.

[27] To use an analogy, the on-diagonal percentages can be thought of as representing the native elements of a mixture, and the off-diagonal percentages as representing foreign elements. The greater the proportion of native elements, the greater the purity of the mixture. Applied to the term "compositional integrity" as used here, this analogy means that the greater the on-diagonal percentage, the greater the compositional integrity of the category.

percent + 34.8 percent = 66.8 percent) to screen *not*-unemployed people out of its unemployed category) **.**

2.5.a.ii CPS-Based Percentage Distributions

The CPS-based percentage distributions, as shown in Table 1B (section 4.5), provide an indication of the capability of the census classification system to filter, or screen, *into* a category those people who belong in the category. In this sense, they are the complements of the census-based distributions, which, as described above, are related to the screening *out* capacity of the census. Under the supposition that the CPS classification represents a person's true category, the percentages in the on-diagonal (shaded) cells of the CPS-based distributions indicate the success rate of the census classification system in directing people to their true category; the off-diagonal percentages reflect census failures. (For example, Table 1B indicates that the census succeeded about 90 times out of 100 (or 90.6 percent) in classifying employed people to their true category according to the CPS: see the "16 years and over, Employed" row in Table 1B for "All Races, Both Sexes;" the corresponding census success rate for unemployed people in the CPS was 40.2 percent: see the "16 years and over, Unemployed" row in Table 1B for "All Races, Both Sexes:).

2.5.b Summary Measures of Response Error and Variability

The two summary measures presented in this report, the net difference rate and the index of inconsistency, describe, respectively, the amount of bias in the data and the impact of response errors on the variability of the data. Appendix H presents the formulas for computing the measures. All summary measures of response error have been multiplied by 100 so that the computed values can be discussed as percentages.

2.5.b.i Measure of Bias

Response bias reflects a systematic pattern or direction in the difference between the respondents' answers to a question and the "correct" or " true " answers [28]. The measure of bias presented in this report is the *net difference rate.* For categorical variables like employment status, the net difference rate for a particular category describes the difference between the census proportion of persons in the category and the CPS proportion of persons in that category. A positive value of the net difference rate indicates that the proportion of persons in the category according to the census is greater than the corresponding CPS proportion, whereas a negative value indicates that the census proportion is less than the corresponding CPS proportion. A difference between the census and CPS estimates that is beyond what is expected from sampling variability may indicate the presence of bias in the census statistic when, as is assumed for employment status, the CPS data are considered to be more accurate. The use of the net difference rate as a measure of bias, however, is fully justified only if the CPS estimates

---

[28] Bias is the difference between the expected value of a statistic and its true value.

themselves are free of bias, a condition not likely to be generally true [29] [30]; hence, the interpretation of the results of such an application of the net difference rate must be made cautiously. For a given category, the index tables displayed in this report show the proportion of persons in the category according to the CPS (in the "Percent in class CPS" column in Table 1C) as well as the net difference rate. The sum of these two values equals the proportion of persons in the category according to the census (shown in the "Percent in class: Census" column).

Another measure of bias for a given category can also be derived. This measure, referred to as the *net shift*, is obtained by dividing the net difference rate for the category by the best estimate of the proportion of persons in that category – considered to be the CPS estimate for this study. The net shift, however, is not shown in this report since the net difference rate, having a smaller sampling error than the net shift, provides a somewhat more reliable estimate of bias.

2.5.b.ii Measure of Response Variability

The measure of response variability presented in this report is the index of inconsistency. An oversimplified but nontechnical definition of the index is that it is the ratio of the *simple response variance* – a measure of the average variability, across units, of responses to the same question over repeated trials – to the total variance, a quantity that includes the sampling variance [31]. The index is a relative measure of response variance, showing the comparative effect that the simple response variance has on an estimate.

There are various ways of interpreting the index of inconsistency. Although each interpretation uses different terms, they are closely related. For this report, the index of inconsistency is interpreted as the complement of a measure of agreement between the census and the CPS responses. Viewed in this way, the index is the ratio of the observed number of response differences to the number that would have occurred if the cell counts had been formed by a random agreement mechanism based on the observed marginal distributions (census and CPS) . Under this interpretation, the index measures inconsistency (lack of agreement) on a scale from zero (perfect consistency or agreement) to 100 (complete lack of consistency or agreement) [32] .

---

[29] See the discussion concerning bias in the CPS in Appendix A.

[30] If corresponding CPS and census estimates are biased in the same direction (lower or higher than the true value), then the net difference rate understates the amount of bias in the census estimate and provides a lower bound on it. Conversely, if the corresponding estimates are biased in opposite directions, then the net difference rate overstates and provides an upper bound on the census bias.

[31] The sampling variance is the variability in the population of the characteristic being measured.

[32] Strict adherence to this interpretation requires acceptance of the unrealistic assumption that the index itself is free from error.

When the second observation is not an attempt to repeat the original interview procedure, but may represent an "improved" data source – as is presumed to be true for the CPS, the estimated index of inconsistency is almost sure to be an understatement of the ratio of the simple response variance of the original interview procedure to the sum of the sampling variance and simple response variance. The interpretation of the index given here is appropriate, however, even when the second observation is not an attempt to repeat the original interview procedure identically.

Values of the index of inconsistency are computed and displayed for each of the three major employment status categories: employed, unemployed, and not in labor force. An index of inconsistency for the entire distribution of people by these three categories, referred to as the *aggregate index of inconsistency* [33], is also displayed . This index is a weighted average of the individual indices computed for each category of the distribution. It indicates whether an entire variable has a problem, against, say, just one category in a multi-category variable. Conceptually, this measure is similar to the indices computed for individual categories. That is, it expresses the ratio of the observed number of differences in the entire distribution to the number of response differences that would be expected to result from a random association between the aggregate-index classifications on the first and second observations.

The index of inconsistency optimally estimates the ratio of simple response variance to the sum of the sampling variance and the simple response variance only when the census and the CPS meet the assumptions that they are independent replications of the same survey procedure under the same general conditions. The user is cautioned that the values for the index of inconsistency in this report may not fully meet the first of these assumptions – independence, and definitely do not meet the second – replication. Independence means that the response errors are not correlated between the census interview and the matched CPS interview. If the respondents remembered their answers to the census when they responded to the CPS, or vice versa, and consciously repeated them, the independence assumption would be violated. Lack of independence generally results in underestimates of response variance. Replication means that both observations for a matched case were obtained under the same conditions, an assumption clearly violated in this CPS-Census match study, although the extent of the violation is not known. Replication flaws lead to an underestimate of the value of the index that would result from a duplication of the census, and to an overestimate of the value from a duplication of the CPS. The magnitudes of any effects from violations of either the independence or replication assumptions on the estimates for the index of inconsistency in this report are unknown [34] [35].

---

[33] This index was formerly known as the "L-fold Index of Inconsistency."

[34] Lack of independence probably would make the net difference rate closer to zero than it would otherwise be. Perfect replication should yield a net difference rate of zero; to the extent that replication is imperfect, the net difference rate is likely to differ from zero.

[35] The net difference rate helps to indicate how well the census meets the model assumptions. A statistically significant NDR (i.e., statistically different from zero) suggests that the census may not replicate the original survey conditions as well as desired.

It should also be recognized that the level of the index is sensitive to the detail of the categories in which the data are collected or tabulated. As the detail of the categories is decreased, the index cannot increase and will most likely decrease. Thus, the response variance associated with a particular distribution may be decreased to some extent by collapsing the categories of that distribution.

2.6 Sampling Variability and Accuracy of the Estimates

The Census 2000 data contained in this report are ultimately based on the sample of households who responded to the Census 2000 long form. Nationally, approximately one out of every six housing units was included in this sample. As a result, the sample estimates may differ somewhat from the100-percent figures that would have been obtained if all housing units, people within those housing units, and people living in group quarters had been enumerated using the same questionnaires, instructions, enumerators, and so forth. The sample estimates also differ from the values that would have been obtained from different samples of housing units, and hence of people living in those housing units, and people living in group quarters. The deviation of a sample estimate from the average of all possible samples is called the sampling error.

In addition to the variability that arises from the sampling procedures, both sample data and 100-percent data are subject to nonsampling error. Nonsampling error may be introduced during any of the various complex operations used to collect and process data. Such errors may include: not enumerating every household or every person in the population, failing to obtain all required information from the respondents, obtaining incorrect or inconsistent information, and recording information incorrectly. In addition, errors can occur during the field review of the enumerators' work, during clerical handling of the census questionnaires, or during the electronic processing of the questionnaires.

While it is impossible to completely eliminate error from an operation as large and complex as the decennial census, the Census Bureau attempts to control the sources of such error during the data collection and processing operations. The primary sources of error and the programs instituted to control error in Census 2000 are described in detail in *Summary File 3 Technical Documentation* under Chapter 8, "Accuracy of the Data," located at *www.census.gov/prod/cen2000/doc/sf3.pdf* .

Nonsampling error may affect the data in two ways: (1) errors that are introduced randomly will increase the variability of the data and, therefore, should be reflected in the standard errors; and (2) errors that tend to be consistent in one direction will bias both sample and 100-percent data in that direction. For example, if respondents consistently tend to underreport their incomes, then the resulting estimates of households or families by income category will tend to be understated for the higher income categories and overstated for the lower income categories. Such biases are not reflected in the standard errors.

All comparisons made in this report have undergone statistical testing (Bonferroni Method) and are significant at the 90-percent confidence level, *unless otherwise noted*. Except as noted, a 90-percent confidence interval has been constructed and is shown in the tables for each of the estimates. If all possible samples were selected, each of them surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated for each sample, then approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average value of all possible samples. The average value of all possible samples may or may not be contained in any *particular* computed interval, but for a particular sample, one can say with specified confidence that the average of all possible samples is included in the constructed interval. These confidence intervals have been estimated from the sample results and provide a rough approximation of the extent of sampling error associated with each estimate. [36]

2.7 Use of Response Error Measures in Evaluating the Quality of Data

Of the two summary response error measures used in this report, the index of inconsistency probably provides the most information on the accuracy of the data collected, whereas the net difference rate can be used to adjust published census distributions. For categories in a distribution where the CPS-census comparisons suggest the presence of bias and the CPS data are assumed to be more accurate, the net difference rate can be added to the published census percent in the class to correct for the perceived bias (or more strictly, for the bias of the census estimate from the CPS representation of the truth). The index of inconsistency cannot be used to correct census distributions, but it provides insights into the reliability of the data presented in the published distributions (both one-way frequency distributions and cross-tabulations).

Both the index of inconsistency and the net difference rate capture the effects of response errors that occurred in the field stage of enumeration as well as the effects of subsequent clerical and computer processing operations. Thus, these summary measures indicate the amount of inconsistency and bias associated with the published census data, and provide valuable information about the quality of the data collected.

2.7.a Simple Distributions [37]

The net difference rate and its 90-percent confidence interval indicate whether systematic errors in reporting have introduced biases into the census distribution of people by employment status (provided, as assumed here, that the CPS data are more accurate than the census data). A bias in a particular category of a distribution is indicated when the 90-percent confidence interval of the net

---

[36] Further information on the accuracy of published Census 2000 data is located at *www.census.gov/prod/cen2000/doc/sf3.pdf* .

[37] Simple distributions are also known as "one-way frequency distributions."

difference rate does not include zero as a possible value. The sign on the limits of the interval indicates the direction of the bias – a positive value indicates that the estimated census percent in class is greater than the corresponding CPS percent, whereas a negative value indicates the opposite.

The indices of inconsistency associated with a simple distribution of a characteristic are important in evaluating the adequacy of the entire data collection process for providing valid measures of the characteristic. For the purpose of evaluating the adequacy of a data collection system, indices under 20 are considered small or low, those between 20 and 50 are moderate, and those over 50 are large or high. Large values of the index for a particular category or for an entire distribution are an indication that (1) improvements are required in the method used to collect the data, (2) the concept itself may not be measurable by a household survey method, or (3) respondents are not able to provide accurate information to the detail desired.

2.7.b Cross-tabulations

For one characteristic presented in a cross-tabulation with another characteristic (for example, employment status by age and race), erroneous classification into or out of the various categories of the distribution of either characteristic could introduce biases into the cross-tabulated data. In addition, the greater the index of inconsistency for each of the characteristics, the more likely it is that relationships between the characteristics are distorted. The expected effect is a reduction of correlation among characteristics. The indices may serve as a guide in making inferences about the quality of the cross-tabulated data. If the indices of inconsistency associated with each of the characteristics involved in the cross-tabulation are large (over 50), it is likely that the cross-tabulated data are subject to serious biases. In such cases, the user is advised to exercise caution when using the data, particularly when inferences regarding the relationships between the characteristics are desired. Conversely, if the indices of inconsistency associated with each of the characteristics are small (under 20), the user can be somewhat more confident about the accuracy of the cross-tabulated data. There are no specific guidelines appropriate for levels between these extremes (that is, for moderate-level indices). For these situations the user should again exercise caution when using the data and recognize that even a moderate degree of inconsistency in one or all of the characteristics can produce serious distortions in cross-tabulated data.

3. LIMITATIONS OF THE DATA

The match dataset used for this report is the one for the combined-month CPS sample. This dataset was designed to investigate differences between estimates. The data cannot be used to define "errors" without some additional assumptions or evidence from outside their scope. They do, however, throw light on some limitations of estimates designated as "official" (see U.S. Office of Management and Budget, 1978) .

There are certain differences between the estimates from the "official" Current Population Survey and estimates from the combined-month sample, which arise from its construction. Most notably, the combined-month sample should produce estimates which differ from those of any of the months which comprise it. For some purposes, e.g., the comparison of race and Hispanic origin responses in the Census and survey, the combined-month sample offers the advantage of more cases in sparse cells. For others, e.g., the comparison of reports on employment status in the two surveys, the difference in the week to which the question about activities last week refers can only be a disadvantage. One can, however, produce estimates from all (matched and unmatched) cases in the combined-month sample and compare them with a single-month estimate from the official dataset in order to gain some sense of the effect of the combination of months.

The rate at which interviewed survey addresses are matched in the census is high – 98 percent. The rate at which members of interviewed survey households are matched in the census (93.0 percent) is about the level achieved in earlier attempts to match the CPS and Census, and leaves room for uncertainty about the magnitude and source of CPS/Census differences for small groups. This uncertainty is *not* represented in the variances provided as guides to inference.

Table D – Matching experience in previous CPS-Census match studies

Year	Match rate	Comments
1950	98%	Matched people.
1960	92%	Only attempted to match people at CPS addresses which received the Census long form
1970	75%	Only attempted to match people at CPS addresses which received the Census long form
1980	94%	Calculated from data weighted to population estimates from P-sample data in the 1980 Post Enumeration Program.

Sources: Bancroft, 1958; U.S. Census Bureau, 1964; U.S. Census Bureau, 1975, p.20; Fay, 1988b

The match study was originally designed with a field follow-up phase to resolve ambiguous matches and unmatched addresses and people. For budgetary reasons, this phase was not carried out, and the match suffers accordingly, relative to the CPS-Census match in other years. [38]

---

[38] The March match file focuses on cases from the March Current Population Survey Annual Social and Economic Supplement, in order to maximize the observations available for analyzing data collected only in that survey, e.g., income and poverty. This choice might compromise use of these data to estimate Census "coverage", but there are far superior vehicles for that purpose, e.g., the Accuracy and Coverage Evaluation Survey. (See Petroni and Childers, 2003 and references cited there.) In any case, the temporal difference between the CPS interview date and the Census reference date – April 1, 2000 – provides an interval in which households or people might move, and legitimately have different addresses in the two surveys, thus confounding mobility and match failure. Choice of the April instead of the March CPS would have slightly lengthened this interval, since 90 percent of the 64,944 March

This study uses the match dataset for the combined sample to evaluate the employment-status item in Census 2000. Several assumptions underlie the use of this dataset for this purpose. To the extent that the assumptions are unfounded, the methods and analysis based on them may be flawed or weakened. Discussions supporting the claims to reasonableness of the assumptions, examining their basis in fact or theory, or explicating their implications, are distributed throughout the main text and the appendixes. The following chart briefly catalogs these assumptions and directs the reader to the sections of the text where they are discussed:

Assumption	Location of Discussion
Records for some members of the CPS households originally included in the matching operations for the CPS-Census 2000 Match could not be linked to records from Census 2000. If the response error distributions of these unmatched cases are generally different from those for the matched population, the distributions and summary measures shown in this report could be biased. The full extent of such differences is unknown, and the assumption was made that nonmatch bias does not appreciably affect the validity of the statistics shown in this report.	Section 2.2.a
The elements of the operational definition of the employment status concept used in the CPS and the census are objectively observable	Section 1; Box 2 in section 2.3
The CPS-Census 2000 Match can be used to measure bias and response variability (at least the impact of simple response variance) on the Census 2000 estimates of employment status	Section 2.5
As a means of measuring employment status, the CPS methodology is superior to the census methodology	Appendix A

2000 CPS household interviews were completed within 10 days of April 1, while only 25 of the 60,729 April 2000 CPS household interviews were completed before April 15.

The CPS classification of an individual's employment status is more likely to be accurate (to reflect the truth) than the census classification, given that their reference periods are identical	Appendix A
The reference period for the census classification of an individual's employment status may not be the same as that for the CPS classification	Section 4.1, Appendixes B and F
The reference period of the census observation of an individual's employment status can be reasonably modeled from administrative data associated with the observation	Appendix B
The reference-period modeling procedure for census observations can be used to control for reference-period differences between the census and the CPS	Section 4.2; Appendix B
Census and CPS classifications based on fully reported information are more likely to be accurate than those based on imputations or assignments	Section 4.3, Appendix E
Differences between the weighting procedure for the CPS-Census 2000 Match and that for published Census 2000 estimates do not invalidate the use of weighted data from the Match to provide insights into the accuracy of the published Census 2000 estimates	Section 2.3
With due caution, the net difference rates presented in this report may be interpreted as measures of bias	Section 2.5.b.i
Violations in the data from the CPS-Census 2000 Match of the assumptions of independence and replication do not invalidate the use of the index of inconsistency as a measure of response variability	Section 2.5.b.ii

4. RESULTS

The results of this study are analyzed in this section; the tables referenced here are found in section 4.5 under the heading "Detailed Tables 1A – 4C " [39]. The study looked at CPS-census classification comparisons from two perspectives: (1) for matched cases in general; and, (2) for subsets of matched cases, selected in ways to control for various effects that confound the interpretation of the data as indicators of the capacity of the census to measure employment status. Tables 1A-C represent the first perspective on the matched results; the remaining tables, 2A-C, 3A-C, and 4A-C, represent the second perspective. In these sets of tables, the A and B tables present percentage distributions, while the C table presents the summary measures of response errors corresponding to the data in the A and B tables.

4.1 Employment Status by Age, Race and Hispanic Origin For All People

For all people in Census 2000, Tables 1A and 1B show percent distributions of Census 2000 employment status by CPS employment status (in the first month of the February 2000-to- May 2000 period that they were represented in the CPS), for selected age, race, and Hispanic origin groupings. The data in Table 1C present the summary measures of response error described above, for the three major employment-status categories (employed; unemployed; not in labor force); the measures correspond to the data in Tables 1A and 1B.

An important factor complicating the use and interpretation of these tables, particularly the index data in Table 1C, is that, in both the census and the CPS, a person's employment status is defined in relation to a particular calendar week, the reference period . This time dimension affects the comparability of CPS and census classifications. The census classification relates to the full calendar week (Sunday through Saturday) preceding the date that the person answered the census questionnaire.[40] That week could have been at any time from March 2000 until August 2000 (approximately 90 percent of the people in the census sample responded during March, April, and May). The CPS classification relates to the full calendar week that includes the 12th day of the first month between February and May 2000 when the person was enumerated.[41] Hence, a person's census reference week is not necessarily the same as that person's CPS week; and, because a

---

[39] The estimates in this report are based on responses from a sample of the population. As with all surveys, estimates may vary from actual values because of sampling variation or other factors. All comparisons made in this report have undergone statistical testing and are significant at the 90-percent confidence interval unless otherwise noted.

[40] In the case of the job search question, which is a decisive item for determining whether a person should be classified as unemployed, the reference period includes this week and the three prior ones.

[41] As in the census, the reference period in the CPS for the job search questions includes this week and the three prior ones. Individuals are interviewed in each of four consecutive months by the CPS, so this period spans the range of weeks between interviews.

person's relationship to the labor force, which is what employment status measures, can vary from week to week, a difference between a CPS and census classification may reflect a true change in that relationship between two different weeks.

These considerations mean that some portion of the classification differences shown in Tables 1A-C are likely to be valid, rather than reflections of errors.[42] The index values in Table 1C presumably reflect a combination of response errors and real changes in employment status, meaning that the indices of inconsistency are probably overstated to the extent that they incorporate actual changes in employment status. The effect of actual changes on the index of inconsistency and the net difference rate cannot be exactly determined; hence, these measures must be interpreted cautiously.[43]

Viewed with the above consideration in mind, the percentage distributions in Tables 1A and 1B reveal that, in general, the census did a good job of collecting data for the employed category and a reasonably good one for the not in labor force category, but a fairly poor job for the unemployed category . Table 1A, for example, shows that 92.9 percent of the people in the employed category in the Census were also employed in the CPS ("on the diagonal"), and 83.2 percent of the people in the not in labor force category were on the diagonal; for the unemployed category, only 33.2 percent of the people in the census category were also unemployed in the CPS. This same statement can be made, with more or less precision, for each of the race/Hispanic origin, sex, age groups throughout Table 1A.[44]

Table 1B shows that the census was successful, overall, about 90 percent of the time in placing CPS employed people in the census employment category, and about 86 percent of the time in making the corresponding placement to the not in labor force category, but only 40 percent of the time for making the correct placement to the unemployment category. The relationship among the three categories for people overall is repeated at varying average levels for the race/Hispanic origin, sex, age groups throughout Table 1B; for example, for people 16-19 years old, the on-diagonal percentages for the employed and not in labor force categories, 79.7 percent and 74.5 percent , respectively, though both lower than the corresponding percentages given above for all people, were still much higher than the 29.6 percentage on the diagonal for the unemployed category.

---

[42] Appendix F presents the results of some preliminary research that used the CPS-Census 2000 Match dataset to estimate the effects on the census labor force estimates of the variable nature of the reference period.

[43] As explained in section 4.2 , the tables in that section use modeling techniques to associate a calendar week with each person's census classification and thereby to control for reference-week effects, but the models are based on assumptions whose degree of validity is unknown, so the figures in those tables must be considered hypothetical estimates.

[44] However, the not in labor force category shows much lower values in the 20-54 age groups, while employed was still relatively high.

As an upper limit of variability, the aggregate index of inconsistency in Table 1C for all people (25.7) indicates that employment status as measured in the census was moderately consistent with that measured in the CPS.[45] The level of consistency did not differ appreciably between the sexes in general (aggregate index for men: 28.0; for women: 25.0) . Considerable differences in consistency, however, appear by age. The aggregate indices are at the high end of the moderate range for people under 25 (46.3 for people 16-19; 44.2 for people 20-24); generally decline by age to a level of 21.6 for people 55-64 and then rise to 29.0 for people 65 and over . The aggregate-index pattern by age for women is similar to the one for all people; but in the pattern for men, index values remain in the 45-55 range until they suddenly decline to 33.6 for men 45 to 54 years old. The overall aggregate index also varied considerably by race and Hispanic origin: 20.4 for the non-Hispanic White group; 38.6 for Blacks; and 44.1 for people of Hispanic origin. The same aggregate-index patterns by sex and age that mark the data for all persons, are generally evident within each of these race/Hispanic groups, with decreasing values in the 45-64 age groups.

At the individual category level among the three major categories of the employment status variable, the employed and not in labor force categories had indices of inconsistency (22.6 () and 23.4 , respectively) in the low part of the moderate range. The unemployed category, however, had a very high value of 65.7, indicating a high level of disagreement between the CPS and census measurements. This across-category pattern generally prevailed throughout the race/Hispanic, sex, and age groups in the table. Most noteworthy is that, with few minor exceptions, the index values for the unemployed category were in the high range (above 50), sometimes as large as 100.[46]

For most people, unemployment is a more transitory state than being employed or not in the labor force, and the transition from unemployment to another status can occur on short notice.[47] For this reason, some part of the shadow cast on the census data in the unemployment category by the figures in Tables 1A-C may reflect real changes in unemployment status rather than classification differences, more so than is likely true for the data in the other two classifications. Nevertheless – and this is borne out by the analysis in sections 4.2 and 4.3 – the findings in Tables 1A-C most likely reflect a real problem in the census in collecting accurate unemployment data (or at least unemployment data that are consistent with those from the CPS).

---

[45] As explained previously , for purposes of evaluating the adequacy of a data-collection system, at the category level, values for the aggregate index of inconsistency under 20 are considered low; those between 20 and 50, moderate; and those above 50, high.

[46] Under the unrealistic assumption that the index is without error, an index value of 100 indicates complete inconsistency between the two measuring systems. For the data collected in the CPS and the census, it is assumed that the true value of any index is never greater than 100. Despite this assumption, a computed value of the index above 100 may occur as a result of sampling error.

[47] The median length of a spell of unemployment for the total population was 1.8 months in the 1996 to 1999 period, as shown in the Census Bureau publication *Dynamics of Economic Well-Being: Spells of Unemployment, 1996-1999* (P70-93), available at http://www.census.gov/hhes/www/laborfor .

Some historical perspective on the data in Table 1C is provided by Table E below, whichcompares the index measures for Census 2000 with those from the 1970 and 1960 censuses (no match was done for the 1990 or 1980 censuses):[48]

Table E. Indices of Inconsistency for Employment Status for the United States : Census 2000, 1970 Census, and 1960 Census

Employment Status and Sex	Census 2000		1970 Census		1960 Census	
	Index	90-percent confidence interval	Index	90-percent confidence interval	Index	90-percent confidence interval
Total*						
Aggregate Index	25.7	24.8 to 26.6	17.9	17.1 to 18.7	16.7	16.5 to 17.0
Employed	22.6	21.7 to 23.5	15.5	14.8 to 16.3	14.3	14.0 to 14.5
Unemployed	65.7	62.4 to 69.0	61.1	56.6 to 66.0	56.1	54.5 to 57.7
Not in Labor Force	23.4	22.5 to 24.3	16.0	15.3 to 16.8	15.0	14.7 to 15.3
Male*						
Aggregate Index	28.0	26.6 to 29.3	19.9	18.6 to 21.3	20.9	20.4 to 21.4
Employed	25.3	23.8 to 26.7	17.2	15.9 to 18.5	17.6	17.1 to 18.1
Unemployed	61.2	56.4 to 66.0	58.1	52.3 to 64.6	49.3	47.4 to 51.2
Not in Labor Force	25.0	23.6 to 26.4	16.9	15.6 to 18.3	18.3	17.8 to 18.9
Female *						

---

[48] The universes for the 1960 and 1970 data in the table were restricted to people enumerated as members of households; the 2000 data include people in non-institutional group quarters. The 1960 and 1970 indexes and confidence intervals are based on the data found in the 1960 and 1970 studies cited in footnote 1.

Aggregate Index	25.0	23.9 to 26.1	20.3	19.2 to 21.4	20.2	19.8 to 20.7
Employed	21.4	20.4 to 22.5	17.7	16.7 to 18.9	17.4	17.0 to 17.8
Unemployed	70.7	65.9 to 75.4	65.0	58.1 to 72.7	68.4	65.5 to 71.4
Not in Labor Force	23.2	22.1 to 24.4	19.1	18.0 to 20.3	19.1	18.7 to 19.5

* Persons 14 years old and over for the 1960 and 1970 data; persons 16 years and over for the Census 2000 data.

Source: For the Census 2000 data, Table 1C. For the 1970 Census data: U.S. Bureau of the Census, 1970 Census of Population and Housing, Evaluation and Research Program, *Accuracy of Data for Selected Population Characteristics as Measured by the 1970 CPS-Census Match*, Series PHC(E)-11, U.S. Government Printing Office, Washington, D.C., 1975. For the 1960 census data: U.S. Bureau of the Census, *Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Accuracy of Data on Population Characteristics as Measured by the CPS-Census Match*, Series WER60, No.5., U.S. Government Printing Office, Washington, D.C., 1964.

The census employment questions in 2000 were somewhat similar to those used in 1970; however, the census questions in 1960 differed considerably from those in 1970 and 2000.[49] The data in Table E reveal that the degree of inconsistency for employment status in general, as measured by the aggregate index, has increased from the low range in 1960 and 1970, to the low end of the moderate range in 2000. The same trend appears in the data for the employed and not in labor force categories. Significantly, although the index for the unemployed category in 2000 also increased from 1960 and 1970 levels, these previous levels themselves were already in the high range (see Figure 1). The historical comparisons starkly reveal that the census traditionally has displayed serious shortcomings as a means of measuring unemployment, and that refinements and major revisions to the questions over time have not remedied the problem. The census apparently has been able to collect data for the other two employment-status categories that are reasonably consistent with the CPS, but, even for them, the census moved into the moderately inconsistent range in 2000.

---

[49] See the Introduction to: U.S. Bureau of the Census, Census of Population: 1970, SUBJECT REPORTS, Final Report PC(2)-6A, *Employment Status and Work Experience,* U.S. Government Printing Office, Washington, D.C. 20402, April 1973.

Figure 1. Indexes of Inconsistency between CPS and Census Employment-Status Estimates: 2000, 1970, 1960

Source: Table E.
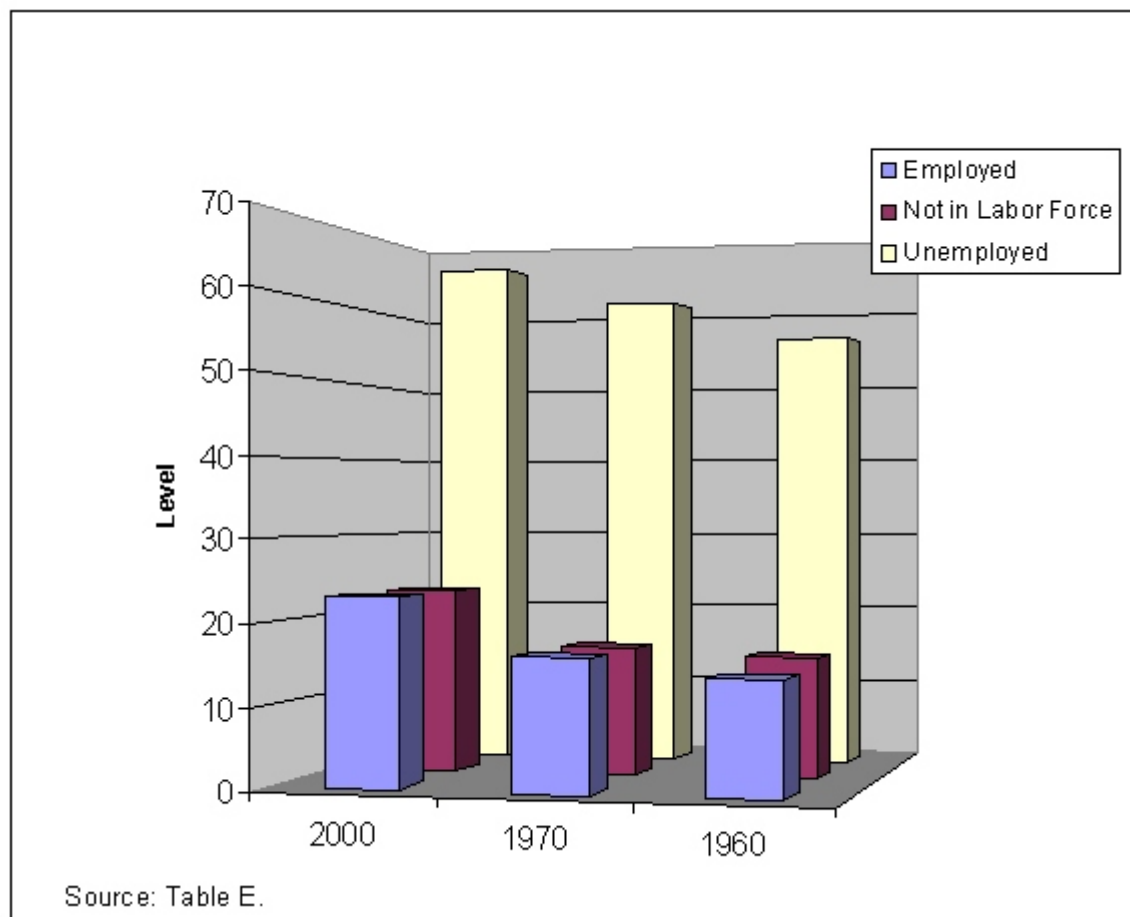
4.2 Employment Status For People With Comparable Reference Weeks

As discussed above, reference-period effects compromise some of the value of the measures in tables 1A-C. This is especially true of their value as indicators of the capacity of the census instrument to collect quality employment-status data . To remove the effect of reference-period differences, it is necessary to restrict the CPS-census comparisons to people whose reference week is, ideally, the same – or almost the same – in both classifications. Unfortunately, the dataset used in this study does not identify the specific dates of a person's census reference week (this information was not collected in the census). Nevertheless, the dataset does contain the date the person's questionnaire was entered into the census processing system, or the "check-in" date. From a person's check-in date, it is possible to estimate, or model, the dates of the person's reference week, and in this way to associate a hypothetical reference week with each person's census employment-status classification. The modeling procedure described in Appendix B was

used to restrict the data in Tables 2A-C to people whose hypothetical census reference week was in March 2000 and whose CPS reference week was also in March 2000 [50]. The data are shown for all such people only, and not by race, sex, or age.

The modeling procedure is subject to errors because it is based on assumptions about the relationship between the check-in date and the reference week whose validity is unknown. For this reason, the data in Tables 2A-C are hypothetical. Even if they were not, they would still be subject to reference-period differences because the census reference weeks for the people in the tables, although being in March 2000, are not all likely to be in the week of March 12-18, which is the CPS reference week for March 2000. This complication, however, does not detract significantly from the usefulness of the data. (Appendix D reproduces tables 2A-C for people whose modeled census reference week is the week of March 12-18, 2000; it also reproduces these tables for all people whose modeled census reference week is in the same month as their CPS week, regardless of the month in question. In both cases, the values of the quality measures do not differ appreciably from those in Tables 2A-C.)

As expected, Tables 2A-C show that the consistency between CPS and census results is improved when the comparisons are controlled for reference-week effects. This improvement is particularly seen in the unemployed category. The on-diagonal percentages in tables 2A and 2B are generally a few percentage points higher than their counterparts in tables 1A and 1B; for the unemployment category, they are about 14 and 18 percentage points higher. The aggregate index of inconsistency, and the indexes of inconsistency for the employed and not in labor force categories, are in the small (low) range in table 2C, down from the moderate range in table 1C; the index for the unemployed category moved slightly into the moderate range in table 2C (49.4) from the high range in table 1C (65.7).

Although the quality measures for the unemployed category show improvement in levels, they are still at such levels to indicate that the quality of the data is problematic and the capacity of the census to collect high quality unemployment data is suspect. The improvements in CPS-census consistency for the category brought about by presumed reductions in reference-period effects is support for the theory that measurements of unemployment are particularly sensitive to timing because of the relatively transitory nature of joblessness.

The census employed category consists of two sub-categories: employed, at work; and employed, not at work (for example, on vacation, ill, or on strike). The first subcategory is particularly important, because it is the major component in the definition of the universe for the place-of-work and journey-to-work data from the census that are widely used in transportation-planning

---

[50] An error in the modeling procedure identified the hypothetical reference week for a small number of people as being in March 2000 when it was actually in February 2000. This error could have an impact on any of the data in this report that use the hypothetical reference week, except for the data in Appendixes D and F, for which the error was corrected. The impact should be negligible, however, because of the small number of people involved; for example, the error affected fewer than one-half of one percent of the people in the universes for Tables 2A-C, 3A-C, and 4A-C.

studies. Tables 2A and B show that Census 2000 was over 90 percent successful in filtering people correctly into or out of the employed-at-work category: the on-diagonal percentage for the census-based distributions (Table 2A) was 93.5 percent ; that for the CPS-based distributions (Table 2B) was 92.0 percent (see Figure 2).[51] Appendix G presents additional analyses of the Census 2000 employed-at-work and employed-not-at-work categories.

Tables 2A and B (and Figure 2) also provide descriptive measures of the quality of the data in the two components of the unemployed category: on layoff; and looking-for-work (labeled as "other" under the "unemployed, total" banner).[52] These categories are primarily useful in measuring total unemployment, rather than in themselves (see the definition of unemployed in Box 2), so the percentages located at their intersections with the "unemployed, total" category are more significant than their strictly on-diagonal percentages. The data show that Census 2000 was moderately successful in funneling people with these characteristics into the unemployment category. According to Table 2A, 48.8 percent and 46.6 percent in the census on-layoff and looking-for-work categories, respectively, were unemployed in the CPS. Table 2B reveals that 65.7 percent of people in the CPS on-layoff category and 57.2 percent of the people in the CPS looking-for-work category were made unemployed in the census. [53]

---

[51] The calculation of the summary measures (indexes of inconsistency and the net difference rate) for this category may be undertaken in future research.

[52] The figures in the "unemployed, layoff", and "unemployed, other"columns in the tables of this report are derived from models. The census does not publish official figures for these categories.

[53] This obvious failure to perform the filtering out function well for these categories may be the cause of the finding that Census 2000 counted a significantly higher number of unemployed people than the CPS for March or April 2000.

Figure 2. Percentage of cases with same employment status classification in CPS and Census 2000 (with modeled census reference week in March 2000)

Source: U.S. Census Bureau, CPS-Census 2000 Combined-month Match.
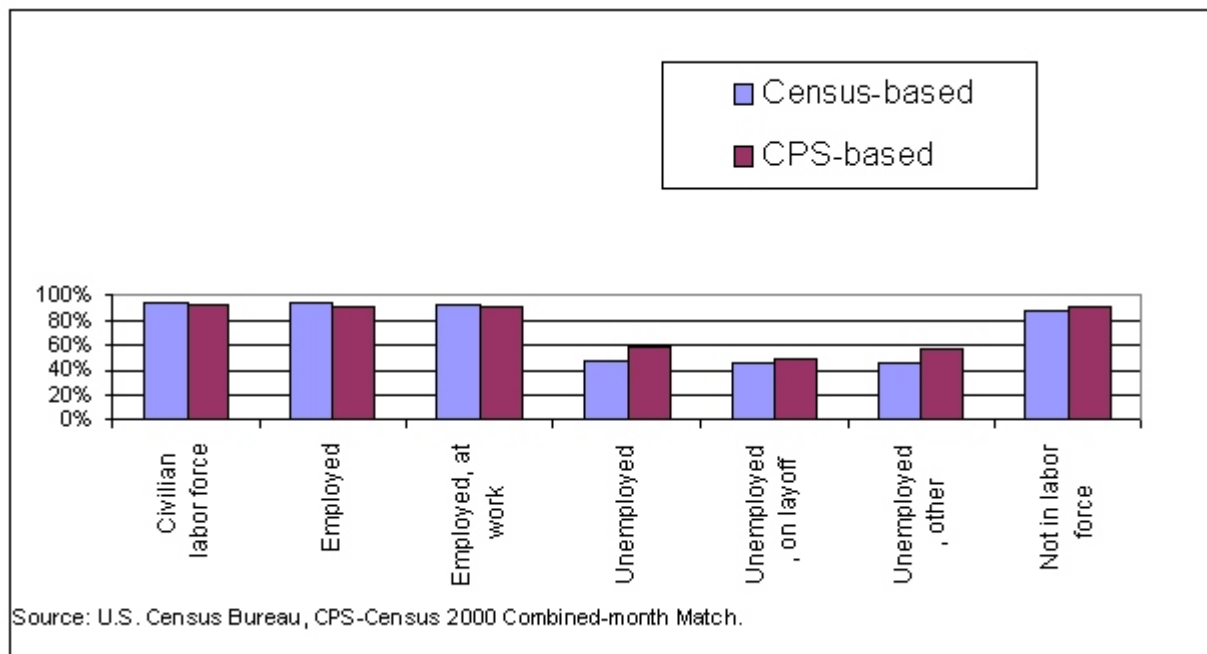
Except for being on layoff from a job, a person can be classified as unemployed, according to the official definition, only if the person conducted an active search for a job (see Box 2). One often-proposed theory to explain why, in both the 1990 and 2000 censuses, the census over-estimated both the number of unemployed people and the unemployment rate relative to the CPS, is that, unlike the CPS, the census is not able to screen out of the unemployment category people who use only "passive" methods to look for work. This theory is supported by the data in Table 2A that show that under half (44.4 percent) of the people who looked for work in the census (and for this reason were classified as unemployed) also looked for work in the CPS (for further analysis of this issue, see Appendix G).

**4.3 Employment Status For People With Comparable Reference Weeks
Whose CPS and Census Employment Status Categories Were Not Imputed**

With respect to their patterns of responses to the census employment questions, people are classified by employment status in the census in one of three ways:

(1) "Fully-reported" people are those who fully and consistently answer all the census employment questions relevant to their labor-market-related labor-market-related activities

39

or situation. They are classified outright to the first category in the census hierarchy (see Box 2) whose criteria they meet ;

(2) "Assigned " people provide only a minimum amount of useable information. They are *placed* in the first category of the hierarchy whose criteria they would most likely meet (in the judgment of the authors of the classification system), if complete information were available for them; and

(3) "Imputed" people are those who either provide no information at all, or provide less than a necessary amount of useable information, and so they are imputed a value through a hot-deck imputation (statistical-match) procedure.

Including the imputed people and the assigned people in the measures of response error in Tables 2A-C detracts from their value as indicators of the capacity of the census questions to collect accurate data, for these people were not necessarily even exposed to the questions.[54] To minimize distortions from this source, Tables 3A-C present data for the subset of the people in Tables 2A-C whose employment status value was not imputed in the census nor in the CPS, although their census value may have been assigned . Tables 4A-C refine the data even more by restricting them to fully reported people only (that is, people who were neither imputed nor assigned a census value, and whose CPS value was not imputed).

The data in Tables 3A-C are very similar to those in their 2A-C counterparts. They do reveal increases in CPS-census consistency, but the differences from Tables 2A-B are mostly marginal, except in the unemployment category. The index of inconsistency measures in Table 3C improve somewhat over those in Table 2C for all the categories, but the index for the unemployment category, at 43.1, remains near the extreme end of the moderate range (see Figure 3).

The "fully reported" people represented in Tables 4A-C provided, at least in theory, the highest quality responses to the employment questions in Census 2000 census 2000, so the data should exhibit the greatest degree of CPS-census consistency. If these data suggest that there are problems with collecting data for an item in the census, or for a category of an item, either because of flaws in the questions themselves or because of how and when they are used, then the case for the existence of such problems (although the converse is not necessarily true) would be considerably strengthened. The percentage and summary measures in Tables 4A-C generally do show a high level of CPS-census agreement, except, again, for the unemployment category. In Table 4C the indexes of inconsistency are lower than those for any of the universes in the prior summary-measure tables, but the index is still in the high end of the moderate range (40.9) for the unemployed category (see Figure 3).

Tables 3A-C and 4A-C indirectly shed some light upon the soundness of the census procedures to impute or assign values. The fact that Tables 3A-C, which are restricted to not-imputed people,

---

[54] This assertion assumes that people who did not respond to the questions, or who did not respond fully and consistently, failed to do so because they chose not to respond to the questions, and not because of factors related to the questions themselves or the context of their use.

are only marginally different from Tables 2A-C, is some indication that the census imputation procedures are likely performing reasonably well in correctly classifying people. That Tables 4A-C represent only marginal improvements over Tables 3A-C can be interpreted as indicating that the census value-assignment procedures are also performing adequately. The discussion and tables in Appendix E take a more direct approach in using the 2000 CPS-Census match classifications to judge the soundness of census imputation and assignment procedures.

Figure 3. Indexes of Inconsistency for Census 2000 Employment Status Categories (with modeled census reference week in March 2000)
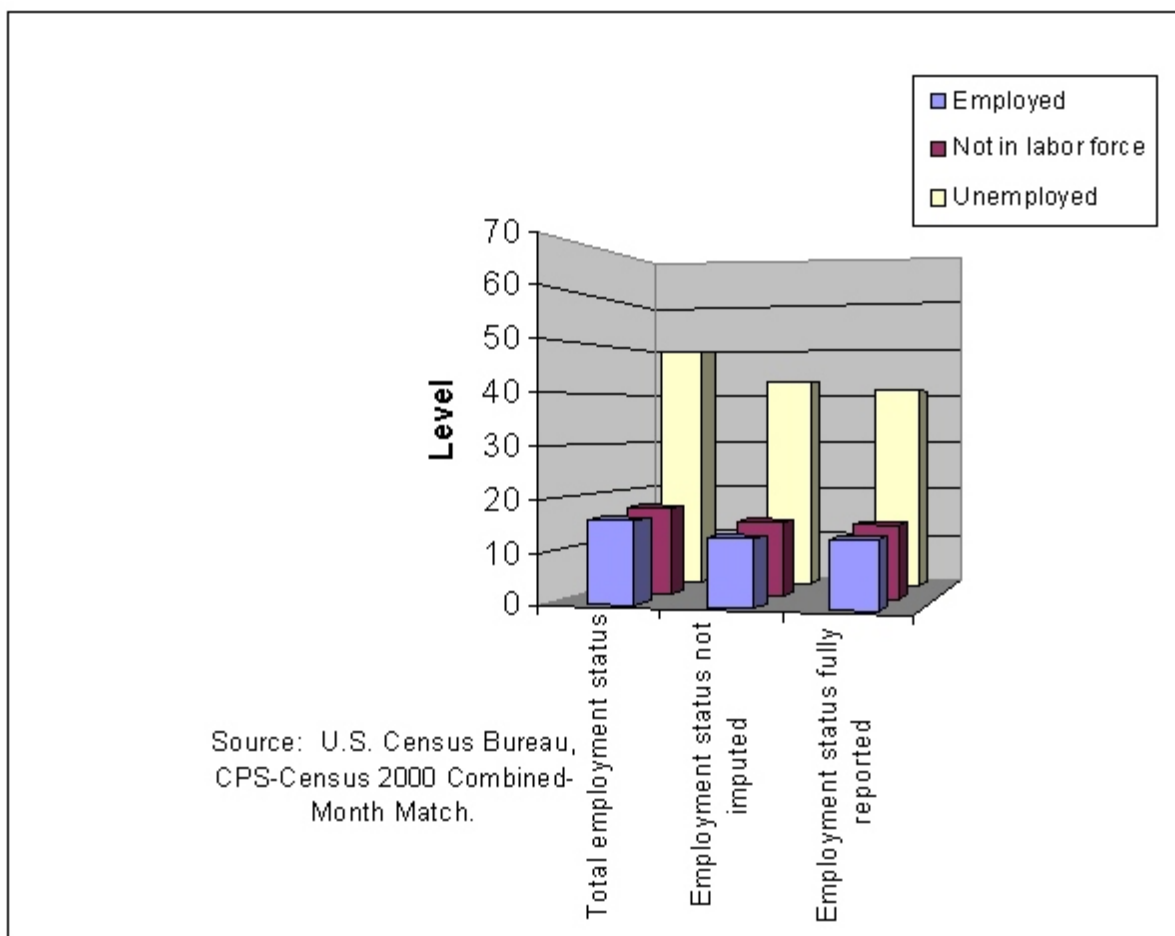
4.4 Using the CPS-Census Match to Explain Differences between Published Estimates from Census 2000 and Official CPS Estimates

The data in Table F show that, relative to the official CPS employment-status estimates for March, April, and May 2000 at the national level, Census 2000 underestimated the number of employed people, and overestimated the number of unemployed people and people not in the labor force:

Table F. Comparison of Published Estimates of Employment Status Between Census 2000 and the Current Population Survey for March, April, and May 2000
(Civilian noninstitutional population. Numbers in thousands)

Employment Status Category	Census 2000	March 2000 CPS	April 2000 CPS	May 2000 CPS
Population 16 Years and Over	212,034	211,772	212,018	212,242
Civilian Labor Force	137,669	142,123	142,138	142,145
Employed	129,722	136, 054	136,927	136,685
Unemployed	7,947	6,069	5,212	5,460
Not in labor force	74,365	69,649	69,879	70,097

At a general level, the data in the detailed tables of this report suggest some of the factors responsible for these CPS-census gaps:

a) The differences between the census and the CPS reference periods are a factor in the gaps, though probably not a primary one. The measures in Tables 2A-C, which attempt to remove the effects of reference-period differences, are similar to those in Tables 1A-C.

b) The underestimate of employment and the overestimate of people not in the labor force are likely related to the failure of the census classification system to filter more employed people out of the not in labor force category and into the employed category. This failure may be related to the change in wording between the 1990 and 2000 census in the "work last week"question, which is the key question in the decision to classify a person to the employed category.[55] Table 1A shows

---

[55] In 1990, this question asked: "Did this person work at all last week?" In 2000, the question asked: "Last week, did this person do any work for either pay or profit?" Perhaps the "pay or profit" addition caused many employed people, who had jobs that were too marginal or irregular to characterize as "pay or profit " jobs, or people who *worked for*, but did not actually *receive* pay or profit in the reference week, to answer "no" to the question in 2000 . The word "profit" may also have confused people who responded to it, to the exclusion of the word "pay ," or who worked for compensation that they may have considered neither pay nor profit, such as commissions, or who thought that profit had to be one of their compensation options. Even though the Census 2000 wording was identical (deliberately) to its CPS counterpart, the method of CPS data collection would have allowed the question to be clarified in a way that was not possible, for the most part, in the census .

that nearly 15 percent of the people in the Census 2000 not in labor force category were in the CPS employed category; Table 1B shows that nearly 8 percent of employed people in the CPS were put into the not in labor force category in the census. The corresponding figures in Tables 2A and 2B, which are controlled for reference period differences, are 11 percent and 7 percent, respectively.

c) Census 2000 may not have been equal to the task of collecting accurate unemployment data. It especially failed to keep employed people and people not in labor force out of the unemployment category (the census unemployment category is made up of about equal percentages of these latter kinds of people). It did a slightly better job of funneling unemployed people into the unemployed category. That it was better at funneling-in than in screening-out probably at least partly explains why Census 2000 overestimated unemployment relative to the CPS.

4.5 Detailed Tables 1A – 4C

The following is a list of the tables presented in this section:

Table 1A.   Census-Based Percentage Distributions--Employment Status of the Civilian Noninstitutional Population 16 years and over in the Combined-month CPS-Census 2000 Match by Race, Hispanic Origin, Sex, and Age for the United States Total: 2000

Table 1B.   CPS-Based Percentage Distributions--Employment Status of the Civilian Noninstitutional Population 16 years and over in the Combined-month CPS-Census 2000 Match by Race, Hispanic Origin, Sex, and Age, for the United States Total: 2000

Table 1C.   Summary Response Measures--Employment Status of the Civilian Noninstitutional Population 16 years and over in the Combined-month CPS-Census 2000 Match by Race, Hispanic Origin, Sex, and Age, for the United States Total:2000

Table 2A.   Census-Based Percentage Distributions--Employment Status of the Civilian Noninstitutional Population 16 years and over in the Combined-month CPS-Census 2000 Match With Reference Week in March 2000, for the United States Total: 2000

Table 2B.   CPS-Based Percentage Distributions–Employment Status of the Civilian Noninstitutional Population 16 years and over in the Combined-month CPS-Census 2000 Match With Reference Week in March 2000, for the United States Total: 2000

Table 2C.   Summary Response Measures–Employment Status of the Civilian Noninstitutional population 16 years and over in the Combined-month CPS-Census 2000 Match With reference Week in March 2000, for the United States Total: 2000

Table 3A.   Census-Based Percentage Distributions–Employment Status of the Civilian Noninstitutional Population 16 years and over in the Combined-month CPS-Census 2000 Match With Reference Week in March 2000 and Employment Status Not Imputed for the United States Total: 2000

Table 3B.   CPS-Based Percentage Distributions–Employment Status of the Civilian Noninstitutional Population 16 years and over in the Combined-month CPS-

(Insert Tables 1A to 4C here)

5. CONCLUSIONS/RECOMMENDATIONS

This study examined micro-level comparisons of the Census 2000 and CPS employment-status classifications of the same individual for the people in the Combined-month CPS-Census 2000 Match. In each survey, the employment classification of a person represented the outcome of the observation of an event, which was the relationship of the person to the labor force at a given period of time. The unit of analysis for this study was a comparison of these dual outcomes.

An individual's Census 2000 classification may differ from the same individual's CPS classification because of collection or processing errors in either or both surveys. A particular goal of this study was to obtain insights concerning the source, nature, and prevalence of such errors in the Census 2000 classifications.

The analysis assumed that the CPS was more likely than the census to make an accurate classification by employment status, given that the two surveys were observing the same event. This assumption permitted the analysis to provide measures of census bias. A major limitation on the interpretation of the results, however, was the inability to vouch for this assumption in any particular case because of possible differences in the time-reference periods of the observed events. Efforts were made to control the confounding effects of this problem by modeling the reference period of the census observations.

The analysis evaluated census–CPS consistency using percentage measures and two response error measures, the net difference rate and the index of inconsistency. The index of inconsistency is especially useful for evaluating the suitability of the census as an instrument for classifying people to particular employment-status categories.

The study showed that the census and the CPS are reasonably consistent in classifying people to the employed and not in labor force categories, but they exhibit considerable variability in classifying people to the unemployed category. The previous studies of census-CPS employment classifications, which were done for the 1970 and 1960 censuses, revealed similar patterns, but, for Census 2000, the consistency for all three categories slipped somewhat from the 1970 levels, in spite of efforts, particularly after the 1990 census, to make the census employment questions conform more closely with the CPS questions.

As was true in the 1970 and 1960 studies, the index of inconsistency measurements for 2000 for the unemployed category were high enough to suggest that major improvements are required in the method used to collect the data, or that the concept itself may not be measurable in a census context (or, more generally, outside of a CPS context). The short-lived nature of many spells of unemployment may be a factor, however, in exaggerating CPS-census inconsistencies. The analysis suggested that a serious deficiency of the census – one that fosters an over-counting of unemployed people – is its inability to distinguish between active and passive methods of searching for a job. The results of this study for the employed and not in labor force categories

indicated that, although the census is able to measure these concepts reasonably well, improvements are needed in the methods used to collect them.

This study also made an effort to relate the general findings above to possible shortcomings of particular Census 2000 questions. This effort led to the following insights:

– After the 1990 census, the census employment questions were redesigned to make them more like the CPS questions. The results of this study suggest that these changes may not have had the desired effect. Of course, they could have worked very well indeed, and prevented other factors from making the employment data even worse, but whether this happened is unknown.

– There was a tendency for employed people in the CPS to be classified as not in labor force in Census 2000. This tendency may be related to shortcomings in the *work last week* and *temporary absence* questions:

1. Work Last Week question
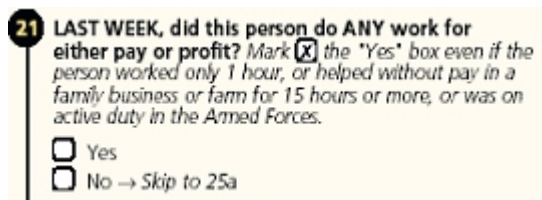
The *work last week* question may have a problem in separating people who have jobs or business from those who do not. For some unknown reason, it appears that respondents -- or their proxies-- too often answer "no" to this question when they have performed what is commonly considered to be economic kinds of work. This mistake usually caused Census 2000 to classify a genuinely employed person as not in the labor force. The problem may be related to confusion about the phrase "either pay or profit", and to misunderstandings concerning contingent, temporary, marginal, or irregular work, self-employment, and unpaid work in a family business or farm.

The study also revealed that the *work last week* could do a better job of separating employed people into those who were at work and those who were temporarily absent from jobs. This is an important distinction for the journey to work data, which are heavily used to do transportation studies. The problem is that people who are temporarily absent tend to misreport in the *work last week* question that they were *at* work.

2. Temporary Absence from Work question
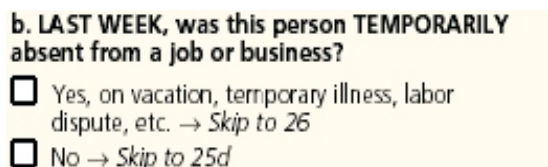
People who have jobs from which they are temporarily absent (on vacation or maternity leave, for example) should be classified as employed. The *temporary absence* question, however, did not explicitly mention family leave, and this omission may have caused many people on such leave to be incorrectly classified as not in the labor force in the census.

– Census 2000 used five questions to classify people as unemployed. The evaluation suggested that there may be some problems with at least two of them:

1. Looking for Work question

d. Has this person been looking for work during the last 4 weeks?

☐ Yes
☐ No → *Skip to 26*

The *looking for work* question may be a chief culprit. Its problem is that it fails to distinguish between active and passive methods searching for work. Only people who are actively looking for work — doing things that in and of themselves could lead to job offers, such as visiting employers – should legitimately answer "yes" to the *looking for work* question, and thus be legitimately classified as unemployed. The question, however, lends itself to misreporting by people who use passive job-search methods only – looking at want ads at the kitchen table, for instance – and they end up being misclassified a unemployed when they are really not in the labor force. For the same reason, the *looking for work* question may also have a tendency to cause misreporting by so-called discouraged workers – people who have given up looking for work because they believe no jobs are available. Again, such misreporting leads to incorrect classification to the unemployed category.

2. Work Last Week Question

The work last week question, already discussed in terms of the employed category, may also have had a significant role in Census 2000 unemployed misclassifications. People who were working at temporary jobs while they were on temporary layoff or looking for permanent jobs may have had a tendency to report "no" to the *work last week* question, and thereby to be misclassified as unemployed in Census 2000.

Several appendixes in this report present the results of attempts to use the CPS-Census 2000 Match to examine the quality of the census edit and imputation system and to explain some of the macro level differences between Census 2000 and the CPS described in Census 2000 Auxiliary Evaluation B.8. Briefly, these additional studies suggest that:
– the Census 2000 edit and imputation system for employment status performed reasonably well– probably as well as can be expected, though more research is needed on this subject;
– several hypothesized factors – such as the shortcomings in the census questions discussed above and differences in census and CPS reference periods – may have had a part in creating the

wide CPS–Census 2000 gaps in aggregate estimates of employment and unemployment, but even collectively, their likely effects explain only a part of the gaps.

 The above conclusions lead to several recommendations:

– The results of this study should be useful in improving the quality of employment status data collected in future demographic surveys and censuses, particularly in the new American Community Survey (ACS), which uses the same employment questions as those used in Census 2000. Preliminary comparisons of aggregate-level ACS labor force estimates with CPS estimates reveal that the ACS has many of the same shortcomings relative to the CPS as Census 2000 does. The results of this Census 2000 evaluation should have considerable applicability to the ACS. In particular, it is likely that the suggested problems with the Census 2000 questions discussed above will also be detrimental to the collection of accurate labor force data in the ACS. Substantial research should be devoted to revising the ACS questions by addressing these issues, though it should not be limited to them.
– Research aimed at improving the accuracy of the ACS employment data through questionnaire improvements must include a large component of cognitive/behavioral research to develop new questions or approaches prior to pre-testing them. This evaluation suggests that the effects of shortcomings in the employment-status questions may be too subtle to detect in pre-tests alone.
– The ACS will have the opportunity to collect labor force data through respondent-enumerator interactions, primarily via computer-assisted instruments, to a much greater extent than was true in Census 2000. The kinds of flaws in the Census 2000 employment-status questions, and by implication in those same questions in the ACS, suggested by this evaluation, may be especially amenable to amelioration or even elimination through the use of such methods. Hence, special attention should be devoted to the development of the enumerator versions of the employment-status questions in the ACS. In this effort, however, consideration must be given to how differences in the effectiveness of various collection modes may differentially impact the quality of the data for various segments of the population.
 – Attempts to revise the ACS employment status questions should proceed by evolutionary or incremental means. The evaluation results suggests that the existing questions, in spite of their likely flaws, likely have many virtues as well.
– Efforts should be made to measure the amount of bias and response variability in the ACS employment status data. It is especially important to make users aware of the potentially serious consequences of response variability on the accuracy of cross-tabulations of employment status data by other characteristics.
– Suggestions for future research:
(a) Use multivariate analytical methods to examine some topics further (such as differences in error tendencies among demographic groups, and the effect f complex skip patterns): This study suggested that many factors are involved in census–CPS classification differences. Multivariate analytical techniques have the benefit of describing the relative influence of separate factors in multi-factor relationships. The match identified rich areas for the application of such techniques. Using them, for example, to look at the correlation between an individual's demographic characteristics and the likelihood of being misclassified in a particular way, may help to detect or

pinpoint shortcomings of the questions or other aspects of the collection or processing of the labor force data.

(b) Study collection-mode effects (paper/enumerator): One topic briefly examined in this study – and which is a potentially rewarding subject for further research – is the relationship between the mode of collection in the census – whether the data were self-reported or collected in the nonresponse followup by enumerators – and the amount of bias and levels of inconsistency in the data.

(c) Use the datasets of the CPS-Census 2000 Match to study other topics: The Match file is a rich resource for assessing the accuracy of the employment-status data in Census 2000, but this use merely scratches the surface of its potential. The two match datasets – the combined-month dataset used in this study, and its March counterpart – could be used to examine many other items collected in Census 2000 (and that continue to be collected in the ACS), and to evaluate the accuracy of CPS data.

REFERENCES

Adams, Tamara. 2003a. *Specifications for the Analyst Matching of CPS/NHIS to the 2000 Decennial Census.* DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter HH-2 (Memorandum for the record from Magdalena Ramos).

Adams, Tamara. 2003. *A Tale of Two Surveys: Measuring Labor Force Status in the Current Population Survey and Census 2000.* Unpublished draft of an internal U.S. Census Bureau paper.

Bancroft, Gertrude. 1958. *The American Labor Force: Its Growth and Changing Composition.* New York, Wiley.

Biemer, Paul P. 1985. *Evaluating Censuses of Population and Housing.* U.S. Department of Commerce, U.S. Census Bureau, Statistical Training Document, Chapter 3 and Appendix, ISP-TR-5.

Brown, Sharon and Flaim, Paul. 1992. *LAUS Technical Memorandum.* U.S. Bureau of Labor Statistics, No. S-93-1 (part).

Clark, Sandra Luckett, John Iceland, Thomas Palumbo, Kirby Posey and Mai Weismantle. 2003. *Comparing Employment, Income, and Poverty: Census 2000 and the Current Population Survey.* U.S. Census Bureau, Census 2000 Auxiliary Evaluation.

Dalaker, Joseph. 2001. *Poverty in the United States: 2000.* U.S. Census Bureau, Current Population Reports, Series P60-214.

Davern, Michael, Lynn A. Blewett, Boris Bershadsky, and Noreen Arnold. 2003. *Missing the Mark? Imputation Bias in the Current Population Survey's State Income and Health Insurance Coverage Estimates.* Unpublished paper under submission to Journal of Official Statistics.

Ennis, Sharon and Phyllis Singer. 2003. *Census 2000 Content Reinterview Survey: Accuracy of Data for Selected Population and Housing Characteristics as Measured by Reinterview.* U.S. Census Bureau, Census 2000 Evaluation B.5.

Fay, Robert E., Jeffrey S. Passel and J. Gregory Robinson. 1988. *The Coverage of Population in the 1980 Census.* 1980 Census of Population and Housing, Evaluation and Research Reports, PHC80-E4.

Fay, Robert E. 1988. *Evaluation of Census Coverage from the 1980 Post Enumeration Program (PEP): P-sample and E-sample Results by Type of Enumeration Area and by Mail Response.* Preliminary Results Memorandum No. 120.

Fay, Robert E. 1989. *An analysis of within-household undercoverage in the Current Population Survey.* U.S. Census Bureau, Proceedings, Annual Research Conference.

Gottschalck, Alfred O. 2003. *Dynamics of Economic Well-Being: Spells of Unemployment, 1996-1999.* U.S. Census Bureau, Current Population Reports, Series P70-93.

Gunnison Consulting Group, Inc. 2002. *CPS Pre-Processing Software Specifications Document (Final).* Chevy Chase, MD.

Gunnison Consulting Group, Inc. 2002. *CPS MaRCS User Reference Guide (Final).* Chevy Chase, MD.

Gunnison Consulting Group, Inc. 2002. *Current Population Survey (CPS) Matching Review and Coding System (MaRCS) Software Specifications Document (Final).* Chevy Chase, MD.

Gunnison Consulting Group, Inc. 2003. *Current Population Survey (CPS) Matching Review and Coding System (MaRCS) Final Software Documentation (Final).* Chevy Chase, MD.

Judson, Bauder, Gorshak, and Bobbitt. 2002. *CPS to Census Match, Computer Match Documentation: HCUF, CPS, and NHIS File Preprocessing, Data Preprocessing, Address Match, Person Match, Quality Assurance, Post Processing.* U.S. Census Bureau, Administrative Records Research/PRED (Final).

Lyberg, Lars and Danniel Kasprzyk, 1991. *Data Collection Methods and Measurement Error: An Overview.* in Paul Biemer,et al., eds., Measurement Error in Surveys. Wiley, New York.

Martin, Elizabeth and Manuel de la Puente. 1993. *Research on sources of undercoverage within households.* American Statistical Association, Proceedings of the Section on Survey Methods Research. Pp 1262-1267.

Mule, Thomas, 2002, *A.C.E. Revision II Results: Further Study of Person Duplication.* DSSD Revision II Memorandum Series #PP-51.

National Center for Education Statistics. 2003. *Digest of Education Statistics, 2002.* Washington, D.C.

National Research Council. 2001. *Small-Area Income and Poverty Estimates: Priorities for 2000 and Beyond.* Panel on Estimates of Poverty for Small Geographic Areas. Washington D.C.: National Academy Press.

Petroni, Rita and Danny R.Childers. 2003. *Coverage Measurement from the Perspective of March 2000 A.C.E.* Census 2000 Testing, Experimentation, and Evaluation Program, Topic Report Series, No 4.

U.S. Census Bureau. 2000. *Money Income in the United States 1999.* Current Population Reports. P60-209.

U.S. Census Bureau. 2003. *Census 2000 Summary File 1 Technical Documentation.*

U.S. Census Bureau. 2003. *Census 2000 Summary File 3 Technical Documentation.*

U.S. Census Bureau. 2002. *Census 2000 Summary File 3 (SF3) - Sample Data* (American FactFinder).

U.S. Census Bureau. 1992. *1990 Census of Population and Housing (STF-3)* (American Fact Finder).

U.S. Census Bureau and Bureau of Labor Statistics. 2000. *Current Population Survey: Design and Methodology,* (Technical Paper 63).

U.S. Census Bureau. 1973. Census of Population: 1970. *Employment Status and Work Experience.* SUBJECT REPORTS, Final Report PC(2)-6A. Washington, D.C. : U.S. Government Printing Office.

U.S. Census Bureau. 2003. *Census 2000, Summary File 3 Documentation, Data Note 4.*

U.S. Census Bureau. 1985. *Evaluating Censuses of Population and Housing.* Statistical Training Document, ISP-TR-5. Washington, D.C.

U.S. Census Bureau. 1964. *Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Accuracy of Data on Population Characteristics as Measured by the CPS-Census Match.* Series WER60, No.5. Washington, D.C.: U.S. Government Printing Office.

U.S. Census Bureau. 1975. *Accuracy of Data for Selected Population Characteristics as Measured by the 1970 CPS-Census Match.* 1970 Census of Population and Housing, Evaluation and Research Program, Series PHC(E)-11. Washington, D.C.: U.S. Government Printing Office.

U.S. Census Bureau and U.S. Bureau of Labor Statistics, 2002. *Design and Methodology*. Current Population Survey Technical Paper 63RV, TP63RV.

U.S. Office of Management and Budget. 1978. *Statistical Policy Directive No. 14: Definition of Poverty for Statistical Purposes.* (http://www.census.gov/hhes/poverty/povmeas/ombdir14.html)

Zbikowski, Andrew. 2003. *Weighting specifications for Current Population Survey (CPS)-Census 2000 matched data file (MATCH2000-01)*. Memorandum for Chester E Bowie, Chief, Demographic Surveys Division from Alan R Tupek, Chief, Demographic Statistical Methods Division.

Appendix A. Major Conceptual and Methodological Differences between the CPS and the Census

1. Differences Supporting the Presumption of Superior CPS Accuracy

(Note: The following discussion was adapted from the paper prepared by Sharon Brown and Paul Flaim, U.S. Bureau of Labor Statistics, as part of LAUS Technical Memorandum No. S-93-1, November 18, 1992.)

There are significant procedural and conceptual differences between the census and the CPS, an analysis of which leads to the conclusion that the CPS data are more accurate and reliable, at both the national and state levels, than those collected through the census.

1) Interviewer-controlled environment versus self-enumeration:
All data from the CPS are gathered by trained field interviewers through personal visits and telephone interviews. For the most part, decennial census data, which were once also collected by interviewers--100 percent in 1950--are now *largely* self-reported; that is, by themselves, individuals fill out a simplified questionnaire mailed to them[56]. For these kinds of respondents, there are generally no interviewers to clarify survey questions and probe for more accurate and detailed responses, as is the case in the CPS.

2) Specific versus general survey instruments:
The CPS currently uses 13 specific, detailed questions to determine an individual's employment status. In the census, the questions are fewer—only six. The enhanced specificity in the CPS is designed to avoid mis-classifications; the relative lack of specificity in the census undoubtedly results in some mis-classifications. For this reason, too, the CPS does a better job of ferreting out marginal work activity than the census. For example, laid-off people who worked at a temporary, perhaps part-time, job in the reference week might totally discount such work and classify themselves in the census as "on layoff" and thus be counted as unemployed. In the CPS, more detailed and probing questions are more likely to prompt respondents to mention the temporary or part-time jobs, in which case they would be officially classified as employed. Indeed, once people report having a job to CPS interviewers, they cannot be asked questions about layoff status or job seeking, whereas in the census such choices could easily be made. Moreover, it is also possible that people classified as *discouraged workers* in the CPS--and thus outside the labor force--would have reported themselves as unemployed in the census.

3) Intensive versus limited quality control of data collection:
CPS data are subject to much more rigorous quality control standards than are the census data . CPS interviewers are trained extensively before going out into the field, and proficiency checks are conducted regularly. In addition, each month, a portion of the households in the

---

[56] In Census 2000, according to calculations performed for this study, the responses for roughly 70 percent of the people in the employment-status universe were collected in this way.

CPS sample are re-interviewed, and the results are used to control and measure the quality of the data. In the census, the extent to which the quality of the data can be controlled or evaluated is much more limited.

4) Definite versus variable reference period:
The CPS questions for determining current employment status relate to a specific reference week, the week including the 12th of the month (or, in the case of job search, the 4 weeks preceding the survey collection week); the census questions relate to the calendar week preceding the date that the questionnaires were completed. In 2000, most of the questionnaires (approximately 96 percent) were completed between March and May, but some were not completed until August.[57] Thus, the reference week for the Census 2000 varies from the first week in March to some week in August. The census employment and unemployment may be biased relative to CPS estimates for any given month in this period because they may somewhat reflect changes in the economy over a longer period of time than a month.

For more information on the CPS, see Current Population Survey: Design and Methodology, Technical Paper (TP63RV), available at http://www.bls.census.gov/cps/tp/tp63.htm .

2. Instrument Differences

The chart below compares the CPS battery of employment status questions with the Census 2000 battery. The number in the note column refers to the note below the table that explains the reasons for any major differences between the CPS question and its corresponding Census 2000 question(s). The CPS and census questions are both products of revisions to earlier questions made in the 1990s. The revised CPS questions were introduced in 1994 as part of the project to convert the CPS collection mode from a paper questionnaire to an automated, or computer-assisted interviewing (CAI), instrument. The census questions were revised as part of the development and testing process for Census 2000 between 1995 and 1998, and were intended to conform as much as practicable with the revised set of CPS questions. The primary reasons for differences between the two batteries of questions is: (1) space and respondent-burden considerations limited the census to six questions; and (2) the difference in collection modes: paper for the census questions; computer-assistance for the CPS questions.

---

[57] The reference week could have been as early as January for remote parts of Alaska.

Chart – Correspondence between CPS and Census 2000 Employment-Status Questions
(Note: question numbers in the chart represent the order of the question within the battery of questions and are not equivalent to the numbering system used in the survey)

CPS Question	Corresponding Census 2000 Question	Note
1. Does anyone in the household have a business or a farm?	No corresponding question	1
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.*	1. LAST WEEK, did you do ANY work for either pay or profit? *If 1 is "no" , ask 2.*	2
3. LAST WEEK, did you do any unpaid work in the family business or farm? *If 2 and 3 are both "no, " ask 4.*	No directly corresponding question; the instruction for census question 1 above asked the respondent to answer "yes" to the question if the respondent "helped without pay in a family business of farm for 15 hours or more"	1
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.*	3. LAST WEEK, were you TEMPORARILY absent from a job or business? –Yes, on vacation, temporary illness, labor dispute, etc – No	3
5. LAST WEEK, were you on layoff from a job? *If 5 is "yes," ask 6. If 5 is "no," ask 8.*	2. LAST WEEK, were you on layoff from a job? *If 2 is "yes" , ask 4; otherwise, ask 3.*	4
6. What was the main reason you were absent from work LAST WEEK? *There are 14 answer categories including: on layoff; slack work; vacation/personal days, etc.*	No directly corresponding question; examples of reasons for temporary absences are associated with the "yes" answer box in question 3	1

7. Has your employer given you a date to return to work? *If "no," ask 8.*	4. (*For people on layoff*) Have you been informed that you will be recalled to work within the next 6 months OR been given a date to return to work?	5
8. Have you been given any indication that you will be recalled to work within the next 6 months? *If "no," ask 9.*	4. (*For people on layoff*) Have you been informed that you will be recalled to work within the next 6 months OR been given a date to return to work?	5
9. Have you been doing anything to find work during the last 4 weeks? *If "yes," ask 10.*	5. Have you been looking for work during the last four weeks?	6
10. What are all of the things you have done to find work during the last 4 weeks?	No corresponding question	1
11.LAST WEEK, could you have started a job if one had been offered ? *If "no," ask 13.*	6. Could you have started a job last week if offered one, or returned to work if recalled? – Yes, could have gone to work – No, because of own temporary illness – No, because of all other reasons (in school, etc.)	7
12.Could you have returned to work LAST WEEK if you had been recalled? *If "no," ask 13.*	6. Could you have started a job last week if offered one, or returned to work if recalled? – Yes, could have gone to work – No, because of own temporary illness – No, because of all other reasons (in school, etc.)	7

13. Why is that?	6. Could you have started a job last week if offered one, or returned to work if recalled? – Yes, could have gone to work – No, because of own temporary illness – No, because of all other reasons (in school, etc.)	7

Notes:

1. This question was not included in Census 2000 because it had lower priority than competing items.

2. The major difference is that the CPS, owing to its CAI capabilities, inserts "either" and "or profit" conditionally, whereas they are both fixtures in the paper-bound census question.

3. The census question is a combination of CPS questions 4 and 6.

4. The census question on layoff (labeled as census question 2 in the chart) was asked before the census question on temporary absences from work (census question 3), but in the CPS the corresponding questions (CPS questions 5 and 4, respectively) were asked in the reverse order. The rationale for the divergence in ordering has to do with the perception, reinforced by experience, that many people on temporary layoff from a job still consider themselves to "have" that job. Thus, if the census had asked such people whether they were temporarily absent from a job (census question 3) before they were asked if they were on layoff from a job (census question 2), they might well have answered "yes" that they were temporarily absent, which would have increased their chances of being misclassified as employed. The only way to avoid the problem – other than the method that was actually used of asking the layoff question (census question 2) before the temporary absence question (census question 3)– would have been to ask all people who reported that they were temporarily absent to answer subsequent questions about layoff (census question 2) and about looking for work (census question 5).This approach was thought to impose an unacceptable response burden on the bulk of the temporarily absent people who were not on layoff, and for this reason it was rejected. The corresponding CPS approach avoids the problem by asking people who answer "yes" in the CPS temporary absence question (CPS question 4) to specify the main reason they were absent from work (CPS question 6). The census did not have the luxury of asking a corresponding additional question.

5. To save space, CPS questions 7 and 8 were combined into the one census question 4.

6. The phrase "doing anything to find work" in CPS question 9 was replaced by "looking for work" in the corresponding census question 5. The CPS used CPS question 10 as a followup for people who answered "yes" in CPS question 9. The categories of CPS question 10 enabled the CPS to ascertain whether the individual's job search had been active or passive (only active searches qualify as a condition of unemployment). The census did not have room for a corresponding followup, yet it needed to convey the message that the respondent should answer

"yes" to census question 5 only if the respondent had used active methods to search for work. It was thought that, in the common parlance, the expression "looking for work" connoted the use of active search methods more forcibly than the rather flat and all-inclusive expression "anything to find work" that begged for a followup unavailable to the census.

7. To save space, CPS questions 11, 12, and 13 were combined into the one census question 6.

3. Bias in the CPS

To contend that the CPS may be a more accurate source of labor force estimates than the census is not to imply that the CPS is error free. In fact, the Current Population Survey Technical Paper referenced above includes a comprehensive discussion of various kinds and sources of errors in the CPS (see Chapters 15 and 16). One kind of error, known as "month-in-sample-bias " or "rotation-group bias," may be especially relevant to the measures of the accuracy of the Census 2000 data presented in this report. This kind of bias is exhibited, among other ways, by the finding that unemployment estimates are generally higher for persons in their first and fifth months in the CPS sample than in their other months (each monthly CPS sample is divided into eight representative subsamples or *rotation groups*; these groups are in the sample for 4 consecutive months, out for the following 8 months, back in for the next 4 months, then retired from the sample). The effects of this kind of CPS bias on the data in this report are not known.

Appendix B. Modeling the Census Reference Week

Two basic kinds of questionnaires were used in the Census 2000: mail-out/mail-back questionnaires (mail forms), which were intended to be completed by respondents themselves; and enumerator questionnaires, which were completed by census Field Representatives during interviews with census respondents. After being completed, the forms were returned to the census collection centers for processing. The date when a completed form first entered into the processing system was captured as a piece of information, called the check-in date, that is available for each person represented on the form, and thus, for each person on the CPS-Census 2000 Match dataset that forms the basis for the estimates in this report.

The reference period for the questions related to an individual's census employment status is intended to be the full calendar week, Sunday through Saturday, prior to the day when the employment-status questions were answered by or for the individual[58]. The identity of this day and of its concomitant reference week were not collected or captured in the census, so they cannot be determined with certainty. Nevertheless, the check-in date for a person can be used to estimate, or model, the reference week, by making a set of reasonable assumptions regarding the relationship between the check-in date for a individual and that individual's reference week.

This study used the following set of assumptions to associate a modeled reference week with each individual on the match dataset [59]:

(1) For mail forms, it was assumed that: a) the completed form was mailed the day after it was completed; b) there was a 3-day delay, on average, between the time the form was put into the mailbox (M day) and the day that the form was given a check-in date (C day) at the census collection center; and c) weekends and holidays had no effect upon the timing of any event related to the value of the check-in date.
(2) For enumerator forms, it was assumed that: a) there was a 7-day delay, on average, between the time the enumerator completed the form (F day) and the day that the form was given a check-in date (C day); and (b) weekends and holidays had no effect upon the timing of any event related to the value of the check-in date.

These assumptions led to the following conclusions:
(1) For mail forms: forms with check-in dates of Friday in week T to Thursday in week T+1 have reference period of week T-1.

---

[58] The questions for many people are answered by someone else in the individual's household – so-called "proxy" respondents.

[59] For the remainder of this discussion, the term "mail forms" excludes forms used to enumerate the population in group quarters (in Census 2000, the long forms used for the quarters population were: Form D-15B, the Individual Census Questionnaire; Form 20B, the Individual Census Report; Form D-21, The Military Census Report; and Form D-23, the Shipboard Census Report) ; the group-quarters forms are included in the term "enumerator forms."

(2) For enumerator forms: forms with check-in dates of Monday in week T+1 to Sunday in Week T+2 have reference period of week T-1.

These conclusions are reflected in the following table of correspondence between a person's check-in date (expressed as MM/DD) and the beginning and ending dates of the modeled reference week (also expressed as MM/DD) for the person:

Table B-1.1 Correspondence between census check-in dates and modeled reference weeks for mail-in forms

Check-in Date Range		Reference-period Week Number	Start and End Dates of Modeled Reference Week for Employment Status	
Start	End		Start	End
NA	03/09	1	02/20	02/26
03/10	03/16	2	02/27	03/04
03/17	03/23	3	03/05	03/11
03/24	03/30	4	03/12	03/18
03/31	04/06	5	03/19	03/25
04/07	04/13	6	03/26	04/01
04/14	04/20	7	04/02	04/08
04/21	04/27	8	04/09	04/15
04/28	05/04	9	04/16	04/22
05/05	05/11	10	04/23	04/29
05/12	05/18	11	04/30	05/06
05/19	05/25	12	05/07	05/13
05/26	06/01	13	05/14	05/20
06/02	06/08	14	05/21	05/27
06/09	06/15	15	05/28	06/03
06/16	06/22	16	06/04	06/10
06/23	06/29	17	06/11	06/17
06/30	07/06	18	06/18	06/24
07/07	07/13	19	06/25	07/01
07/14	07/20	20	07/02	07/08
07/21	07/27	21	07/09	07/15
07/28	08/03	22	07/16	07/22
08/04	08/10	23	07/23	07/29
08/11	08/17	24	07/30	08/05
08/18	08/24	25	08/06	08/12
08/25	NA	26	08/13	NA

NA– Any time prior to end of processing.

Table B-1.2 Correspondence between census check-in dates and modeled reference weeks for enumerator forms

Check-in Date Range		Reference-period Week Number	Start and End Dates of Modeled Reference Week for Employment Status	
Start	End		Start	End
NA	03/12	NA	02/20	02/26
03/13	03/19	1	02/27	03/04
03/20	03/26	2	03/05	03/11
03/27	04/02	3	03/12	03/18
04/03	04/09	4	03/19	03/25
04/10	04/16	5	03/26	04/01
04/17	04/23	6	04/02	04/08
04/24	04/30	7	04/09	04/15
05/01	05/07	8	04/16	04/22
05/08	05/14	9	04/23	04/29
05/15	05/21	10	04/30	05/06
05/22	05/28	11	05/07	05/13
05/29	06/04	12	05/14	05/20
06/05	06/11	13	05/21	05/27
06/12	06/18	14	05/28	06/03
06/19	06/25	15	06/04	06/10
06/26	07/02	16	06/11	06/17
07/03	07/09	17	06/18	06/24
07/10	07/16	18	06/25	07/01
07/17	07/23	19	07/02	07/08
07/24	07/30	20	07/09	07/15
07/31	08/06	21	07/16	07/22
08/07	08/13	22	07/23	07/29
08/14	08/20	23	07/30	08/05
08/21	08/27	24	08/06	08/12
08/28	09/03	25	08/13	08/19
09/04	NA	26	08/20	NA

NA– Any time prior to end of processing.

The following boxes display an excerpt from the computer program that applied the correspondences in the above table to each observation in the match dataset, and the definitions of the variables used in the program:

Box B-1 Excerpt from SAS ™ computer program that modeled the reference period

```
if RFT in ( '02', '04') then CAPDATE = MAILD;
else if NRD not in ("0000", " ") then CAPDATE = NRD;
else if CID not in ("0000", " ") then CAPDATE = CID;
else CAPDATE= REPDATE;

if RFT in ('02','04') and CAPDATE gt "0000" then do;
if    CAPDATE le   "0309"   then REFWEEK =  1       ;
else if CAPDATE le  "0316"   then REFWEEK =  2       ;
else if CAPDATE le  "0323"   then REFWEEK =  3       ;
.
.
.
else if CAPDATE le  "0824"   then REFWEEK =  25      ;
else if CAPDATE gt  "0824"   then REFWEEK =  26      ;
else REFWEEK =     27      ;
end;
else if CAPDATE gt "0000" then do;
if    CAPDATE le "0319"   then REFWEEK =    1       ;
else if CAPDATE le "0326"   then REFWEEK =  2       ;
else if CAPDATE le "0402"   then REFWEEK =  3       ;
.
.
.
else if CAPDATE le "0903"   then REFWEEK =  25      ;
else if CAPDATE gt  "0903"   then REFWEEK =  26     ;
else REFWEEK =     27      ;
end ;

else if CAPDATE eq "0000" then REFWEEK = 0;
else REFWEEK= -1;
```

Box B-2 Definitions of variables (extracted from 2000 Decennial Census SCUF Documentation) used in modeling program

---

RFT FORM TYPE :
01 = D-1 (Short Form MR)
02 = D-2 (Long Form MR)
03 = D-1(UL) (Short Form MR)
04 = D-2(UL) (Long Form MR)
05 = D-1(E) (Short Form EQ)
06 = D-2(E)   (Long Form EQ)
07 = D-10 (Be Counted)
08 = (not used)
09 = D-15A   (ICQ, Short)
10 = D-15B   (ICQ, Long)
11 = D-20A   (ICR, Short
12 = D-20B   (ICR, Long)
13 = (not used)
14 = D-21 (MCR)
15 = (not used)
16 = D-23 (SCR)
17 = D-1(E)SUPP   (Enumerator Supplement, short)
18 = D-2(E)SUPP   (Enumerator Supplement, long)
19 = D-1(E)(ccf)   (Short EQ converted to continuation)
20 = D-2(E)(ccf)   (Long EQ converted to continuation)
MAILD MAIL RETURN CHECK-IN MONTH AND DAY:
0000 = No Mail Return Check-in
0099 = Reverse Check-in
(When it is determined during the data capture process that a form doesn't contain enough data to be considered checked-in, MAILD is set to 0099.)
0101-1231= Check-in Day of 1st return
2000 = Checked-in in 2000 but we do not know the day it was actually checked-in.
 NRD NRFU CHECK-IN MONTH AND DAY (From OCS2000):
(May also be set from UUE or LE. If there is both a late mail return check-in and a NRFU check-in, NRD will contain the NRFU check-in month and day; however, the PSA will determine which   return is selected for the Census.)
0000 = No NRFU Check-in
0101-1231= NRFU Check-in Month and Day
CID CIFU CHECK-IN MONTH AND DAY (From OCS2000):
0000 = No CIFU Check-in
0101-1231= CIFU Check-in Month and Day
REPDATE EARLIEST FORM PROCESSING DATE
(from DCS2000 capture system)
blank = Date not captured
0101-1231= Earliest date (month and day)

Appendix C. Base Data for Detailed Tables

(Insert Appendix C Tables 1- 4 here)

Appendix D. Counterparts to Detailed Tables 2A-C

(Insert Appendix D Tables 1A-C, 2A-C here.)

Appendix E. On Using the CPS-Census 2000 Match to Evaluate the Performance of the Census 2000 Edit and Imputation Procedures for Employment Status

Note: This appendix reports the results of experimental research. It has undergone a Census Bureau review more limited in scope than that given to the main body of this report and to official Census Bureau publications. This appendix is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any comparisons made in this appendix have not undergone statistical testing and may not be significant at the 90-percent confidence level.

1. Background

The CPS-census match classifications can be used to evaluate the soundness of the census procedures that assign or impute values. The tables in this section are intended to take advantage of this capability. For this purpose, two operational definitions of soundness are used: (1) *overall* soundness is defined as the capacity of a census procedure to classify a person to the same category of a variable as the CPS does, regardless of category; it is measured by the proportion of *same* classifications (those where the census and CPS classifications agree) made by a procedure out of all classifications for the variable made by the procedure; (2) *within category* soundness is defined as the capacity of a census procedure to classify people to the same *given* category of a variable as the CPS; it is measured by the proportion of *same* classifications made by a procedure to a given category out of all its classifications to that category.

This appendix focuses on three census procedures: imputations in general; a special kind of imputation known as "MESRB" imputation; and census value-assignments in general. The first and third procedures were described in section 4.3 of the main body of this report. The following paragraph provides the background for the "MESRB" procedure:

In Census 2000, two matrixes were used to impute a person's employment status value. The first, called MESRA, was used when the person did not provide any useable information about whether they worked in the reference period. The donors to MESRA consisted of all people who had a fully reported or assigned employment-status value, regardless of the nature of the value. The nature of the donor pool meant that people imputed a value from MESRA could receive any one of the possible employment status values. The second matrix, MESRB, was used to impute values to people who indicated that they did *not* work in the census reference week, but who gave little or no other information. Donors to MESRB were restricted to people who reported that they *too* did not work last week. This restriction meant that MESRB could impute people only to the unemployed and not in labor force categories (it was possible to be imputed to the "employed, with a job but not at work" category from MESRB, but the chances were slight) .

2. Census Imputations

Table E-1A shows that, overall, the census imputation procedure was successful in making a correct classification nearly three-fourths of the time (72.1 percent of the classifications agreed with the CPS).

Table E-1B presents the data for the "within-category" measures of soundness. They show that the imputation procedures had a success rate of 79.2 percent for the not in labor force category and 69.4 percent for the employed category, but only 1.1 percent for the unemployed category .

(Insert Appendix E Tables 1A and 1B here.)

3. MESRB Imputations

Table E-2A shows that, overall, the census imputation procedure using matrix MESRB was successful in making a correct classification nearly 80 percent of the time (77.7 percent of its classifications agreed with the CPS).

Table E-2B shows that, for the "within-category" measures of soundness, the MESRB procedure had a success rate of 86.7 percent for the not in labor force category .

(Insert Appendix E Tables 2A and 2B here.)

4. Assignments

Table E-3A shows that, overall, the census value-assignment was successful in making a correct classification nearly 85 percent of the time (84.9 percent of its classifications agreed with the CPS).

Table E-3B shows that, for the "within-category" measures of soundness, the assignment procedure succeeded 85.6 percent of the time for the not in labor force category, and 59.5 percent of the time for the unemployed category.

(Insert Appendix E Tables 3A and 3B here.)

Appendix F. On Using the CPS-Census 2000 Match to Quantify the Reference Period Effect on Comparisons of Census 2000 and CPS Estimates

Note: This appendix reports the results of experimental research. It has undergone a Census Bureau review more limited in scope than that given to the main body of this report and to official Census Bureau publications. This appendix is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any comparisons made in this appendix have not undergone statistical testing and may not be significant at the 90-percent confidence level.

The reference period of an estimate is the span of time during which the events associated with the estimate were observed; it is analogous to the *exposure* period in photography. A reference period has the following properties: a *duration* (for example:1 day; 7 successive days; 30 total days); a *framework* (for example: a full calendar week; a calendar month; the first quarter of a particular year); and a calendar *orientation* or *timing* (for example: the full calendar week containing the 12th day of a particular month; the full calendar week prior to some date or action; the week of March 19, 2000 through March 25, 2000).

The duration and framework of the reference period of the Census 2000 labor force concept were the same for all of the observed events: that is, the seven successive days of a full calendar week, from Sunday through Saturday. The timing, however, is marked by considerable indistinctness, related to fact that the labor force estimates are aggregates of individual observations, and, for operational reasons, the reference period for any particular observation is not necessarily the same as that for any other observation.

The Census 2000 labor force questions asked each individual to describe events that occurred in the calendar week prior to when the individual filled out the Census 2000 form. People filled out the forms in a variety of weeks, so the timing of the description for any individual can vary over the approximately 25 full calendar weeks in the Census 2000 data-collection period [60]. This variation means that the aggregates of the individual observations (that is, the published labor force estimates) are associated with a range of calendar weeks, rather than with a particular calendar week as in the CPS, where all observations are connected to the same week. Hence, at the aggregate level, the Census 2000 reference period is a fuzzy concept, possessing the nature of a composite; it is perhaps best expressed by the phrase "at the time of Census 2000" (and left at that).

Since people can change their relationship to the work force – which is what the Census 2000 and CPS labor force concepts measure – from one week to the next, the timing of the Census 2000 and CPS reference periods is a factor in the sizes of their respective labor force estimates. In an attempt to quantify the contribution of this factor to the Census 2000 estimates, the procedure

---

[60] Because of misunderstandings by respondents, it may also vary according to when the respondent considers "last week" to have begun.

described in this appendix defined a quantity called the "Reference-Period Effect" (RPE). The RPE for a given Census 2000 labor force estimate is the difference between the actual estimate and what the estimate would have been if the reference period for each person represented in the estimate had occurred in the same calendar month [61], called the *focus* month. The procedure attempted to estimate the RPE for each of the national-level estimates of the labor force categories using the records of the CPS – Census 2000 Match file (CCM). Two sets of estimates were made, one using March 2000 as the focus month, and the other using April 2000.

The estimates of the RPEs are based on the following assumptions:
1. the subset of people in the CCM whose reference period for the CPS employment-status variable was in the focus month are representative of the corresponding general population;
2. the true reference week for the Census 2000 employment-status variable for each of these people is the one predicted by the modeling methods described in Appendix B;
3. the true employment status *in the focus month* of those people whose modeled Census 2000 reference week was *not* in the focus month, was the employment status recorded for them in the CPS for the focus month;
4. if the Census 2000 reference period for people in assumption 3 had been in the focus month, then their employment status category in the Census 2000 would have been the same as their CPS category for that month.

The procedure to make the estimates, in essence, created a simulated Census 2000 employment-status distribution for the focus month, by (1) accepting the actual Census 2000 value[62] of people whose modeled Census 2000 reference period was in that month, and (2) replacing the actual Census 2000 value with the CPS value for the focus month, for people whose modeled Census 2000 reference period was *not* in that month. The result was a new distribution consisting entirely of either actual or simulated values whose modeled Census 2000 reference period was *in the focus month*. This new distribution was then compared with the published Census 2000 distribution, which consisted entirely of actual values (whose respective reference periods were not necessarily in the focus month). The difference between the published estimate for a category and the corresponding estimate in the new distribution was the RPE for that category.

The following paragraphs describe the steps in the procedure, using March 2000 as the focus month. The description is followed by Tables F-1 and F-2 that show the results from the procedure for the March 2000 and April 2000 focus months, respectively. A brief discussion of the results follows the tables.
Procedure:

---

[61] It would have been preferable to have used the condition that the reference period for all people was the same calendar *week* (in particular, the CPS reference week), but this level of precision was beyond the capacity of the methodology.

[62] That is, the value they actually received in the census and that is reflected in published census figures.

Step 1. Tabulate the weighted Census 2000 employment-status (ESR [63]) distribution for all people in the CCM who have a March CPS record [64] and whose Census 2000 age and CPS age are both greater than 15 years. Label the quantity in a given ESR category of this distribution: *Observed Census 2000 March ESR* quantity.

Step 2. Create a cross-tabulation of Census 2000 ESR by CPS March ESR for people whose modeled Census 2000 reference period is **not** in March and whose Census 2000 ESR is **not** the same as their CPS March ESR, and whose Census 2000 age and CPS age are both greater than 15.

Step 3. Using the above cross tabulation, create the following table:

CPS March ESR	Census 2000 ESR			
	Employed	Unemployed	Not in Labor Force	Total
Employed	Not Applicable	a_1	a_2	$a_1 + a_2 = a$
Unemployed	b_1	Not Applicable	b_2	$b_1 + b_2 = b$
Not in Labor Force	c_1	c_2	Not Applicable	$c_1 + c_2 = c$
Total	$b_1 + c_1 = d$	$a_1 + c_2 = e$	$a_2 + b_2 = f$	$d+e+f = a+b+c$

Step 4. To each of the employment categories in the distribution developed in Step 1, add the quantity in the "Census 2000 ESR, Total" column of the corresponding row of the category, and subtract the quantity in the "March CPS ESR, Total" row of the corresponding column of the category (for example, to the employed category of the step 1 distribution, add quantity a and subtract quantity d). Label the quantity in a given ESR category of the new distribution developed by this procedure: *CPS-Modeled Census 2000 March ESR* quantity, or *Modeled Census 2000 March ESR* quantity, for short.

Step 5. For each ESR category, express the Modeled Census 2000 March ESR quantity as a ratio of the Observed Census 2000 March ESR quantity. Label these ratios *Adjustment Coefficients*.

Step 6. Multiply each published Census 2000 ESR quantity by its corresponding adjustment coefficient from Step 4. Label the ESR distribution formed by these quantities *Adjusted Published (AP) Distribution* .

Step 7. Subtract the quantities in the AP Distribution from the corresponding categories of the published Census 2000 distribution for the civilian noninstitutional population 16 years and over.

---

[63] The employment status variable in the census and the CPS is commonly labeled "ESR," the acronym for Employment Status Recode, since it represents a recode of the values from other variables.

[64] Not all the people on the CCM have a record for the March CPS: the file contains the record for the first month on or after February 2000 in which the person's household was in the CPS sample.

The resulting figure for each category represents the effect that the non-uniformity of the Census 2000 reference-period had on that category, using March 2000 as the frame of reference.

The following tables are worksheets presenting the results of the procedure:

Table F-1. Estimates of Reference Period Effects Using March 2000 as the Focus Month

Labor Force Category	Published Census 2000 Data (p)	Step 1 Output: Observed Census 2000 March ESR (a)	Step 4 Input: March CPS ESR (b)	Step 4 Input: Census 2000 ESR (c)	Step 4 Output: Modeled Census 2000 March ESR (d = (a) + (b-c))
Employed	129,722,000	79,369,254	3,702,198	2,931,139	80,140,313
Unemployed	7,947,000	4,381,229	1,137,502	1,259,152	4,259,579
Not in Labor Force	74,365,000	43,273,320	2,789,651	3,439,060	42,623,911
	Adjustment Coefficients (d / a)	Adjusted Published Distribution (e = (p) * (d/a))	Reference Period Effects (p - e)	Labor Force Category	
	1.01	130,982,227	-1,260,227	Employed	
	0.97	7,726,342	220,658	Unemployed	
	0.98	73,248,994	1,116,006	Not in Labor Force	
	Step 2 and Step 3 Outputs:				

		Census 2000 ESR			
		Employed	Unemployed	Not in Labor Force	Total
March CPS	Employed	0	688,001	3,014,197	3,702,198
	Unemployed	712,639	0	424,863	1,137,502
	Not in Labor Force	2,218,500	571,151	0	2,789,651
	Total	2,931,139	1,259,152	3,439,060	7,629,351

72

Table F-2. Estimates of Reference Period Effects Using April 2000 as the Focus Month

Labor Force Category	Published Census 2000 Data	Step 1 Output: Observed Census 2000 April ESR	Step 4 Input: April CPS ESR	Step 4 Input: Census 2000 ESR	Step 4 Output: Modeled Census 2000 April ESR
	(p)	(a)	(b)	(c)	(d = (a) + (b-c))
Employed	129,722,000	18,494,205	1,675,829	1,263,594	18,906,440
Unemployed	7,947,000	1,022,148	646,168	580,953	1,087,363
Not in Labor Force	74,365,000	10,613,969	1,202,455	1,679,905	10,136,519

Adjustment Coefficients	Adjusted Published Distribution	Reference Period Effects	Labor Force Category
(d / a)	(e = (p) * (d/a))	(p - e)	
1.02	132,613,498	-2,891,498	Employed
1.06	8,454,034	-507,034	Unemployed
0.96	71,019,826	3,345,174	Not in Labor Force

Step 2 and Step 3 Outputs:

April CPS	Census 2000 ESR Employed	Unemployed	Not in Labor Force	Total
Employed	0	275,641	1,400,188	1,675,829
Unemployed	366,451	0	279,717	646,168
Not in Labor Force	897,143	305,312	0	1,202,455
Total	1,263,594	580,953	1,679,905	3,524,452

Discussion of results:

The RPE figures for March 2000 in the above table indicate that, if the Census 2000 reference week had been in March 2000 for all people, the Census 2000 estimate of the number of employed people would have been about 1.3 million higher than the published figure, the number of unemployed about 200,000 fewer, and the number not in the labor force 1.1 million less. The parallel figures for April 2000 are: employed – 2.9 million higher ; unemployed – 500,000 higher ; and not in labor force – 3.3 million lower.

The validity of comparisons of published figures from the Census 2000 and the CPS suffers from the presence of RPEs in the Census 2000 figures. By supposedly eliminating the reference-period effects from the Census 2000 figures for the focus month, the "adjusted published (AP) distributions" in the above tables permit one to make Census 2000–CPS comparisons free from the distortions of these effects. This is done by comparing the Census 2000 AP figures for a focus month with the CPS figures for that same month. The results of such comparisons are presented in Tables F-3 and F-4:

Table F-3. Employment Status Estimates: Published Census 2000 figures, Adjusted Published Census 2000 figures, and Current Population Survey figures for March 2000 : United States, Total (numbers in thousands)

Employment Status	Published Census 2000 data (col 1)	Adjusted Published Census 2000 Data for Focus Month of March 2000 (col 2)	March 2000 CPS data (col 3)	Difference: col 1 - col 3	Difference: col 2 - col 3
Employed	129,722	130,982	136, 054	-6,332	-5,072
Unemployed	7,947	7,726	6,069	1,878	1,657
Not in Labor Force	74,365	73,249	69,649	4,716	3,600

Table F-4. Employment Status Estimates: Published Census 2000 figures, Adjusted Published Census 2000 figures, and Current Population Survey figures for April 2000 : United States, Total (numbers in thousands)

Employment Status	Published Census 2000 data (col 1)	Adjusted Published Census 2000 Data for Focus Month of April 2000 (col 2)	April 2000 CPS data (col 3)	Difference: col 1 - col 3	Difference: col 2 - col 3
Employed	129,722	132,613	136, 927	-7,205	-4,314
Unemployed	7,947	8,454	5,212	2,735	3,242
Not in Labor Force	74,365	71,020	69,879	4,486	1,141

The RPE estimates in Tables F-1 to F-4 are merely first approximations. The estimates for both comparison months are surprisingly high, and the ones for April 2000 are especially suspect, given the results shown in Table F-5:

Table F-5. Differences between estimates from Census 2000 and from the Current Population Survey for March, April, and May 2000 : United States, Total (numbers in thousands)

Employment Status	March 2000 CPS	April 2000 CPS	May 2000 CPS	Weighted Average CPS March-May 2000
Employed	-6,332	-7,205	-6,963	-7,084
Unemployed	1,878	2,735	2,487	2,367
Not in Labor Force	4,716	4,486	4,268	4,490

The rightmost column of Table F-5 shows the difference between the Census 2000 published figures and the corresponding weighted average CPS figures for March-May 2000. Like the figures in the rightmost column of Table F-3 and of Table F-4, these differences represent the outputs of a method – less refined, but likely effective – to eliminate reference-period effects from Census 2000 – CPS comparisons. That they are so different from their counterparts in Tables F-3 and F-4 may be an indication of the presence of flaws in the procedure used in Tables F-1 and F-2 to estimate RPEs.[65]

---

[65] Possible flaws include weaknesses in the validity of the underlying assumptions, especially the first one (particularly for the April focus month).

Appendix G. Using the CPS-Census 2000 Match to Develop or Examine Hypotheses About the Census 2000 Employment Status Categories

Note: This appendix reports the results of experimental research. It has undergone a Census Bureau review more limited in scope than that given to the main body of this report and to official Census Bureau publications. This appendix is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any comparisons made in this appendix have not undergone statistical testing and may not be significant at the 90-percent confidence level.

1. Hypotheses concerning the Employed, With a Job, Not at Work category in Census 2000
a. Classification as Employed in Census 2000
As explained in Box 2 in the main text, the employed category has two subcategories: (1) Employed, at work (the "at-work" subcategory) ; and (2) With a job, not at work (the "with-job" subcategory) . Both the census and the CPS provide counts for each of these subcategories. A comparison of Census 2000 estimates with CPS estimates for these categories, for March 2000, April 2000, and the combined March-April 2000 period, is shown in Table G-1:

Table G-1.Comparison of Census 2000 and CPS Estimates for March 2000, for April 2000, and for March-April 2000 Averages, for the At-Work and With-Job Subcategories of Employed People (numbers in thousands):

Employed category	Census 2000	March 2000 CPS	April 2000 CPS	March-April 2000 Average	Difference Census 2000 – Average CPS
Total Employed	129,722	136,054	136,927	136,490	-6,768
At work	127,156	131,206	132,877	132,041	-4,885
With a job, not at work	2,565	4,848	4,050	4,449	-1,884

The table shows that the census estimate in the with-job category was about 40 percent lower than the average March-April 2000 CPS estimate. Perhaps more significantly, the difference between the two surveys' estimates in the with-job category made up slightly over 25 percent of the difference between their estimates in the overall employed category, even though the with-job category made up only 2 .0 percent of the Census 2000 employed category and 3.3 percent of the CPS employed category. The highlighted cell in the table, representing approximately 1.9 million people, shows the absolute difference between the Census 2000 and average CPS counts for this category.

An individual who is in the with-job category in the CPS, but who is not in that category in

Census 2000 (as could be true for any of the people represented in the highlighted cell) may still be classified as employed in Census 2000 if the individual is in the Census 2000 at-work category. In other words, because shifts between the employed subcategories have no effect on the overall employed category, a CPS-Census 2000 discrepancy in the with-job category does not necessarily imply a discrepancy in the employed category. The actual contribution of the difference in the with-job category to the difference in the overall CPS-Census 2000 employment category depends upon the proportion of the people in the highlighted cell who are classified as at-work in the census[66]. The CPS-Census 2000 Match provides a means to test the hypothesis that this proportion is high, or, alternately stated, that most of the with-job people in the CPS who were missed by the Census 2000 with-job category were still classified as employed in Census 2000 because they fell into the Census 2000 at-work category.

The data in Table G-2 support this hypothesis. They show that, among the cases in the CPS-Census 2000 Match in general, 81.4 percent of the people in the CPS with-job category were classified as employed in Census 2000. This high proportion came about because, even though only 11.1 percent of these people were classified as with-job in Census 2000, 70.3 percent were captured by the Census 2000 at-work category.

Table G-2. CPS-Based Percentage Distribution– CPS Employed Categories by Employment-Status Category in Census 2000, for All People in the CPS-Census 2000 Combined-month Match

CPS Category	**Census 2000 Category**					
	Total	**Employed**			**Unemployed**	**Not in labor Force**
		Total	**Employed, At Work**	**Employed, With Job, Not at Work**		
Employed, Total	100%	90.6	89.6	1.0	1.7	7.7
At Work	100%	90.9	90.2	0.7	1.7	7.4
With a job, not at work	100%	81.4	70.3	11.1	2.9	15.7

Further support for the hypothesis is provided by the data in Tables G-3 and G-4. Table G-3 attempts to lessen the impact of reference-period effects on the analysis (see Section 3.1 in the

---

[66] Ignoring those people in the Census 2000 with-job category who are not classified as with-job in the CPS.

77

main text, and Appendix F) by restricting the comparisons in Table G-2 to people whose modeled Census 2000 reference week was in March 2000 (see Section 3.2 and Appendix B) and whose CPS reference week was in March 2000. Table G-3 shows that 83.9 percent of people in the CPS with-job category were classified as employed in Census 2000: 65.4 percent as at-work and 18.5 percent as with-job.

Table G-3. Percentage Distribution– CPS Employed Categories by Employment-Status Category in Census 2000, for People in the CPS-Census 2000 Combined-month Match With Modeled Census 2000 Reference Week in March 2000 and CPS Reference Week in March 2000

CPS Category	Census 2000 Category					
	Total	Employed			Unemployed	Not in labor Force
		Total	Employed, At Work	Employed, With Job		
Employed, Total	100%	92.3	91.2	1.1	1.2	6.5
At Work	100%	92.5	92.0	0.5	1.2	6.3
With a job, not at work	100%	83.9	65.4	18.5	2.7	13.3

Table G-4 further restricts the comparison to people who gave a complete report to the employment questions on the Census 2000 questionnaire. Again, the data show that a high proportion of people in the CPS with-job category, 87.7 percent, were classified as employed in Census 2000: 67.9 percent as at-work and 19.8 percent as with-job.

Table G-4. Percentage Distribution– CPS Employed Categories by Employment-Status Category in Census 2000, for People in the CPS-Census 2000 Combined-month Match With Modeled Census 2000 Reference Week in March 2000 and CPS Reference Week in March 2000, Whose Employment Status Items Were Fully-Reported in Census 2000

CPS Category	Census 2000 Category					
	Total	Employed			Unemployed	Not in labor Force
		Total	Employed, At Work	Employed, With Job		
Employed, Total	100%	95.1	94.0	1.0	0.9	4.1
At Work	100%	95.3	94.9	0.4	0.8	3.9
With a job, not at work	100%	87.7	67.9	19.8	2.9	9.3

The data in Tables G-2, G-3, and G-4 indicate that, at most, about 20 percent of the people in the CPS with-job category were **not** classified as employed in Census 2000. Applying this percentage to the highlighted figure of -1.9 million in Table G-1 implies that, at the maximum, factors related to the Census 2000 with-job category may have contributed about 400,000 people (that is, approximately .2 multiplied by -1,894,000) to the 6.8 million gap between the average March-April 2000 CPS estimate and the Census 2000 estimate of total employed (about 6 percent).

The CPS collects information on the reason people in the with-job category are not at work. These data, available on the CPS-Census 2000 Match, can be used to gain some insights into why Census 2000 likely failed to classify a significant number of people in the CPS with-job category to one of the Census 2000 employed categories. The census questionnaire is an obvious starting point from which to seek the sources of any such failure; and the most useful data for examining questionnaire issues are those that are restricted to people who fully reported the employment items in the census, for these data are theoretically free of confounding effects from census edit or imputation factors.

The universe of Table G-5A consists of the people in the CPS with-job category who fully reported the employment questions in Census 2000. The table distributes these people by the main reason they were not at work in the CPS; then it shows the percentage distribution for the people in each reason category by whether they were employed in Census 2000 . For the same universe, Table G-5B presents percentage distributions of people in the Census 2000 employed/not employed categories, by reason for not working in the CPS. Tables G-5A and G-5B suggest that people who were not at work because of the following reasons: "maternity/paternity leave, " "weather affected job," "school/training, " and "other reasons", were most likely not to be classified as employed in Census 2000. The data support the hypothesis that the absence of these reasons from among the list

of examples in the Census 2000 "temporary work " question[67] could have been a significant source of Census 2000 misclassifications, for the answers to this question determined whether an individual in the census was classified as with-job (and therefore as employed) or as not employed (unemployed or not in labor force) .

---

[67] The Census 2000 question listed "on vacation, temporary illness, labor dispute, etc." only as examples of reasons for answering "yes" to the question: "LAST WEEK, was this person TEMPORARILY absent from a job or business?

Table G-5A. Percentage Distribution – CPS Employed With Job, Not At Work Category, by Reason Not At Work in CPS, by Employed/Not Employed Status in Census 2000, for People in the CPS-Census 2000 Combined-month Match With Modeled Census 2000 Reference Week in March 2000 and CPS Reference Week in March 2000, Whose Employment Status Items Were Fully-Reported in Census 2000

Reason Not At Work in CPS	Total		Employed in Census 2000	Not Employed In Census 2000
	Number (in thousands)	Percent		
With a job, not at work, Total	1,434	100%	87.7	12.3
Illness	347	100%	90.1	9.9
Vacation	751	100%	91.5	8.5
Weather Affected Job	9	100%	8.4	91.6
Labor Dispute	4	100%	100.0	-
Child Care Problems	-	-	-	-
Family/Personal Obligation	24	100%	98.5	1.5
Maternity/Paternity Leave	158	100%	79.7	20.3
School/Training	23	100%	69.9	30.1
Civic/Military Duty	44	100%	100.0	-
Does not work in Business	-	-	-	-
Other reason	73	100%	59.4	40.6

- Zero or rounds to zero.

Table G-5B. Percentage Distribution – CPS Employed With Job, Not At Work Category, by Employed/Not Employed Status in Census 2000, by Reason Not At Work in CPS, for People in the CPS-Census 2000 Combined-month Match With Modeled Census 2000 Reference Week in March 2000 and CPS Reference Week in March 2000, Whose Employment Status Items Were Fully-Reported in Census 2000

Reason Not At Work in CPS	Employed in Census 2000	Not Employed In Census 2000
With a job, not at work, Number (in thousands)	1,258	176
Percent	100%	100%
Illness	24.8	19.6
Vacation	54.6	36.2
Weather Affected Job	--	4.9
Labor Dispute	0.3	--
Child Care Problems	--	--
Family/Personal Obligation	1.9	0.2
Maternity/Paternity Leave	10.0	18.3
School/Training	1.3	3.9
Civic/Military Duty	3.5	--
Does not work in Business	--	--
Other reason	3.5	16.9

- Zero or rounds to zero.

b. Classification to the With-Job category in Census 2000

A survey like the CPS or Census 2000 takes measurements of a variable, such as employment status, at the person level (micro-level measurements) to produce two kinds of measurements at the aggregate level (macro-level estimates): measurements of aggregate levels of the variable (for example, how many people are employed; the unemployment rate in a given place); and measurements of the aggregate relationships between that variable and other variables (for example, how many females between 20 and 44 years of age are in the labor force). Errors at the micro level (for example, classifying a person whose characteristics meet the criteria the for with-job category to the at-work category) do not necessarily affect the accuracy of the first kind of aggregate estimate. If one member of a group who truly belongs in the with-job category is erroneously

classified as at-work, and vice-versa for another member of the group, the overall counts of with-job and at-work for the group are unaffected. Such errors, however, very likely affect the accuracy of the second kind of aggregate estimate.

In addition to its extrinsic use in the census in classifying people to the employed category, the with-job category is intrinsically crucial to collecting accurate data on the at-work population. The at-work population is the basis of the census data on journey-to-work, which are important in transportation-planning studies. Response errors in the with-job category at the micro-level, even if they do not have a major impact on macro-level census estimates of total employed (or any impact at all in the case of off-setting errors between the with-job and the at-work categories) may seriously distort measurements of the correlations between variables, which are critical in such studies.

The data in Tables G-2,G-3, and G-4 indicate that Census 2000 probably did a poor job of making the at-work/with-job distinction for employed people. This assertion must be made cautiously because the difference between the CPS and Census reference periods probably has its greatest effect on characteristics that tend to be short-lived, and the at-work/with-job distinction, which involves such states as being on vacation or short illnesses, is likely to be the most fleeting of all labor force relationships for most people. Nevertheless, it does appear that a substantial proportion of people classified in the Census 2000 as at-work probably should have been classified to the with-job category.

Tables G-6A and G-6B are the counterparts of Tables G-5A and G-5B, the difference being that, in the former tables, a with-job/not-with-job dichotomy replaces the employed/not-employed dichotomy of the latter tables. Like the G-5 tables, the G-6 tables hint at problems in the census questionnaire as the source of census misclassifications in the with-job category. It appears that regardless of the reason that people in the CPS with-job category were not at work, they had a high propensity to be in some other category in Census 2000, with most reporting that they were at work; people in the "vacation" category had the greatest numerical impact. To be classified to the at-work category in the census, an individual must answer "yes" to the census at-work question: "LAST WEEK, did this person do ANY work for either pay or profit"; if the answer to this question is "yes", the person is not asked the question about temporary absences that is the determining factor in making the with-job classification. Tables 6A and 6B, especially Table 6A, suggest that there was considerable misunderstanding of the at-work question in Census 2000 among people who were temporarily absent from a job.

Table G-6A. Percentage Distribution – CPS Employed With Job, Not At Work Category, by Reason Not At Work in CPS, by With Job/Not With Job Status in Census 2000, for People in the CPS-Census 2000 Combined-month Match With Modeled Census 2000 Reference Week in March 2000 and CPS Reference Week in March 2000, Whose Employment Status Items Were Fully-Reported in Census 2000

Reason Not At Work in CPS	Total		In the With-Job Category in Census 2000	Not in the With-Job Category in Census 2000	
	Number (in thousands)	Percent		Total	At Work
With a job, not at work, Total	1,434	100%	19.8	80.2	67.9
Illness	347	100%	45.5	54.5	44.6
Vacation	751	100%	3.5	96.5	88.0
Weather Affected Job	9	100%	-	100.0	8.4
Labor Dispute	4	100%	-	100.0	100.0
Child Care Problems	-	-	-	-	-
Family/Personal Obligation	24	100%	-	100.0	98.5
Maternity/Paternity Leave	158	100%	63.4	36.6	16.3
School/Training	23	100%	-	100.0	69.9
Civic/Military Duty	44	100%	-	100.0	100.0
Does not work in Business	-	-	-	-	-
Other reason	73	100%	-	100.0	59.4

-- Zero or rounds to zero.

Table G-6B. Percentage Distribution – CPS Employed With Job, Not At Work Category, by With Job/Not With Job Status in Census 2000, by Reason Not At Work in the CPS, For People in the CPS-Census 2000 Combined-month Match With Modeled Census 2000 Reference Week in March 2000 and CPS Reference Week in March 2000, Whose Employment Status Items Were Fully-Reported in Census 2000

Reason Not At Work in CPS	With Job in Census 2000	Not With Job in Census 2000	
		Total	At Work
With a job, not at work, Number (in thousands)	284	1,150	974
Percent	100%	100%	100%
Illness	55.5	16.5	15.9
Vacation	9.2	63.0	67.9
Weather Affected Job	-	0.8	0.1
Labor Dispute	-	0.4	0.4
Child Care Problems	-	-	-
Family/Personal Obligation	-	2.1	2.5
Maternity/Paternity Leave	35.3	5.0	2.6
School/Training	-	2.0	1.6
Civic/Military Duty	-	3.8	4.5
Does not work in Business	-	-	-
Other reason	-	6.4	4.5

- Zero or rounds to zero.

2. Hypotheses concerning the Employed, at Work category in Census 2000

Table 2B in the main body of the text showed that there was considerable agreement between the census and the CPS classifications for people who were at work in the CPS. Nevertheless, about 7 percent of the people in this category in the CPS were not employed in the census (controlling for reference period effects). Because this category contains a relatively large number of people, even a

small percentage difference here between the census and the CPS can lead to large absolute differences between their respective estimates of employed people.

This section discusses an effort to use the CPS-Census 2000 Match to search for reasons why a significant proportion of people who were at-work in the CPS were classified into one of the not employed (unemployed/not in labor force) categories in Census 2000 [68]. The effort focused on the Census 2000 questionnaire as a source of these discrepancies, so the universe of the study was restricted to fully reported persons in the census whose Census 2000 and CPS reference weeks were in March 2000. The research tried to identify characteristics related to a high propensity of people in the CPS at-work category to be classified as not employed in the census, in the hope that such relationships could help reveal problems with the census questions. For the at-work people in the CPS, there is no equivalent information to the "reasons for not working" data available for the CPS with-job population, so greater reliance must be placed on inferences about relationships than is the case for the CPS with-job category.

Table G-7A reveals that the listed categories of the following characteristics are associated with a high propensity among people in the CPS at-work category to be classified in Census 2000 as not employed:
– Age: 16 to 19 years; 20 to 24 years; 65 years and over;
– Class of worker: self-employed, unincorporated; without-pay worker;
– Educational Attainment: High School or less, no diploma .

Table G-7B reveals that people with the characteristics in the above list are over-represented in the census not employed category, compared with their representation in the employed category.

Adams (2003) hypothesizes that differences between the census and the CPS in how they collect labor force information from self-employed people, multiple jobholders, and retired people may be a factor in differences between their labor force estimates. Another theory is that the increasing difficulty of the census to accurately measure labor force status may be related to the growing presence in the workforce of people with nontraditional work arrangements, such as so-called contingent workers, for whom many of the terms used in the census questions (such as "last week," "at-work," "temporarily absent," "layoff," "looking for work") have ambiguous, nontraditional, or even ambivalent meanings, and for whom the official concept of employment status may be too rigid to describe their relationship to the labor market. The findings in Tables 7A and 7B appear, at least superficially, to be consistent with these hypotheses. For the census employed category, the findings indicate that problems with the "work last week" question may be a major source of misclassifications, for this question is almost the sole factor in determining whether to classify a person as employed or not employed. The problems that this question poses to people with the high-propensity characteristics would be a potentially fruitful area for further research.

---

[68] The ideas in this section borrow heavily from those in Adams (2003).

Table G-7A. Percentage Distribution – Selected CPS-Based Characteristics of People in the CPS Employed At Work Category By Employed/Not Employed Category in Census 2000, for People in the CPS-Census 2000 Combined-month Match Whose Modeled Census 2000 Reference Week Was in March 2000 and Whose CPS Reference Week Was in March 2000, and Whose Age Is Greater than 15 in Both the CPS and Census

Selected Characteristics in CPS	Percent	Employed in Census 2000	Not Employed In Census 2000
Total	100%	95.3	4.7
SEX			
Male	100%	95.2	4.8
Female	100%	95.3	4.7
AGE			
16-19 years of age	100%	91.5	8.5
20 -24 years of age	100%	91.9	8.1
25-34 years of age	100%	95.4	4.6
35-44 years of age	100%	96.3	3.7
45-54 years of age	100%	96.6	3.4
55-64 years of age	100%	95.7	4.3
65 years and over	100%	89.0	11.0
CLASS OF WORKER OF JOB 1			
Federal Government	100%	99.0	1.0
State Government	100%	95.9	4.1
Local Government	100%	98.0	2.0
Private	100%	95.6	4.4
Self-Employed, Unincorporated	100%	86.9	13.1

Without Pay	100%	41.4	58.6
EDUCATIONAL ATTAINMENT			
High School or less, no diploma	100%	86.9	13.1
High School diploma	100%	94.7	5.3
Some college, Associate Degree	100%	95.3	4.7
Bachelor's Degree	100%	98.1	1.9
Master's Degree or Doctorate Degree	100%	98.9	1.1
MULTIPLE JOBHOLDING			
Single Jobholder	100%	95.1	4.9
Multiple Jobholder	100%	97.3	2.7

Table G-7B Percentage Distribution – Selected CPS-Based Characteristics of People in the CPS Employed At Work Category By Employed/Not Employed Category in Census 2000, for People in the CPS-Census 2000 Combined-month Match Whose Modeled Census 2000 Reference Week Was in March 2000 and Whose CPS Reference Week Was in March 2000, and Whose Age Is Greater than 15 in Both the CPS and Census

Selected Characteristics in CPS	Employed in Census 2000	Not Employed In Census 2000
Total, Number (in thousands)	162,506	8,104
Percent	100%	100%
SEX		
Male	53.0	53.7
Female	47.0	46.3
AGE		
16-19 years of age	4.2	7.9
20 -24 years of age	8.4	14.7
25-34 years of age	21.7	20.7
35-44 years of age	27.1	20.7
45-54 years of age	24.1	17.1
55-64 years of age	11.0	10.0
65 years and over	3.6	8.9
CLASS OF WORKER OF JOB 1		
Federal Government	2.8	0.5
State Government	3.9	3.4
Local Government	9.2	3.8
Private	78.0	71.7

Self-Employed, Unincorporated	6.0	18.1
Without Pay	-	2.5
EDUCATIONAL ATTAINMENT		
High School or less, no diploma	9.5	28.7
High School diploma	28.5	32.2
Some college, Associate Degree	28.3	28.2
Bachelor's Degree	21.6	8.2
Master's Degree or Doctorate Degree	12.2	2.7
MULTIPLE JOBHOLDING		
Single Jobholder	94.6	97.0
Multiple Jobholder	5.4	3.0

- Zero or rounds to zero.

3. Hypotheses concerning the Unemployed category in Census 2000

At the national level, the Census 2000 count of unemployed people was considerably higher than the corresponding counts from the CPS for the March through May 2000 period . Several hypotheses have been advanced to account for the gap. One is that the census questionnaire does not clearly distinguish between active and passive methods of job search. Only active search methods qualify a person to be unemployed, and so, by failing to make the active/passive distinction, the census could be erroneously inflating the unemployed count by including in it people who used passive methods only. A second hypothesis is that the census may be classifying as unemployed people who are considered by the CPS to be so-called discouraged workers: that is, people without jobs, who want to work, but who did not recently look for work because they believed that no work was available for which they were qualified. A third hypothesis is that the census may be classifying as unemployed some people who looked for work while they worked at jobs; according to the hierarchical criteria for the employment status classification, these people should be classified as employed. This kind of error may have been especially prevalent for many people in the so-called contingent workforce, which, as described in the above section, consists of people with

nontraditional work arrangements; why the error occurred, if it did, is unknown. In this section, the CPS-Census 2000 Match is used to looked for evidence to support these hypotheses.

For evidence supporting the first two hypotheses, the people in the universe of Tables G-5, G-6, and G7 who were classified in Census 2000 as unemployed, looking for work, but as not in the labor force in the CPS, were tabulated to see how many were either discouraged workers in the CPS, or were people who had looked for worked in the CPS, but had used only passive methods to search [69]. The results indicated that of the approximately 450,000 people in the tabulation, about 66,000 (approximately 15 percent) were either discouraged workers or passive job searchers. This finding suggests that confusion about active and passive methods of job search in the census, and issues related to discouraged workers, are important, but not decisive, factors in creating the census overcount of unemployed people compared with the CPS.[70]

Some admittedly-weak support for the third hypothesis is provided by the data in the following table, which is extracted from Table 2A of the main text:

Table G-8 Census-Based Percentage Distribution – People with modeled census reference week in March 2000 in the Unemployed, Looking for Work category in Census 2000, by Employment status in the CPS

CPS Classification	Unemployed, Looking for Work in Census 2000
16 years and over	100.0%
Civilian Labor Force	68.8%
Employed	22.2%
At Work	20.7%

The data show that, for people classified in the Census 2000 as unemployed because they met the job search criteria, slightly over 20 percent were considered employed, at work, in the CPS. The meaning and implications of this finding are subjects for further research.

4. A hypothesis concerning the effect of collection mode in the Census 2000 on the differences between published Census 2000 and CPS employment-status estimates

It has been suggested that a major reason for census-CPS gaps in employment status estimates is that the census results are based, to a high degree, upon self-reporting of respondents, whereas the

---

[69] In the CPS, these people had (PRDISC=1) or (PRLKMD10=1 or PRLKMD11=1 PRLKMD13=1) or (PRWNTJOB=1 and PRJOBSEA in (1 or 2)).

[70] Rough calculations applying the 15-percent coefficient derived from the exercise described in this paragraph to data in Tables 1 and 2 of Appendix C indicate that these problems may have erroneously increased the Census 2000 count of unemployed people, which was approximately 7.9 million, by 300,000 to 400,000.

CPS results are based on responses collected by a professional staff of trained, experienced enumerators (see, for example, item 1 in the discussion in Appendix A). Census 2000 primarily used two data- collection modes. It relied heavily on self-reporting by respondents or "mail returns", but followup interviews were conducted by personal visit for people who failed to respond by mail by a designated cut-off date – these are called "enumerator returns". That Census 2000 employment-status are based on both mail and enumerator returns, and that the employment-status data can be distinguished by their mail/enumerator origins, provides an opportunity to compare the data by origin. If the enumerated data are more in line with corresponding CPS data than the mail-return data, and if the quality of the enumerated data, as measured by exact-match CPS-Census comparisons, is as good as or better than the quality of the mail-return data, these findings would support the hypothesis that self-reporting in the census is a significant factor in the differences between CPS and Census 2000 estimates [71]. They would suggest that, had the census employment data been entirely collected by enumerators, the census employment-status estimates would have more closely matched corresponding CPS estimates . The presentation below describes the results of experimental research to examine these issues; no conclusions are reached, and the data and the analysis are merely meant to suggest some potentially rewarding paths for future research.

Tables G-9 and G-10 compare percentage distributions of Census 2000 employment status data based on mail and enumerator returns, respectively, against CPS distributions for March and April 2000. The Census 2000 data are based on a rough-and-ready 1-in-500 sample of the cases in the Census 2000 Sample Edited Detail File (SEDF) and they exclude people in group quarters.

Table G-9. Percentage Distributions – Experimental Census 2000 Employment Status Estimates Based on Mail Returns (excluding Group Quarters Population), compared with CPS Estimates for March and April 2000

Employment Status	Experimental Census 2000 Estimates from Mail Returns	March 2000 CPS data	April 2000 CPS data
Total 16 years and over	100.0 %	100.0%	100.0%
Employed	60.5	64.2	64.6
Unemployed	3.0	2.9	2.5
Not in Labor Force	36.5	32.9	33.0

---

[71] The converse is not necessarily true because it is likely that , on average, Census 2000 enumerators were not as extensively trained nor as experienced as CPS enumerators.

Table G-10. Percentage distributions – Experimental Census 2000 Employment Status Estimates Based on Enumerator Returns (excluding Group Quarters Population), compared with CPS Estimates for March and April 2000

Employment Status	Experimental Census 2000 Estimates from Enumerator Returns	March 2000 CPS data	April 2000 CPS data
Total 16 years and over	100.0 %	100.0%	100.0%
Employed	64.4	64.2	64.6
Unemployed	4.2	2.9	2.5
Not in Labor Force	31.4	32.9	33.0

The data in these tables seem to suggest that universal use of enumerator returns in Census 2000 may have closed the gap between the CPS and Census estimates of employment (see Table F in section 4.4 of the body of the text), but widened the gap between their estimates of unemployment.

Tables G-11, G-12, G13, and G-14 display the results of tabulations of fully-reported Match cases by Census 2000 collection mode.

Table G-11. Employment Status of the Civilian Noninstitutional Population For Fully-Reported Match Cases, for the United States, Total: Mail-Form Respondents

CPS classification	Under 16 years	Census classification				
		16 years and over				
		Civilian labor force				Not in
		Total	Total	Employed	Unemployed	labor force
ESTIMATES						
Under 16 years						
16 years and over		135,192,645	96,838,038	93,033,519	3,804,519	38,354,607
Civilian labor force		96,863,837	92,409,612	89,889,386	2,520,226	4,454,225
Employed		93,905,671	89,946,280	89,045,743	900,537	3,959,391
Unemployed		2,958,166	2,463,332	843,643	1,619,689	494,834
Not in labor force		38,328,808	4,428,426	3,144,133	1,284,293	33,900,382
STANDARD ERRORS						
Under 16 years						
16 years and over		1,639,782	1,315,248	1,293,762	233,358	766,926
Civilian labor force		1,329,330	1,265,248	1,254,246	171,297	249,696
Employed		1,289,178	1,232,235	1,236,445	113,644	237,460
Unemployed		204,858	180,818	123,550	138,281	80,725
Not in labor force		755,312	232,332	196,563	149,147	742,299
(X) Not available						

93

Table G-12. Census-Based Percentage Distributions – Employment Status of the Civilian Noninstitutional Population For Fully-Reported Match Cases, for the United States, Total: Mail-Form Respondents

CPS Classification	Census Classification				
	16 years and over				
		Civilian Labor Force			Not in
	Total	Total	Employed	Unemployed	labor force
16 years and over	100.0%	100.0%	100.0%	100.0%	100.0%
Civilian labor force	71.6%	95.4%	96.6%	66.2%	11.6%
Employed	69.5%	92.9%	95.7%	23.7%	10.3%
Unemployed	2.2%	2.5%	0.9%	42.6%	1.3%
Not in labor force	28.4%	4.6%	3.4%	33.8%	88.4%
STANDARD ERRORS					
16 years and over	0.0%	0.0%	0.0%	0.0%	0.0%
Civilian labor force	0.4%	0.2%	0.2%	2.9%	0.6%
Employed	0.5%	0.3%	0.2%	2.7%	0.6%
Unemployed	0.1%	0.2%	0.1%	3.0%	0.2%
Not in labor force	0.4%	0.2%	0.2%	2.9%	0.6%

Table G-13. Employment Status of the Civilian Noninstitutional Population For Fully-Reported Match Cases, for the United States, Total: Enumerator-Form Respondents

CPS classification	Census classification					
	Under 16 years	16 years and over				
			Civilian labor force			Not in
		Total	Total	Employed	Unemployed	labor force
ESTIMATES						
Under 16 years						
16 years and over		47,184,109	33,429,567	31,308,921	2,120,645	13,754,542
Civilian labor force		33,850,779	30,797,564	29,348,698	1,448,866	3,053,215
Employed		31,657,040	29,224,837	28,474,186	750,651	2,432,203
Unemployed		2,193,740	1,572,727	874,512	698,215	621,013
Not in labor force		13,333,330	2,632,003	1,960,224	671,779	10,701,327
STANDARD ERRORS						
Under 16 years						
16 years and over		1,298,119	975,041	928,542	182,764	561,492
Civilian labor force		1,004,135	935,311	903,045	150,126	249,119
Employed		945,588	893,313	880,075	99,445	210,912
Unemployed		199,347	169,785	114,252	110,847	113,731
Not in labor force		574,355	192,101	155,452	101,486	521,069
(X) Not available						

Table G-14. Census-Based Percentage Distributions – Employment Status of the Civilian Noninstitutional Population For Fully-Reported Match Cases, for the United States, Total: Enumerator-Form Respondents

CPS Classification	Census Classification				
	16 years and over				
		Civilian Labor Force			Not in
	Total	Total	Employed	Unemployed	labor force
16 years and over	100.0%	100.0%	100.0%	100.0%	100.0%
Civilian labor force	71.7%	92.1%	93.7%	68.3%	22.2%
Employed	67.1%	87.4%	90.9%	35.4%	17.7%
Unemployed	4.6%	4.7%	2.8%	32.9%	4.5%
Not in labor force	28.3%	7.9%	6.3%	31.7%	77.8%
STANDARD ERRORS					
16 years and over	0.0%	0.0%	0.0%	0.0%	0.0%
Civilian labor force	0.9%	0.6%	0.5%	3.9%	1.7%
Employed	1.0%	0.7%	0.6%	4.0%	1.5%
Unemployed	0.4%	0.5%	0.4%	4.1%	0.8%
Not in labor force	0.9%	0.6%	0.5%	3.9%	1.7%

Tables G-11 and G-13 indicate that, in terms of the overall soundness measure defined in Appendix E, the mail returns had a higher level of soundness than the enumerator returns: for the mail returns, 92 percent census classifications agreed with their CPS counterparts; for the enumerator returns, the comparable figure is 85 percent. Tables G-12 and G-14 reveal that the within-category measures of soundness, also defined in Appendix E, were higher for the mail returns than for the enumerator returns.

Appendix H. Computation of Response Variance Measures

This appendix presents the computational forms of the response variance measures used in this report, along with a numerical example (using an illustrative variable called "temperature"). The discussion centers around two cross-tabulations: the first, shown in Table H.1, is in general form; the second, shown in Table H.2, is for the numerical example.

Table H.1 Display of cross-tabulated data - General procedure

[Display of cross-tabulated data for characteristic with L categories (L ≥ 2). The general term X_{ij} represents the number of weighted or unweighted sample elements in the ith category in Survey-B and the jth category in Survey-A.]

Survey-B	Total	Not reported	**Reported**	1	2	...	i	...	L
			Survey-A classification						
Total	n'..[1]								
Not reported									
Reported			n..[2]	$X_{.1}$	$X_{.2}$	...	$X_{.i}$	...	$X_{.L}$
Item responses:									
1. Category 1			$X_{1.}$	X_{11}	X_{12}	...	X_{1i}	...	X_{1L}
2. Category 2			$X_{2.}$	X_{21}	X_{22}	...	X_{2i}	...	X_{2L}
.			.	.	.	.	.	.	.
.			.	.	.	.	.	.	.
.			.	.	.	.	.	.	.
i. Category i			$X_{i.}$	X_{i1}	X_{i2}	...	X_{ii}	...	X_{iL}
.			.	.	.	.	.	.	.
.			.	.	.	.	.	.	.
.			.	.	.	.	.	.	.
L. Category L			$X_{L.}$	X_{L1}	X_{L2}	...	X_{Li}	...	X_{LL}

[1] n'.. is the total number of sample cases.

[2] n.. is the total number of sample cases for which there was a report in both the census and the reinterview. That is, n.. is the number of the sample cases minus the "not reported" cases.

Table H.2 Example of procedure: Temperature (Mock Data)

Survey-B classification	Total	Not reported	**Reported**	1	2	3	4
			Survey-A classification				
Total	19897	1454	18443	9143	4514	4364	422
Not reported	31	8	23	9	7	5	2
Reported	19866	1446	**18420**	**9134**	**4507**	**4359**	**420**
Item response:							
1. Hot	9861	634	**9227**	**8226**	882	90	29
2. Warm	4720	364	**4356**	752	**3498**	32	74
3. Cool	4872	418	**4454**	120	56	**4192**	86
4. Cold	413	30	**383**	36	71	45	**231**

A. Computing the net difference rate and the index of inconsistency.

1.*Net difference rate (NDR)*

For category I

$$NDR = \frac{X_{\cdot i} - X_{i \cdot}}{n} \times 100$$

For Temperature category " Warm"

$$NDR = \frac{4507 - 4356}{18420} \times 100 \approx 0.8$$

2. Index of inconsistency

For category I

$$I = \frac{X_{\cdot i} + X_{i \cdot} - 2X_{ii}}{\frac{1}{n_{\cdot \cdot}}[X_{\cdot i}(n_{\cdot \cdot} - X_{\cdot i}) + X_{i \cdot}(n_{\cdot \cdot} - X_{i \cdot})]} \times 100 .$$

For Temperature category "Warm"

$$I = \frac{4507 + 4356 - 2(3498)}{\frac{1}{18420}[4507(18420 - 4356) + 4356(18420 - 4507)]} \times 100$$

$$= \frac{8863 - 6996}{\frac{1}{18420}[4507(14064) + 4356(13913)]} \times 100 \approx 27.7$$

3. Aggregate index of inconsistency (I_A)

General formula

$$I_A = \frac{n_{..} - \sum_{i=1}^{L} X_{ii}}{n_{..} - \sum_{i=1}^{L} X_i X_{i.}} \times 100$$

For Temperature

$$I_A = \frac{18420 - (8226 + 3498 + 4192 + 231)}{18420 - \frac{1}{18420}[9134(9227) + 4507(4356) + 4359(4454) - 420(383)]} \times 100$$

$$= \frac{18420 - 16147}{18420 - \frac{1}{18420}(123487756)} \times 100 \approx 19.4$$

B. Computing the variances for the net difference rate and the index of inconsistency.

Variances for these measures were formed by using replicate weights representing a 160 independent samples from the match datasets. (See U.S. Bureau of the Census and U.S. Bureau of Labor Statistics, 2000, Chapter 14.) The replicates for the dataset for the combined-month sample

98

were adjusted for nonmatch and then had a second-stage ratio adjustment applied; the nonmatch adjusted CPS sample weights were then multiplied by the Census 2000 sample weights to represent the effect of Census 2000 sampling. The estimated sampling variance of an estimate is obtained by using the adjusted replicate weights to make 160 separate estimates, and estimating their variance as $4 * \dfrac{\sum_{i=1}^{160} (X_i - X_0)^2}{160}$, where X_0 is the statistic of interest estimated on the full sample, X_i is the estimate formed using the ith set of replicate weights, and the fraction 4/160 represents the treatment of self-representing and non-self-representing primary sampling units. (See U.S. Bureau of the Census and U.S. Bureau of Labor Statistics, 2002, chapter 14.)