

## DATA QUALITY IN THE 1990 CENSUS - THE CONTENT REINTERVIEW SURVEY

Kathryn F. Thomas, Tamara L. Dingbaum, U.S. Bureau of the Census  
Kathryn F. Thomas, Bur. of the Census, DSSD Rm 1641-3, Wash., DC 20233

**KEY WORDS:** Reinterview, Response Bias, Response Variance

### I. INTRODUCTION

A great deal of effort goes into planning the content of each decennial census. Many topics are considered, but only a limited number of items are selected for inclusion. Integrated into this content development process is the Content Reinterview Survey (CRS). The CRS produces valuable information on data quality that tell data users which data can be used and which data should be used with caution. As part of the continuous content development process, these same CRS data indicate to census planners where improvements may be necessary.

### II. BACKGROUND

The Content Reinterview Survey (CRS), the largest content evaluation conducted by the Census Bureau, is a part of the 1990 Research, Evaluation and Experimental (REX) program. A similar survey has been conducted after each decennial census since 1950. The CRS is designed to measure response error associated with selected population and housing items. The CRS sample was restricted to long form census households. Census households responding by mail and enumerator return households were reinterviewed.

### III. MEASURES OF RESPONSE ERROR

Measures of the bias and variance components of the total mean square error due to response error are used as indicators of data quality. Response errors may be introduced in the data collection phase and in the processing phase of the census. Response errors may be due to the questionnaire design, the data collection methods, the specific question asked, or other causes. These errors may be accidental, may be introduced purposely, or they may arise from insufficient information on the part of respondents to answer the questions.

The measure used to analyze the impact of simple response variance on estimates is the index of inconsistency (I). The index of inconsistency may be interpreted as that portion of the total variance accounted for by simple response variance.

The objective of a response bias type reinterview is to measure the "true" characteristic of every individual or housing unit in the sample; however, at best, the reinterview provides only "better" responses than those obtained in the original survey. The reinterview uses improved measurement procedures in an attempt to get the most accurate responses possible. The improved procedures took several forms in the CRS. In some cases, a more detailed sequence of probing questions was used. They were designed to clarify areas that may have been unclear in the census. Self-response was used to the extent possible to improve the accuracy of response.

The net difference rate, under the assumption that the reinterview is a perfect second trial and represents the "truth", gives an estimate of the amount of response bias in the distribution or category. The net difference rate for a category is defined as the expected difference between the census and the reinterview for the proportion of cases in that category. However, in most cases, the reinterview is only an improved procedure and provides more accurate data on average than the original interview, but not necessarily the "true" value in all cases. In this case, the net difference rate provides an underestimate of the bias in the distribution or category.

### IV. HOUSING CHARACTERISTICS

#### A. Response Variance Type Reinterviews

Only three housing items were reinterviewed by asking the same question as was asked in the census. The three items, description of building, size of lot and agricultural sales, showed moderate inconsistency in response between the two surveys.

Data were collected for the description of building item in the 1980 Reinterview. This item was in the moderate consistency range just as it was found to be in the 1990 CRS. Although in the same range, the trend for the index is downward, indicating that the newer data is a little more consistent.

The item on size of lot, asking respondents in single units if their house is on less than 1 acre, was reported more consistently on mail return questionnaires than on enumerator returns. There was no difference in mail returns and enumerator returns for the agricultural sales question which asked for the dollar value of actual sales of all agricultural products from the property.

**B. Response Bias Type Reinterviews**

For the remainder of the housing questions, a probing set of questions was asked in a response bias type reinterview.

- Tenure - Overall, the item on tenure was consistently reported between the surveys. This item was found to be consistently reported in both the 1980 and the 1970 reinterviews also. However, analysis showed that there was bias in all categories of the distribution except "occupied without payment of cash rent." "Rented for cash rent" and "owned by you or someone in this household with a mortgage or loan"

appear to be overreported in the census while "owned by you or someone else in this household free and clear (without a mortgage)" is underreported.

- Plumbing Facilities - Only one of these response bias type items, plumbing facilities, was reported with high inconsistency. This was not significantly different from the results of the 1980 reinterview. Collapsing the 1990 CRS plumbing data to the "Yes" and "No" categories of the census indicates no bias in the distribution, although the data are inconsistent.

The lack of consistency in reporting for this item is not so much due to reporting problems in the category "Yes, have all three facilities," as it is in the lack of consistency of reporting in the category, "No." Almost 99 percent of the households reporting "Yes" in the census also did so in the reinterview. However, only 44 percent of those answering "No" in the census reported the same in the reinterview.

- Meals Included in Rent - Households which rent their units for cash rent were asked "Does the monthly rent include any meals?" to help identify congregate housing. 1990 is the first time this item was included in the census. There were not enough housing units responding positively to this question to allow response error measures to be precisely computed.

**Table A. Response Bias Type Reinterviews (Housing Data)**

Characteristic	1990		1980	
	L-Fold Index of Inconsistency	90% Conf. Int.	L-fold Index of Inconsistency	90% Conf. Int.
Tenure	13.3	12.6-14.0	8.0	7.2-9.1
Number of Autos, Vans, and Trucks	32.1	31.1-33.1	34.0 autos 26.0 trucks/vans	32.1-35.1 24.6-28.3
Monthly Rent	34.7	33.2-36.4		
Meals Included in Rent	...	... ..		
Year Built	27.6	26.7-28.6		
Single Units	26.2	25.2-27.2	25.0	23.3-25.9
Multi-Units	36.9	34.3-39.8	43.0	39.0-48.4
Plumbing Facilities	53.8	45.1-64.2	47.0	39.6-55.7

... Not sufficient data to compute response error measures

• Other Housing - The other housing items, monthly rent, number of autos, vans and trucks, and year built were all reported with moderate inconsistency between the census and the CRS. (See Table A.) Year built was asked in reinterview in both 1970 and 1980 with results similar to those gathered in 1990. Number of automobiles and number of vans and trucks were asked separately in 1980 with similar results in consistency as was found in 1990. Monthly rent was not evaluated in 1980, but 1970 reinterview data appear to be more consistent with that census.

Monthly rent was collected in 26 categories in the census. The data collected in the reinterview show no real bias in this data. Only one category, \$600 to \$649, shows slight underreporting in the census.

Number of autos, vans, and trucks was asked as one question in the census. Automobiles was asked separately from vans and trucks in the CRS. There is some bias in these data. The category "None" shows significant overreporting in the census, while the category "4" is underreported.

Two of the eight categories in the year built item (excluding the "Don't Know" category) show some evidence of bias. There is evidence of underreporting in the category "1989 or 1990," the most recently built homes. There is also significant underreporting in the category "1970 to 1979."

## V. POPULATION CHARACTERISTICS

### A. Response Variance Type Reinterviews

Several population items were repeated in the CRS just as they were asked in the Census. These response variance type reinterviews produced indices in the low range for data on Spanish origin and school enrollment. More moderate inconsistency was found between the two data sources for the information on year of immigration and employment-history. (See Table B.)

• Hispanic Origin - The item on Hispanic origin received a response variance type reinterview in the 1980 CRS and exhibited a low level of response variability there, comparable to these 1990 results. Four of the

**Table B. Response Variance Type Reinterviews (Population Data)**

Characteristic	L-Fold Index of Inconsistency	90% Confidence Interval
Spanish Origin	12.2	11.2-13.2
School Enrollment	17.3	16.6-18.0
Year of Immigration	23.0	21.1-25.2
Employment History	32.6	31.6-33.7

five response categories for this item in 1990 showed very low levels of variability. Only the category "Yes", other Spanish" displayed moderate variability.

Analysis of this inconsistency showed that almost 96 percent of those reporting "No, not Spanish" in the census then "Yes", other Hispanic" in the reinterview and providing a write-in response reported an actual Hispanic group indicating they truly are of Hispanic descent. The CRS results for the Hispanic origin item for 1990 are almost identical to those obtained from the 1980 Content Reinterview. This suggest that the reporting problems have not changed over the decade.

• School Enrollment - School enrollment, whether a person has attended school or college, public or private, just prior to the census, was consistently reported. Prior data is not available on this item. 1990 results indicate this is not a troublesome item for respondents.

• Year of Immigration - Year of immigration, asking when those born outside the United States came to the United States to stay, shows variability in the low moderate range. The same question was asked in 1980 and was reported very consistently. However, caution should be used in comparing these two results since there were only 6 possible response categories in 1980 and the 1990 item had 10 response choices. Differences in the response errors may well be attributable to the difference in the number of response categories rather than any true response differences. It is of interest to note that much of the variability in the 1990 data comes from the new, more specific intervals.

All other intervals are reported more consistently.

- Employment (Industry) - An item on type of industry was asked in the CRS as a control for the class of worker item. This item was reported with moderate inconsistency and should, therefore, be used with caution.

## B. Response Bias Type Reinterviews

Many population items were asked in the CRS using a probing sequence of questions aimed at making the census question more clear or gathering more supporting data for responses given. With this data it is then possible to measure response bias using the net difference rate.

The first four items shown in Table C were reported with very low levels of inconsistency between the two sources. The last four items had indexes of inconsistency in the more moderate range. One disability item on self-care limitations was highly inconsistent.

Four of the items, place of birth, citizenship, other languages and educational attainment, were all asked in the 1980 Reinterview. For each of these items the measure of accuracy between the two decades was comparable. The question on citizenship was also asked in the 1970 Reinterview with very similar results.

- Place of Birth - Place of birth is reported with considerable accuracy and with virtually no bias.

- Military Service - The census asked all those aged 15 and over about active-duty military service. There is a great deal of bias in the category "Yes, now on active duty" which is overreported in the census. Further analysis revealed that persons contributing to this situation have been in the military service and gave dates in the reinterview ranging between 1943 and 1988 for their ending date for their period of service. This indicates the error may be in the census.

- Citizenship - In 1980 citizenship was asked of foreign born persons only. However, many people born in the United States erroneously answered the question, making for very

inconsistent data since there were no appropriate categories for them to choose. When the data was edited and these native-born people were removed, the remaining data for foreign-born persons was reported quite accurately. For 1990, the citizenship question was revised to address the total population rather than only the foreign born. Again, as reported previously, these data were quite consistent between the 1990 census and the CRS.

- Race - The race question is asked of all persons in the census. The concept of race as used by the Census Bureau reflects self-identification by respondents. Race was not evaluated in the 1970 nor the 1980 CRS. In the 1990 Reinterview, race was asked as a response bias type question.

For initial analysis, the race data were collapsed into six major categories. These data show that "Whites" are underreported in the census while "Other Race" is overreported. 229 people reported in the census that they were "White", but told the CRS enumerator that they were "Other Race." Forty-one percent did not report what that "Other Race" was, so no further analysis is possible for them. For the 134 people who did report what their "Other Race" was, the vast majority were of Hispanic origin. The Hispanic population appears to have difficulty in classifying themselves into the race categories presented.

The data for race was divided between those who said they were Hispanic in the census and those who reported themselves as Non-Hispanic. The Non-Hispanics represent over 94 percent of the total population and the race data for the Non-Hispanics contained no bias at all. The race data for all people showed significant bias in the categories "White" and "Other Race". Looking at the same race data for just Hispanics, there is significant bias in all categories except APIs. This indicates that the Hispanic population are contributing most of the bias in the race data in the census.

- Other Language - There is no bias in the data asking if a person speaks a language other than English at home. If a person reported speaking a

language other than English, they were asked how well they speak English. There was not much consistency between trials on this item.

- **Class of Worker** - Each person in the census who had worked anytime since 1985 was asked to describe their most recent job activity and to select their class of work. Three of these eight categories showed some degree of bias. There was a small amount of bias in the first category, "Employee of PRIVATE FOR PROFIT company or business...", which was overreported in the census. Two other categories, "state government employee" and "SELF-EMPLOYED in own NOT INCORPORATED business, professional practice, or farm" were underreported in the census and showed a large amount of bias.

- **Educational Attainment** - The census asked how much school a person had COMPLETED. That was compared to a recode of questions from the Reinterview asking about school attended, completed and degrees received. Almost all of the categories (13 of 17) showed some bias, most of them highly biased.

- **Disability** - The census asked

questions on work disability, mobility limitation and self-care limitation of persons 15 and over. The work disability question was revised from 1980 while the mobility and self-care limitations are new questions for 1990. The CRS made extensive inquiry into the ability to perform specific functional activities and the use of special aids. There was moderate inconsistency in reporting for the first 3 items and the self-care limitation was highly inconsistent between the two sources. Only the question on mobility limitation was reported with no bias. The first part of the question on work disability, limits the kind or amount of work this person can do, is highly biased and underreported in the census for those reporting "Yes". The second part of that question, prevents persons from working at a job, is very highly biased in both categories. The question on self-care limitations is highly biased in the "Yes" category and more moderately biased in the "No" category. All this indicates that the census data on disability is highly suspect and should only be used with great caution.

**Table C. Response Bias Type Reinterviews (Population Data)**

Characteristic	1990		1980	
	L-Fold Index of Inconsistency	90% Conf. Int.	L-Fold Index of Inconsistency	90% Conf. Int.
Place of Birth	4.9	4.6-5.2	6.0	6.2-6.8
Military Service	8.5	7.9-9.2		
Citizenship	10.9	10.0-12.0	73.0	70.9-74.9
Race	16.3	15.5-17.1	13.0*	12.0-14.8*
Other Language	26.9	25.6-28.3	25.0	23.2-26.5
Industry & Occupation				
Class of Worker	27.5	26.5-28.5		
Educational Attainment	32.3	31.7-32.9	35.0	30.8-39.0
Disability	43.0	40.9-45.1		
Work Disability	45.7	42.4-49.3		
Mobility Limitation	47.1	44.2-50.2		
Self-care Limitation	73.6	69.5-78.0		

\*Edited Data

## VI. DATA LIMITATIONS

### A. SAMPLING VARIABILITY

The data presented here are shown as numbers of sample housing units and sample persons, not as inflated national estimates. The sample was selected using equal probability methods so that each sample housing unit had the same weight.

### B. SOURCE OF RESPONSE ERROR

This report concentrates on comparisons with census data before consistency edits and imputations for missing data were made as part of the computer processing operations. Thus, the response error measures reflect mainly errors which occurred during the field data-collection process.

### C. NONINTERVIEWS

Initially, an oversample was drawn so that housing units vacant in the census could be eliminated from the sample. A sample of 12,891 housing units was sent to CATI. 83.1 percent of those were returned as completed interviews. The refusal rate was 4.5 percent. In addition, 7.2 percent of the housing units had moved since the census and 1.0 percent were vacant at the time of the reinterview. 4.1 percent were in the "Other" category representing housing units which had language barriers, household members who were ill, hospitalized or temporarily absent, deceased, and so forth.

### D. MATCHING

In the past, matching in reinterview studies has been a problem which has sometimes caused severe data limitations. Efforts were made in planning this evaluation to keep matching problems to a minimum. The housing unit sample was selected from the ACF (Address Control File) making it possible to track a housing unit through the census process. Then, names of the census occupants of the sample units and their corresponding census line numbers were keyed and maintained on both CATI and personal

visit questionnaires. Using this information, all housing units were matched back to the census.

### E. TIMING

The CRS was conducted immediately following the completion of the decennial census field work. This was more than five months after the first census data was collected and as much as nine months before the final CRS data was collected. Timing of a reinterview survey is important because going to the field too soon after the original data collection can increase the effects of conditioning. There is a potential loss of accuracy in the data collected if too much time passes. The magnitude of the effect of the time lag cannot be determined. It exists in varying degrees for many of the items included in the survey. There was also some data loss because some respondents had moved before the reinterview was conducted.

This paper reports the general results of research undertaken by Census Bureau Staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

### Bibliography

- [1] "CRS: Accuracy of Data for Selected Population and Housing Characteristics as Measured by Reinterview," HPC 80 - E2, 1980 Census of Population and Housing, US Department of Commerce, Bureau of the Census, Sept. 1986.
- [2] "Accuracy of Data for Selected Population Characteristics as Measured by Reinterviews," PHC(E)-9, US Bureau of the Census, Aug. 1974.
- [3] "Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews," PHC(E)-10, US Bureau of the Census, April 1975.
- [4] "Evaluating Censuses of Population and Housing," US Department of Commerce, Bureau of the Census, Statistical Training Document, ISP-TR-5, September 1985.