

Census 2000 Topic Report No. 12

*Census 2000 Testing, Experimentation,
and Evaluation Program*

Issued March 2004

TR-12

Content and Data Quality in Census 2000

U S C E N S U S B U R E A U

Helping You Make Informed Decisions

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

United States
**Census
2000**

Acknowledgments

The Census 2000 Evaluations Executive Steering Committee provided oversight for the Census 2000 Testing, Experimentation, and Evaluations (TXE) Program. Members included **Cynthia Z. F. Clark**, Associate Director for Methodology and Standards; **Preston J. Waite**, Associate Director for Decennial Census; **Carol M. Van Horn**, Chief of Staff; **Teresa Angueira**, Chief of the Decennial Management Division; **Robert E. Fay III**, Senior Mathematical Statistician; **Howard R. Hogan**, (former) Chief of the Decennial Statistical Studies Division; **Ruth Ann Killion**, Chief of the Planning, Research and Evaluation Division; **Susan M. Miskura**, (former) Chief of the Decennial Management Division; **Rajendra P. Singh**, Chief of the Decennial Statistical Studies Division; **Elizabeth Ann Martin**, Senior Survey Methodologist; **Alan R. Tupek**, Chief of the Demographic Statistical Methods Division; **Deborah E. Bolton**, Assistant Division Chief for Program Coordination of the Planning, Research and Evaluation Division; **Jon R. Clark**, Assistant Division Chief for Census Design of the Decennial Statistical Studies Division; **David L. Hubble**, (former) Assistant Division Chief for Evaluations of the Planning, Research and Evaluation Division; **Fay F. Nash**, (former) Assistant Division Chief for Statistical Design/Special Census Programs of the Decennial Management Division; **James B. Treat**, Assistant Division Chief for Evaluations of the Planning, Research and Evaluation Division; and **Violeta Vazquez** of the Decennial Management Division.

As an integral part of the Census 2000 TXE Program, the Evaluations Executive Steering Committee chartered a team to develop and administer the Census 2000 Quality Assurance Process for reports. Past and present members of this team include: **Deborah E. Bolton**, Assistant Division Chief for Program Coordination of the Planning, Research and Evaluation Division; **Jon R. Clark**, Assistant Division Chief for Census Design of the Decennial Statistical Studies Division; **David L. Hubble**, (former) Assistant Division Chief for Evaluations and **James B. Treat**, Assistant Division Chief for Evaluations of the Planning, Research and Evaluation Division; **Florence H. Abramson**, **Linda S. Brudvig**, **Jason D. Machowski**, and **Randall J. Neugebauer** of the Planning, Research and Evaluation Division; **Violeta Vazquez** of the Decennial Management Division; and **Frank A. Vitrano** (formerly) of the Planning, Research and Evaluation Division.

The Census 2000 TXE Program was coordinated by the Planning, Research and Evaluation Division: **Ruth Ann Killion**, Division Chief; **Deborah E. Bolton**, Assistant Division Chief; and **Randall J. Neugebauer** and **George Francis Train III**, Staff Group Leaders. **Keith A. Bennett**, **Linda S. Brudvig**, **Kathleen Hays Guevara**, **Christine Louise Hough**, **Jason D. Machowski**, **Monica Parrott Jones**, **Joyce A. Price**, **Tammie M. Shanks**, **Kevin A. Shaw**,

George A. Sledge, **Mary Ann Sykes**, and **Cassandra H. Thomas** provided coordination support. **Florence H. Abramson** provided editorial review.

This report was prepared under contract by **Paula J. Schneider**, an independent contractor. The project managers were **Joan M. Hill** and **Jason D. Machowski** of the Planning, Research and Evaluation Division. The following authors and project managers prepared Census 2000 experiments and evaluations that contributed to this report:

Decennial Statistical Studies Division:

Kevin J. Zajac

Demographic Statistical Studies Division:

Sharon R. Ennis

Phyllis Singer

Planning, Research and Evaluation Division:

Sherri J. Norris

Greg Carroll and **Everett L. Dove** of the Administrative and Customer Services Division, and **Walter C. Odom**, Chief, provided publications and printing management, graphic design and composition, and editorial review for print and electronic media. General direction and production management were provided by **James R. Clark**, Assistant Division Chief, and **Susan L. Rappa**, Chief, Publications Services Branch.

Census 2000 Topic Report No. 12

*Census 2000 Testing, Experimentation,
and Evaluation Program*

Issued March 2004

TR-12

CONTENT AND DATA QUALITY IN CENSUS 2000



U.S. Department of Commerce
Donald L. Evans,
Secretary

Vacant,
Deputy Secretary

Economics and Statistics Administration
Kathleen B. Cooper,
Under Secretary for Economic Affairs

U.S. CENSUS BUREAU
Charles Louis Kincannon,
Director

Suggested Citation

Paula J. Schneider
*Census 2000 Testing,
Experimentation, and Evaluation
Program Topic Report No. 12, TR-12,
Content and Data Quality
in Census 2000,*
U. S. Census Bureau,
Washington, DC 20233



Economics and Statistics Administration

Kathleen B. Cooper,
Under Secretary for Economic Affairs



U.S. CENSUS BUREAU

Charles Louis Kincannon,
Director

Hermann Habermann,
Deputy Director and Chief Operating Officer

Cynthia Z. F. Clark,
Associate Director for Methodology and Standards

Preston J. Waite,
Associate Director for Decennial Census

Teresa Angueira,
Chief, Decennial Management Division

Ruth Ann Killion,
Chief, Planning, Research and Evaluation Division

For sale by the Superintendent of Documents, U.S. Government Printing Office
Internet: bookstore.gpo.gov Phone: toll-free 866-512-1800; DC area 202-512-1800
Fax: 202-512-2250 Mail: Stop SSOP, Washington, DC 20402-0001

Contents

Foreword	v
Executive Summary	1
1. Introduction	3
2. Background	5
3. Scope	7
4. Definitions	9
5. Limitations	11
6. Research Questions	13
7. Results	15
7.1 100 percent item nonresponse rates	15
7.2 100 percent item imputation rates	16
7.3 Sample item imputation rates	17
7.4 Comparison of allocation rates between the Census 2000 sample and the Census 2000 Supplementary Survey (C2SS)	18
7.5 Results from the Content Reinterview Survey	18
7.6 Summary assessment of quality measures	20
7.7 Validation against external benchmarks	21
7.8 Indicators of coding quality	21
8. Recommendations/Conclusions	23
References	25
Appendix	27
LIST OF TABLES	
Table A 100 Percent Item Nonresponse Rates by Form Type and Response Mode: Census 2000	15
Table B 100 Percent Item Imputation Rates by Form Type, Response Mode, and Household/Proxy Response: Census 2000	17
Table C 100 Percent Item Imputation Rates by Tenure, Form Language, and Selected Form Source: Census 2000	17
Table 1. Summary of Quality Measures for Populations Items: Census 2000 and 1990	27
Table 2. Summary of Quality Measures for Housing Items: Census 2000 and 1990	28

This page intentionally left blank.

Foreword

The Census 2000 Testing, Experimentation, and Evaluation Program provides measures of effectiveness for the Census 2000 design, operations, systems, and processes and provides information on the value of new or different methodologies. By providing measures of how well Census 2000 was conducted, this program fully supports the Census Bureau's strategy to integrate the 2010 planning process with ongoing Master Address File/TIGER enhancements and the American Community Survey. The purpose of the report that follows is to integrate findings and provide context and background for interpretation of related Census 2000 evaluations, experiments, and other assessments to make recommendations for planning the 2010 Census. Census 2000 Testing, Experimentation, and Evaluation reports are available on the Census Bureau's Internet site at: www.census.gov/pred/www/.

This page intentionally left blank.

Executive Summary

The purpose of this report is to synthesize results from Census 2000 evaluations and other assessments of the quality of census data on population and housing characteristics and to make recommendations for planning the 2010 census.

The formal studies of content and data quality included in the Census 2000 Testing, Experimentation, and Evaluation Program include five studies covering—

- Nonresponse and imputation for the 100 percent census items,
- Response variance for most items included in Census 2000, and
- Consistency of information from selected census items with similar information from other data sources.

The analysis included here also is based on a set of tabulations prepared for the National Academy of Sciences, Committee on National Statistics. These tabulations show information on allocation for missing responses for the Census 2000 sample (long form) questionnaire in comparison with the 1990 Census and the Census 2000 Supplementary Survey. Also included is some limited information on the quality of Census 2000 coding operations.

Evaluation report B.1.b, Analysis of Item Nonresponse Rates for the 100 Percent Housing and Population Items from Census 2000, provides the base set of information on the quality of cen-

sus data. Consideration of patterns of nonresponse is critical for improving question design, training, and procedures and thereby improving quality in future censuses. Nonresponse rates for the 100 percent items in Census 2000 were fairly low overall, ranging from 1.1 to 4.1 percent.

In report B.1.a, Analysis of Imputation Rates for the 100 Percent Person and Housing Unit Data Items from Census 2000, imputation refers to assignment and allocation. The data show that, as expected, patterns of imputation varied in a manner similar to nonresponse rates. Imputation rates ranged from a low of 2.0 percent for sex to 5.1 percent for age and 5.5 percent for tenure. Imputation rates are higher than nonresponse rates shown in these studies partially because of universe differences. Also, responses that were not meaningful or that were not consistent with other information for the person were treated as blank and were imputed. Short forms had somewhat lower imputation rates than the same items on census long forms for all items except race. The biggest difference between short and long form imputation rates was for the item on housing tenure, which was most likely related to the design of the long form questionnaire.

Except for the item on Hispanic origin, self response questionnaires had lower imputation rates than enumerator filled questionnaires.

Questionnaires received via the Internet or Telephone Questionnaire Assistance had quite low imputation rates. These forms of response required a proactive effort on the part of the household and, therefore, are likely to reflect a segment of the population highly motivated to participate in the census.

The tabulations prepared for the National Academy of Sciences show that only a few sample items in Census 2000 had lower allocation rates than in 1990. For one of these, the item on length of service in the Armed Forces, a questionnaire design change made to improve reporting appears to have been successful.

Many sample items in Census 2000 had at least double the rate of allocation that occurred in the 1990 Census. At least part of this increase can be attributed to procedural changes between 1990 and 2000.

The Census 2000 Supplementary Survey was included as part of Census 2000 to determine whether such an independent sample survey (the American Community Survey) with essentially the same content as the census sample form could be accomplished successfully during the same time period as the census itself. One might expect the quality of these data to be better than the Census 2000 sample data since the exclusive focus of the supplementary sample and the American Community Survey is to provide high quality measures of social, economic and

housing variables. This expectation was realized in terms of allocation rates.

The purpose of the Content Reinterview Survey was to measure the consistency of response to questions asked in the census. Response variance can result from a number of factors, but high measures of variability usually indicate that the question needs improvement, the concept is difficult to measure in a setting that is primarily self response, and/or that the respondent was unable to provide the information desired. The measures of simple response variance from the Content Reinterview Survey also reflect methodological limitations of the reinterview and differences between the Census 2000 and Content Reinterview Survey methods.

For the long form census questions evaluated in the Content

Reinterview Survey, the index of inconsistency (measure of response variance) ranged from quite low levels under ten to some very high levels over 75. Some questions with extensive and critical government uses like race, ability to speak English, and selected income types had rather high measures of inconsistency. Possible reasons behind response variability include the subjective nature of some questions, the effect of rare occurrences on the calculation of the index, recall problems, and privacy concerns.

To those individuals planning the 2010 Census, questions with serious quality problems signal the need for further work in question design, procedures, and concept clarification. To users of Census 2000 data, quality problems should be a warning to use data from these items with caution.

A Census 2000 auxiliary assessment study compared employment, income, and poverty data from Census 2000 to such data from the Current Population Survey. This was conducted by the Census Bureau in partnership with the Bureau of Labor Statistics. Since the Current Population Survey is designed specifically to provide the measures of employment and income at the national and state levels for the federal government, these measures are viewed as standards against which the census results are compared. The differences between the two data sources are such that additional work is needed to explain the reasons behind them. Some answers may be forthcoming from Study B.7, not completed in time for this report. That study matches individual responses from the Current Population Survey with the responses for the same person in Census 2000.

1. Introduction

This report summarizes key results on content and data quality from the Census 2000 Testing, Experimentation, and Evaluation Program. It also includes assessments of data quality from tabulations and reports not part of the formal program.

The formal studies include analyses of nonresponse and imputation for the 100 percent census items, response variance for most census data items, and consistency of selected census information with similar data from other sources.

The other assessments cover allocation for missing responses for sample (long form) Census 2000 data items, comparison of sample data completeness with the 1990 Census and the Census 2000 Supplementary Survey, and limited information on the quality of Census 2000 coding operations.

This page intentionally left blank.

2. Background

Census 2000 differed from the 1990 Census in a number of ways that directly or indirectly affected the quality of the resulting statistics. This report focuses on the quality of the characteristics data produced from the census, not the completeness of the population counts. The studies summarized herein cannot directly measure the effect of any one factor on the quality of a particular statistic or set of numbers but it is often possible to theorize about such relationships.

Among changes designed to affect census content results were changes to the wording, format or placement of questions between 1990 and 2000. For example, the question on Hispanic origin was placed before the question on race in an effort to improve the level and quality of response. Details on such changes are contained in a number of formal documents on testing done in preparation for Census 2000, which are cited in the References section of this report.

There were also changes with a possible indirect effect on the content quality. For the first time in Census 2000, names of persons entered on the questionnaire were captured in computer-readable form and could be used in processing. This was done primarily to improve the population counts but it also allowed the "assignment" of sex for individuals for whom this characteristic was not reported.

Other procedural changes in Census 2000 (as compared to 1990) also had potential impact on content quality. For example, in 1990 there was an operation designed to improve the completeness of information on the sample questionnaires through a follow-up interview for questionnaires with excessive blanks. There was no similar operation in Census 2000. This may have served to increase nonresponse for questions on the 2000 sample long form.

In an attempt to encourage mail response in Census 2000 the basic mail questionnaire was made as

short as possible and allowed room to record information for only six household members. Since it was often difficult to obtain the information for persons in larger households, this too may have increased nonresponse for the questions.

Perhaps the most important content-related improvement in Census 2000 was the inclusion of the Census 2000 Supplementary Survey (C2SS). The C2SS was conducted to demonstrate the feasibility of collecting long form information at the same time as, but in a separate process from the census. It used the American Community Survey questionnaire and procedures and covered 1,203 counties nationwide surveying 58,000 households each month. As a result, we have a comparison of results from the Census 2000 and the C2SS for the same content areas. This will be a critical planning tool for the 2010 Census.

This page intentionally left blank.

3. Scope

The primary purpose of this topic report is to summarize the findings and recommendations from the formal studies of content and data quality included in the Census 2000 Testing, Experimentation, and Evaluation Program. These studies include-

- Census 2000 Evaluation B.1.a, Analysis of Imputation Rates for the 100 Percent Person and Housing Unit Data Items from Census 2000
- Census 2000 Evaluation B.1.b, Analysis of Item Nonresponse Rates for the 100 Percent Housing and Population Items from Census 2000

- Census 2000 Evaluation B.5, Census 2000 Content Reinterview Survey: Accuracy of Data for Selected Population and Housing Characteristics as Measured by Reinterview
- Census 2000 Evaluation B.7, Current Population Survey (CPS)-Census 2000 Match Study (not available at the time this report was prepared)
- Census 2000 Auxiliary Assessment, Comparing Employment, Income, and Poverty: Census 2000 and the Current Population Survey

The analysis also is based on a set of tabulations prepared for the National Academy of Sciences, Committee on National Statistics, showing information on allocation for missing responses for the Census 2000 sample (long form) questionnaire in comparison with the 1990 Census and with the C2SS. In addition, information on the quality of the coding processes for selected Census 2000 content items has been extracted from selected documents not part of the formal evaluation program.

There is a related topic report focusing on the race and ethnicity data from Census 2000.

This page intentionally left blank.

4. Definitions

There are a number of ways to characterize the quality of the content results from a census or survey. Although sampling error is often the primary source of error that can be measured for statistical results, sampling error is minimal for census results at the national level.

Types of nonsampling error include coverage error, household and data item nonresponse, errors introduced in the processing of the information (for example, coding of responses), and measurement error. This last type of error is quantified by studies that measure the consistency of responses to an item from successive reports (that is, the census response and the content reinterview), and the consistency of the resulting statistics across separate measurement systems (for example, the Census and Current Population Survey results for the same topic.) This report contains to varying degrees information on each of these types of measures, with the exception of coverage error.

In reading the sections that follow, it is important to understand the meaning of the various quality measures. Therefore, a set of definitions is included here for easy reference.

Item Nonresponse —This refers to whether there is an entry for a data item, regardless of whether it is an acceptable response. This includes appar-

ent responses that are not valid answers or are inconsistent with other information for the person. In such cases, the entry is not accepted and the item is treated as a blank in the imputation process.

Imputation —Three components comprise the imputation process for Census 2000—substitutions, assignments, and allocations, as defined below. For the analysis of individual data items in Report B.1.a (Section 7.2 of this report) the term "imputation" includes assignments and allocations. For the analysis in section 7.3, only allocations are included.

Substitution —In any large statistical operation like a census there are cases for which no information can be obtained (that is, noninterviews.) In Census 2000 approximately 3.4 million persons in households were noninterviews and, therefore, had all their 100 percent characteristics imputed by the replication of data from a household of the same size with fully reported 100 percent population information. Such cases, called "whole household substitutions" are excluded from the analysis of the imputation measures for individual subject items.

Assignment —When a response is missing or inconsistent with other responses and the value for that item can be

determined from other information reported for the person or housing unit, an "assignment" is made. A prime example is when sex is not reported but can be determined from the person's name.

Allocation —When the value for a data item is missing or inconsistent with other responses and the value cannot be determined from other information for that person or housing unit, "allocation" of a value for the item occurs. This is done using a reported value from another person in the household or from a nearby housing unit.

Index of Inconsistency —The Census 2000 Content Reinterview Survey (CRS) used a test-retest methodology in which a sample of households from Census 2000 long form respondents were contacted a second time and re-asked most of the long form questions. The intent was to measure the simple response variance. The measure used to summarize this response variance is the index of inconsistency. The higher the index value, the more problematic is the interpretation of the data from the census item. Historically, an index value less than 20 has been viewed as low or good level of response variance, an index between 20 and 50 as moderate, and an index over 50 as high.

This page intentionally left blank.

5. Limitations

The universe for analysis in this report is restricted to the household population; that is, residents of group quarters are excluded. The quality of reporting for the group quarters population is generally less complete than for the household population, especially for sample questions included only on the census long forms.

The definition of universes for the calculation of nonresponse and imputation rates differs between the formal evaluation studies discussed in this report and other sources for similar data. For example, households contacted during a coverage edit followup operation in Census 2000 were classified as "enumerator returns" in the evaluation reports but as "self responses" in the tabulations prepared for the National Academy

of Sciences. Other slight differences can make the imputation rates shown in formal evaluations differ from the Census 2000 rates included in the American FactFinder tabulations.

The Content Reinterview Survey measured only response variance. There is no measure of response bias other than comparison of totals to independent sources such as the Current Population Survey.

Apparent response variance as measured by the index of inconsistency is affected by differences in methodology between Census 2000 and the CRS. For example, the universe for the CRS contained a high proportion of households that mailed back their Census 2000 questionnaires. Data from households that mailed back their

questionnaires are generally more accurate than responses from households that required enumeration by a temporary census employee. Other factors that could have affected the measures of response variance are discussed in section 7.5.

Comparisons of imputation rates between the census 100 percent and sample estimates are affected by the fact that sample (long form) questionnaires with no or minimal sample information were excluded from the calculation of the rates for sample estimates and the rates are based on different universes. Also, comparisons between the Census 2000 long form and the Census 2000 Supplementary Survey (C2SS) are affected by differences in methodology and in question design.

This page intentionally left blank.

6. Research Questions

There were several predefined research questions concerning data quality that the evaluation studies were designed to address. These questions are summarized as follows:

How complete are census data and what is their quality, as measured by –

- Item nonresponse rates, item assignment rates, item allocation rates, and substitution rates for census data (evaluations B.1.a and B.1.b)
- Rate of proxy response (information obtained from someone other than a household member) for enumerator returns (evaluation B.1.a)
- Reliability as compared to independent results from the Content Reinterview Survey (evaluation B.5)
- Validation against external benchmarks (auxiliary assessment)

How did the new race and Hispanic origin questions affect the content and quality of the data compared to the questions asked in previous censuses? How can race data collected using the new questions be compared with race data from previous censuses? These questions are addressed in the Race and Ethnicity Topic Report.

Original plans for evaluation of content and data quality included a research question on the accuracy of data edits and imputations. However, there were no studies undertaken to address this issue.

In addition to the predefined questions, this report also addresses the following content/quality topics:

- The formal evaluation study on imputation deals only with the questions asked of the entire population, those questions appearing on both the short and long census questionnaires. However, Census 2000 summa-

ry file 3 and the tables prepared for the National Academy of Sciences contain information on the level of imputation for the sample questions, those only on the long form. These results are included here.

- The Census 2000 Supplementary Survey results contain information on imputation for the same items as on the Census 2000 long form. A comparison across these two sources is also provided in this report.
- There are no formal evaluations addressing the accuracy of the coding operations used to translate textual responses in Census 2000 into numeric codes (for example, responses to the industry and occupation questions.) There are, however, Census Bureau reports and memoranda that provide useful information. Such information is summarized in this report.

This page intentionally left blank.

7. Results

7.1 100 percent item nonresponse rates

The evaluation report B.1.b, Analysis of Item Nonresponse Rates for the 100 Percent Housing and Population Items from Census 2000, provides the base set of information on the quality of census data. The primary reason for item imputation in the census is because no response was provided. There is some additional contribution from processing error and inconsistent reporting but most imputations result from blanks on the incoming questionnaires. Therefore, studying patterns of nonresponse is critical for improving question design, training, and procedures and thereby improving quality in future censuses. This analysis is based on the universe of "data-defined persons," that is, those people with at least two items reported in the census. Valid people with no or only one item reported are excluded. These rates, therefore, understate the true level of item nonresponse.

Nonresponse rates for the 100 percent items in Census 2000 were

fairly low overall, ranging from 1.1 to 4.1 percent. As could be expected, the nonresponse rates among households that mailed back their forms, called self respondents, were generally better than for households whose forms were completed by an enumerator (enumerator returns.) The largest differences between self and enumerator forms were for the items on tenure (owner/renter) and on age. For tenure, the placement of the question on the form, especially the long form, may have contributed to enumerators not asking the question. For age, enumerators were instructed to ask only the date of birth question, rather than both age and date of birth, in many instances. The calculation of nonresponse for this evaluation considered only the age item, not date of birth. One question, the item on Hispanic origin, had slightly better response on enumerator forms than on self response forms. One might speculate that some respondents did not understand the difference between Hispanic origin and race, whereas enumera-

tors were trained to ask both questions.

In comparing item nonresponse levels between short and long form questionnaires, the problem with the long form placement of the tenure question (after all the questions for persons in the household) is obvious. When considering both form type and response mode, one sees that self response long forms had somewhat better response levels to the sex, Hispanic origin and race questions than did self response short forms.

This relationship may need more analysis, especially by whether the respondent was a household member or a proxy.

Major recommendations from the study on nonresponse include the following.

- Clarify training for interviewers/enumerators on whether it is necessary to ask both age and date of birth.
- Test improved form design and question placement to avoid problems like the placement of

Table A.
100 Percent Item Nonresponse Rates by Form Type and Response Mode: Census 2000

Item	Total	Short Form (SF)	Long Form (LF)	Self Response (Self)	Enumerator Response (Enum)	SF Self	SF Enum	LF Self	LF Enum
Relationship	1.3	1.3	1.7	1.1	2.1	1.0	2.1	1.4	2.2
Sex	1.1	1.2	1.0	0.9	1.7	1.0	1.7	0.5	1.9
Age	3.7	3.7	4.1	1.9	8.8	1.9	8.9	1.9	8.4
Hispanic origin	3.1	3.1	3.2	3.2	2.7	3.3	2.6	3.1	3.2
Race	2.9	3.0	2.6	2.8	3.2	3.0	3.1	2.3	3.4
Tenure	4.1	3.1	9.6	2.6	8.8	2.1	6.3	5.7	17.8

Source: Evaluation report B.1.b, Analysis of Item Nonresponse Rates for the 100 Percent Housing and Population Items from Census 2000.

the tenure question and the effect that placement had on nonresponse.

7.2 100 percent item imputation rates

There are three components of imputation for Census 2000—substitution, assignment, and allocation. These are analyzed in the report, Census 2000 Evaluation B.1.a, Analysis of Imputation Rates for the 100 Percent Person and Housing Unit Data Items from Census 2000. The effects of substitution, which are not reflected in the analysis of individual data items in the B.1.a report, occurs when the census process obtained no usable data for a household. In these cases, an entire set of characteristics from a neighboring household was "substituted" for the missing values. This occurred for 1.5 million households and 3.4 million people or 1.3 percent of the total household population in 2000. By definition, the substitution rate doesn't vary by item since each item is imputed when a substitution occurs. For the remaining portion of this report, imputation will refer only to assignments and allocations.

As expected, the pattern of imputation among the 100 percent questions/items varied in a manner similar to nonresponse rates. Imputation rates and nonresponse rates, however, are not strictly comparable for several reasons. First, some apparent responses are not valid answers and, therefore, are treated as blanks in the imputation process. Also, there are differences in the universes used to calculate the two rates. The universe for nonresponse rates includes only "data defined people," those people with at least two characteristics reported. The universe for imputation rates

includes, in addition, people in non-substituted households who had no or only one reported characteristic. As a result of these differences, the imputation rates in Table B are larger than the item nonresponse rates in Table A.

Imputation rates (including assignments and allocations) ranged from a low of 2.0 percent for sex to 5.1 percent for age and 5.5 percent for tenure. Assignments occur when a response for an item is missing but the value can be determined by other information reported for that person. Assignments were an important component of the imputation rate for age (assigned from date of birth) and for sex (assigned from name.)

Short form items had somewhat lower imputation rates than the same items on census long forms, except that the race item had slightly higher imputation on short forms. The biggest difference between short and long form imputation rates was for the tenure item (3.7 percent vs. 13.2 percent). As mentioned above, it is likely that the placement of this housing question after all the person questions on the long form contributed to the rather high rate of imputation.

Except for the question on Hispanic origin, self response questionnaires had lower imputation rates than enumerator filled questionnaires. As mentioned in the analysis of nonresponse rates, it is possible that respondents occasionally left blank the Hispanic origin question when they did not consider it relevant to them. This was the case to a greater extent in the 1990 Census when the question on Hispanic origin followed the question on race. The order was reversed in Census 2000 to

reduce the level of nonresponse on Hispanic origin.

Not surprisingly, imputation rates on English language forms in general were lower than the rates on the forms in other languages. Households completing the non-English forms were presumably recent immigrants or others with difficulty in English and possibly difficulty with understanding the census itself. It is noteworthy, however, that Korean language forms had relatively low imputation rates for most items.

Households who owned their homes had lower imputation rates than renters. For questionnaires completed by enumerators, those with a household respondent had lower imputation rates than those with a non-household proxy respondent.

Questionnaires received via the Internet (classified as self response) or Telephone Questionnaire Assistance (TQA) (classified as enumerator response) had quite low imputation rates. These types of response required a proactive effort on the part of the household and, therefore, are likely to reflect a segment of the population highly motivated to participate in the census. On the other end of the scale, forms completed in operations designed to enumerate some of the most difficult to reach households had relatively high imputation rates.

These operations include the Be Counted campaign and coverage improvement followup (CIFU) operations that occurred later in the census process.

This same pattern of relative cooperation and level of imputation is seen when imputation rates are viewed by the check-in date of the questionnaire. The earlier a mail

Table B.
100 Percent Item Imputation* Rates by Form Type, Response Mode, and Household/Proxy Response: Census 2000

Item	Total	Short Form	Long Form	Self Response	Enumerator Response (Enum) Filled	Enum Household Respondent	Enum Proxy Respondent
Relationship	2.6	2.5	3.2	2.0	4.1	3.3	9.1
Sex	2.0	2.0	2.1	1.7	2.8	2.2	6.1
Age	5.1	5.0	5.7	3.0	10.9	6.7	42.7
Hispanic origin	4.4	4.3	4.6	4.5	4.1	2.6	11.6
Race	4.0	4.0	3.8	3.8	4.3	2.5	11.0
Tenure	5.5	3.7	13.2	3.0	12.2	7.3	30.1

* Imputation refers to assignments and allocations.

Source: Evaluation report B.1.a, *Analysis of Imputation Rates for the 100 Percent Person and Housing Unit Data Items from Census 2000*.

Table C.
100 Percent Item Imputation* Rates by Tenure, Form Language, and Selected Form Source: Census 2000

Item	Own	Rent	English Form	Other Language Form	Self Total	Internet	Be Counted	Enumerator Total	TQA	CIFU	Remote Alaska
Relationship	2.1	3.5	2.5	3.6-10.1	2.0	1.3	18.0	4.1	2.9	6.3	12.6
Sex	1.7	2.6	1.9	2.4-7.1	1.7	1.2	7.8	2.8	0.5	4.3	12.2
Age	4.1	7.2	5.0	3.0-10.3	3.0	1.5	9.2	10.9	2.9	21.6	15.3
Hispanic origin	3.9	5.5	4.3	7.1-20.7	4.5	3.0	12.5	4.1	1.7	6.7	12.7
Race	3.3	5.5	3.8	1.8-17.5	3.8	3.3	11.7	4.3	2.2	6.3	12.0
Tenure	X	X	5.2	4.8-26.7	3.0	1.4	3.6	12.2	0.8	22.8	3.7

*Imputation refers to assignments and allocations.

Source: Evaluation report B.1.a, *Analysis of Imputation Rates for the 100 Percent Person and Housing Unit Data Items from Census 2000*.

return form was received, the lower the level of imputation.

Overall, 88.1 percent of the non-substituted household population had no imputation in any of the five population items and 97.3 percent had none or only one item imputed. The key recommendations from the imputation study include the following.

- Test whether encouraging use of the Internet and telephone to respond to the census would be effective and result in quality improvements. Additional information about these response modes can be found in the Response Rates and Behavior Analysis Topic Report.

7.3 Sample item imputation rates

The summary tables of quality measures in the Appendix of this report show allocation rates for both 100 percent and sample items from tabulations prepared at the request of the National Academy of Sciences, Committee on National Statistics. In addition to the rates for Census 2000, the tables show rates for the 1990 Census and Census 2000 based on a comparable set of definitions, and rates from the Census 2000 Supplementary Survey. The measures are called "allocation rates" since "assignments" are not included.

Only a few sample items had lower

allocation rates in Census 2000 than in 1990. These include language spoken at home, ability to speak English, migration (place level), length of service in the Armed Forces, and year structure built. For the item length of service, there was a question design revision for Census 2000 designed to improve response and hence lower the allocation rate. This seems to have worked. For migration-place level, one might assume there were improvements in the coding operations.

Many sample items in Census 2000 had at least double the rate of allocation that occurred in the 1990 Census. A listing of these items follows.

Sex
 Age
 Race
 Marital status
 Work Disability
 Employment status
 Vehicle occupancy
 Wage or salary income
 Interest or dividend income
 Social Security income
 SSI/public assistance income
 Retirement income
 Other income
 Tenure
 Units in structure
 Year moved in
 Number of rooms
 Plumbing facilities
 Telephone availability
 Heating fuel
 Vehicles available
 Business on property
 Cost of electricity, gas,
 water/sewer
 Monthly rent
 Mortgage and mortgage payment
 Second mortgage and payment
 Real estate taxes
 Insurance
 Value of home

There were procedural changes between 1990 and 2000 that presumably affected the sample allocation rates. For example, the questionnaire in Census 2000 had room for reporting the characteristics of six household residents, as compared to seven residents in 1990. Since it was often difficult to obtain the information for additional household members in large households, this increased the imputation level to some extent.

Also, in the 1990 Census there was a procedure in which census enumerators telephoned households whose sample questionnaires had insufficient information. This operation was not conducted in Census 2000.

7.4 Comparison of allocation rates between the Census 2000 sample and the Census 2000 Supplementary Survey (C2SS)

The C2SS was included as part of Census 2000 to determine whether an independent sample survey (the American Community Survey or the ACS) with essentially the same content as the census sample form could be accomplished successfully during the same time period as the census itself. As a result, we have measures of the data quality, in terms of allocation rates, that can be compared between the Census 2000 sample and the C2SS.

One would expect the quality of the C2SS data to be better than the Census 2000 sample data since the exclusive focus of the C2SS and the ACS is to provide high quality measures of social, economic, and housing variables needed by federal, state and local governments. This expectation was realized in terms of allocation rates. Only the items listed below had Census allocation rates the same as or lower than the C2SS rates.

Hispanic origin (3.6 percent in both)

Migration:

State (8.6 percent in census; 14.0 percent in C2SS)

County (8.6 percent in census; 14.5 percent in C2SS)

Place (8.8 percent in census; 14.9 percent in C2SS)

Responsible for grandchild (15.3 percent in census; 17.6 percent in C2SS)

Months responsible (17.8 percent in census; 19.7 percent in C2SS)

Period of service in Armed Forces (9.8 percent in census; 13.4 percent in C2SS)

Year structure built (11.7 percent in census; 13.4 percent in C2SS)

The positive results from the American Community Survey methods are encouraging since the plans for the 2010 census include using the American Community Survey data in place of the census sample long form.

7.5 Results from the Content Reinterview Survey

The purpose of this Census 2000 evaluation study was to measure the consistency of response to questions. This was accomplished by completing a survey of census respondents that re-asked the sample or long form questions thereby allowing the calculation of a measure of response variance called the index of inconsistency. The lower the value of the index the more consistent were respondents' answers to the question and the more confidence users can have in the resulting statistics. A general guideline used to categorize levels is that index values less than 20 historically have been considered as low, between 20 and 50 as moderate, and over 50 as high.

Response variance can result from a number of factors, but high measures of variability usually indicate that the question needs improvement, the concept is difficult to measure in a survey setting that is primarily self response, and/or that the respondent was unable to provide the information desired. Also, the index of inconsistency is sensitive to the amount of detail the question is designed to obtain. The finer the detail and the greater number of categories, the more likely is a high measure of response variance.

Apparent response variance can be partially attributed to differences in census and CRS methods and limi-

tations in the CRS methodology. Of special note is that Census 2000 had a large mail back component whereas the CRS was conducted entirely through interviewing, either face-to-face or by telephone. The questions on personal characteristics were asked in the CRS of only one person in each household. Only 48 percent of the CRS respondents were the same self respondents who had answered the census. Twenty percent of the cases were the same proxy respondent in the CRS and census. Thus, 32 percent of the CRS respondents were different from the persons who responded in the census. Many of the differences in census and CRS methods that limited the 2000 results were also present in 1990.

The CRS used unedited and unallocated data in all analysis except for the race and Hispanic origin items where there was minimal editing prior to analysis.

The initial sample for the CRS was 30,000 persons, and 20,000 interviewed persons were included in the final results. Three quarters of the CRS sample had mailed back their Census 2000 long form. This is a high percentage compared to the general response to the Census 2000 long form. The CRS interviews were conducted fairly close to the census operations and the number of completed interviews exceeded expectations.

Among the long form items asked of each person in the census household and included in the CRS, the index of inconsistency values were relatively high (over 50) for more than one third of the items. For housing questions, index values were over 50 for close to half the items. (See table 2 on Summary of Quality Measures in the Appendix.)

The questions on whether the person was responsible for the care of a grandchild living in the home and, if so, the number of months responsible had rather high indexes of inconsistency (46.1 for the former and 53.7 for the latter). This series of questions was added to Census 2000 at a late date without the benefit of testing. If the topic is to be included in the future, testing will be needed to attempt to improve the results.

Some questions with extensive and critical government uses had rather high measures for the index of inconsistency and/or increases in the index value over the 1990 Census content reinterview study. For instance, the race question had an index of 23.1, an increase over the 1990 index of 16.3. Since 2000 was the first census to allow reporting of more than one race, this may be part of the reason for this increase. Also, the question on ability to speak English, asked of persons who speak a non-English language at home, had a high index of inconsistency in both 2000 and 1990. This may be partly because of the subjective nature of the question.

Some items of potential concern reflect the attempt to measure "rare occurrences." For example, some types of income show high indexes of inconsistency. When the receipt of income from a less common source like public assistance is associated with any discrepancy in reporting between the census and the CRS, this tends to inflate the value of the index more than would be true for more common income sources like wages and salaries. The same is true for items on whether a housing unit has complete plumbing facilities and complete kitchen facilities (since it is relatively rare not to have these.)

Census questions that measure a current status, like school enrollment, have less response variance than questions that require recall, like educational attainment. This might also be true for employment, but the CRS couldn't measure current employment since the reference period of "last week" could not be replicated in the CRS interview.

Other patterns that may be more broadly applicable to self response survey design were also observed. Subjective or opinion questions, such as some parts of the series on disability and the question on ability to speak English, have relatively high response variability. Also, questions associated with memorable events (like military service during a period of war) are reported more consistently than similar questions about less memorable dates (peacetime military service). Finally, questions with which the public may have privacy concerns (for example, types of income) tend to have high response variability.

Key recommendations from the Content Reinterview Survey include the following:

1. Use cognitive experts as part of the testing of revised questions for new and/or troublesome topics, that is, those with high response variance.
2. Include a reinterview program as part of the American Community Survey to allow a continuing evaluation tool.
3. Use separate yes/no questions for each response category when the respondent can select more than one category. An example is the period of military service question in Census 2000.

4. Review and test the respondent instructions for problematic questions. The CRS report specifically mentions the question on Hispanic origin.
5. Determine characteristics related to high inconsistency measures and perform multivariate analysis with respect to these characteristics in the attempt to develop better ways to ask the questions.
6. Issue guidance to users on the types of uses to avoid for questions with high levels of inconsistency. For instance, data from such questions probably should not be used for small areas or in detailed cross tabulations with other variables.
7. Consider not publishing detail from questions with high variability. For example, the questions on type of income may be useful for determining total income but they may not be appropriate to use as individual data items.
8. If research and testing cannot improve the question design, consider whether some questions with persistent high response variability should be dropped from the census/ACS.
9. Investigate whether records on such topics as utility costs and real estate taxes could be substituted for or used to supplement direct inquiry.
10. Research whether modeling in conjunction with independent records might be an appropriate way of providing improved data on some topics such as second mortgage payment and insurance payment.

7.6 Summary assessment of quality measures

The summary tables in the Appendix display information on allocation rates for Census 2000, the 1990 Census, and the Census 2000 Supplementary Survey, plus information on response variability (index of inconsistency) from the Content Reinterview Studies associated with Census 2000 and the 1990 Census. It is included to provide readers with an easy reference to the basic quality measures associated with census data items, especially the sample or long form items. To those planning the 2010 census, questions with serious quality problems signal the need for further work in question design, procedures or concept clarification. To users of Census 2000 data, quality problems should be a warning to use data from these items with caution.

Following is a list of the items with two poor quality measures, that is, allocation rates above 10 percent and indexes of inconsistency over 50.

- Work disability
- Months responsible for grandchild
- Weeks worked
- Income receipt
 - interest or dividend
 - public assistance
 - other income
- Income amount
 - Social Security
 - Supplemental Security
 - public assistance
- Agricultural sales
- Cost of electricity
- Cost of gas
- Second mortgage/payment

- Insurance payment
- Value of home
- Mobile home cost

Several additional items have quite high indexes of inconsistency but have allocation rates less than 10 percent.

- Ability to speak English
- Disability, mental
 - self-care
 - mobility
- Selected periods of service
- Number of rooms
- Plumbing facilities
- Kitchen facilities
- Telephone available
- Business on property

Finally, the following items had fairly substantial increases from 1990 to 2000 in the level of the index of inconsistency.

- Hispanic origin
- Race
- Educational attainment
- Ancestry
- Language spoken at home
- Disability
 - mobility
 - work
- Veteran status
- Tenure
- Plumbing Facilities
- Vehicles available
- Business on property
- Agricultural sales

7.7 Validation against external benchmarks

The Census 2000 auxiliary assessment that compares employment, income, and poverty estimates between Census 2000 and the Current Population Survey contains the only information available on comparisons to external benchmarks at the time this report was prepared.

The Current Population Survey (CPS) is conducted by the Census Bureau in partnership with the Department of Labor, Bureau of Labor Statistics (BLS) to provide official monthly estimates of employment and unemployment at the national and state levels and official annual measures of poverty. Since the survey is designed specifically for these purposes, the resulting measures are considered standards against which Census 2000 results are compared.

As historically has been true, the census estimate of employment was five percent lower than the comparable CPS estimate. However, the number of unemployed persons measured in Census 2000 was 50 percent higher than the CPS estimate (7.9 million versus 5.2 million) and the unemployment rate in the census was 2.1 percentage points higher. Before the 1990 census, the CPS traditionally measured higher unemployment than the census. Although the 1990 census had higher unemployment measures than the CPS, the gap widened greatly in 2000. This occurred in spite of changes to the Census 2000 employment status questions designed to narrow the differences with the CPS. The Census Bureau and the BLS are continuing to examine the differences and potential contributing factors.

The CPS measures of income and poverty are based on a more detailed series of questions than in Census 2000 and the CPS is conducted in an interview setting with experienced interviewers. The income measures from the CPS are, therefore, expected to be more complete. However, Census 2000 measured a significantly higher median household income than did the CPS. The same was true of the median income measures for families, married-couple families, and families with a female householder and no spouse present. For families with a male householder and no spouse present, the Census 2000 median income measure was lower than the CPS. Again, these findings require more study.

Poverty measures in Census 2000 and the CPS were relatively consistent, at 12.4 percent and 11.9 percent, respectively. Also encouraging was the fact that the Census 2000 Supplementary Sample measure of poverty was 12.2 percent.

7.8 Indicators of coding quality

Several of the items in Census 2000 were answered at least in part by narrative responses that had to be translated into a code that classified the responses into categories for tabulation. The 100 percent items on race and Hispanic origin had a relatively small number of narrative or write in responses, whereas all responses were narrative for the sample questions on ancestry, language spoken at home, place of birth, place of work, industry and occupation. Although there were no formal studies of these coding operations available at the time this report was prepared, there is some analysis presented by Census Bureau staff at professional association meetings and in inter-

nal memoranda. This information is summarized here.

The assignment of codes to narrative responses in the race (including American Indian tribe), Hispanic origin, and ancestry questions was accomplished in a multi-phase operation. First, responses were matched via computer to a set of dynamic coded master files, containing write in responses that had been coded, from the 1990 Census and ongoing survey work in the American Community Survey. Write in responses from Census 2000 were matched against these master files and, if a match was found, the response was automatically coded. If the write in response could not be coded automatically, the response was sent to expert coders who had additional reference materials. The expert or clerical coding operation was subjected to quality assurance procedures designed to provide feedback to the expert coders during the process. The documentation of the coding process mentions some problems in the coding operations, but there is no quantitative assessment of any error resulting from the problems. One problem noted was that some of the cases assigned a status "uncodeable" early in the process probably could have been coded later when the learning curve had progressed. Also, errors discovered in the master files could not be corrected during the coding process.

The procedures to assign codes to the industry and occupation responses began with an automated coding system using a database from the National Institute of Occupational Safety and Health, supplemented by coded responses from the American Community Survey. This database was validated by expert review. It assigned

codes to 59 percent of the industry responses and 56 percent of the occupation responses. The accuracy levels were 94 and 92 percent, respectively. Responses that could not be coded by the automated system were sent to clerical coders with automated reference

materials. These coders could "refer" to a group of expert coders any cases they could not code. There was a quality assurance process with adjudication on the clerical coding. This process involved independent assignment of codes by two additional coders.

The early results showed that the code assigned to the census record was a "minority code" of the three codes assigned in 13 percent of industry cases and 16 percent of occupation cases.

8. Recommendations/Conclusions

In general, the evaluations show the Census 2000 data are of reasonably high quality. However, improvements should be pursued through the use of the American Community Survey as an integral part of 2010 Census planning and through other innovations. The recommendations that follow comprise a subset of the recommendations from the individual evaluation reports plus additional recommendations based on the tabulations on sample data allocation prepared for the National Academy of Sciences. They are grouped into three categories for ease of reference.

Data Collection Operations and Procedures

- Continue with plans to use the American Community Survey (ACS) to obtain sample data in place of a long form in the 2010 Census, assuming that continuing studies of data quality yield acceptable results.
- Test whether increased use of the Internet and telephone response would encourage greater self response and yield quality improvements.
- Include an evaluation project, like content reinterview, as part of the American Community Survey program to allow continuing measures of quality.
- Investigate whether records on such topics as utility costs and real estate taxes could be substituted for or used to supplement direct inquiry.

- Research whether modeling might be a useful way of improving data on some topics such as second mortgage payment and insurance payment.

Questionnaire Design and Testing

- Test form design and question placement to avoid problems like the placement of the tenure question and the effect that placement had on item nonresponse.
- Use cognitive experts as part of the testing of revised questions for new and/or troublesome topics.
- Use separate yes/no questions for each response category when the respondent can select more than one category. An example is the period of military service question in Census 2000.
- Review and test the respondent instructions for problematic questions. The Content Reinterview Report specifically mentions the question on Hispanic origin.
- Determine characteristics related to high inconsistency measures and perform multivariate analysis with respect to these characteristics in the attempt to develop better ways to ask the questions.
- For the items with serious quality problems noted in section 7.6 conduct additional research and testing to determine whether the ACS procedures provide improved data or whether ques-

tion design and/or concept clarification improvements are needed.

- The Census Bureau and the BLS should continue work on the employment and unemployment questions to be included in the American Community Survey to ensure they complement the BLS program on local area unemployment statistics.

User Guidance

- Issue guidance to users on the types of uses to avoid for questions with high levels of inconsistency. For instance, data from such questions should probably not be used for small areas or in detailed cross tabulations with other variables.
- Consider not publishing detail from questions with high variability. For example, the questions on type of income may be useful for determining total income but they may not be appropriate to use as individual data items.
- Continue to study the differences between the Census 2000 and the CPS measures of employment and income and provide findings to users.

This page intentionally left blank.

References

- Kirk, Mary; Buckles, Ester; Mims, Wynona; Apple, Marty; and Johnson, Patricia, Preliminary Results from the Census 2000 Industry and Occupation Coding, July 18, 2001.
- Love, Susan, Memorandum for Daniel Weinberg, Comparison of Item Imputation Levels by Mode: Census 2000, Census Supplementary Survey, and 1990 Census, April 24, 2003.
- Norris, Sherri, Census 2000 Evaluation B.1.b, Analysis of Item Nonresponse Rates for the 100 Percent Housing and Population Items from Census 2000, September, 2003.
- Palumbo, Tom; Weismantle, Mai; Clark, Sandra; Posey, Kirby; Iceland, John, Comparing Employment, Income, and Poverty: Census 2000 and the Current Population Survey, September, 2003.
- Singer, Phyllis and Ennis, Sharon R., Census 2000 Evaluation B.5, Census 2000 Content Reinterview Study: Accuracy of Data for Selected Population and Housing Characteristics as Measured by Reinterview, September, 2003.
- U.S. Census Bureau, Documentation of the Race, Hispanic Origin, and Ancestry Coding Operations from Census 2000, March, 2003.
- U.S. Census Bureau, Population Division Working Paper No. 18, Results of the Race and Ethnic Targeted Test, May, 1997.
- U.S. Census Bureau, 1996 National Content Survey Findings on Questions on Race and Hispanic Origin, December, 1996.
- Zajac, Kevin J., Census 2000 Evaluation B.1.a, Analysis of Imputation Rates for the 100 Percent Person and Housing Unit Data Items from Census 2000, September, 2003.

This page intentionally left blank.

Appendix

Table 1.
Summary of Quality Measures for Populations Items: Census 2000 and 1990

Person Item (Sample Household Population)	Allocation Rate for Census 2000	C2SS	Census 2000 (1990 defn)	1990 Census	CRS Index of Inconsistency, Census 2000	CRS Index of Inconsistency, 1990 Census
Relationship	2.3	1.6	2.7	1.9	X	X
Sex	0.9	0.5	1.6	0.8	1.7	X
Age	2.6	2.4	2.6	0.9	7.8	X
Hispanic origin	3.6	3.6	4.0	3.4	17.2	12.2
Race	3.2	2.4	3.2	1.1	23.1	16.3
Marital Status	2.2	1.8	3.4	0.9	5.8	X
School enrollment	6.2	4.0	6.2	4.2	13.5	17.3
Grade attending	9.0	5.5	X	X	9.0	X
Educational attainment	7.2	4.8	7.2	4.5	36.5	32.3
Ancestry	¹ 19.3	X	X	¹ 11.7	30.7	26.5
Speaks non-English language	5.2	4.3	5.8	4.8	22.7	26.9
Language spoken	11.4	8.9	11.4	11.9	17.9	5.2
English ability	7.6	6.0	7.6	8.5	59.5	60.3
Place of birth	9.2	6.4	9.2	5.1	3.2	4.9
Citizenship	0.8	0.5	5.2	4.2	9.8	10.9
Year of entry	14.7	7.8	14.7	8.9	18.9	23.0
Mobility status	5.8	4.0	8.6	5.2	22.2	X
Migration:						
State	8.6	14.0	8.6	5.7	4.4	X
County	8.6	14.5	8.6	7.5	4.4	X
Place	8.8	14.9	8.8	9.6	X	X
Disability:						
Sensory	6.9	5.0	X	X	47.2	X
Physical	7.6	5.2	X	X	42.0	X
Mental	7.5	4.7	X	X	54.4	X
Self-Care	7.9	4.8	7.9	5.8	51.7	73.6
Mobility	9.9	5.6	10.0	5.1	64.5	47.1
Work	11.4	5.9	11.4	4.0	80.5	43.0-45.7
Grandchildren in home (persons 15+)	4.5	3.4	X	X	25.8	X
Responsible for grandchild	15.3	17.6	X	X	46.1	X
Months responsible	17.8	19.7	X	X	53.7	X
Veteran status	7.5	4.7	7.5	4.8	18.7	8.5
Period of service	9.8	13.4	10.3	6.1	² 7.8-93.0	² 3.4-93.7
Length of service	9.1	6.7	9.1	17.5	41.6	58.8
Employment Status	11.1	6.0	11.1	3.8	X	X
Place of work:						
State	9.7	5.8	9.7	7.2	X	X
County	10.1	6.2	10.1	7.9	X	X
Place	10.6	6.7	10.6	9.7	X	X
Means of transportation	7.6	4.6	8.2	4.6	X	X
Vehicle occupancy	10.0	5.8	10.0	4.9	X	X
Time left home	15.0	11.3	15.8	10.8	X	X
Travel time	11.8	8.7	12.3	6.9	X	X
Industry	14.9	9.4	15.2	8.0	X	X
Occupation	14.9	9.5	16.1	9.1	X	X
Class of worker	17.0	8.3	17.6	9.0	X	X
Work last year	9.4	X	13.6	13.5	24.3	45.9
Weeks worked	19.3	9.6	20.2	14.7	57.5	56.8
Usual hours worked	17.4	9.1	18.1	14.5	34.3 ³	40.1
Personal Income						
Wage/salary	20.0	16.4	20.0	10.0	21.2,48.0	X
Self-employment	9.9	6.3	9.9	6.4	44.4,45.3	X
Interest, dividend	20.8	13.3	20.8	8.1	58.0,44.7	X
Social Security	19.3	11.7	19.3	8.0	13.4,60.4	X
Supplemental Security	19.0	10.2	19.0	⁴ 7.5	48.2,55.6	X
Public Assistance	18.2	10.5	18.2	⁴	53.9,61.7	X
Retirement, pension	18.8	11.0	18.8	7.7	36.8,42.0	X
Other income	18.3	10.6	18.3	7.6	60.7,49.6	X
All income allocated	24.5	20.0	24.5	11.7	X	X
Some income Allocated	29.7	23.9	29.7	13.4	X	X

X means not applicable

¹Since imputed values were not allocated for blank ancestry items, this represents a nonresponse rate rather than an allocation rate. It also includes as nonresponse persons whose response could not be classified into an ancestry category.

²Each period of service was analyzed separately. The numbers reflect the range of values across all periods of service.

³The first number represents the index of inconsistency for receipt of the income type; the second number represents the index for the dollar amount.

⁴In the 1990 census, the receipt and amount of Supplemental Security Income and public assistance income were collected as a combined category.

Table 2.
Summary of Quality Measures for Housing Items: Census 2000 and 1990

Housing Item (Sample Household Population)	Allocation Rate for Census 2000	C2SS	Census 2000 (1990 defn)	1990 Census	CRS Index of Inconsistency, Census 2000	CRS Index of Inconsistency, 1990 Census
Tenure	4.3	1.4	8.0	1.4	19.4	13.3
Units in structure	4.4	1.4	4.4	1.6	20.8	21.9
Year built	11.7	13.4	11.7	23.0	29.3	40.6
Year moved in	6.2	3.7	6.2	2.9	21.2	X
Rooms	6.2	2.6	6.2	0.4	57.1	X
Bedrooms	8.9	4.2	10.2	7.5	20.4	X
Plumbing facilities	3.4	1.0	3.4	1.7	85.2	53.8
Kitchen facilities	3.4	0.9	3.4	1.8	75.8	X
Telephone	4.3	1.0	4.3	1.9	54.7	X
Heating fuel	7.4	2.1	7.4	2.9	17.7	14.0
Vehicles available	6.2	1.6	6.2	2.2	37.1	32.1
Business on property	8.2	2.8	8.2	2.4	65.8	50.0
Lot size	10.6	3.6	X	X	20.9	27.8
Agricultural sales	14.3	4.7	14.3	13.7	52.0	41.7
Cost: Electricity	17.1	6.9	18.5	5.5	68.8	X
Gas	23.9	11.5	24.7	10.7	54.9	X
Water/sewer	19.6	8.3	21.8	7.3	43.8	X
Other fuel	28.7	13.4	31.9	17.5	46.0	X
Monthly rent	15.6	5.3	15.6	1.3	23.2	34.7
Meals in rent	7.9	4.5	7.9	5.1	38.2	71.6
Mortgage	6.0	2.0	18.6	5.7	17.2	X
payment	19.6	10.1	22.4	5.5	27.6	X
Second mortgage	11.8	X	16.0	5.1	48.6	X
payment	23.9	18.9	23.9	8.1	93.7	X
Real estate taxes	32.0	20.8	32.0	12.2	¹ 18,644.0	X
Insurance	36.6	24.5	36.6	16.8	¹ 26,665.6	X
Value	13.3	9.7	13.3	3.3	59.1	X
Mobile home costs	63.5	37.8	63.5	41.8	² 60,682.2	X

X means not applicable.

¹The first number represents the index of inconsistency for whether real estate taxes and insurance premiums were included with mortgage payments; the second number represents the index for the amount of such payments.

²The first number represents the index of inconsistency for having an installment loan or contract on the mobile home; the second number represents the index for the dollar amount.

