Effects of Coders

Introduction and Summary

This report, one of a series of evaluation and research reports providing measures of the quality of the 1960 Censuses of Population and Housing, is concerned with the contribution of census coders to the total mean-square error of census statistics.

Project A of the 1960 Evaluation and Research Program was composed of three studies to measure the variance in census statistics introduced by interviewers, crew leaders, respondents, and coders. The largest study in this group was Response Variance Study I to measure the response variance due to interviewers and crew leaders. Results of this study were reported in Evaluation and Research Program report, Series ER 60 No. 7, *Effects of Interviewers and Crew Leaders* [6]. Project A also contained Response Variance Study II, a study to measure the contribution of respondents to census statistics. Data from that study will be in a later report. The third study in Project A was Response Variance Study III, a study to measure the effect of census coders on census statistics. This latter study is the subject of this report. It will be referred to as the "coder variance study."

The method used to estimate the between-coder component of the mean-square error was to compare the work of pairs of coders, each of whom had been randomly assigned to work on the schedules for a predetermined geographic area. An estimate was made for each of these areas based on the comparison of the two coders. These estimates were averaged over all pairs of coders in the study so that averages of between-coder variances for different items are presented for assignments of about 4,200 persons. The between-coder variance reflects the correlation among deviations in coded entries within a single coder's work. This term contains the impact of the coder's "personal equation"—his tendency to neglect to code difficult items and leave blanks, his tendency to code certain entries to a common occupation or industry, his possible misunderstanding of the training on how to code certain items, etc.

We were also able to estimate a term which reflects the basic trial-to-trial variability in coding over all units. In other contexts, this term has been called the "simple response variance." In order that readers be able to relate this term to the simple response variance mentioned in other Evaluation and Research reports, we shall call this term the "simple coding variance." In a coding context, it means the variability that one would get if an entry for the same unit were coded many times and then the results averaged over all units.

The findings of the study relate to both the correlated component of the coding variance and the simple coding variance. They are:

1. The correlated component of coding variance is very small in comparison with the correlated component of response variance due to enumerators.

2. The simple coding variance is about one-fourth to one-sixth the size of the simple response variance. This result is based on comparing the data with reinterview studies based on census data. However, the simple coding variances for industry items are about one-half the size of the simple response variances derived from a reinterview after the Current Population Survey interview.

These data tell little about the error that arises from errors or misunderstandings that all coders have in common. They deal only with the variability among coders. The rigorous quality control of the coding operation probably reduced this variability considerably.

The data reported here relate only to census items, some of which are more difficult to code than other census items. For example, place of birth is probably easier to code than occupation. Researchers working in medical coding, coding of attitudes, or other fields which may involve coding more difficult items may find different results. Also, the data relate only to items coded from census data collected in areas where a two-stage census was conducted (areas which contained about 80 percent of the country's population), and only for data collected from persons living in regular housing units. Persons living in group quarters, such as jails, dormitories, hospitals, etc., were not included in the study. However, the data collected were the same and the coding procedures identical.

Background of the Coder Variance Study

At the time of the 1960 census, there were about 54 million housing units and 180 million persons in the United States. Persons in 25 percent of all housing units were asked to fill out a long census questionnaire which contained questions on mobility, birthplace, education, number of children, employment status, occupation, industry, and veteran status, in addition to the basic census items such as age, relationship, etc. Many of these items were coded by the census enumerator directly onto the enumeration schedule. However, more complex items were left for coding by trained census coders.

The items to be coded were divided into two groups. A group of general coders entered codes for the following items:

Relationship to head of household of nonrelatives and "other relatives"

Family and household composition

Type of institution (for group quarters)

Entries for "other races" and for Spanish surnames in five Southwestern States

Place of birth

Mother tongue

Parents' place of birth

Residence 5 years prior to the census

Place of work

Income

The second group of items related to occupation and industry entries. These were coded by a specialized group of industry and occupation (I & O) coders.

The I & O coder first inspected the age of the sample person, since he entered codes only for persons of age 14 years and over. He then looked at a question which would indicate whether the sample person was in the labor force now or had been in the labor force during the last 10 years. Respondents had been asked to give the occupation and industry of their last job if they had been employed within the last 10 years. If the sample person had been employed within the last 10 years, the I & O coder was to enter occupation and industry codes on the census schedule.

In addition to coding instructions, the coder had indexes of occupations and industries and lists of company names. A Company Name List was prepared for each county from the returns of the 1958 Economic Censuses. On the list were the names and industry codes for all manufacturing establishments employing at least 50 people and all businesses employing at least 100 people. If a coder found on the Company Name List the company name which was written on the census schedule, he could use the code shown on the list. Otherwise, the coder had to make a more complicated determination of the code from the written descriptions of the industry and the occupation.

The peak strength of the general coding unit was 335 clerks in August 1960 and that of the industry and occupation unit was 365 clerks in September 1960. General coding began late in May 1960 and was completed in April 1961. On the average, each general coder made entries for about 1,300 persons a day. I & O coding began late in May 1960 as the schedules were released from the general coding unit, and was completed in April 1961. Each I & O coder made entries for about 1,300 persons a day. There was an elaborate quality control operation on both kinds of coding. This quality control operation has been described in [1] and [2].

The results on the between-coder variance and the simple coding variance are based on the work of a sample of both general and I & O coders, for a selected group of statistics. After the field enumeration was completed and all schedules were received in the processing office in Jeffersonville, Ind., a sample of enumeration books was selected from each of the 50 study areas in which the Response Variance Study took place. (For a description of the Response Variance Study, see [6].) These books were photocopied before they were sent through the coding process. As the regular census enumeration books were received in the general coding unit (they were processed by State), two general coders were selected at random from all available coders in the section. A restriction was that anyone who had been selected when another State was processed was not selected again. States which contained more than one of the 50 study areas also had more than two selections since each of the 50 study areas was to be processed by different coders. When the books were received in the I & O section, two I & O coders were selected at random from all the available coders in the section, with the same restrictions as for the general coders.

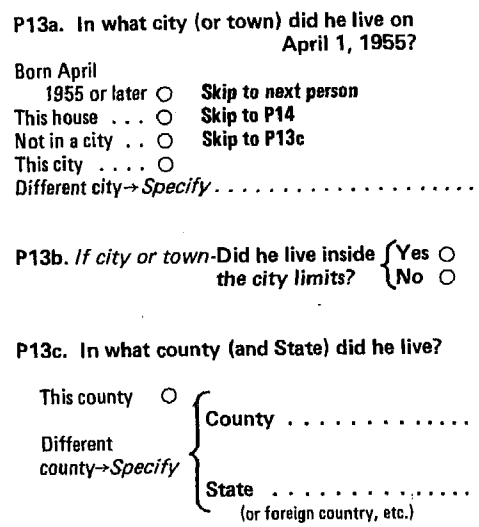
Thus, for each of 50 sets of enumeration books, there were two general coders and two I & O coders. These coders were called "transcoders" since their task was somewhat different from regular coders. The regular coders darkened appropriate circles signifying the code directly on the enumeration schedules. The transcoders were to work with photocopies of the enumeration books and instead of darkening the appropriate circle in the photocopied book, they were to determine the correct code and darken the appropriate circle in a blank enumeration book. Their task was somewhat more difficult than that of the regular census coders.

At the same time as the regular enumeration books were processed by the regular census coders, the photocopied books were processed by the transcoders. Each general transcoder was given a set of four or five photocopied enumeration district (ED) books. For each ED book, he then took a blank ED book and copied the ED number from the cover of the photocopy to the cover of the blank ED book. He then copied the page number for the first page in the photocopied book to the blank book and the age entries for each person in the household. He then determined the code for every written entry that he would have coded in the regular census coding and transcoded it into the blank book.

After the general transcoder finished his set of books, he turned them in and received the set the other general transcoder had finished. Thus, for every ED in the study, we had one original ED book, one photocopied ED book, and two "transcoded" books with only page number and age entries as well as the codes for each person in the ED.

The books then went to the I & O section, where each of the two I & O transcoders worked from the photocopied ED book and coded entries into the two "transcoded" books. They had to be very careful to make their entries for the correct person, matching on page number and age. It was easier to make an error in this transcoding work than on the regular coding assignment.

The transcoded books were different from the regular census coded books in several ways. Some of the items for which the enumerator did the coding were not coded in the transcoded books. For example, the question on mobility appeared as follows on the census schedule:



Thus, the enumerator filled the circles for every case except for those people who had lived in a different city, county, or State in 1955. Any circles the enumerator filled in P13a, b, or c would not show in the transcoded books. This, of course, complicated the tabulation and made the tabulation categories for the transcoded books different from the regular census categories. Thus, the categories for this item for the transcoded books are:

Blank

Different county, same State

Different State

Abroad

Included in the "blank" category are all those who had lived

in the same county in 1955 as well as all those the transcoder neglected to code.

Similar differences exist for other items. Included in a "not coded" category are those for whom an enumerator filled a code, those under 14 years of age who did not require a code, other persons who did not require a code, and those whom the transcoder neglected to give a code.

There was a quality control operation on the regular census coding to insure that the census coders were not making too many errors. Error rates for each coder were calculated weekly and those coders who made too many errors were released. Error rates for the transcoders were also calculated. The error rates for each transcoder were compared with his error rate on regular census coding for the 2-week period immediately preceding his transcoding work, with his error rate on regular census coding for the 2-week period immediately following his transcoding work, and with his error rate on regular census coding cumulated over 6 months. In addition, his error rate was compared with the error rate of the regular census coder for the time when he coded these same ED's.

It was found that in about 10 areas out of the 50, the transcoders were performing with higher error rates than their regular census work and than the regular census coder for the same ED's. Some possible reasons for this were:

1. Transcoding from a photocopied book to a different enumeration book was a harder task than that of the regular census coder.
2. The photocopies were hard to read.
3. The transcoders were not on production standards while they were doing the transcoding and so did not perform as well.

Because there was this difference, the estimates of "between-coder" variance may not adequately reflect the variance in the actual census process. It does reflect a "between-coder" variance, perhaps for a slightly harder task than the census coding, namely transcoding. However, it was impractical to replicate the census coding exactly, due to the fact that the census coder was instructed to code directly onto the enumeration schedule. If the second trial was to be an independent replication of the census process, the result of the first coding should not be apparent to the second coder. The final data are based on the matched records of 214,422 sample persons.

Estimation of Coding Variance

The general model of response error [3] decomposes the mean-square error of a statistic produced by a survey into components associated with various sources of error in the survey. We are here concerned with the variations introduced by differences in the ways in which a given coder would code a questionnaire entry at different times, and by differences in the ways in which two coders would code the same entry. If x denotes some statistic and Q denotes a fixed set of questionnaire entries as recorded by respondents and enumerators, the variance of x may be written according to a well-known theorem as

$$\text{Var}\,(x) = E\,\text{Var}\,(x|\,Q) + \text{Var}\,E(x\,|Q) \qquad (1)$$

In these expressions, the expectation operation E is understood to range over all possible sets Q of questionnaire entries that could have been presented to a coder.

The second term of (1) includes among its components the sampling variance, the simple response variance, the correlated response variance arising from differences among enumerators or crew leaders, and other components. Estimates of these have been given elsewhere. In particular, the very significant enumerator contribution has been described in considerable detail in report No. 7 of this series. The second term also includes other contributions which are related to coding (for example, the covariance between coding deviations and response deviations), but we do not attempt to estimate these here.

The first term of (1), which we may call the coding variance, is the component whose estimation is the object of this report. We estimate the magnitude of the quantity for various statistics, and compare the magnitude with the corresponding correlated component of response variance reported in ER 60, No. 7. We also decompose the coding variance into two components which we call the simple coding variance and the correlated component of coding variance, and provide estimates of these components.

Consider a fixed set of N lines on questionnaires which might be processed by any one of a universe of coders. Each line contains one person enumerated in the census. Let x_{jt} denote the value given to the j-th person at trial t by a random coder. If $\bar{x}_j$ denotes the average value of x_{jt} for all possible trials, we may write:

$$x_{jt} = (x_{jt} - \bar{x}_j) + \bar{x}_j$$

An estimate of the average value of x obtained by having a random coder process the whole set of questionnaires is

$$\bar{x}' = \frac{1}{N}\sum_{j}^{N}(x_{jt} - \bar{x}_j) + \frac{1}{N}\sum_{j}^{N}\bar{x}_j$$

Since the first of the two terms has a zero expectation and the second term is a constant, then

$$\begin{aligned}\text{Var}\,\bar{x}' &= E\left[\frac{1}{N}\sum_{j}^{N}(x_{jt} - \bar{x}_j)\right]^2 \\ &= \frac{1}{N^2}\left\{\sum_{j}^{N}E(x_{jt} - \bar{x}_j)^2 + \sum_{j\neq k}^{N}E(x_{jt} - \bar{x}_j)(x_{kt} - \bar{x}_k)\right\} \\ &= \frac{1}{N}\left\{\sigma_c^2 + (N-1)\,\rho_c\sigma_c^2\right\} \qquad (2)\end{aligned}$$

where we have defined

$$\sigma_c^2 = E(x_{jt} - \bar{x}_j)^2$$

and

$$\rho_c = \frac{E_{j\neq k}\,(x_{jt} - \bar{x}_j)(x_{kt} - \bar{x}_k)}{\sigma_c^2}$$

We call σ_c^2 the simple coding variance and $\rho_c\sigma_c^2$ the correlated coding variance. As can be seen from the definition, σ_c^2 is a measure of the variability in coding from trial to trial for a given unit, averaged over all units. The quantity ρ_c is the correlation between coding deviations for distinct units within the same coding trial.

We have here conducted an experiment in which two transcoders processed the same set of N persons. Consider the statistic

$$C = \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^2$$

where $\bar{x}_1$ and $\bar{x}_2$ are the means (per person) for the first and second transcoders respectively. Assuming the trials of the two transcoders to be independent,

$$E(C) = \text{Var}\,\bar{x} = \frac{\sigma_t^2}{N}\left\{1 + \rho_t (N - 1)\right\} \qquad (3)$$

where $\bar{x}$ denotes the mean estimated by a single coder and σ_t^2 refers to the simple transcoding variance.

Now consider the statistic

$$g = \frac{1}{N}\sum_{i=1}^{N}(x_{i_1} - x_{i_2})^2$$

where x_{i_1} and x_{i_2} denote the values assigned to the i-th person by the first and second transcoders respectively. It is easily shown that

$$E(g) = 2\sigma_t^2$$

assuming that the trials are independent, so that g/2 is an unbiased estimate of σ_t^2.

It follows that

$$\frac{N}{N-1}C - \frac{g}{2(N-1)} \qquad (4)$$

is an unbiased estimate of $\rho_t\sigma_t^2$. Thus we are able to estimate both σ_t^2 and $\rho_t\sigma_t^2$, for possible substitution into the total coding variance given by equation (2).

If we now consider that the first coder is a regular census coder and the second coder is a transcoder, we may no longer assume that their results have the same expectation or variance. We now find, still assuming independence, that

$$E(C) = \frac{1}{2}\left[\sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2 + B^2\right]$$

where $B = E(\bar{x}_1 - \bar{x}_2)$ and the subscripts refer to the first and second coders. Each of these variances may be written in the same form, with parameters of different value:

$$\sigma_{\bar{x}_1}^2 = \frac{\sigma_c^2}{N}\left\{1 + \rho_c(N - 1)\right\}$$

$$\sigma_{\bar{x}_2}^2 = \frac{\sigma_t^2}{N}\left\{1 + \rho_t(N - 1)\right\}$$

We also find that

$$E(g) = \sigma_c^2 + \sigma_t^2 + B^2$$

Let us now denote by C_{tt} and g_{tt} the statistics obtained from a comparison of the two transcoders, and by C_{tc} and g_{tc} the statistics derived from a comparison of the census coder with one of the transcoders. We find that

$$E(g_{tt} - g_{tc}) = 2\sigma_t^2 - (\sigma_t^2 + \sigma_c^2 + B^2)$$
$$= \sigma_t^2 - \sigma_c^2 - B^2$$

where σ_t^2 and σ_c^2 denote the simple coding variances for transcoders and census coders respectively. We also find that

$$E\left[\frac{N}{N-1}C_{tc} - \frac{g_{tc}}{2(N-1)}\right]$$

$$= \frac{N}{N-1}\cdot\frac{1}{2}\left[\frac{\sigma_t^2}{N}\left\{1 + \rho_t(N-1)\right\} + \frac{\sigma_c^2}{N}\left\{1 + \rho_c(N-1)\right\} + B^2\right]$$

$$- \frac{1}{2(N-1)}\left[\sigma_t^2 + \sigma_c^2 + B^2\right]$$

$$= \frac{1}{2}\left[\rho_t\sigma_t^2 + \rho_c\sigma_c^2 + B^2\right]$$

Clearly, we can now obtain estimates of B and σ_c^2. First we note that

$$E(\bar{x}_c - \bar{x}_t) = B$$

Hence

$$E\left\{g_{tc} - \frac{1}{2}g_{tt} - (\bar{x}_c - \bar{x}_t)^2\right\}$$

$$= E\left(g_{tc}\right) - \frac{1}{2}E\left(g_{tt}\right) - 2E\left(C_{tc}\right)$$

$$= \sigma_t^2 + \sigma_c^2 + B^2 - \frac{1}{2}\cdot 2\,\sigma_t^2 - \left\{\frac{\sigma_t^2}{N}\left[1 + \rho_t(N-1)\right]\right.$$

$$\left. + \frac{\sigma_c^2}{N}\left[1 + \rho_c(N-1)\right] + B^2\right\}$$

$$= \sigma_c^2 - \frac{\sigma_t^2}{N}\left[1 + \rho_t(N-1)\right] - \frac{\sigma_c^2}{N}\left[1 + \rho_c(N-1)\right]$$

$$= \frac{N-1}{N}(1 - \rho_c)\sigma_c^2 - \frac{\sigma_t^2}{N}\left[1 + \rho_t(N-1)\right]$$

so that

$$g_{tc} - \frac{1}{2}g_{tt} - (\bar{x}_c - \bar{x}_t)^2$$

is an underestimate of σ_c^2 by the small quantity shown above. We already have an estimate of

$$\frac{\sigma_t^2}{N}\left[1 + \rho_t(N-1)\right]$$

namely C_{tt}, so that we are led to construct the statistic

$$g_{tc} - \frac{1}{2}g_{tt} - 2C_{tc} + C_{tt}$$

whose expected value is

$$\frac{N-1}{N}(1 - \rho_c)\sigma_c^2$$

We shall therefore take

$$s_c^2 = g_{tc} - \frac{1}{2}g_{tt} - 2C_{tc} + C_{tt} \qquad (5)$$

to be an estimate of σ_c^2 with the understanding that it is a slight underestimate, on the average. In the use of this formula, whenever g_{tc} or C_{tc} are used, the average of the values from comparing each transcoder with the census coder is used.

In the presentation of estimates of simple response variances, we also show estimates of indexes of inconsistency, I. The index of inconsistency is defined as the ratio of the simple response variance to the sum of the simple response variance and the sampling variance of a given item. Thus, a value of I=.05 means that the simple response variance accounts for only about 5 percent of the total variance of the item.

As an estimator of the index of inconsistency, we use

$$I = \frac{s_c^2}{pq}$$

where p is the estimate of the proportion of the population in the given category and q=1-p. It has been shown by Hansen, Hurwitz, and Bershad [3] that pq, under certain general conditions, is the sum of the simple response variance and the sampling variance. If we are estimating I for the census coder, we use the estimates of pq from the census coder; if we are estimating I for the transcoders, we use the average of the estimates of the two transcoders.

By making use of the estimates for transcoders, we are also able to estimate the values of ρ_t, the intraclass correlation among coder deviations in the same transcoder assignment.

In addition, estimates of the bias of the transcoding experiment are presented. In most cases, the bias was negligible. The estimates of bias were weighted averages of the biases in each of the 50 areas made by averaging the difference between each transcoder and the census coder.

Estimates of the simple coding variance and the correlated component of the coding variance were made for each of the 50 study areas. The estimates shown in the tables in this report are weighted averages of these individual estimates, the weights reflecting the probability of selection of the areas.

Estimates of Simple Coder Variances and Indexes of Inconsistency

Tables 1 through 7 at the end of this report show the L x L tables for the two transcoders for the following characteristics:

Place of birth
Residence in 1955
Occupation
Industry
Wage and salary income
Income from own business
Income from other sources

Notice that the total number of persons shown in these tables is of the order of 138 million. This is considerably below the 1960 population of about 180 million since, as explained earlier, it excludes about 36 million people who lived in areas not covered by the two-stage census and about 5 million people who lived in group quarters—colleges, rooming houses, institutions, etc. Neither of these two latter groups were included in the Response Variance Study.

In the tables, the transcoder 1 classification is shown in the boxhead of the table and the transcoder 2 classification is shown in the stub of the table. However, this designation of "1" or "2" was purely arbitrary in each of the 50 areas. Thus, there were 2^{50} different ways these tables could have been formed by considering all possible ways of assigning "1" and "2". The tables shown are the averages of these 2^{50} possible arrangements. One must keep in mind when looking at tables 1 to 7 that the two transcoders worked from the same basic document so that the differences are due only to coding, not to reporting by the respondents or enumerators. Thus, we see in table 1 that of the 98,814,000 persons the first transcoder coded as having their birthplace in the same State as their 1960 residence, 263,000 were coded by the second transcoder as having their birthplace in a different State.

The income tables have fewer off-diagonal entries than the other tables. The occupation and industry tables seem to have the most off-diagonal entries. From table 3 on occupation, we see that persons who were coded as operatives and kindred workers by one transcoder were frequently coded as something else by the other transcoder. Similar findings can be observed in table 4 for construction workers, workers in nondurable manufacturing industries, retail trade workers, and others.

Table 8 shows the estimates of simple coder variance and indexes of inconsistency for both the transcoders based on tables 1 to 7, and for census coders, using equation (5). Thus, we see that s_c^2 for birthplace in the same State is 0.00539 while s_{tt}^2 for the same item is 0.00270. This is one of the few items for which the estimate for the census coder is much larger than for the transcoders. This is entirely due to the census coder not leaving as many cases uncoded as the transcoders did. The census coder neglected to code about 1.8 percent of all persons in the study, while the transcoders neglected to code about 2.3 percent of all persons. The transcoders may have left blank the cases for which they were not sure about the proper code. In most cases, the estimate of the simple coder variance for census coders is somewhat larger than that for the transcoders. This may be because the transcoders left blank the hardest cases. However, the estimates are, in most cases, very similar.

For only a few items was the estimate of bias of any substantial magnitude. Table 9 shows the estimates of bias. The largest biases estimated were for "not coded" items. In every one of these cases the sign of the bias was negative, indicating that the transcoders neglected to enter a code for more persons than did the census coder.

Estimates of simple response variances are available from reinterview and record check studies in the 1960 evaluation program. We should expect these estimates to be larger than the estimates based on coders, since they involve respondents, enumerators, and coders, thus giving more opportunities for different entries for the same person. For occupation items, we can compare estimates of the simple variance with estimates from three other sources. These are shown in table A.

The estimates from the CPS-Census match [4] and the Employer Record Check [5] are comparable, while those from the CPS-Reinterview Program [7] are much smaller. The reason for this is that in the CPS-Reinterview Program, the responses on the reinterview are compared with the original responses and the householders are asked to explain any discrepancies. This has the effect of biasing the estimate of simple response variance downward. Even though the estimates are subject to sampling error (see footnotes, table A) we may conclude that the simple coder variances are roughly one-sixth the size of the simple response variances for the CPS-Census Match and the Employer Record Check, and about one-third the size of those for the CPS-Reinterview Program.

Table B shows the estimates of simple coding and simple response variance for industry items. In this table, we do not have comparable data from the Employer Record Check. The estimates are subject to sampling error as shown in the footnotes. Since the standard errors are small, we conclude that the simple coder variances are roughly one-fourth the size of the simple response variances for industry items based on the CPS-Census Match and one-half of those from the CPS-Reinterview Study.

From an examination of tables A and B and a rough comparison of the magnitudes of the simple coding variances and simple response variances, we can conclude that data collection and coding of industry items is more accurate than data collection and coding of occupation items. Industry items seem to have smaller simple response variances and smaller simple coding variances than the occupation items. However, simple coding variance as a proportion of simple response variance is higher for industry than for occupation items.

Table A.—ESTIMATES OF SIMPLE CODING VARIANCE COMPARED WITH ESTIMATES OF SIMPLE RESPONSE VARIANCE FOR OCCUPATION ITEMS

Occupation	Simple coding variance[1] (1)	Simple response variance		
		CPS--Census Match[2] (2)	Employer Record Check[3] (3)	CPS--Reinterview Program[4] (4)
Professional, technical, etc.	0.0028	0.0129	0.0174	0.0090
Farmers, farm managers	0.0006	0.0071	([5])	0.0010
Managers, officials, proprietors	0.0033	0.0298	0.0219	0.0172
Clerical workers	0.0046	0.0192	0.0256	0.0122
Sales workers	0.0021	0.0151	0.0132	0.0070
Craftsmen, foremen, etc.	0.0045	0.0318	0.0297	0.0182
Operatives	0.0063	0.0339	0.0275	0.0189
Private household workers	0.0004	0.0022	([5])	0.0007
Service workers	0.0024	0.0109	0.0092	0.0053
Farm laborers	0.0009	0.0058	([5])	0.0014
Laborers	0.0024	0.0203	0.0153	0.0103

[1]The standard errors of these estimates range from approximately 0.00003 for the smallest estimate to 0.0002 for the largest estimate.

[2]These data were based on the CPS-Census Match in which data for about 22,000 persons in the Current Population Survey were matched to the census data for the same persons. The standard errors of these estimates range from 0.0004 for the smallest estimate to 0.0015 for the largest.

[3]These data were based on data for two-stage census areas only from the Employer Record Check. In this study data for about 1,600 persons in the census were matched to employers' records for the same persons. The standard errors of these estimates range from 0.0017 for the smallest estimate to 0.0030 for the largest.

[4]These data were based on quarterly averages over the 6-year period 1960 to 1966 of reconciled responses from a reinterview following the CPS. The standard errors of these estimates range from 0.00006 for the smallest estimate to 0.0001 for the largest.

[5]These groups were excluded from the Employer Record Check.

Table B.—ESTIMATES OF SIMPLE CODING VARIANCE COMPARED WITH ESTIMATES OF SIMPLE RESPONSE VARIANCE FOR INDUSTRY ITEMS

Industry	Simple coding variance[1] (1)	Simple response variance	
		CPS--Census Match[2] (2)	CPS--Reinterview Program[3] (3)
Agriculture, forestry, fisheries	0.0009	0.0067	0.0018
Mining	0.0004	0.0024	0.0012
Construction	0.0019	0.0101	0.0041
Manufacturing--durable goods	0.0039	0.0176	0.0075
Manufacturing--nondurable goods	0.0041	0.0165	0.0081
Transportation	0.0010	0.0045	([4])
Communication	0.0002	([4])	([4])
Utilities	0.0003	([4])	([4])
Wholesale trade	0.0038	0.0163	0.0081
Retail trade	0.0050	0.0237	0.0097
Finance, insurance, real estate	0.0007	0.0038	0.0016
Business and repairs	0.0022	0.0094	0.0051
Personal services	0.0016	0.0056	0.0023
Entertainment and recreation	0.0007	0.0014	0.0010
Professional services	0.0022	0.0078	0.0034
Public administration	0.0012	0.0048	0.0025

[1]The standard errors of these estimates range from 0.00001 for the smallest estimate to 0.0001 for the largest estimate.

[2]The standard errors of these estimates range from 0.0003 for the smallest estimate to 0.0013 for the largest estimate.

[3]The standard errors of these estimates range from 0.00007 for the smallest estimate to 0.0002 for the largest estimate.

[4]Estimates for these categories were not available from this study.

Estimates of the Total Coding Variance and the Correlated Component of Coding Variance

As shown earlier, the statistic C in equation (3) is an estimate of the total coding variance. We shall look at C_{tt} which is the coding variance for transcoders. By subtracting a multiple of g_{tt} from C_{tt}, as shown in equation (4), we get an estimate of the correlated component of coding variance.

Table 10 shows estimates of the relative coding variances for the transcoders, based on the average transcoding assignment of 4,288 sample persons. The coding coefficients of variation are generally very small. Columns (3) and (4) of table 10 show the magnitudes of the total coding variance and the correlated component of coding variance. From these data it is obvious that for many of the characteristics studied, the correlated component was the major part of the coding variance, often accounting for more than 80 percent of the total coding variance.

Column (7) of table 10 gives the estimates of ρ_t, the intraclass correlation among units within the same transcoder's assignment. The largest estimate of ρ_t was 0.0032 for "residence in 1955 in a different State from the 1960 residence." Even this small number, however, can increase the coding variance by a factor of 10. These estimates of ρ_t are especially small for income items.

A question of interest is how the size of the correlated component of coding variance compares with the correlated component of enumerator variance. Unfortunately, there are not many items for which we can make a direct comparison.

In ER 60 No. 7 of this series, the relative response variances due to enumerators were given for selected items for an area containing 3,900 persons in the assignment of a single enumerator. Since a coder's assignment typically included all of an enumerator's work, our estimates of the coding variance also apply to an enumerator's assignment. Comparisons of these correlated components are shown in table C.

The income items are not directly comparable, since the variances were computed separately for males and females in the enumerator study, and the class intervals are not always the same.

Table C shows that the coder effect, at least for occupation items, was usually less than one-seventh of the enumerator effect. However, for other items, such as mobility, the coder effect may be somewhat larger.

Limitations of the Data

There are several things one must keep in mind when looking at the data in this report. Some of these are as follows:

1. The estimates were produced from a sample of households and a sample of coders and thus are subject to sampling variability.

2. During the census processing, the data were edited by computer. Entries were imputed for missing items and inconsistent responses were resolved. The data on which these coder variances are based were obtained before these edits were performed. Thus, they do not reflect the coder effect on the published census statistics.

3. Unless a record existed for each of the transcoders as well as the census coder and a match of the three records could be made, the record was not included in the tabulation. Approximately 10 percent of all units originally included in the study were not included in the tabulations because the records could not be matched.

Table C.—COMPARISON OF CORRELATED COMPONENTS OF RELVARIANCES DUE TO CODERS AND ENUMERATORS, FOR A PLACE OF 3,900 PERSONS ENUMERATED BY A SINGLE ENUMERATOR AND CODED BY A SINGLE CODER

Characteristic	Relvariances		Coefficients of variation		Ratio of enumerator to coder relvariance
	Enumerator	Coder	Enumerator	Coder	
Residence in 1955[1]	0.00324	0.00067	0.057	0.003	4.8
Professional, technical workers	0.00216	0.00029	0.046	0.017	7.4
Farmers, farm managers	0.00868	0.00110	0.093	0.033	7.9
Clerical workers	0.00247	0.00025	0.050	0.012	9.9
Sales workers	0.00000	0.00018	0.000	0.014	0.0
Craftsmen, foremen	0.00397	0.00018	0.063	0.013	22.1
Operatives	0.00285	0.00013	0.053	0.011	21.9
Farm laborers	0.04845	0.00500	0.220	0.071	9.7

[1]The enumerator worked with three categories of "residence in 1955": Different county, same State; different State; or abroad. The coder worked with one--same State.

Table 1.—ESTIMATES OF DIFFERENCES IN U.S. CENSUS CODING FOR PLACE OF BIRTH

[Thousands of persons. Figures may not add to totals due to rounding]

Transcoder 2 coding classification	Transcoder 1 coding classification						
	Total persons	Birthplace not coded[1]	Birthplace coded				
			Total	Same State	Different State	U.S. possession	Abroad or at sea
Total persons	138,318	3,249	135,069	98,814	27,747	310	8,198
Birthplace not coded[1]	3,249	3,166	83	70	8	2	4
Birthplace coded	135,069	83	134,985	98,744	27,739	308	8,194
Same State	98,814	70	98,744	98,441	263	1	39
Different State	27,747	8	27,739	263	27,430	--	45
U.S. possession	310	2	308	1	--	307	--
Abroad or at sea	8,198	4	8,194	39	45	--	8,110

--Represents zero.

[1]These cases include many persons coded by enumerators as having been born in the same State.

Table 2.—ESTIMATES OF DIFFERENCES IN U.S. CENSUS CODING FOR RESIDENCE IN 1955 FOR PERSONS 5 YEARS OF AGE AND OVER

[Thousands of persons. Figures may not add to totals due to rounding]

Transcoder 2 coding classification	Transcoder 1 coding classification						
	Total persons	Persons under 5 years of age	Persons 5 years of age and over				
			Total	Residence not coded[1]	Different county, same State	Different State	Abroad
Total persons	138,318	16,322	121,996	96,574	14,537	9,765	1,120
Under 5 years of age	16,322	16,294	28	23	4	1	--
5 years and over	121,996	28	121,968	96,551	14,533	9,764	1,120
Residence not coded[1]	96,574	23	96,551	96,031	309	173	38
Different county, same State	14,537	4	14,533	309	14,136	87	1
Different State	9,765	1	9,764	173	87	9,502	2
Abroad	1,120	--	1,120	38	1	2	1,079

-- Represents zero.

[1]This category includes all those who lived in the same county in 1955 as in 1960, as well as the cases the transcoders neglected to code.

Table 3.—ESTIMATES OF DIFFERENCES IN U.S. CENSUS CODING FOR MAJOR

[Thousands of persons. Figures may

	Transcoder 2 coding classification	Transcoder 1 coding classification					
		Total persons	Persons under 14 years of age	Persons 14 years of age and over			
				Total	Occupation not coded[1]	Occupation coded	
						Total	Professional, technical
1	Total persons, all ages.....	138,318	42,602	95,716	29,268	66,448	6,515
2	All persons under 14 years of age.....	42,602	42,574	28	20	8	--
3	All persons 14 years of age and over.....	95,716	28	95,688	29,248	66,440	6,515
4	Occupation not coded[1].....	29,268	20	29,248	28,606	642	60
5	Occupation coded.....	66,448	8	66,440	642	65,798	6,455
6	Professional, technical, etc.....	6,515	--	6,515	60	6,455	6,224
7	Farmers, farm managers.....	1,729	2	1,727	11	1,716	--
8	Managers, officials, proprietors.....	5,007	--	5,007	49	4,958	35
9	Clerical workers.....	12,573	1	12,572	121	12,451	83
10	Sales workers.....	5,892	1	5,891	51	5,840	8
11	Craftsmen, foremen.....	8,714	--	8,714	63	8,651	41
12	Operatives.....	13,528	1	13,527	148	13,379	12
13	Private household workers.....	1,577	1	1,576	18	1,558	--
14	Service workers.....	6,390	--	6,390	62	6,328	50
15	Farm laborers.....	1,466	2	1,464	20	1,444	--
16	Laborers, except farm.....	3,057	--	3,057	39	3,018	2

-- Represents zero.

[1]Includes persons 14 years of age and over who were not in the labor force and who had no occupation, as well as persons transcoders neglected to code.

OCCUPATION GROUPS FOR PERSONS 14 YEARS OF AGE AND OVER

not add to totals due to rounding]

Transcoder 1 coding classification--Continued										
Persons 14 years of age and over--Continued										
Occupation coded--Continued										
Farmers, farm managers	Managers, officials, proprietors	Clerical workers	Sales workers	Craftsmen, foremen	Operatives	Private household workers	Service workers	Farm laborers	Laborers	
1,729	5,007	12,573	5,892	8,714	13,528	1,577	6,390	1,466	3,057	1
2	--	1	1	--	1	1	--	2	--	2
1,727	5,007	12,572	5,891	8,714	13,527	1,576	6,390	1,464	3,057	3
11	49	121	51	63	148	18	62	20	39	4
1,716	4,958	12,451	5,840	8,651	13,379	1,558	6,328	1,444	3,018	5
--	35	83	8	41	12	--	50	--	2	6
1,660	7	--	--	--	--	--	--	49	--	7
7	4,578	93	53	83	55	--	28	1	25	8
--	93	12,020	91	39	76	--	22	--	26	9
--	53	91	5,616	11	41	--	9	1	10	10
--	83	39	11	8,175	225	--	23	--	54	11
--	55	76	41	225	12,794	5	26	18	127	12
--	--	--	--	--	5	1,521	26	3	3	13
--	28	22	9	23	26	26	6,127	--	17	14
49	1	--	1	--	18	3	--	1,342	30	15
--	25	26	10	54	127	3	17	30	2,725	16

Table 4.—ESTIMATES OF DIFFERENCES IN U.S. CENSUS CODING FOR

[Thousands of persons. Figures may

	Transcoder 2 coding classification	Transcoder 1 coding classification							
		Total persons	Persons under 14 years of age	Persons 14 years of age and over					
				Total	Industry not coded[1]	Industry coded			
						Total	Agriculture, forestry, fisheries	Mining	Construction
1	Total persons, all ages...	138,318	42,602	95,716	28,383	67,333	3,555	484	3,818
2	All persons under 14 years of age....	42,602	42,574	28	20	8	4	--	--
3	All persons 14 years of age and over..	95,716	28	95,688	28,363	67,325	3,551	484	3,818
4	Industry not coded[1].....	28,383	20	28,363	27,915	448	23	2	29
5	Industry coded.....	67,333	8	67,325	448	66,877	3,528	482	3,789
6	Agriculture, forestry, fisheries	3,555	4	3,551	23	3,528	3,458	--	5
7	Mining.....	484	--	484	2	482	--	456	6
8	Construction.....	3,818	--	3,818	29	3,789	5	6	3,596
9	Manufacturing-durable goods.....	11,052	--	11,052	76	10,976	1	10	49
10	Manufacturing-nondurable goods..	9,447	1	9,447	53	9,593	11	3	11
11	Transportation.....	2,628	--	2,628	13	2,615	3	2	9
12	Communication.....	1,065	--	1,065	5	1,060	--	--	1
13	Utilities.....	816	--	816	2	814	--	--	11
14	Wholesale trade.....	2,448	--	2,448	16	2,432	12	1	12
15	Retail trade.....	11,895	1	11,895	70	11,824	12	3	30
16	Finance, insurance, real estate.	3,237	--	3,237	21	3,216	1	1	6
17	Business and repair services....	1,604	--	1,604	11	1,593	1	--	12
18	Personal services.....	3,997	2	3,995	31	3,964	13	--	2
19	Entertainment and recreation....	678	--	678	6	672	2	--	2
20	Professional services.....	7,621	--	7,621	64	7,557	4	1	15
21	Public administration.....	2,988	--	2,988	23	2,965	6	--	22

-- Represents zero.

[1]Includes persons 14 years of age and over who were not in the labor force and had no industry classification as well as persons transcoders neglected to code.

MAJOR INDUSTRY CLASSIFICATION FOR PERSONS 14 YEARS OF AGE AND OVER

not add to totals due to rounding]

Transcoder 1 coding classification--Continued													
Persons 14 years of age and over--Continued													
Industry coded--Continued													
Manufacturing-durables	Manufacturing-nondurables	Transportation	Communication	Utilities	Wholesale trade	Retail trade	Finance, insurance, real estate	Business and repair	Personal services	Entertainment and recreation	Professional services	Public administration	
11,052	9,447	2,628	1,065	816	2,448	11,895	3,237	1,604	3,997	678	7,621	2,988	1
--	1	--	--	--	--	1	--	--	2	--	--	--	2
11,052	9,446	2,628	1,065	816	2,448	11,894	3,237	1,604	3,995	678	7,621	2,988	3
76	53	13	5	2	16	70	21	11	31	6	64	23	4
10,976	9,393	2,615	1,060	814	2,432	11,824	3,216	1,593	3,964	672	7,557	2,965	5
1	11	3	--	--	12	12	1	1	13	2	4	6	6
10	3	2	--	--	1	3	1	--	--	--	1	--	7
49	11	9	1	11	12	30	6	12	2	2	15	22	8
10,623	88	10	5	4	63	55	2	39	2	1	16	8	9
88	8,966	12	2	2	122	114	4	28	9	3	12	6	10
10	12	2,517	1	2	28	10	2	4	1	2	2	10	11
5	2	1	1,037	2	--	4	1	4	2	1	--	--	12
4	2	2	2	788	1	--	--	--	--	--	1	4	13
63	122	28	--	1	2,047	113	4	20	2	1	5	1	14
55	114	10	4	--	113	11,324	9	65	30	14	35	6	15
2	4	2	1	--	4	9	3,137	15	6	--	12	15	16
39	28	4	4	--	20	65	15	1,356	10	3	32	5	17
2	9	1	2	--	2	30	6	10	3,834	8	40	4	18
1	3	2	1	--	1	14	--	3	8	619	11	5	19
16	12	2	--	1	5	35	12	32	40	11	7,331	40	20
8	6	10	--	4	1	6	15	5	4	5	40	2,832	21

Table 5.—ESTIMATES OF DIFFERENCES IN U.S. CENSUS CODING FOR

[Thousands of persons. Figures may

	Transcoder 2 coding classification	Transcoder 1 coding classification							
		Total persons	Persons under 14 years of age	Persons 14 years of age and over					
				Total	Income not coded[1]	Wage and salary income coded			
						Total	$1 to $499	$500 to $999	$1,000 to $1,499
1	Total persons, all ages.....	138,318	42,602	95,716	45,078	50,638	6,266	3,638	3,140
2	All persons under 14 years of age.....	42,602	42,574	28	20	8	4	4	--
3	All persons 14 years of age and over....	95,716	28	95,688	45,058	50,630	6,262	3,634	3,140
4	Wage and salary income not coded[1].....	45,078	20	45,058	44,582	476	85	53	44
5	Wage and salary income coded.....	50,638	8	50,630	476	50,154	6,177	3,581	3,096
6	$1 to $499.....	6,266	4	6,262	85	6,177	6,114	13	8
7	$500 to $999.....	3,638	4	3,634	53	3,581	13	3,541	1
8	$1,000 to $1,499.....	3,140	--	3,140	44	3,096	8	1	3,057
9	$1,500 to $1,999.....	2,379	--	2,379	29	2,350	4	4	9
10	$2,000 to $2,499.....	3,071	--	3,071	34	3,037	6	1	1
11	$2,500 to $2,999.....	2,697	--	2,697	27	2,670	6	--	1
12	$3,000 to $3,499.....	3,610	--	3,610	26	3,584	4	--	--
13	$3,500 to $3,999.....	3,195	--	3,195	24	3,171	6	--	--
14	$4,000 to $4,499.....	3,707	--	3,707	27	3,680	5	--	2
15	$4,500 to $4,999.....	3,015	--	3,015	20	2,995	4	1	1
16	$5,000 to $5,999.....	6,072	--	6,072	41	6,031	4	7	--
17	$6,000 to $6,999.....	3,776	--	3,776	23	3,753	--	5	6
18	$7,000 to $9,999.....	4,343	--	4,343	26	4,317	2	7	3
19	$10,000 to $14,999.....	1,259	--	1,259	11	1,248	--	--	6
20	$15,000 to $24,999.....	346	--	346	4	342	--	--	--
21	$25,000 and over.....	124	--	124	2	122	--	--	--

-- Represents zero.

[1]Includes persons 14 years of age and over who were coded by the census enumerator as having no wage and salary income, as well as persons the transcoder neglected to code.

WAGE AND SALARY INCOME OF PERSONS 14 YEARS OF AGE AND OVER

not add to totals due to rounding]

Transcoder 1 coding classification--Continued													
Persons 14 years of age and over--Continued													
Wage and salary income coded--Continued													
$1,500 to $1,999	$2,000 to $2,499	$2,500 to $2,999	$3,000 to $3,499	$3,500 to $3,999	$4,000 to $4,499	$4,500 to $4,999	$5,000 to $5,999	$6,000 to $6,999	$7,000 to $9,999	$10,000 to $14,999	$15,000 to $24,999	$25,000 and over	
2,379	3,071	2,697	3,610	3,195	3,707	3,015	6,072	3,776	4,373	1,259	346	124	1
--	--	--	--	--	--	--	--	--	--	--	--	--	2
2,379	3,071	2,697	3,610	3,195	3,707	3,015	6,072	3,776	4,373	1,259	346	124	3
29	34	27	26	24	27	20	41	23	26	11	4	2	4
2,350	3,037	2,670	3,584	3,171	3,680	2,995	6,031	3,753	4,317	1,248	342	122	5
4	6	6	4	6	5	4	4	--	2	--	--	--	6
4	1	--	--	--	--	1	7	5	7	--	--	--	7
9	1	1	--	--	2	1	--	6	3	6	--	--	8
2,314	--	3	1	2	2	3	1	4	3	--	--	--	9
--	3,001	11	6	1	--	--	1	1	9	--	--	--	10
3	11	2,642	--	4	--	--	1	1	2	--	--	--	11
1	6	--	3,556	6	1	--	6	--	3	--	--	--	12
2	1	4	6	3,142	2	--	4	--	3	--	--	--	13
2	--	--	1	2	3,642	12	4	--	5	4	--	--	14
3	--	--	--	--	12	2,974	--	--	1	--	--	--	15
1	1	1	6	4	4	--	5,993	6	4	--	--	--	16
4	1	1	--	--	--	--	6	3,727	2	--	--	--	17
3	9	2	3	3	5	1	4	2	4,274	--	--	--	18
--	--	--	--	--	4	--	--	--	--	1,236	2	2	19
--	--	--	--	--	--	--	--	--	--	2	340	1	20
--	--	--	--	--	--	--	--	--	--	2	1	119	21

Table 6.—ESTIMATES OF DIFFERENCES IN U.S. CENSUS CODING FOR

[Thousands of persons. Figures may

	Transcoder 2 coding classification	Transcoder 1 coding classification							
		Total persons	Persons under 14 years of age	Persons 14 years of age and over					
				Total	Income not coded[1]	Self-employment income coded			
						Total	$1 to $499 or loss	$500 to $999	$1,000 to $1,499
1	Total persons, all ages.....	138,318	42,602	95,716	88,553	7,163	1,452	816	636
2	All persons under 14 years of age.....	42,602	42,574	28	22	6	6	--	--
3	All persons 14 years of age and over....	95,716	28	95,688	88,535	7,157	1,446	816	636
4	Self-employment income not coded[1].....	88,553	22	88,535	88,401	134	30	16	13
5	Self-employment income coded.....	7,163	6	7,157	134	7,023	1,416	800	623
6	$1 to $499 or loss.....	1,452	6	1,446	30	1,416	1,400	7	--
7	$500 to $999.....	816	--	816	16	800	7	794	--
8	$1,000 to $1,499.....	636	--	636	13	623	--	--	620
9	$1,500 to $1,999.....	401	--	401	6	395	--	--	2
10	$2,000 to $2,499.....	443	--	443	8	435	2	--	1
11	$2,500 to $2,999.....	318	--	318	8	310	2	--	--
12	$3,000 to $3,499.....	403	--	403	10	393	1	--	--
13	$3,500 to $3,999.....	250	--	250	3	247	1	--	--
14	$4,000 to $4,499.....	360	--	360	4	354	1	--	--
15	$4,500 to $4,999.....	207	--	207	3	204	1	--	--
16	$5,000 to $5,999.....	486	--	486	10	476	--	--	--
17	$6,000 to $6,999.....	304	--	304	7	297	--	--	--
18	$7,000 to $9,999.....	448	--	448	7	441	--	--	--
19	$10,000 to $14,999.....	337	--	337	5	332	--	--	--
20	$15,000 to $24,999.....	205	--	205	1	204	--	--	--
21	$25,000 and over.....	97	--	97	2	95	--	--	--

-- Represents zero.

[1]Includes persons 14 years of age and over who were coded by the census enumerator as having no self-employment income, as well as persons the transcoder neglected to code.

SELF-EMPLOYMENT INCOME OF PERSONS 14 YEARS OF AGE AND OVER

not add to totals due to rounding]

Transcoder 1 coding classification--Continued													
Persons 14 years of age and over--Continued													
Self-employment income coded--Continued													
$1,500 to $1,999	$2,000 to $2,499	$2,500 to $2,999	$3,000 to $3,499	$3,500 to $3,999	$4,000 to $4,499	$4,500 to $4,999	$5,000 to $5,999	$6,000 to $6,999	$7,000 to $9,999	$10,000 to $14,999	$15,000 to $24,999	$25,000 and over	
401	443	318	403	250	360	207	486	304	448	337	205	97	1
--	--	--	--	--	--	--	--	--	--	--	--	--	2
401	443	318	403	250	360	207	486	304	448	337	205	97	3
6	8	8	10	3	4	3	10	7	7	5	1	2	4
395	435	310	393	247	354	204	476	297	441	332	204	95	5
--	2	2	1	1	1	1	--	--	--	--	--	--	6
--	--	--	--	--	--	--	--	--	--	--	--	--	7
2	1	--	--	--	--	--	--	--	--	--	--	--	8
392	--	2	--	--	--	--	--	--	--	--	--	--	9
--	433	--	--	--	--	--	--	--	--	--	--	--	10
2	--	304	--	1	--	--	--	--	1	--	--	1	11
--	--	--	387	2	--	--	--	--	--	2	--	--	12
--	--	1	2	244	--	--	--	--	--	--	--	--	13
--	--	--	--	--	350	-2	1	--	--	--	--	--	14
--	--	--	--	--	2	202	--	--	--	--	--	--	15
--	--	--	--	--	1	--	475	--	--	--	--	--	16
--	--	--	--	--	--	--	--	295	--	--	--	1	17
--	--	1	--	--	--	--	--	--	438	1	--	--	18
--	--	--	2	--	--	--	--	--	1	327	1	--	19
--	--	--	--	--	--	--	--	--	--	1	200	2	20
--	--	1	--	--	--	--	--	1	--	--	2	91	21

Table 7.–ESTIMATES OF DIFFERENCES IN CENSUS CODING FOR

[Thousands of persons. Figures may

	Transcoder 2 coding classification	Transcoder 1 coding classification							
		Total persons	Persons under 14 years of age	Persons 14 years of age and over					
				Total	Other income not coded[1]	Other income coded			
						Total	$1 to $499	$500 to $999	$1,000 to $1,499
1	Total persons, all ages.....	138,318	42,602	95,716	71,141	24,575	11,218	6,432	3,029
2	All persons under 14 years of age.....	42,602	42,574	28	22	6	4	2	--
3	All persons 14 years of age and over....	95,716	28	95,688	71,119	24,569	11,214	6,430	3,029
4	Other income not coded[1].....	71,141	22	71,119	70,769	350	140	82	44
5	Other income coded.....	24,575	6	24,569	350	24,219	11,074	6,348	2,985
6	$1 to $499.....	11,218	4	11,214	140	11,074	11,001	33	12
7	$500 to $999.....	6,432	2	6,430	82	6,348	33	6,296	--
8	$1,000 to $1,499.....	3,029	--	3,029	44	2,985	12	--	2,952
9	$1,500 to $1,999.....	1,397	--	1,397	24	1,373	7	5	9
10	$2,000 to $2,499.....	871	--	871	13	858	4	1	2
11	$2,500 to $2,999.....	429	--	429	10	419	3	--	2
12	$3,000 to $3,499.....	346	--	346	12	334	2	1	2
13	$3,500 to $3,999.....	180	--	180	4	176	6	--	--
14	$4,000 to $4,499.....	166	--	166	5	161	2	--	2
15	$4,500 to $4,999.....	69	--	69	4	65	--	--	--
16	$5,000 to $5,999.....	124	--	124	7	117	2	3	1
17	$6,000 to $6,999.....	83	--	83	2	81	1	--	--
18	$7,000 to $9,999.....	111	--	111	3	108	1	7	1
19	$10,000 to $14,999.....	70	--	70	--	70	--	--	1
20	$15,000 to $24,999.....	28	--	28	--	28	--	--	--
21	$25,000 and over.....	22	--	22	--	22	--	1	--

-- Represents zero.

[1]Includes persons 14 years of age and over who were coded by the census enumerator as having no other income, as well as persons the transcoder neglected to code.

INCOME FROM OTHER SOURCES FOR PERSONS 14 YEARS OF AGE AND OVER

not add to totals due to rounding]

Transcoder 1 coding classification--Continued													
Persons 14 years of age and over--Continued													
Other income coded--Continued													
$1,500 to $1,999	$2,000 to $2,499	$2,500 to $2,999	$3,000 to $3,499	$3,500 to $3,999	$4,000 to $4,499	$4,500 to $4,999	$5,000 to $5,999	$6,000 to $6,999	$7,000 to $9,999	$10,000 to $14,999	$15,000 to $24,999	$25,000 and over	
1,397	871	429	346	180	166	69	124	83	111	70	28	22	1
--	--	--	--	--	--	--	--	--	--	--	--	--	2
1,397	871	429	346	180	166	69	124	83	111	70	28	22	3
24	13	10	12	4	5	4	7	2	3	--	--	--	4
1,373	858	419	334	176	161	65	117	81	108	70	28	22	5
7	4	3	2	6	2	--	2	1	1	--	--	--	6
5	1	--	1	--	--	--	3	--	7	--	--	1	7
9	2	2	2	--	2	--	1	--	1	1	--	--	8
1,352	--	--	--	--	--	--	--	--	--	--	--	--	9
--	844	1	1	--	--	--	2	--	2	--	--	--	10
--	1	412	--	--	--	--	--	--	--	--	--	--	11
--	1	--	329	--	--	--	--	--	--	--	--	--	12
--	--	--	--	168	2	--	--	--	--	--	--	--	13
--	--	--	--	2	155	--	--	--	--	--	--	--	14
--	--	--	--	--	--	65	--	--	--	--	--	--	15
--	2	--	--	--	--	--	110	--	--	--	--	--	16
--	--	--	--	--	--	--	--	80	--	--	--	--	17
--	2	--	--	--	--	--	--	--	97	--	--	--	18
--	--	--	--	--	--	--	--	--	--	69	--	--	19
--	--	--	--	--	--	--	--	--	--	--	28	--	20
--	--	--	--	--	--	--	--	--	--	--	--	21	21

Table 8.—ESTIMATES OF SIMPLE CODER VARIANCE AND INDEXES OF INCONSISTENCY FOR SELECTED ITEMS FOR CENSUS CODERS AND TRANSCODERS

Classification	Simple coder variance		Index of inconsistency	
	Census coder	Trans-coder	Census coder	Trans-coder
PLACE OF BIRTH				
Same State	0.0054	0.0027	0.0267	0.0132
Different State	0.0024	0.0023	0.0147	0.0145
U.S. possession	0.0001	0.0000	0.0416	0.0099
Abroad or at sea	0.0011	0.0006	0.0198	0.0114
RESIDENCE 5 YEARS AGO				
Blank[1]	0.0038	0.0039	0.0178	0.0186
Different county, same State	0.0030	0.0029	0.0320	0.0308
Different State	0.0013	0.0019	0.0202	0.0290
Abroad	0.0003	0.0003	0 0412	0.0369
OCCUPATION				
Professional, technical, etc.	0.0028	0.0021	0.0629	0.0468
Farmers, farm managers	0.0006	0.0005	0.0510	0.0404
Managers, officials, proprietors	0.0033	0.0031	0.0936	0.0889
Clerical	0.0046	0.0040	0.0552	0.0484
Sales workers	0.0021	0.0020	0.0526	0.0489
Craftsmen, foremen, etc.	0.0045	0.0039	0.0763	0.0660
Operatives	0.0063	0.0053	0.0710	0.0601
Private household workers	0.0004	0.0004	0.0357	0.0359
Service workers	0.0024	0.0019	0.0555	0.0431
Farm laborers	0.0009	0.0009	0.0948	0.0855
Laborers, except farm	0.0024	0.0024	0.1085	0.1111
INDUSTRY				
Agriculture, forestry, fisheries	0.0009	0.0007	0.0358	0.0364
Mining	0.0004	0.0002	0.1039	0.0580
Construction	0.0019	0.0016	0.0680	0.0598
Manufacturing--durables	0.0039	0.0031	0.0534	0.0422
Manufacturing--nondurables	0.0041	0.0035	0.0648	0.0546
Transportation	0.0010	0.0008	0.0549	0.0431
Communication	0.0002	0.0002	0.0207	0.0265
Utilities	0.0003	0.0002	0.0501	0.0345
Wholesale trade	0.0038	0.0029	0.2145	0.1667
Retail trade	0.0050	0.0041	0.0631	0.0525
Finance, insurance, real estate	0.0007	0.0007	0.0315	0.0316
Business and repair	0.0022	0.0018	0.1835	0.1564
Personal services	0.0016	0.0012	0.0549	0.0420
Entertainment and recreation	0.0007	0.0004	0.1421	0.0875
Professional services	0.0022	0.0021	0.0418	0.0403
Public administration	0.0012	0.0011	0.0570	0.0534
WAGE AND SALARY INCOME				
$1 to $499	0.0011	0.0011	0.0253	0.0254
$500 to $999	0.0008	0.0007	0.0306	0.0274
$1,000 to $1,499	0.0007	0.0006	0.0323	0.0270
$1,500 to $1,999	0.0006	0.0005	0.0315	0.0278
$2,000 to $2,499	0.0007	0.0005	0.0356	0.0233
$2,500 to $2,999	0.0006	0.0004	0.0425	0.0208
$3,000 to $3,499	0.0009	0.0004	0.0341	0.0154
$3,500 to $3,999	0.0007	0.0004	0.0300	0.0170
$4,000 to $4,499	0.0006	0.0005	0.0231	0.0180
$4,500 to $4,999	0.0005	0.0003	0.0262	0.0139
$5,000 to $5,999	0.0008	0.0006	0.0179	0.0136
$6,000 to $6,999	0.0006	0.0004	0.0209	0.0133
$7,000 to $9,999	0.0007	0.0005	0.0227	0.0164
$10,000 to $14,999	0.0002	0.0002	0.0221	0.0184
$15,000 to $24,999	0.0000	0.0000	0.0174	0.0174
$25,000 and over	0.0000	0.0000	0.0222	0.0402
SELF-EMPLOYMENT INCOME				
$1 to $499 or loss	0.0004	0.0004	0.0378	0.0365
$500 to $999	0.0002	0.0002	0.0282	0.0271
$1,000 to $1,499	0.0001	0.0001	0.0148	0.0253
$1,500 to $1,999	0.0001	0.0001	0.0257	0.0225
$2,000 to $2,499	0.0001	0.0001	0.0364	0.0227
$2,500 to $2,999	0.0001	0.0001	0.0352	0.0441
$3,000 to $3,499	0.0001	0.0001	0.0223	0.0400
$3,500 to $3,999	0.0000	0.0000	0.0218	0.0241
$4,000 to $4,499	0.0001	0.0001	0.0207	0.0279
$4,500 to $4,999	0.0000	0.0000	0.0148	0.0241
$5,000 to $5,999	0.0001	0.0001	0.0214	0.0228
$6,000 to $6,999	0.0000	0.0001	0.0222	0.0296
$7,000 to $9,999	0.0001	0.0001	0.0261	0.0227
$10,000 to $14,999	0.0001	0.0001	0.0342	0.0302
$15,000 to $24,999	0.0000	0.0000	0.0201	0.0193
$25,000 and over	0.0000	0.0000	0.0453	0.0413
OTHER INCOME				
$1 to $499	0.0017	0.0016	0.0230	0.0273
$500 to $999	0.0011	0.0010	0.0250	0.0222
$1,000 to $1,499	0.0004	0.0006	0.0191	0.0260
$1,500 to $1,999	0.0002	0.0003	0.0238	0.0325
$2,000 to $2,499	0.0002	0.0002	0.0254	0.0312
$2,500 to $2,999	0.0001	0.0001	0.0339	0.0398
$3,000 to $3,499	0.0001	0.0001	0.0405	0.0493
$3,500 to $3,999	0.0001	0.0001	0.0673	0.0668
$4,000 to $4,499	0.0001	0.0001	0.0557	0.0603
$4,500 to $4,999	0.0000	0.0000	0.0262	0.0723
$5,000 to $5,999	0.0000	0.0001	0.0426	0.1126
$6,000 to $6,999	0.0001	0.0000	0.0487	0.0241
$7,000 to $9,999	0.0001	0.0001	0.0965	0.1267
$10,000 to $14,999	0.0000	0.0000	0.0000	0.0181
$15,000 to $24,999	0.0000	0.0000	0.0000	0.0000
$25,000 and over	0.0001	0.0000	0.0449	0.0072

[1]Blank includes persons who were living in the same house, in a different house in the same county, and the persons the transcoders neglected to code.

Table 9.–ESTIMATES OF BIAS IN TRANSCODING EXPERIMENT CONTRASTED WITH CENSUS CODING EXPERIMENT

Classification	Bias[1]
PLACE OF BIRTH	
Not coded	-0.0044
Coded	0.0044
Same State	0.0028
Different State	0.0014
U.S. possession	0.0001
Abroad or at sea	0.0001
RESIDENCE IN 1955	
Not coded	-0.0004
Coded	0.0004
Different county, same State	-0.0001
Different State	0.0004
Abroad	0.0001
OCCUPATION	
Not coded	-0.0007
Coded	0.0007
Professional, technical	0.0002
Farmers, farm managers	-0.0001
Managers, officials, proprietors	-0.0002
Clerical workers	0.0003
Sales workers	-0.0001
Craftsmen, foremen	0.0002
Operatives	-0.0001
Private household workers	0.0001
Service workers	0.0001
Farm laborers	0.0001
Laborers	0.0003
INDUSTRY	
Not coded	-0.0011
Coded	0.0011
Agriculture, forestry, fisheries	-0.0002
Mining	0.0000
Construction	0.0004
Manufacturing--durable	0.0003
Manufacturing--nondurable	-0.0003
Transportation	0.0000
Communication	0.0000
Utilities	0.0000
Wholesale trade	0.0003
Retail trade	0.0003
Finance, insurance, real estate	0.0001
Business and repair	0.0000
Personal services	0.0000
Entertainment and recreation	0.0001
Professional services	0.0003
Public administration	-0.0001
WAGE AND SALARY INCOME	
Not coded	-0.0003
Coded	0.0003
$1 to $499	-0.0002
$500 to $999	-0.0001
WAGE AND SALARY INCOME--Continued	
Coded--Continued	
$1,000 to $1,499	0.0001
$1,500 to $1,999	0.0003
$2,000 to $2,499	0.0000
$2,500 to $2,999	0.0002
$3,000 to $3,499	-0.0002
$3,500 to $3,999	0.0001
$4,000 to $4,499	0.0000
$4,500 to $4,999	0.0001
$5,000 to $5,999	-0.0002
$6,000 to $6,999	-0.0001
$7,000 to $9,999	-0.0002
$10,000 to $14,999	0.0000
$15,000 to $24,999	0.0000
$25,000 and over	0.0000
SELF-EMPLOYMENT INCOME	
Not coded	-0.0001
Coded	0.0001
$1 to $499 or loss	-0.0001
$500 to $999	0.0000
$1,000 to $1,499	0.0000
$1,500 to $1,999	0.0000
$2,000 to $2,499	0.0000
$2,500 to $2,999	0.0000
$3,000 to $3,499	0.0000
$3,500 to $3,999	0.0000
$4,000 to $4,499	0.0000
$4,500 to $4,999	0.0000
$5,000 to $5,999	0.0000
$6,000 to $6,999	0.0000
$7,000 to $9,999	0.0000
$10,000 to $14,999	0.0000
$15,000 to $24,999	0.0000
$25,000 and over	0.0000
OTHER INCOME	
Not coded	-0.0006
Coded	0.0006
$1 to $499	0.0000
$500 to $999	0.0002
$1,000 to $1,499	0.0002
$1,500 to $1,999	0.0001
$2,000 to $2,499	0.0001
$2,500 to $2,999	0.0000
$3,000 to $3,499	0.0000
$3,500 to $3,999	0.0000
$4,000 to $4,499	0.0000
$4,500 to $4,999	0.0000
$5,000 to $5,999	0.0000
$6,000 to $6,999	0.0000
$7,000 to $9,999	0.0000
$10,000 to $14,999	0.0000
$15,000 to $24,999	0.0000
$25,000 and over	0.0000

[1]A negative estimate of bias means that the transcoders had more persons in the category than the census coder.

Table 10.—ESTIMATES OF TOTAL CODING RELVARIANCE, CORRELATED COMPONENT OF CODING RELVARIANCE, AND INTRACLASS CORRELATION OF DIFFERENT UNITS IN SAME CODER ASSIGNMENT (ρ_t) IN AN AVERAGE CODING ASSIGNMENT OF 4,288 PERSONS BY ONE CODER

Classification	Number of persons	Percent of total	Coder relvariance		Coefficient of variation		Intraclass correlation
			Total	Correlated	Total	Correlated	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Total persons	4,288	100.0	-	-	-	-	-
BIRTHPLACE							
Not coded	99	2.3	0.001528	0.001252	0.039	0.035	0.0011
Coded	4,189	97.7	0.000001	0.000001	0.001	0.001	0.0011
Same State	3,065	71.5	0.000011	0.000010	0.003	0.003	0.0017
Different State	862	20.1	0.000071	0.000058	0.008	0.008	0.0010
U.S. possession	9	.2	0.002112	0.001150	0.046	0.034	0.0028
Abroad or at sea	253	5.9	0.000188	0.000145	0.014	0.012	0.0008
RESIDENCE IN 1955							
Not coded	3,500	81.6	0.000020	0.000019	0.005	0.004	0.0003
Different county, same State	451	10.5	0.000735	0.000674	0.027	0.026	0.0026
Different State	303	7.1	0.001314	0.001226	0.036	0.035	0.0032
Abroad	34	.8	0.001898	0.000785	0.044	0.028	0.0002
OCCUPATION							
Not coded	2,228	52.0	0.000025	0.000020	0.005	0.004	0.0001
Coded	2,060	48.0	0.000029	0.000024	0.005	0.005	0.0001
Professional, technical	202	4.7	0.000511	0.000285	0.023	0.017	0.0003
Farmers, farm managers	54	1.3	0.001898	0.001101	0.044	0.033	0.0003
Managers, officials, proprietors	155	3.6	0.002154	0.001601	0.046	0.040	0.0007
Clerical	390	9.1	0.000259	0.000248	0.016	0.012	0.0003
Sales workers	183	4.3	0.000430	0.000183	0.021	0.014	0.0000
Craftsmen, foremen	270	6.3	0.000410	0.000178	0.020	0.013	0.0000
Operatives	419	9.8	0.000258	0.000129	0.016	0.011	0.0002
Private household workers	49	1.1	0.000903	0.000106	0.031	0.010	0.0000
Service workers	198	4.6	0.000416	0.000204	0.020	0.014	0.0002
Farm laborers	45	1.1	0.006816	0.005004	0.083	0.071	0.0006
Laborers	95	2.2	0.002614	0.001492	0.051	0.039	0.0003
INDUSTRY							
Not coded	2,199	51.3	0.000018	0.000015	0.004	0.004	0.0001
Coded	2,089	48.7	0.000020	0.000016	0.004	0.004	0.0001
Agriculture, forestry, fisheries	110	2.6	0.000520	0.000494	0.023	0.022	0.0005
Mining	15	.4	0.008465	0.003431	0.092	0.059	0.0002
Construction	118	2.8	0.000749	0.000276	0.027	0.017	0.0001
Manufacturing-durables	343	8.0	0.000231	0.000119	0.015	0.011	0.0002
Manufacturing-nondurables	293	6.8	0.000585	0.000411	0.024	0.020	0.0006
Transportation	82	1.9	0.000747	0.000233	0.027	0.015	0.0001
Communication	33	.8	0.001128	0.000472	0.034	0.022	0.0002
Utilities	25	.6	0.002369	0.000739	0.049	0.027	0.0001
Wholesale trade	76	1.8	0.008173	0.006037	0.090	0.078	0.0007
Retail trade	369	8.6	0.000471	0.000341	0.022	0.019	0.0007
Finance, insurance, real estate	100	2.3	0.000249	0.000000	0.016	0.000	0.0000
Business and repair services	50	1.2	0.008151	0.005064	0.090	0.071	0.0004
Personal services	124	2.9	0.000339	0.000021	0.018	0.005	0.0000
Entertainment and recreation	23	.5	0.007780	0.003339	0.088	0.058	0.0002
Professional services	236	5.5	0.000283	0.000125	0.017	0.011	0.0002
Public administration	92	2.2	0.001194	0.000664	0.034	0.026	0.0003

Table 10.–ESTIMATES OF TOTAL CODING RELVARIANCE, CORRELATED COMPONENT OF CODING RELVARIANCE, AND INTRACLASS CORRELATION OF DIFFERENT UNITS IN SAME CODER ASSIGNMENT (ρ_t) IN AN AVERAGE CODING ASSIGNMENT OF 4,288 PERSONS BY ONE CODER–Continued

Classification	Number of persons	Percent of total	Coder relvariance		Coefficient of variation		Intraclass correlation
			Total	Correlated	Total	Correlated	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
WAGE AND SALARY INCOME							
Not coded	2,719	63.4	0.000017	0.000015	0.004	0.004	0.0002
Coded	1,569	36.6	0.000050	0.000044	0.007	0.007	0.0002
$1 to $499	194	4.5	0.000240	0.000114	0.016	0.011	0.0002
$500 to $999	113	2.6	0.000337	0.000153	0.018	0.012	0.0002
$1,000 to $1,499	97	2.3	0.000403	0.000139	0.020	0.012	0.0001
$1,500 to $1,999	74	1.7	0.000431	0.000080	0.021	0.009	0.0001
$2,000 to $2,499	95	2.2	0.000474	0.000204	0.022	0.014	0.0002
$2,500 to $2,999	84	2.0	0.000406	0.000160	0.021	0.013	0.0002
$3,000 to $3,499	112	2.6	0.000224	0.000087	0.015	0.009	0.0001
$3,500 to $3,999	99	2.3	0.000221	0.000061	0.015	0.008	0.0001
$4,000 to $4,499	115	2.7	0.000225	0.000095	0.015	0.010	0.0002
$4,500 to $4,999	93	2.2	0.000172	0.000037	0.013	0.006	0.0001
$5,000 to $5,999	188	4.4	0.000116	0.000051	0.011	0.007	0.0002
$6,000 to $6,999	117	2.7	0.000208	0.000096	0.014	0.010	0.0002
$7,000 to $9,999	134	3.1	0.000155	0.000047	0.012	0.007	0.0001
$10,000 to $14,999	39	.9	0.000765	0.000540	0.028	0.023	0.0006
$15,000 to $24,999	11	.3	0.002945	0.002199	0.054	0.047	0.0007
$25,000 and over	4	.1	0.019161	0.013403	0.139	0.116	0.0005
SELF-EMPLOYMENT INCOME							
Not coded	4,065	94.8	0.000001	0.000000	0.001	0.000	0.0000
Coded	223	5.2	0.000142	0.000054	0.012	0.007	0.0001
$1 to $499 or loss	45	1.0	0.000654	0.000000	0.026	0.000	0.0000
$500 to $999	25	.6	0.001629	0.000381	0.040	0.020	0.0001
$1,000 to $1,499	20	.5	0.002072	0.000351	0.046	0.019	0.0000
$1,500 to $1,999	13	.3	0.003626	0.001009	0.060	0.032	0.0001
$2,000 to $2,499	14	.3	0.002704	0.000774	0.052	0.028	0.0001
$2,500 to $2,999	10	.2	0.003564	0.000000	0.060	0.000	0.0000
$3,000 to $3,499	13	.3	0.002387	0.000000	0.049	0.000	0.0000
$3,500 to $3,999	9	.2	0.002326	0.000000	0.048	0.000	0.0000
$4,000 to $4,499	11	.3	0.001292	0.000000	0.036	0.000	0.0000
$4,500 to $4,999	6	.2	0.004330	0.000781	0.066	0.028	0.0001
$5,000 to $5,999	15	.4	0.001754	0.000095	0.042	0.010	0.0000
$6,000 to $6,999	9	.2	0.003159	0.000000	0.056	0.000	0.0000
$7,000 to $9,999	14	.3	0.002411	0.000665	0.049	0.026	0.0001
$10,000 to $14,999	10	.2	0.003454	0.002037	0.059	0.045	0.0003
$15,000 to $24,999	6	.2	0.002694	0.001140	0.052	0.033	0.0002
$25,000 and over	3	.1	0.016521	0.009382	0.128	0.097	0.0003
OTHER INCOME							
Not coded	3,527	82.3	0.000006	0.000005	0.002	0.002	0.0011
Coded	761	17.8	0.000119	0.000099	0.011	0.010	0.0011
$1 to $499	348	8.1	0.000210	0.000155	0.015	0.012	0.0007
$500 to $999	199	4.6	0.000203	0.000096	0.014	0.010	0.0002
$1,000 to $1,499	94	2.2	0.000350	0.000090	0.019	0.010	0.0001
$1,500 to $1,999	43	1.0	0.001185	0.000450	0.034	0.021	0.0001
$2,000 to $2,499	27	.6	0.001348	0.000210	0.037	0.014	0.0000
$2,500 to $2,999	13	.3	0.002735	0.000000	0.052	0.000	0.0000
$3,000 to $3,499	11	.2	0.004805	0.000527	0.069	0.023	0.0000
$3,500 to $3,999	5	.1	0.008666	0.000000	0.093	0.000	0.0000
$4,000 to $4,499	5	.1	0.012736	0.002910	0.113	0.054	0.0001
$4,500 to $4,999	2	.1	0.020052	0.000000	0.142	0.000	0.0000
$5,000 to $5,999	4	.1	0.027282	0.004450	0.165	0.067	0.0000
$6,000 to $6,999	3	.1	0.033207	0.000000	0.182	0.000	0.0000
$7,000 to $9,999	3	.1	0.035207	0.005930	0.188	0.077	0.0000
$10,000 to $14,999	2	.0	0.013585	0.006297	0.117	0.079	0.0002
$15,000 to $24,999	1	.0	0.057810	0.000000	0.240	0.000	0.0000
$25,000 and over	1	.0	0.045914	0.000000	0.214	0.000	0.0000