ACCURACY OF 1970 CENSUS POPULATION AND HOUSING CHARACTERISTICS AS MEASURED BY REINTERVIEWS

Charles Jones, Henry Woltman, Kathryn Thomas and Stanley Cullimore
U.S. Bureau of the Census

Introduction

Reinterview as an evaluation method may be characterized in general terms as the selection of a sample of persons (or households), who, after answering the questions in the basic census or survey, are contacted at a later time and asked the same or similar questions again. The two responses for each person (or household) are compared to estimate the variability of responses in repeated interviews, and to estimate biases in the distribution of census or survey statistics. 1/

In the evaluation work of the Census Bureau, two general types of reinterview studies have been used. In the first type of study, each person (or household) is viewed as having an infinite set of responses to a specific question which can be generated by independent repetitions of the same survey procedure under the same general conditions. The initial census or survey obtains one of these responses while the reinterview obtains a second, by applying the same survey procedures under the same general conditions as existed in the initial interview. The two responses are assumed to have been randomly selected and are compared to produce estimates of the average trial-to-trial response variability, which is commonly referred to as simple response variance.

The second type of reinterview study is designed to obtain more accurate data than was feasible in the initial interview. These data are used as a standard of comparison for the initial census or survey responses. Here, the initial responses are viewed as being possibly defective because the enumerator may have been inadequately trained, the person answering the questions may not have been the most knowledgeable respondent, the questions and instructions may have been ambiguous, or other things of that nature. It is assumed that these deficiencies can be minimized in the reinterview by applying survey procedures such as the use of well-trained, highly qualified interviewers, choosing the most knowledgeable respondents to provide the data, applying detailed questioning sequences to probe those areas where the questions or instructions may have been ambiguous or inadequate, and reconciliation of differences in responses collected in the two interviews.

It is clear that neither of these two types of studies, in application, can meet their theoretical objectives. In both cases the estimates of response error have a tendency to be understated. In the first case, the conditions of the original interview cannot be duplicated in the reinterview to yield an independent response under the same general survey conditions. For example, the respondent, having answered the question once, is likely to be conditioned in his response in the second interview. The second type of reinterview study is unlikely to obtain the truth in all cases since the respondent may deliberately falsify his responses, or he may simply not know the answer to a particular question. Further, Census shares with all demographic survey organizations the problems of noninterviews. In our reinterviews we usually are not able to complete the study plans for all sample cases. For example, in carrying through the 1970 reinterview study we were unable to complete the study plan for about 25 percent of the persons and 20 percent of the housing units selected for reinterview.

Even with these limitations, however, reinterview is a valuable evaluation methodology. For example, following the 1950, the 1960, and the 1970 decennial censuses, reinterview studies were major components of the evaluation and research programs and these have produced useful data on response errors and their distributions.

Data were collected in the 1970 reinterview study for some 29 of the 100 plus questions included in the census. The questions selected were those for which the reinterview seemed to be an adequate vehicle for collecting reasonably accurate response error data, and for which another type of study would not produce the data more accurately or at less cost.

In the reinterview study, both of the reinterview techniques just described were used to evaluate the quality of the data collected in the 1970 census. For one set of characteristics, the reinterview questions approximated those used in the census. For another set of characteristics, including some of the new questions which lacked precise definitions on the specific subgroups of the population to be identified, more detailed questioning sequences were applied.

Comparison of the census and reinterview responses to a specific question for each person (or housing unit) yields the 2x2 table shown on the first page of the handout. In this table, the cells denoted as a and d represent sample counts where the census and reinterview responses agreed (i.e., either in or out of category) while the b and c cells represent sample counts where responses differed.

The effect of response errors on the quality of data collected for a particular category of a classification system is reflected by the levels of gross and of net error associated with that category. The gross error associated with a category represents the total number of response differences associated with the category (b+c), while the net error is the difference between the number erroneously included in the category and those erroneously omitted from the category (c-b). We have selected two summary response error measures to describe the level of gross and net error associated with the data collected in the census.

The summary measure of gross error - the index of inconsistency - is approximately the complement of the correlation between the census and reinterview responses. For example, if there is perfect positive correlation, the index is zero; if there is zero correlation, the index is 100. The index is

interpreted as the ratio of the observed response differences between two interviews to the response differences that would be expected if there was no correlation between interviews (i.e., if the two sets of responses were randomly associated). In terms of the handout table, the observed response differences are given by b and c, while the expected response differences would be estimated for those cells by the cross products of the marginal proportions in the two interviews. Under the conditions of the first type of reinterview discussed above, the index is also interpreted as the proportion of total variance accounted for by simple response variance. 2/ An index of inconsistency is estimated for each category of a distribution. In addition, a weighted average of the individual indices - referred to as the L-Fold index of inconsistency - provides a measure of the amount of inconsistency in the entire distribution. (When there are only two categories in the distribution, the L-Fold index is identical to the indices for the individual categories.) The range of the index (both for individual categories and for the entire distribution) is zero to 100. As a rule of thumb, we interpret an estimated index between 0 and 20 as indicating low inconsistency, between 20 and 50 moderate inconsistency, and above 50 high inconsistency.

The summary measure of net error - the net difference rate - provides a measure of bias in the census distribution when the reinterview responses are assumed to be more accurate than the census responses. This rate is simply the difference between the census and reinterview estimates of the proportion of housing units or persons in a given category of the distribution. A negative estimate indicates the census proportion is smaller than the reinterview; a positive estimate indicates the census proportion is larger than the reinterview proportion in the category.

Response Variance Reinterviews

For 11 population and housing characteristics the questions and procedures used in the reinterview were similar to those used in the census. That is, the methods used in the reinterview were not an attempt to collect an "improved" response, but simply to obtain responses to questions similar to those used in the census. 3/ The distributions for five of these 11 characteristics had, on the average, fairly low levels of response variance or inconsistency as the estimated L-fold indices were all under 20 (Table 1). The distribution for five of the characteristics had moderate levels of inconsistency, with L-fold indices between 24 and 45, while for one characteristic - Value of Home - the census and reinterview responses were highly inconsistent, as reflected by the estimated L-fold index of 58.

The estimated level of the index of inconsistency for a characteristic is sensitive to the detail of the classification system. Given the same set of responses, the estimate will usually increase as the detail increases and decrease as the detail decreases. For example, the index for Value of Home is 58 when the detail is the 11 categories in which the responses were collected on the census questionnaire. When the detail is reduced by forming fewer, broader categories, the index is also reduced. For example, looking at one of our published distributions for Value of Home which had only 6 categories, the index is estimated at 47. The indices shown in the handout were all estimated at the detail in which data were collected in the census. These would not apply to those published distributions where the detail is collapsed to broader categories. Additional evaluation of the Value of Home by use of a record check type study is being conducted. Those data will provide further insight into the accuracy of census responses for this question.

Response distributions for housing items were analyzed by occupancy status, tenure and size of structure. Generally, owners reported housing data more consistently than renters, responses for occupied units were more consistent than those for vacant units, and respondents in single unit structures reported more consistently than those in multi-unit structures.

Response error data for two of the 11 characteristics were available from the 1960 census - Bathtub or Shower Facilities and Flush Toilet Facilities. Comparison of 1970 and 1960 data for these items indicates that the response variability was of about the same order of magnitude in the two censuses (Table 3 - see notice on page 6 about table availability).

One question often asked is whether the use of a mail census increased or decreased the simple response variance in comparison with an enumerative canvass. Although the mail census technique is expected to have reduced the correlated component of response error contributed by enumerators, it is assumed that there may be differences in the simple response variance as well. A study included in the 1960 census estimated simple response variance for the two types of procedures in independent samples. The results indicated slight differences in simple response variances for the two types of procedures with the differences as observed tending to favor the mail procedure as having lower simple response variances. 4/ The 1970 reinterview study sheds little additional light on this question. Slight differences in response variance were observed for the two types of censuses, but these differences may reflect, in addition to procedural differences, differences in simple response variance for population groups covered by each type of census in 1970.

Response Bias Reinterviews

For 15 population and housing characteristics, a response bias type of reinterview was attempted. 5/ The reinterview involved the use of a detailed questioning sequence designed to probe areas where the question or instructions may have been unclear, and/or a reconciliation to obtain the "best" response when the census and reinterview classed the person or housing unit in different categories. The reinterviews were conducted in personal visits and the reinterviewer was instructed, when feasible, to observe the housing facilities, in addition to questioning the respondent. For population characteristics each adult member of the household responded for himself in the reinterview. The reinterviews were conducted by the permanent staff

of interviewers who work on the Bureau's Current Population Survey, after special training on reinterview techniques.

For 10 of these 15 characteristics the inconsistency of responses between the census and reinterview was low, the indices all falling below 20. The other five characteristics fell in the moderately inconsistent range with indices estimated between 21 and 45 (Table 2). The index for Number of Rooms appears to be particularly high relative to the other characteristics. Again, this partially reflects the detail in which these data were collected - one room intervals. We found that in about 80 percent of the cases with response differences, the census and reinterview differed by only one room. Collapsing this distribution to 5 categories, which was a published census distribution, the index is estimated at 38.

The consistency of reporting of housing characteristics by occupancy status, tenure, and size of structure generally followed the pattern discussed earlier; that is, owners were more consistent than renters, the reports for occupied units were more consistent than those for vacants, and data for single-units were more consistently reported than data for multi-unit structures. There was one notable exception - reports for Number of Rooms - where the owner/renter and single/multi-unit relationships were reversed. Considering the definitional problems associated with the number of rooms question, this reversal of relationship seems reasonable. Single family homes are more likely than multi-unit structures to contain space for which there may be confusion about whether it qualifies as a room according to the census definitions; for example, utility rooms or basements (finished or unfinished), enclosed porches, knee-walled attic space, etc.

There was corresponding response error data available from the 1960 census for four of these fifteen characteristics (Table 3). For Number of Children Ever Born, the L-Fold index associated with the 1960 census data was 10, indicating about the same level of inconsistency in both censuses. For Number of Bedrooms, the 1970 data had more consistency in reporting than in 1960. The 1970 L-Fold index was estimated at 18 and the 1960 L-Fold index was estimated at 26. Piped Water was also more consistently reported in 1970 than in 1960 with L-Fold indices respectively of 18 and 35. Number of Rooms was less consistently reported in 1970 than in 1960. The 1960 L-Fold index was estimated at 35 while in 1970 it was estimated at 44.

For five of these 15 characteristics moderate sized biases in census distribution were estimated. For Number of Children Ever Born, the "none" category was overstated in the census; that is, more women have had children than the census data indicate (Table 4). A review of the detailed reinterview responses, which we plan to undertake, may shed light on the source of this bias.

In Citizenship reports for the foreign born, the "naturalized" category was estimated as overstated in the census, while the "alien" category was estimated as understated. This seems to be the logical direction of the bias. Detailed review of the reinterview data will show whether persons undergoing the naturalization process, but not having completed it, were confused about the delineation between the two categories.

For Number of Bedrooms, there seemed to be some confusion by census respondents between the "none" and "one bedroom" categories; the "none" category was estimated to be understated in the census, the "one bedroom" category overstated. Here we think we know the source of the bias. For this characteristic, a review of the reinterview data indicates this confusion, for the most part, was associated with one-room efficiency apartments. The census definitions consider such units to have no bedroom. In a fair number of cases these units were reported by the census respondent as having one bedroom. It should be noted that as a part of the computer edits, a one room unit was edited as having no bedroom. Thus, the response errors which occurred in the field phase of enumeration probably were corrected as a result of that edit. We plan to follow through on this to learn how the processing edit may have affected the published distribution.

For Heating Equipment, the reinterview data indicate that biases exist in the census distribution for most of the heating equipment categories. There is some evidence that a fair proportion of the response errors are due to the respondents lack of knowledge, especially among householders in multi-unit structures. Errors in reporting heating equipment, as well as in other structural characteristics, may be reduced by collecting the data from a structure respondent (e.g. building manager, janitor, etc.). Data collected on the reinterview, but not yet analyzed, will indicate to what extent a structure respondent can improve response accuracy for these characteristics.

The response error data on Year Moved Into House reflects, in addition to respondent errors, a difference in the way the responses were recorded and edited in the two interviews. In the census, several year intervals were presented as possible answer categories with a final category of "always lived in this house or apartment." In the reinterview the specific year moved in was obtained and these were then coded to the appropriate category. In the edit of the reinterview responses, a child born after his parents had moved into a house had his response edited to the "always lived in the same house" category. In the census many of these children were reported as moving into the house in their birth year, rather than being reported in the "always lived in same house" category. The census responses for these children were not edited to the "always lived in same house" category during processing. The reinterview estimate of the percent classified in the "always lived in same house" category is on the order of 7 to 9 percentage points higher than the corresponding census percent (Table 5). Most of the understatement in the census is concentrated in the son or daughter of head population group. These data suggest that the consistency check, if applied to census responses, would appreciably improve the accuracy of the published census distribution.

Reinterview For New Population Questions

A third type of reinterview analysis was made a part of the 1970 reinterview study. Data for Mother Tongue (for the total population), Spanish Origin or Descent, and Vocational Training, were obtained for the first time in the 1970 Census. These questions or concepts were not precisely defined in the census and, as a result, the questions were subject to broad ranges of interpretation. For these characteristics, the development of a "correct" response in the reinterview did not appear to be fruitful. Thus, the reinterview focused on collecting detailed data to provide insights into how respondents interpreted the census questions.

Mother Tongue

The census question from which data on Mother Tongue were derived inquired as to "What language other than English was spoken in the person's home when he was a child?" Five answer categories of "Spanish", "French," "German," "Other foreign" (with a write in entry for the specific language) and "English only," were provided in the body of the question. In the reinterview, each person was questioned in detail to determine if any foreign language was used in his childhood home. For persons reporting use of a foreign language in the childhood home, the reinterview obtained data on the frequency of foreign language usage relative to English and on who spoke the language in the childhood home (i.e., person himself, parents, grandparents, etc.).

These reinterview data were used to stratify persons reporting a foreign language usage in their childhood home by the degree or intensity of usage. Table 6 provides the definition of each of the strata or levels and shows the distribution of census responses within each level. Level 1, for example, represents those persons for whom the foreign language was the only language used in the childhood home; English was reportedly not used in the childhood home. At the other extreme, level 7, the foreign language was not spoken by the sample person himself, but was spoken by other family members. Moving from level 1 to level 7, the reporting completeness in the census drops off from 97 percent to about 52 percent (Table 6, column 5).

Summary measures of response error, which result when alternative definitions of the population to be identified by the question are considered, are provided in Table 7. Using the broadest definition of Mother Tongue, levels 1 through 7, the proportion of persons reporting the use of a foreign language in the childhood home in the census is understated between 4 and 5 percentage points. The index of inconsistency for this reinterview definition is the lowest among the alternatives considered, estimated at 18. As the definition becomes more restrictive, the proportion of persons reporting use of a foreign language in their childhood home in the census exceeds the corresponding reinterview proportion. The index of inconsistency also increases as the definition becomes more restrictive. These data indicate that respondents tended to apply a literal interpretation to the census question, reporting any foreign language usage in the childhood home regardless of the intensity of usage.

Both errors in reporting use of a foreign language in the childhood home as well as errors in reporting the specific foreign language spoken affects the Mother Tongue distribution of languages. Table 8 shows that when both sources of error are considered, the census reports for 8 language categories were fairly consistent with those derived from the reinterview data, using levels 1 through 7 to identify persons having a foreign language spoken in their childhood home. The indices of inconsistency for these categories range from 11 to 24. The majority of the inconsistencies result from differences in reporting use of a foreign language in the childhood home, and, to a much lesser extent, the inconsistencies in reporting the specific foreign language spoken. Evidence of this is provided in Table 9 which shows the indices of inconsistency for the language categories which reflect only differences in reporting the specific foreign language spoken (i.e., these indices are based on persons reporting use of a foreign language in both the census and reinterview). These indices range from 1 to 8.

Spanish Origin or Descent

The census question on Spanish Origin or Descent asked if the person's origin or descent was "Mexican," "Puerto Rican", "Cuban," "Central or South American," "Other Spanish," or "none of these". The reinterview probed in detail about Spanish ancestry on either side of the family, and if such ancestry was detected, questioned about who the ancestors were (e.g., father, grandmother, great grandparent) and the country from which they came.

The census reports for persons of Spanish Origin or Descent were moderately inconsistent with those obtained from the detailed reinterview data; the index of inconsistency is estimated at 22 (Table 10). The proportion of persons reporting Spanish Origin or Descent in the census is slightly lower than the level estimated in the reinterview (-0.3 percent).

Analysis of the data for selected subgroups of the population indicate that major differences in the consistency of reporting exist, as might be expected, by Spanish surname, by nativity, by race, and by major geography.

A review of responses for persons of Spanish origin in the reinterview by their census responses indicates that reporting of Spanish origin in the census was correlated with (a) whether the origin was on both sides of the family or only on one side of the family, (b) which ancestors were from a Spanish speaking country, and (c) the country of Spanish origin (Table 11).

For example, of the persons with Spanish origin on both sides of the family according to the reinterview, about 97 percent reported themselves as of Spanish origin in the census, while only 21

percent of the persons with Spanish origin on only one side of the family reported themselves as of Spanish origin in the census. When the sample person himself was from a Spanish speaking country, Spanish Origin or Descent was almost always reported in the census (estimated at 97 percent). When the Spanish ancestry was a parent or grandparent, the reporting of Spanish origin in the census was about 80 percent (estimates of 83 and 73 percent). Less than 50 percent reported Spanish origin in the census when the ancestry was further back than grandparent. Reporting of Spanish origin in the census for persons with Spanish origin from Mexico, Cuba, Puerto Rico, or a Central or South American country was estimated at about 90 percent. On the other hand, only about 30 percent of the persons with Spanish origin from some other Spanish speaking country (mostly Spain) reported Spanish origin in the census.

We also observed that among persons not of Spanish origin according to the reinterview, a small number were reported as of Spanish origin in the census. In a third interview (reconciliation) with these persons we learned that these were primarily persons born in southern or midwestern states of the U.S. who had misinterpreted the "Central or South American" response category.

Vocational Training

The census question on vocational training had two parts. The first part asked the respondent, "Has this person ever completed a vocational training program?", while the second part asked the main field of vocational training for those persons reporting completion of a vocational training program. Some examples of programs which were to be reported as vocational were included in the body of the question and additional instructions were provided the respondents and enumerators as to types of programs which were not to be reported.

In the reinterview a detailed battery of questions was used to identify any training experience that might be considered as vocational. These detailed data also provided a basis for identifying persons having training experiences which were clearly not to be reported as vocational according to the census instructions.

The comparison of the reinterview and census responses shows a large number of response differences associated with the question. Of persons who had completed a vocational training program according to the reinterview (Table 12; 1415 cases), about 39 percent did not report completing a vocational training program on the census questionnaire (Table 12; 556 cases). Conversely, of persons who did not complete a vocational training program according to the reinterview (Table 12; 6152 cases), some 8½ percent reported completing a program in the census (Table 12; 524 cases). These relatively large numbers of response differences tended to offset each other so that, on a net basis, the proportion of persons reporting vocational training was estimated to be approximately the same in the reinterview and census. However, the consistency in reporting vocational training between the census and reinterview was quite low as evidenced by the estimated index of inconsistency of 47.

We expected that reporting completeness in the census would be higher for some types of training programs than others - for example, lengthy programs, training actually used on a job, etc. These expectations, for the most part, were not realized. The reinterview data (Table 13) indicate that the proportion of completed training programs classified as vocational in the reinterview which were reported as completed in the census, was not highly correlated with the field of training, the year the program was completed, where the training was received (i.e., trade or technical school, other type of school, not in school) man hours spent in the program, or the usefulness of the training (i.e., currently being used on the job, previously used on the job, or never used). In general, each of these categories was subject to substantial incompleteness of reporting in the census, although minor differences in reporting completeness were observed between some categories.

The reasons for persons erroneously reporting completion of a vocational training program in the census were fairly diverse but they seem to be related to a failure to follow instructions. The data given in Table 14 indicate that the major source of these errors included those resulting from persons reporting completion of a vocational training program when in fact, they had attended a program but had not completed it, and the reporting of academic training, on-the-job training, and training taken in a company school. These were types of training which the respondent and enumerators were specifically instructed not to report in the census.

For persons reporting completion of a vocational training program in both the census and reinterview, the census responses to the Field of Training questions were highly consistent with those obtained in the reinterview. The L-Fold index of inconsistency for the main field of training distribution was estimated at 9 (Table 15.)

1/ The data presented in this report describe the quality of responses recorded on the census questionnaires at the field stage of enumeration. They do not reflect the effect of errors, corrections or additions made during clerical and computer processing, in preparing the data for publication.

2/ The simple response variance is the between trial variation in responses averaged over all persons. Another component of response variance reflects the correlation of response deviations within trial. This component may be introduced by the enumerator, coder, crew leader, etc. The correlated component due to enumerators is being estimated in another study in the 1970 Evaluation Program.

3/ There were, naturally, some differences between the two surveys that may have affected the expected values of the responses, such as the use of well-trained permanent staff of

current survey interviewers to collect the reinterview data rather than temporary enumerators who collected the census data. On balance, however, we believe the reinterview more closely approximates a response variance type reinterview than a response bias type reinterview.

4/ U.S. Bureau of the Census. Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Effects of Different Reinterview Techniques on Estimates of Simple Response Variance, Series ER60, No. 11 U.S. Government Printing Office, Washington, D.C. 1972.

5/ In this type of reinterview, the index of inconsistency is best interpreted as approximately the complement of the correlation between the census and reinterview responses. The alternate interpretation, as the ratio of simple response variance to the total variance, is of questionable validity as the reinterview techniques used tend to introduce a downward bias in the estimate of simple response variance.

NOTICE

Due to space limitations, all the detailed tables on the handout cannot be reproduced here. Those tables essential for understanding the discussion in the text are reproduced below. A complete copy of the handout may be obtained by writing any of the authors at the following address: Statistical Methods Division, Bureau of the Census, Washington, D.C. 20233

HANDOUT

Analysis of response accuracy for a given characteristic between two trials is illustrated as follows: Each element (person or housing unit) is treated as distributed (0,1) in each trial. Responses for a given element in the two trials are compared and the element is placed in one of the internal cells of the 2X2 table. Comparison of responses for each element, over a sample of n elements, generates the entire 2X2 table. (For a characteristic with more than two categories the distribution is collapsed into a series of 2X2 tables, one for each category).

Reinterview Response (TRIAL 2)	Census Response (TRIAL 1)		
	In Category (1)	Not in Category (0)	Total
In Category (1)	a	b	np_2
Not in Category (0)	c	d	nq_2
Total	np_1	nq_1	n

The attached tables present summary measures of response errors for Population and Housing Characteristics as estimated from the reinterview studies. Two summary measures are presented. One describes the gross error while the second describes the net error in the distributions between the two surveys.

The measure of gross error - The Index of Inconsistency - is approximately the complement of the correlation between the census and reinterview responses. It is interpreted as the proportion of total population variance accounted for by simple response variance or as the ratio of the observed response differences between the two trials to the response differences that would be expected if there was no correlation between the two trials (i.e., if the two sets of responses were randomly associated). The index ranges from 0 to 100 and as a rule of thumb an estimated index between 0 and 20 indicates low inconsistency, between 20 and 50 moderate inconsistency, and above 50 high inconsistency. An index is estimated for each category of a distribution.

In the notation of the above table the index is estimated as: $\hat{I} = \left[\frac{\frac{b+c}{n}}{p_1q_2+p_2q_1}\right] X100$

In addition, a weighted average of the individual indices is estimated - referred to as the L-Fold Index of Inconsistency - which describes the amount of inconsistency in the entire distribution, and is interpreted in the same way as the indices for individual categories. (When there are only two categories in the distribution the L-Fold index is identical to the indices for the individual categories.)

In the notation given above the L-Fold index is estimated as: $\hat{I}_L = \sum_{i=1}^{L} \hat{I}_i \cdot \left[\frac{(p_{i1}q_{i2} + p_{i2}q_{i1})}{\sum_{i=1}^{L} (p_{i1}q_{i2} + p_{i2}q_{i1})}\right]$

The summary measure of net error - The Net Difference Rate - provides an estimate of the bias in the census distribution when the reinterview responses are assumed to be more accurate than the census responses. This rate is simply the difference between the reinterview and census estimates of the proportion of persons or housing units in a given category of the distribution.

In the notation of the above table the net difference rate is estimated as: $NDR = (p_1-p_2) X100$

TABLE 1 - L-Fold Index of Inconsistency for Population And Housing Characteristics Estimated From a Response Variance Type Reinterview, 1970 Census

(All estimates have been multiplied by 100.)

Characteristic (No. of Categories in Distribution)	L-Fold Index of Inconsistency @ (1)	95-Percent Confidence Interval for L-Fold Index (2)
Heating Fuel, Occupied Units (7)	12	10.2 to 13.3
Owner Occupied	8	6.8 to 10.0
Renter Occupied	19	15.8 to 22.4
Units in Single Unit Structures	9	7.5 to 10.7
Units in Multi-Unit Structures	20	16.6 to 24.6
Renters Paying Extra for Electricity, All Units (2)	15	11.7 to 18.5
Renters Paying Extra for Gas, All Units (2)	18	15.1 to 21.9
Bathtub or Shower Facilities, All Units (3)	18	15.3 to 21.3
Occupied Units	17	14.3 to 20.7
Vacant Units	24	16.9 to 34.2
Units in Single Unit Structures	15	11.8 to 18.1
Units in Multi-Unit Structures	29	21.4 to 39.1
Flush Toilet Facilities, All Units (3)	18	15.4 to 21.7
Occupied Units	16	13.4 to 20.0
Vacant Units	30	21.2 to 41.1
Units in Single Unit Structures	14	11.4 to 18.0
Units in Multi-Unit Structures	31	22.3 to 42.2
Telephone Availability, Occupied Units (2)	24	21.1 to 26.6
Owner Occupied	22	18.8 to 26.8
Renter Occupied	26	22.5 to 30.6
Year Structure Built, All Units (6)	25	24.0 to 26.5
Owner Occupied	22	20.2 to 23.0
Renter Occupied	36	33.0 to 38.6
Occupied Units in Single Unit Structures	25	23.1 to 26.0
Occupied Units in Multi-Unit Structures	29	25.8 to 31.9
Vacant Units	50	40.8 to 63.1
Vacancy Status (2)	31	24.4 to 39.3
Renters Paying Extra for Water, All Units (2)	39	31.4 to 49.7
Renters Paying Extra for Other Fuels, All Units (2)	45	34.1 to 59.3
Value of Home, Occupied Units (11)	58	56.5 to 60.2

@ The level of the L-fold index is sensitive to the detail of the classification system. For example if Value of Home data were collected in $100 class intervals we would expect to observe many more response diffences between trials, and to obtain a higher estimated L-fold index, than if the data were collected in $10,000 class intervals. The indices shown here were estimated at the detail to which data were collected in the census. These indices would not apply to published distributions where the detail data, as collected, were collapsed to broader categories.

TABLE 2 - L-Fold Index of Inconsistency for Population And Housing Characteristics Estimated From a Response Bias Type of Reinterview, 1970 Census

(All estimates have been multiplied by 100.)

Characteristic (No. of Categories in Distribution)	L-Fold Index of Inconsistency @ (1)	95-Percent Confidence Interval for L-Fold Index (2)
Nativity (2)	1	0.8 to 2.2
Nativity-Father(2)	3	2.1 to 3.4
Nativity-Mother(2)	3	2.1 to 3.4
Tenure, Occupied Units(4)	4	3.6 to 5.1
Contract Rent, Occupied Units(14)	11	9.5 to 12.8
Units in Single Unit Structures	9	6.8 to 11.7
Units in Multi-Unit Structures	12	10.5 to 14.6
Citizenship; Foreign Born (3)	12	8.6 to 17.2
Number of Children Ever Born*(13)	12	11.2 to 13.2
Year Came to U.S. To Stay, Foreign Born(9)	13	10.4 to 17.4
Number of Units in Structure, All Units (10)	15	13.5 to 15.9
Number of Bedrooms, All Units (6)	18	17.0 to 20.1
Occupied Units	18	16.7 to 19.8
Vacant Units	33	23.1 to 48.8
Piped Water, All Units (3)	21	18.2 to 25.0
Occupied Units	18	14.7 to 21.5
Vacant Units	41	31.7 to 53.5
Units in Single Unit Structures	20	17.0 to 24.1
Units in Multi-Unit Structures	41	23.6 to 70.7
Year Moved Into House (6)	25	24.6 to 26.5
Kitchen Facilities, Occupied Units (3)	25	20.4 to 30.6
Units in Single Unit Structures	21	16.7 to 26.8
Units in Multi-Unit Structures	46	30.6 to 68.9
Heating Equipment, All Units (8)	27	26.0 to 28.5
Occupied Units	27	25.3 to 27.9
Vacant Units	53	45.0 to 64.7
Units in Single Unit Structures	25	23.8 to 26.7
Units in Multi-Unit Structures	35	32.0 to 37.7
Owner Occupied	25	23.1 to 26.3
Renter Occupied	32	29.6 to 34.1
Number of Rooms, All Units (9)	45	44.1 to 46.8
Owner Occupied	51	49.6 to 53.0
Renter Occupied	37	35.0 to 39.6
Units in Single Unit Structures	50	48.6 to 51.9
Units in Multi-Unit Structures	34	31.7 to 37.1
Occupied Units	45	43.9 to 46.6
Vacant Units	55	47.4 to 65.0

* Based on reports for ever married females 14 to 64 years old in both Census and Reinterview.

@ The level of the L-fold index is sensitive to the detail of the classification system. For example if Value of Home data were collected in $100 class intervals we would expect to observe many more response differences between trials, and to obtain a higher estimated L-fold index, than if the data were collected in $10,000 class intervals. The indices shown here were estimated at the detail to which data were collected in the census. These indices would not apply to published distributions where the detail data, as collected, were collapsed to broader categories.

TABLE 6 - Comparison of Census and Reinterview Responses For Foreign Language Usage in Childhood Home, 1970 Census

(Data shown as numbers of sample persons reinterviewed and matched to census questionnaires, not as inflated estimates.)

Reinterview Classification	Census Response					
	Number			Percent Distribution		
	Total Persons (1)	Foreign Language Spoken in Childhood Home (2)	English Only Spoken in Childhood home (3)	Total Persons (4)	Foreign Language Spoken in Childhood Home (5)	English Only Spoken in Childhood home (6)
Total Persons	11,102	1,655	9,447	100.0	14.9	85.1
Foreign Language Spoken	2,170	1,627	543	100.0	75.0	25.0
Spoken by Person Himself	1,383	1,216	167	100.0	87.9	12.1
Foreign Language Only Spoken (Level 1)	412	399	13	100.0	96.8	3.2
Foreign Language Predominant, English Also Spoken (Level 2)	445	410	35	100.0	92.1	7.9
Foreign Language Spoken Equally With English (Level 3)	41	38	3	100.0	92.7	7.3
English Predominant, Foreign Language Spoken Frequently 1/ (Level 4)	350	276	74	100.0	78.9	21.1
English Predominant, Foreign Language Spoken Occasionally 2/ (Level 5)	95	72	23	100.0	75.8	24.2
English Predominant, Foreign Language Spoken Seldom 3/ (Level 6)	40	21	19	100.0	52.5	47.5
Not Spoken by Sample Person But Spoken by Other Family Members (Level 7)	787	411	376	100.0	52.2	47.8
English Only Spoken	8,932	28	8,904	100.0	0.3	99.7

1/ For example, spoken daily in the home.

2/ For example, spoken when relatives visited or to keep outsiders from understanding conversation.

3/ For example, used for slang, phrases, expressions.

TABLE 10 - Summary Measures of Response Error for Reporting Spanish Origin or Descent, by Selected Characteristics, 1970 Census

(All estimates have been multiplied by 100)

Characteristic	Index of Inconsistency (1)	95 Percent Confidence Interval for Index of Inconsistency (2)	Percent in Class in Reinterview (3)	Net Difference Rate @ (4)	95 Percent Confidence Interval for Net Difference Rate (5)
Total	22	18.6 to 25.7	4.0	-0.3	-0.6 to -0.1
Spanish Surname	11	7.3 to 17.0	56.9	-3.5	-5.9 to -1.1
No Spanish Surname	49	41.1 to 58.5	1.6	-0.2	-0.4 to -0.1
Native born	27	22.9 to 32.1	3.3	-0.4	-0.6 to -0.1
Foreign born	5	2.5 to 10.9	17.2	0.2*	-1.0 to 1.4
Age 0 to 19	25	20.0 to 31.9	5.1	-1.0	-1.5 to -0.4
Age 20 to 44	17	12.4 to 23.2	4.5	0.0*	-0.5 to 0.4
Age 45 or older	23	16.8 to 32.8	2.4	0.1*	-0.3 to 0.5
Male	21	16.8 to 26.8	4.2	-0.3*	-0.7 to 0.1
Female	22	17.8 to 28.2	3.8	-0.4*	-0.7 to 0.0
Son or Daughter of Head	25	19.6 to 32.2	4.8	-1.0	-1.6 to -0.5
Not Son or Daughter of Head	20	16.2 to 24.9	3.6	0.1*	-0.3 to 0.4
White	20	16.6 to 24.0	4.0	-0.4	-0.7 to -0.1
Negro	88	59.7 to 100.0	1.4	0.7*	-0.5 to 1.9
Other Races	11	4.2 to 28.4	17.9	-3.0*	-7.8 to 0.4
1/ Southwest	11	8.4 to 15.6	14.6	-2.0	-2.9 to -1.2
East	28	21.4 to 37.5	2.8	-0.1*	-0.5 to 0.3
Midwest	35	22.0 to 56.0	1.0	0.9	0.4 to 1.4
Balance of U.S.	69	50.3 to 95.2	1.4	-0.4*	-1.0 to 0.1
Conventional census areas	28	21.1 to 37.7	2.6	-0.1*	-0.5 to 0.3
Mail census areas	20	16.3 to 24.2	4.7	-0.4	-0.8 to -0.1

* Indicates net difference rate is not significantly different from zero at the 95 percent confidence level.

@ Difference between census and reinterview estimates of percent in class. Netative estimate indicates census less than reinterview; positive estimate indicates census larger than reinterview.

1/ Southwest: Arizona, California, Colorado, New Mexico, Texas
East: Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, Pennsylvania, Delaware, Maryland, Washington,D.C., Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida
Midwest: Illinois, Indiana, Ohio, Michigan
Balance of U.S.: All other states not mentioned above.

TABLE 19 - Evaluation of Census Report By Type of Vocational Training Program Completed, 1970 Census

(Data shown as sample counts of persons 14 years and older who, according to the reinterview had completed any training program which might be considered vocational, other than those which were not to be reported according to the census instructions.)

Category	Total Persons	Reported as Completed in the Census		95 Percent Confidence Interval For Percent
		Number	Percent	
Persons classified as completing vocational training program on the basis of the reinterview, total.....	1,415	859	61	58.4 to 63.6
Field of Training				
Business.....	515	305	59	54.6 to 63.4
Nursing.....	183	114	62	55.0 to 69.0
Trades.....	551	339	62	57.8 to 66.2
Engineer.....	68	42	62	50.2 to 73.8
Agriculture.....	43	25	58	43.0 to 73.0
Other.....	15	11	73	50.2 to 95.8
Field not reported	37	-	-	-
Year Program Completed				
1969 or later.....	89	52	58	47.6 to 68.4
1965-1968.....	252	139	55	48.8 to 61.2
1960-1964.....	170	111	65	57.8 to 72.2
1950-1959.....	295	185	63	57.4 to 68.6
1940-1949.....	307	198	64	58.6 to 69.4
1939 or earlier.....	233	141	60	53.6 to 66.4
Year not reported.....	69	-	-	-
Where or How Training Received				
Trade or technical school	636	425	67	63.2 to 70.8
Other school or not in school	682	372	54	50.2 to 57.8
High School.....	249	114	46	39.6 to 52.4
College.....	147	79	54	45.8 to 62.2
Other School.....	127	81	64	55.6 to 72.4
Not in school 1/....	159	98	62	54.2 to 69.8
Where or how received not reported	97	-	-	-
Man Hours in Program				
Less than 100 Man Hours	188	80	42	34.8 to 49.2
Under 25 Man-Hours	40	12	30	15.6 to 44.4
25-99 Man-Hours....	148	68	46	37.8 to 54.2
100 or More Man Hours...	1,119	715	64	61.2 to 66.8
100-249 Man-Hours..	176	86	49	41.6 to 56.4
250-499 Man-Hours..	137	73	53	44.6 to 61.4
500-999 Man-Hours..	209	134	64	57.4 to 70.6
1,000-1,999 Man-Hours	312	210	67	61.8 to 72.2
2,000 or More Man-Hours	285	212	74	69.2 to 79.6
Man Hours Not Reported	108	-	-	-
Usefulness of Training 2/				
Used in Current Job.....	618	410	66	62.2 to 69.8
Previously Used or Never Used in Job	779	438	56	52.4 to 59.6
Previously Used.....	504	295	58	53.6 to 62.4
Never Used, training sufficient to qualify for civilian job in that field.	223	121	54	47.4 to 60.6
Never Used, training not sufficient to qualify for civilian job in that field.	52	22	42	28.4 to 55.6
Usefulness not reported	18	-	-	-

1/ Includes training received through the Job Corps, or an apprenticeship, and military training which could be used in a civilian job.

2/ Respondent's assessment