

HOW GOOD ARE THESE DATA ANYWAY? THE 1990 CONTENT REINTERVIEW SURVEY

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

INTRODUCTION

In the realm of nonsampling error at the Census Bureau, coverage errors and their measurement are of prime importance. Missed persons or households, duplicates and even those erroneously included are of interest to both data users and critics alike. But, what about the other aspects of nonsampling error -- what about the content errors? The Bureau places great importance on eliminating or controlling these response errors and on estimating any response error that remains in the data. Throughout each decade, as the content of the upcoming census unfolds, the Census Bureau tests and analyzes numerous ways of wording and asking the items selected for inclusion. In the following, the content development process for the decennial census is outlined. Then the Content Reinterview Survey (CRS), the Bureau's largest content evaluation, is viewed in more depth. Some actual results from the 1990 CRS are presented in an attempt to answer "How Good Are These Data Anyway?"

I. Overview of the Content Development Process

The Twenty-first Decennial Census was conducted on April 1, 1990. Long before Census Day, the Census Bureau had to make important decisions about the subjects and questions to be included in the census. By law, the Census Bureau had to submit the subjects planned for inclusion in the 1990 Census questionnaires to Congress by April 1, 1987; specific questions had to be submitted to Congress by April 1, 1988. The content development process determines which subjects and questions are included on the census questionnaires. There are two versions of the census questionnaire: a short form containing a limited number of basic population and housing questions, and a long form containing these basic questions as well as a number of additional questions.

The purposes of the census are to provide the population counts needed to apportion seats in the House of Representatives and determine state legislative district boundaries and to meet critical national data needs for the next 10 years. No part of

census planning is more important than selecting the content, or questions, to be included on the questionnaires. While it is impossible to anticipate all uses of data from the 1990 Census, many federal programs, state and local government programs, academic researchers, businessmen and marketing researchers, national, regional, and local organizations and groups, and some individuals will use the data collected.

The National Content Test (NCT) was the single most important 1990 Census content testing program conducted by the Census Bureau. The primary objective of the NCT was to test new and revised question wording, formatting, and sequencing. The decisions concerning the items to be tested in the NCT were based on the recommendations gathered from the major content development programs.

The 1990 Census was the bicentennial anniversary of census-taking in the USA. Many difficult choices about census content were made to complete 1990 Census plans. This important task of selecting the census subject content was the result of careful review and testing of recommendations from a large variety of users of census data, including federal and other governmental agencies, advisory committees, professional groups, and members of the general public.

II. The Content Reinterview Survey (CRS)

The Bureau is continually conducting evaluations for improving data quality. The Content Reinterview Survey is one of many "snapshots" along the way to improving data quality. Traditionally, the Content Reinterview Survey (CRS) has been conducted following the Decennial Census and is designed to measure the response error associated with selected population and housing items. The 1990 study focused on new and revised data items and had innovations in field data collection techniques.

The CRS sample was restricted to long form households. Both households responding by mail and nonmail return households were reinterviewed. A

single stage systematic sample of 12,800 occupied units was selected.

For some items in the Content Reinterview Survey a set of detailed probing questions allows gathering data with a degree of accuracy not possible in the census. That is, the Content Reinterview Survey may be viewed as the "preferred" measurement technique. Comparison of the reinterview data with the census will provide estimates of response bias present in the census data. Items which are being evaluated using a response-bias (probing) type reinterview include Race, Place of Birth, Citizenship, Education Level, Ancestry, Other Languages Spoken, Military Service Status, Employment Status, Tenure, Monthly Rent, Meals in Rent, Number of Autos and Vans, Plumbing Facilities and Year Structure Built.

Simple response variance is estimated for some items by asking the same question(s) asked in the census. Items which are being evaluated using a response-variance type reinterview include Spanish Origin, Description of Building, Year of Immigration, Size of Lot, School Enrollment, Agricultural Sales and Employer: Kind of Business and Type of Company.

III. Measures of Response Error

The 1990 Content Reinterview Survey was designed to provide estimates of simple response variance for some questions and response bias for others. 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 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 estimates of the proportion of cases in that category from the census and from the reinterview. 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.

Although neither of these two types of reinterview studies, in application, can meet its theoretical assumptions, if carefully done, both techniques can provide useful information for content evaluation purposes.

IV. Results of the Content Reinterview Survey

A. Spanish/Hispanic Origin

A question on Spanish/Hispanic origin was added in the 1970 census. It was asked of a 5-percent sample of the households. In the 1970 reinterview a more detailed set of probing questions, a response bias type of reinterview, was conducted. The estimated proportion of persons reporting Spanish origin in the 1970 Census, 3.7 percent, was slightly less than the corresponding proportion in the reinterview, 4.0 percent. This bias of 0.1 to 0.6 percentage points was found to be significant at the 95 percent confidence level.

Spanish origin data, which received a response variance type reinterview in the 1980 CRS, exhibited a low level of response variability.

The Spanish origin question was asked of everyone in the 1990 census; that is, it was a 100 percent item. For 1990, a space was added in this item for persons of "Other Spanish/Hispanic" origin to specify their background. Examples of "Other Spanish/Hispanic" origin groups were also added. This inquiry was reworded and instructions were expanded.

In the 1990 reinterview this question received a response variance type treatment; that is, the same basic question was asked as was asked in the census.

The data from the 1990 CRS for Spanish origin exhibited low levels of response variability. Four of the five response categories showed very low levels of variability. Only the category "Yes, other Spanish" displayed moderate variability.

Table of Indexes for Spanish Origin

	Estimate of Index	Lower Limit	Upper Limit
No (not Spanish/Hispanic)	9.3	8.4	10.2
Yes, Mexican, Mexican-American, Chicano	8.5	7.5	9.7
Yes, Puerto Rican	8.6	6.4	11.5
Yes, Cuban	13.6	9.4	19.5
Yes, other Spanish/Hispanic	34.1	30.8	37.7
L-Fold Index	12.2	11.2	13.2

The inconsistent reports in "Yes, other Spanish/Hispanic" category derive from two sources: persons reporting "No, not Spanish/Hispanic" versus "Yes, other Spanish/Hispanic" in the census and reinterview, respectively, and vice versa and persons switching between the "Yes, Mexican, Mexican-American, Chicano" and "Yes, other Spanish/Hispanic" categories. These types of incorrect responses constitute almost two-thirds of all inconsistent responses.

Those people who reported "No, not Spanish" in the census and then reported "Yes, other Spanish" in the CRS contributed to this variability. In looking at the write-in responses for these people, almost one third did not provide a write-in response. Of those who did provide a write-in response after indicating they were "Other Spanish," almost 96 percent reported an actual Spanish/Hispanic group indicating they truly are of Hispanic descent. The unanswered question is why would they report "not Spanish" in the census? The results presented below are almost identical to those obtained from the 1980 Content Reinterview. This suggests that the reporting problems have not changed over the decade.

Table 1. SPANISH ORIGIN

CRS Item Responses	SPANISH ORIGIN					
	Total Reported	Census Item Responses				
	23,967	1	2	3	4	5
1. No (not Spanish)	22174	31	4	5	70	
2. Yes, Mexican, Mexican-Am., Chicano	59	917	1	0	38	
3. Yes, Puerto Rican	11	0	167	0	4	
4. Yes, Cuban	5	1	0	84	4	
5. Yes, other Spanish	100	33	11	5	263	

B. Race

The 1980 census race item underwent extensive revisions for the 1990 census. A response category for "Other API" was added along with a space for respondents to write in their "Other API" subgroup. In addition, the spanner "Asian or Pacific Islander (API)" was placed over the API response categories. New instructions were added and others expanded, and the term "Race" was used as a label.

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; it does not denote any clear-cut scientific definition of biological stock. The data represent self-classification by people according to the race with which they identify themselves.

Race was not evaluated in the 1970 nor the 1980 CRS. In the 1990 Reinterview, race was asked as a response bias type question. Additional categories were added to the list of Asian and Pacific Islanders. "Other API's" and "Other Race" persons could list more than one group, but, if they gave a multiple responses, they were asked with which group they most closely identified. For American Indians, additional inquiries were made regarding their enrolled or principal tribal affiliation.

For initial analysis, the race data were collapsed into the six major categories shown in the following table.

Table 2. RACE

CRS Item Response	RACE						
	Total Reported	Census Item Responses					
	24,539	1	2	3	4	5	6
1. White	20584	30	43	12	38	349	
2. Black/Negro	59	2308	2	2	3	34	
3. Indian/Eskimo/Aleut	37	5	68	0	0	5	
4. API	24	0	0	410	6	9	
5. Other API	8	7	0	21	14	32	
6. Other Race	229	27	4	8	22	169	

These data show that "Whites" are under reported in the census (from 0.3 to 0.7 percentage points) while "Other Race" is over reported (from 0.4 to 0.7 percentage points). All other categories showed no bias. Most of the whites who were under reported in the census were reported as "Other Race." That other racial group was collected as a write-in in the census and was later coded. Tabulations of those codes from the census are not available to us as this paper is being written, but they will be included in later analysis of the data. There is also a group of 229 people who 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, 14.2 percent actually were "White," as they had originally reported in the census. In the reinterview they reported themselves as "Other Race" and then listed their ancestry - - such as Irish, Polish, German or Italian. A small percentage, 2.2, of the people in this category were American Indians. But the vast majority of these people who reported they were "White" in the census, then reported "Other Race" in the reinterview were Hispanic. 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. Remember that 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.

NET DIFFERENCE RATE FOR RACE

	Non-Hispanic	Hispanic	Total Population
White	0.1	-10.4 *	-0.5 *
Black/Negro	0.1	-3.5 *	-0.1
Indian/Eskimo/Aleut	0.0	0.7 *	0.0
API	-0.1	-0.4	0.0
Other API	0.0	2.4 *	0.0
Other Race	-0.1	11.2 *	0.6 *

* denotes significance at the 10% level

C. Plumbing Facilities

A full-fledged housing census began in 1940. The 1940 census included questions on source of water, flush toilet and bathtub or shower. The question on complete plumbing facilities is a major element in determining the quality and value of housing. Data from this question are used in measuring housing needs and delineating areas in need of assistance. Complete plumbing facilities are defined as consisting of hot and cold piped water, a flush toilet, and a bathtub or shower.

For 1990, the complete plumbing facilities item was placed on the sample questionnaire with responses limited to "yes" and "no."

In 1980, the question was asked on a 100 percent basis and had additional response categories to report the presence of some but not all plumbing facilities and to indicate whether use was exclusively by household members or was shared.

Results from the test censuses and the 1986 National Content Test showed that a question simply asking about the presence or absence of complete plumbing facilities and listing the components in the question itself produced reliable data and required less space on the questionnaire. Data on the presence of some but not all facilities and on exclusive or shared use were less reliable; therefore, these latter parts of the 1980 items were not included on the dress rehearsal and 1990 census questionnaires.

The plumbing facilities question on the 1990 CRS form was a response bias type of interview. Inquiry on each of the components of complete plumbing facilities was made separately.

Looking at Table 3, a comparison of the 1990 CRS data, collapsed to compare it to the "Yes" and "No" categories of the census, indicates no bias in the distribution but these data are somewhat inconsistent.

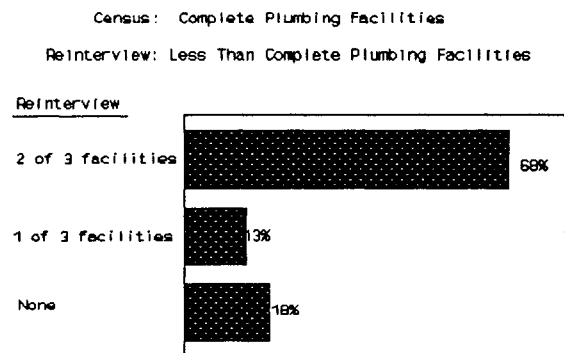
Table 3. PLUMBING FACILITIES

CRS Item Responses	PLUMBING FACILITIES	
	Census Item Responses	
Total Reported	1	2
10,035		
1. Yes, have all three facilities	9910	49
2. No	38	36

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.

Looking at persons (38) who said they had complete plumbing facilities in the census, but reporting less than complete facilities in the Reinterview, the majority of them, 68 percent, had two of the three necessary facilities when the items were asked individually in the CRS.



The net difference rates for this item showed no significant bias. Since a response-bias type interview was conducted in the reinterview, the gross difference rate underestimates the simple response variance for the survey and, thus, the index of inconsistency underestimates the true level of variability. The index of inconsistency in 1990 was not significantly different from that in 1980. Caution should be used in making this comparison because the plumbing item had four response categories in the 1980 census and only two in 1990. The 1990 data appear to be an improvement over 1980 when one of the four categories, "Yes, but also used by another household," showed a large upward bias. There is no significant bias in the 1990 data.

V. Summary

The Census Bureau is dedicated to maintaining and improving the quality of the data it collects and

publishes in the decennial census. We are in a continuous cycle of evaluating the data collected, making improvements, testing the options, and then evaluating again. The Content Reinterview Survey is one of many "snapshots" along the way to improving data quality. The CRS data are just now becoming available and extensive analysis of each of the data items mentioned earlier is planned.

This paper has briefly described the content planning process leading up to the 1990 Census. More details of the CRS and the measures of response error of which it makes use were given. Finally, preliminary data for several items were presented for analysis. The items were selected to represent a little of both the population data and the housing data collected by the CRS. The items also represented both response variance and response bias type interviews.

For the final Content Reinterview Survey report, which is scheduled for printing in the summer of 1993, all items will be analyzed in depth taking the preceding content development (reasons for question wording, placement, etc.) into consideration. The data will be analyzed by various demographic subgroups (where appropriate) such as male-female, black householder-nonblack householder, Spanish/Hispanic householder-nonSpanish/Hispanic householder, by geography, e.g. inside MSA-outside MSA, and for census mail returns versus enumerator returns.

The results will be published so that data users can recognize where there may be problems with the data. They will also know which data are reliable and are appropriate for use and cross-tabulation with other data. The users will then know "how good these data are anyway" for each of the individual topics included in the CRS. The data will also be used by Census planners in the continuing process of evaluating and improving the content data of the next decennial census.

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] Woltman, Henry F. and Thomas, Kathryn F., "Measurement of Content Data Quality in the 1990 Census," presentation at Symposium 90: Measurement and Improvement of Data Quality - Statistics Canada, Oct. 29 - 31, 1990.
- [2] "The Content Development Process for the 1990 Census of Population and Housing," US Department of Commerce, Bureau of the Census, Sept. 1987.
- [3] "Accuracy of Data for Selected Population Characteristics as Measured by Reinterviews," PHC (E)-9, US Bureau of the Census, Aug. 1974.
- [4] "Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews," PHC(E)-10, US Bureau of the Census, April 1975.
- [5] "Content Reinterview Study: 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.
- [6] "1990 Census of Population and Housing - Content Determination Reports - Housing: Plumbing, Equipment, and Fuels, " 1990 CDR - 12, US Department of Commerce, Bureau of the Census, Jan. 1989.
- [7] "1990 Census of Population and Housing - Content Determination Reports - Race and Ethnic Origin," 1990 CDR -6, U.S. Department of Commerce, Bureau of the Census, Feb. 1991.
- [8] "Evaluating Censuses of Population and Housing", Statistical Training Document, ISP-TR-5, U.S. Bureau of the Census, 1985.