Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-40
Selection of sample areas. . . . App-40
Designation of sample housing units for the 1981 survey. . . . App-41
Selection of the 1973 sample housing units. . . . App-41
Splitting of the sample App-41
CEN-SUP sample App-41
Selection of supplemental sample housing units in rural areas. . . . App-41
Selection of sample housing units for the 1976 Coverage Improvement Program App-42
1977 and 1981 sample reductions App-42
Supplemental sample from nonresidential conversions App-42
1981 telephone interviewing equipment App-42
1970 Census of Population and Housing. . . . App-43
ESTIMATION. . . . App-43
AHS national sample App-43
1981 housing inventory App-43
1973-1981 lost units App-44
Ratio estimation procedure of the 1970 Census of Population and Housing App-44
RELIABILITY OF THE ESTIMATES App-44
Sampling errors App-44
Standard errors of estimates of levels App-45
Standard errors of estimates of percentages. . . . App-45
Standard errors of ratios App-45
Illustration of the use of the standard error tables. Illustration I App-45
Illustration II App-46
Differences. . . . App-46
Illustration of the computation of the standard error of a difference App-46
Medians App-47
Illustration of the computation of the 95-percent confidence interval for a median App-47
Nonsampling errors. . . . App-47
1970 census App-47
Reinterview program App-47
Coverage errors App-48
Effects of ratio estimation on estimates of change App-49
Possible effects of telephone interviewing on the data App-49
Rounding errors. . . . App-49

SAMPLE DESIGN

The 1981 estimates are based on data collected from September 1981 through December 1981 for the Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. The sample for this survey was spread over 461 sample areas (called primary sampling units), comprising 923 counties and independent cities with coverage in each of the 50 States and the District of Columbia.

Approximately 56,800 sample housing units (both occupied and vacant) were eligible for interview in the 1981 Annual Housing Survey. Of this number, 3,200 interviews were classified as "noninterview" for various reasons. Occupied housing units were classified as "noninterview" mainly, because the occupants refused to be interviewed after repeated calls. For vacant housing units, interviews were not obtained because an informed respondent was not found after repeated visits. In addition to the 56,800 eligible housing units, there were also 3,600 sample units which were visited but were ineligible for interview for the AHS in terms of collecting information relevant to the 1981 housing inventory.

Selection of sample areas—The United States was divided into areas made up of counties and independent cities referred to as primary sampling units (PSU's). These PSU's were then grouped into 376 strata, 156 of which consisted of only one PSU which was in sample with certainty. These 156 strata were mostly the larger SMSA's and were called self-representing (SR) since the sample from the sample area represented just that PSU. Each one of the other 220 strata consisted of a group of PSU's and were referred to as non-self-representing (NSR), since the sample of housing units from the sample PSU in a stratum represented the other PSU's in the stratum as well.

One PSU was selected from each NSR stratum with probability proportionate to the 1970 census population of the PSU. (This resulted in 220 NSR sample PSU's.) In additon, the NSR strata were grouped into 110 pairs and one stratum was picked at random from each pair. From this stratum, an additional PSU was selected independently of the other PSU selected from this stratum. Since the two PSU's were independently selected, it was possible for the same PSU to be selected twice. This occurred in 25 instances, producing an additional 85 NSR sample PSU's, thus giving a grand total of 461 PSU's.

Designation of sample housing units for the 1981 survey—The sample housing units designated to be interviewed in the 1981 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1980 survey (which included all sample housing units that were selected as part of the 1976 Coverage Improvement Program) and were not part of the 1981 sample reduction.
2. All sample housing units that were either type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of survey but which could become eligible in the future) in the 1980 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1981 AHS questionnaire, page App-16).
3. All sample housing units that were selected from the list of building permits issued since the 1980 survey. (This sample represented the housing units built in permit-issuing areas, since the 1980 survey.)
4. Housing units added as the result of the updated listings in selected areas which do not issue building permits.

Selection of the 1973 sample housing units—The overall sampling rate used to select the sample for the 1973 AHS was about 1 in 1,366. The within-PSU sampling rate for the AHS was determined so that the overall probability of selection for each sample housing unit was the same (e.g., if the probability of selecting a NSR PSU was 1 in 10, then the within-PSU sampling rate would be 1 in 136.6).

Within the sample PSU's, a sample of the housing units enumerated in the 1970 Census of Population and Housing was selected for the AHS. In addition, a sample of new construction building permits was also selected to represent the units constructed since the 1970 census. These samples were selected at about twice the rate mentioned previously (i.e., at 2 in 1,366), thereby producing a sample twice as large as needed. This sample was split into two equal-sized samples—one to be used for the AHS, and one to be held in reserve for possible future use for the AHS. The procedure used to split this sample into equal-sized samples is described in the next section.

The sample of 1970 census units was selected in several stages. Within the sample PSU's, the first step was the selection of a sample of census enumeration districts (ED's), administrative units used in the 1970 census. The probability of selection for an ED was proportional to the following 1970 census counts of housing units (HU's) and persons in group quarters, combined in the following formula:

$$\frac{\text{Number of HU's in the ED} + \dfrac{\text{Number of group quarters persons in the ED}}{3}}{4}$$

The next step was to select an expected cluster of about four neighboring housing units within each sample ED. For most of the ED's, the selection was accomplished using the list of addresses for the ED as compiled in the 1970 census (these ED's are referred to as address ED's). However, in those ED's where addresses were incomplete or inadequate (mostly rural areas), the selection process was accomplished using area sampling methods. These ED's were divided into segments (i.e., small land areas with well-defined boundaries having an expected size of four, or a multiple of four, housing units) and a segment was selected. Those selected segments with an expected size which was a multiple of four were further subsampled at the time of interview so that an expected four housing units were chosen for interview.

The sample of new construction housing units was selected from building permits issued since January 1970. Within each sample PSU, the building permits were chronologically ordered by month issued, and compact clusters of approximately four housing units were created. These clusters were then sampled at the rate of 2 in 1,366. Housing units constructed since the 1970 census in areas which do not issue building permits were brought into the sample as a result of the area sample described above.

Splitting of the sample—The described sample selection procedure produced clusters (or segments) of size-four housing units for the sample taken from the census address frame, the new construction frame, and the area sampling frame (mainly rural areas). Clusters of this size should result in a minimum loss in precision for estimates of housing characteristics in rural areas because of the heterogeneity of neighboring units. However, clusters of size-two housing units were considered to be more optimum within those areas where the housing characteristics of neighboring units tend to be very similar (i.e., urban areas and new construction units). A splitting operation was then carried out for clusters selected from the census address and the new construction frames. This consisted of halving each sample cluster from these frames. Thus, two housing units from each of these clusters were included in the survey and two housing units were held in reserve. No splitting operation was carried out within the clusters selected from the area sampling frame; every other area sample cluster of four housing units was used for the survey and the remaining clusters were assigned to the reserve sample.

CEN-SUP sample—Housing units at addresses missed in the census or inadequately described in the 1970 census address register did not have a chance of being selected for the AHS sample in address ED's. Consequently, a special operation was undertaken to develop a separate sample called the census supplemental (CEN-SUP) sample to represent these units. Due to time constraints on this operation, only about 40 percent of the CEN-SUP units were ready to be interviewed in 1973. The rest were interviewed for the first time in the 1974 AHS.

Selection of supplemental sample housing units in rural areas—In 1974, it was decided to increase the reliability of the AHS estimates of rural housing characteristics by doubling the number of sample housing units from rural areas. This was accomplished by reactivating the reserve sample, selected in the original sampling operations in 1973, from rural areas only. For

the reserve sample selected in census address and new construction frames, the other half of each rural cluster (an expected two housing units) was reactivated in 1974. Similarly, for the area sampling frame, the entire reserve cluster (an expected four housing units) was reactivated in 1974 if the cluster was rural. This supplementation increased the overall probability of selection for sample housing units in rural areas to about 2 in 1,366; whereas, the overall probability of selection for sample housing units in urban areas remained at 1 in 1,366.

Selection of sample housing units for the 1976 Coverage Improvement Program—The 1976 Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS national sample from the census address and new construction frames. The coverage deficiencies included the following types of housing units:

1. New construction housing units from building permits issued prior to January 1970 for which construction had not been completed at the time of the 1970 census.
2. Housing units converted to residential use in structures totally nonresidential at the time of the 1970 census.
3. Houses that had been moved onto their present site since the 1970 census.
4. Mobile homes and trailers placed in parks either missed in the 1970 census or established since the 1970 census.
5. Mobile homes and trailers placed outside parks since the 1970 census or vacant at the time of the 1970 census.

A sample of new construction housing units whose permits were issued before January 1970 was selected in two stages. In the first stage, the units whose permits were issued before January 1970, but which were completed after the census, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. In the second stage, these units were then sampled so that the overall probability of selection was about 1 in 1,320.

A sample of mobile homes and trailers placed in a park that was missed by the census or established after the census was also selected in two stages. During the first stage, a list of mobile home parks was obtained from commercial listings. This list was then supplemented by additional parks identified by a canvassing operation similar to that performed in ED's where area sampling methods are used. The second stage consisted of dividing the parks into clusters of an expected size of four sites. These clusters were then sampled so that the overall probability of selection was about 1 in 1,366.

For the remaining housing units (i.e., mobile homes and trailers placed outside parks since the 1970 census, mobile homes and trailers vacant at the time of the 1970 census, units converted from nonresidential to residential use since the 1970 census, and houses that had been moved onto their present site since the 1970 census), the sampling was done in three stages. First, a subsample of the regular AHS sample housing units from the census address frame was selected. Second, succeeding structures that had been eligible to be selected from the census address frame were then listed until eight structures (excluding mobile home parks) were found. Finally, the intervening structures that had been listed which did not have a chance of selection in the AHS were identified and the units within these structures were interviewed.

1977 and 1981 sample reductions—By 1977, the addition to the sample, primarily from new construction and the coverage improvements, had increased the total sample size (interviews plus noninterviews) to about 81,000 housing units. The sample was reduced by about 7 percent to approximately 75,000 in 1977. This reduction did not include any CEN-SUP units or units which were selected as part of the 1976 Coverage Improvement Program. Initially, the overall sample was further reduced in 1981 by about 5 percent. Subsequently, the rural sample was reduced by about 25 percent in 125 large SMSA's[1] and by about 50 percent in small SMSA's[1] and outside of SMSA's. These reductions brought the sample size down to approximately 60,000 housing units in 1981.

Supplemental sample from nonresidential conversions—The purpose of this sample was to improve the coverage in address ED's (i.e., ED's in which 1970 census address listings were used for selecting the AHS sample) of housing units created in structures that were completely nonresidential at the time of the 1970 census. This sample was derived from listings created for the Commercial Victimization Survey (CVS), a nationally representative area sample survey conducted by the Census Bureau for the Department of Justice. Each of the CVS area segments was initially canvassed either in 1972 or 1975 and eligible business establishments were listed. These listings were updated every 6 months with the last updating in the last half of 1977. Each basic address containing a business establishment listed for CVS was visited for the 1980 Components of Inventory Change Survey (CINCH), which was conducted in conjunction with the 1980 AHS, to identify HU's in structures which were completely nonresidential at the time of the 1970 census. These cases were later matched to the 1970 census address listings to identify those cases in address ED's and as an additional check to see if housing units existed in these structures at the time of the 1970 census. Due to the timing restraints associated with these operations, these cases were not included for the 1980 AHS but were included for the 1981 survey.

1981 telephone interviewing experiment—A large scale telephone interviewing experiment was conducted in conjunction with the 1981 AHS national sample in order to provide more definitive information about the effect of telephone interviewing on AHS data and cost. A random sample of about 12,500 AHS national sample units was selected for the experiment. Among the cases assigned for telephone interviewing, only those sample housing units that had been interviewed in 1980 and had a telephone number available were eligible to be interviewed by telephone. Since a large portion of the total AHS national sample was assigned for telephone interviewing, its possible

[1] A large SMSA is one with a 1970 population of 250,000 persons or more and a small SMSA is one with a 1970 population of less than 250,000 persons.

effect on the data is mentioned in the section on "Reliability of the Estimates" of this appendix.

1970 Census of Population and Housing—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-, 15-, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design can be obtained in the 1970 census report, HC(1)-B1, *Detailed Housing Characteristics, United States Summary.*

ESTIMATION

AHS national sample—The AHS national sample produced estimates of two types: Estimates of the 1981 housing inventory and estimates of units removed from the housing inventory between 1973 and 1981 (i.e., 1973-1981 lost units). Each type of estimate employed a separate, though similar, estimation procedure as described below.

1981 housing inventory—In 1981, the AHS estimates employed a three-stage ratio estimation procedure. However, prior to implementation of the procedure, the basic weight (i.e., the inverse of the probability of selection) was adjusted to account for the type A noninterview housing units encountered in the AHS. This noninterview adjustment was done separately for occupied and vacant units. The noninterview adjustment was equal to the following ratio:

$$\frac{\text{Interviewed housing units + Noninterviewed housing units}}{\text{Interviewed housing units}}$$

The first-stage ratio estimation procedure was employed for sample housing units from non-self-representing (NSR) PSU's only. This procedure was designed to reduce the contribution to the variance arising from the sampling of PSU's. The first-stage ratio estimation procedure takes into account the differences that existed at the time of the 1970 census in the distribution by tenure and residence of the housing population estimated from the sample NSR PSU's and that of the NSR housing population in each of the four census regions of the country.

The first-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{The 1970 census housing population in the residence-tenure category for all NSR strata in a census region}}{\text{Estimates of the housing population category using 1970 census housing counts for sample NSR PSU's in a census region}}$$

The numerators of the ratios were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR stratum and summing these counts across the NSR strata in each census region. The denominators were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR sample PSU, weighting these counts by the inverse of the probability of selecting that PSU, and summing these weighted counts across the NSR sample PSU's in each census region. The computed first-stage ratio estimation factor was then applied to the existing weight for each NSR sample unit in each first-stage ratio estimation category.

The second-stage ratio estimation procedure was designed to adjust the AHS sample estimate of conventional new construction housing units, i.e., sample units built April 1, 1970 or later, to independently derived current estimates where a known deficiency in the AHS sample exists (see the section on nonsampling error) for each of the four regions. These estimates were considered to be the best estimates available for the number of conventional new construction units in these categories.

The second-stage ratio estimation factor was as follows:

$$\frac{\text{Current best estimate of new construction in the category}}{\text{AHS sample estimate of new construction units in the category}}$$

The numerators of the ratios were derived from data based on the Survey of Construction (SOC). The denominators of the ratios were obtained from the weighted estimates for the AHS sample units using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in each second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all the AHS sample units. This procedure was designed to adjust the AHS sample estimates of housing (i.e., the estimates employing the noninterview, first-stage, and second-stage adjustments) to current vacant housing estimates for four categories of vacant housing units and to independently derived current housing estimates for 24 categories of occupied housing units. Each of these categories is a combination of the characteristics of residence, tenure, race of householder, and sex of householder.

The third-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{Current independent estimate of housing units in the category}}{\text{AHS sample estimate of housing units in the category}}$$

The numerators of the ratios for occupied housing units were derived from data based on the Current Population Survey (CPS), a sample household survey conducted monthly by the Bureau of the Census. This is the first time that CPS household estimates based on the 1980 census were used as the independent controls in this adjustment rather than the 1970 census-based controls. The 1970-based estimates were about 2 percent smaller than the 1980-based estimates. This difference should be taken into consideration when comparing the 1981 AHS estimates to the AHS estimates for 1980 or earlier years. The numerators of the ratios for vacant housing units were derived from data based on the Housing Vacancy Survey (HVS), a quarterly vacancy survey also conducted by the Bureau of the Census. The denominators of the ratios were obtained from the weighted estimates for the AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio esti-

mation factor was then applied to the existing weight for each sample unit in each third-stage ratio estimation category.

The second-stage and the third-stage ratio estimation procedures were iterated in order to bring the AHS estimates into close agreement with both sets of independent estimates. The second-stage was modified so that the estimates for all 27 categories of new construction would be identical to the estimates before the third stage. Hence, the repeated second stage had the effect of controlling the AHS sample estimates of new construction units to the unbiased sample estimates for nine categories of new construction mobile homes and trailers for each of the four regions, and of adjusting the AHS sample estimate of 18 categories of conventional new construction units to the independently derived current estimates.

The numerators were either the unbiased weighted estimates for the AHS sample housing units, using the existing weight after the first-stage ratio estimation procedure (i.e., the estimates employing the noninterview and first-stage adjustments) or the independent estimate derived from data based on the Survey of Construction (SOC).

The denominators of the ratios in this iterative process were obtained from the weighted estimates for the AHS sample units after the previous stage of ratio estimation. The factors resulting from this iterative process were then applied to the existing weight on the appropriate records, and the resulting product was used as the final weight for tabulation.

The third-stage ratio estimation procedure, as well as the overall estimation procedure, reduced the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. The distribution of the housing population selected for the sample differed somewhat, by chance, from that of the Nation as a whole in such basic housing characteristics as tenure, vacancy status, residence, race of householder, and sex of householder. These characteristics are probably closely correlated with other housing characteristics measured for the AHS. Therefore, through the use of the three-stage ratio estimation procedure one can expect the sample estimate to be improved substantially.

1973-1981 lost units—The 1973-1981 lost unit estimates employed the three-stage ratio estimation procedure used to produce the AHS national estimates of the 1973 housing inventory, as was described in the 1973 Current Housing Report, Series H-150-73A, *General Housing Characteristics for the United States and Regions.* These 1973-1981 lost units do not include the HU's from the 1976 Coverage Improvement Program. Since the 1973-1981 lost units existed, by definition, in the 1973 housing inventory, there was a 1973 housing inventory weight associated with each 1973-1981 lost unit. This weight, adjusted for the 1977 and 1981 sample reductions, was used to tabulate the estimates of the characteristics of the 1973-1981 and the 1980 census-based controls lost units. The general effect of this estimation procedure was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing—This report presents data on the housing characteristics of the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be obtained in the 1970 census report, HC(1)-B1, *Detailed Housing Characteristics, United States Summary.*

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS national sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census appears in the 1970 census report, HC(1)-B1, *Detailed Housing Characteristics, United States Summary.* The sampling errors for 1970 census data are much smaller than for the AHS data. Therefore, in making comparisons between the two data sources, it can be safely assumed that the census data are subject to zero sampling errors.

Sampling errors—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same questionnaires, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The variability between estimates from all possible samples is defined as sampling error. One common measure of sampling error is the standard error which measures the precision with which an estimate from a sample approximates the average result of all possible samples. In addition, the standard error, as calculated for this report, also partially reflects the variation in the estimates due to some nonsampling errors, but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, and biases and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates so that the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, and each of these samples was surveyed under essentially the same general conditions and an estimate and its estimated standard error were calculated for each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples;
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples;

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples either is or is not contained in any particular computed interval. However, for a particular sample one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables (pages App-49 to App-55) are approximations to the standard errors of various estimates shown in this report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than the precise standard error for any specific item.

Standard errors of estimates of levels—Tables I and II present the standard errors applicable to the 1981 national housing inventory estimates in this report. Tables III and IV present the standard errors applicable to 1973-1981 lost housing unit estimates in this report. Table V presents the standard errors applicable to estimates for the Northeast, North Central, South, and West Regions. Linear interpolation should be used to determine standard errors for levels of estimates not specifically shown in tables I through V.

Tables VI through X present the standard errors of estimated percentages. Table VI shows the approximate standard errors of all national estimated percentages of housing units except those pertaining to the specified items in table II. The standard errors shown in table VII should be used for those specified items. Tables VIII and IX show the approximate standard errors of the estimated percentages of 1973-1981 lost housing units. Table X shows the approximate standard error of all regional estimated percentages of housing units and 1973-1981 lost housing units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables VI through X.

Included in tables I through X are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered to be overestimates of the true standard errors and should be used primarily for construction of confidence intervals for characteristics when an estimate of zero is obtained.

Standard errors of estimates of percentages—The reliability of an estimated percentage, computed by using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more.

Standard errors of ratios—For ratios of the form (100) (x/y), where x is not a subclass of y, tables VI through X, underestimate the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be obtained by letting the standard error of the ratio be approximately equal to:

$$(100)\left(\frac{x}{y}\right)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

where: x = the numerator of the ratio
y = the denominator of the ratio
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

Illustration of the use of the standard error tables. Illustration I—Table A-1 of this report shows that inside SMSA's in the United States there were 11,211,000 owner-occupied housing units with two persons in 1981. Interpolation in standard error table I (page App-49) shows that the standard error of an estimate of this size is approximately 140,000. The following procedure was used in interpolating.

The information presented in the following table was extracted from standard error table I. The entry for "x" is the one sought.

Size of estimate (000)	Standard error (000)
10,000	135
11,211	x
25,000	193

By vertically interpolating between 135 and 193, the entry for "x" is determined to be 140.

$$11{,}211 - 10{,}000 = 1{,}211$$
$$25{,}000 - 10{,}000 = 15{,}000$$

$$135 + \frac{1{,}211}{15{,}000}(193 - 135) = 140$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 11,071,000 to 11,351,000 housing units. Therefore, a conclusion that the average estimate of 1981 housing units of this type lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all possible samples, lies within the interval from 10,987,000 to 11,435,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 10,931,000 to 11,491,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 11,211,000 owner-occupied housing units with two persons inside SMSA's, 4,103,000, or 36.6 percent, were in central cities. Interpolation in standard error

table VI (i.e., interpolation on both the base and percent) (page App-51) shows that the standard error of the above percentage is 0.6 percentage points. The following procedure was used in interpolating.

The information presented in the following table was extracted from table VI. The entry for "p" is the one sought.

Base of percentage (000)	Estimated percentage		
	25	36.6	50
10,000	0.6	a	0.7
11,211		p	
25,000	0.4	b	0.5

1. By horizontal interpolation between 0.6 and 0.7, the entry for cell "a" is determined to be 0.6.

$$36.6-25.0 = 11.6$$
$$50.0-25.0 = 25.0$$

$$0.6 + \frac{11.6}{25.0}(0.7-0.6) = 0.6$$

2. By horizontal interpolation between 0.4 and 0.5, the entry for cell "b" is determined to be 0.4.

$$36.6-25.0 = 11.6$$
$$50.0-25.0 = 25.0$$

$$0.4 + \frac{11.6}{25.0}(0.5-0.4) = 0.4$$

3. By vertical interpolation between 0.6 and 0.4, the entry for "p" is determined to be 0.6.

$$11{,}211-10{,}000 = 1{,}211$$
$$25{,}000-10{,}000 = 15{,}000$$

$$0.6 + \frac{1{,}211}{15{,}000}(0.4-0.6) = 0.6$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 36.0 to 37.2 percent; the 90-percent confidence interval is from 35.6 to 37.6 percent; and the 95-percent confidence interval is from 35.4 to 37.8 percent.

Illustration II—Table A-1 of this report shows that in the United States in 1981 there were 152,000 housing units in structures with four floors or more (see "Elevator in Structure" item) that were outside of SMSA's. Interpolation in standard error table I (page App-49) shows that the standard error of an estimate of this size is approximately 17,000. Consequently, the 68-percent confidence interval is from 135,000 to 169,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1981 housing units in structures with four floors or more that were outside of SMSA's lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all possible samples, lies within the interval from 125,000 to 179,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 118,000 to 186,000 housing units with 95 percent confidence.

In 1981, table A-1 also shows that of the 152,000 housing units in structures with four floors or more that were outside SMSA's, 115,000, or 75.7 percent, were in structures that contained elevators. Interpolation in table VI (i.e., interpolation on both the base and percent; page App-51) shows that the standard error of the above percentage is 5.3 percentage points. Consequently, the 68-percent confidence interval is from 70.4 to 81.0 percent; and the 90-percent confidence interval is from 67.2 to 84.2 percent; and the 95-percent confidence interval is from 65.1 to 86.3 percent.

Differences—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristics in two different areas or the difference between separate and uncorrelated characteristics in the same area. If there is a high positive correlation between the two characteristics, the formula will overestimate the true error. If there is a high negative correlation between the two characteristics, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference—Table A-1 shows that inside SMSA's in the United States there were 6,377,000 owner-occupied housing units with three persons in 1981. Thus, the apparent difference between the number of 1981 owner-occupied housing units with two persons and those with three persons is 4,834,000. The standard error of 11,211,000 is approximately 140,000 as shown above. Interpolation in standard error table I (page App-49) shows that the standard error on an estimate of 6,377,000 to be approximately 110,000. Therefore, the standard error of the estimated difference of 4,834,000 is about 178,000.

$$178{,}000 = \sqrt{(140{,}000)^2 + (110{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 4,834,000 difference is from 4,656,000 to 5,012,000 housing units. Therefore, a conclusion that the average estimate of this difference, derived from all possible samples, lies within a range computed in this way and would be correct for roughly 68 percent of all possible samples. Similarily, the 90-percent confidence interval is from 4,549,000 to 5,119,000 housing units, and the 95-percent confidence interval is from 4,478,000 to 5,190,000. Thus, we can conclude with 95 percent confidence that the number of 1981 owner-occupied housing units inside SMSA's with two persons is greater than the number with three persons.

Medians—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data.

1. From the appropriate standard error table, determine the standard error of a 50-percent characteristic on the base of the median.
2. Add to and subtract from 50 percent the standard error determined in step 1.
3. Using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval for a median—Table A-1 of this report shows the median number of persons in owner-occupied housing units inside SMSA's was 2.6 in 1981. The base of the distribution, from which this median was determined, is 34,788,000 housing units.

1. From table VI, the standard error of a 50-percent characteristic on the base of 34,788,000 is 0.4 percentage points.
2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 49.2 and 50.8.
3. From table A-1, it can be seen by cumulating the frequencies for the first two categories that 16,569,000 owner-occupied housing units inside SMSA's, or 47.6 percent, had one and two persons (actually, for purposes of calculating the median, the category of two persons is considered to be from 1.5 to 2.5 persons) and that an additional 6,377,000 owner-occupied housing units, or 18.3 percent, had three persons (i.e., 2.5 to 3.5 persons). By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about 2.6.

$$2.5 + (3.5-2.5)\,\frac{(49.2-47.6)}{18.3} = 2.6$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about 2.7.

$$2.5 + (3.5-2.5)\,\frac{(50.8-47.6)}{18.3} = 2.7$$

Thus, the 95-percent confidence interval ranges from 2.6 to 2.7 persons. Although it appears that this confidence interval has the sample estimate as the lower limit, it actually is a reflection of the rounding error associated with the median (see paragraph on rounding error in the nonsampling error section of this appendix).

Nonsampling errors—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties; differences in the interpretation of questions; inability or unwillingness to provide correct information on the part of respondents; mistakes in recording or coding the data; and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for the 1970 Census of Population and Housing and the 1981 AHS national sample.

1970 census—A number of studies were conducted to measure two types of general errors associated with 1970 census estimates: "Coverage" and "content" errors. The "coverage" errors determined how completely housing units were counted in the census and the extent to which occupancy status was erroneously reported. The "content" errors measured the accuracy of the data collected for surveyed housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies on coverage and content errors, as well as the methodology employed, can be found in the 1970 Census of Population and Housing Evaluation and Research Program series reports PHC(E)-5, *The Coverage of Housing in the 1970 Census;* and PHC(E)-10, *Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.*

Reinterview program—For the AHS national sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a subsample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and thus were the basis for the measurement of the "content" error of these AHS estimates.

As part of the reinterview, an additional check was carried out for interviewer evaluation and quality control. This check was made at each of these households to determine if the following was done during the original interview.

1. The correct housing unit was visited.
2. The correct number of housing units were interviewed at that address.

3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

The results of the 1981 and 1980 reinterview studies were not available at the time of publication. In addition, the 1979 reinterview was carried out as an interviewer quality control only. Thus, the results of the 1977 and 1978 reinterview studies which were presented in the Census Bureau memoranda, "Reinterview Results for the Annual Housing Survey—National Sample 1977" and "Reinterview Results for the Annual Housing Survey—National Sample 1978" are presented here.

In 1978, a substantial portion of the reinterview questionnaire was devoted to testing the new questions 1 through 7 (parts a and b). The questions (part a), which were asked only at housing units interviewed in the previous year, determined whether there had been a change since last year in selected nonattitudinal items. If a change had been recorded or the respondent did not know if a change had occurred, part b of the question, which collects the value of the item, was asked. The reinterviewers asked these items using the questions as formatted in 1977. Comparing the responses from the differently formatted questions, the 1978 reinterview found that 80 percent of the questions showed low levels of inconsistency with the remainder showing moderate levels.

The 1977 reinterview program showed moderate to high levels of inconsistency with about 21 percent of the nonattitudinal and 56 percent of the attitudinal items showing high levels of inconsistency. But a large proportion (43 percent) of the nonadditudinal items showed a low level of inconsistency. Moderate levels indicate that there are some problems with inconsistent reporting and high levels indicate that improvements are needed in the data collection methods or that the category concepts themselves are ambiguous.

Cross-tabulations involving those items, which are subject to substantial levels of inconsistency may be subject to a large distortion as a consequence of the associated high level of response variance, and thus, are considered to be less reliable than comparable cross-tabulations which do not involve these data. As a consequence, the tables in this report, which contain such cross-tabulations, have been footnoted with a cautionary statement.

The 1970 census reinterview results provide illustrations of possible nonsampling errors for some of the items which also appear in the AHS. For example, median value of homes was consistently underestimated by about 5 percent, and the average monthly cost of electricity and utility gas were consistently overestimated although the net effect on average gross rent was fairly small.

A possible explanation for the results of the AHS and census reinterview studies, as well as the surveys themselves, is that the data are based on the answers given by the respondents who may lack precise information. Also, because the results of the reinterview studies are derived from sample surveys, there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of these studies.

Coverage errors—A deficiency in the representation of conventional new construction (non-mobile home or trailer) for the AHS new construction sample (mentioned previously in the section on estimation) is an example of coverage errors. During the sampling of building permits, only those issued more than 4 months before the survey began were eligible to be selected to represent conventional new construction. Due to time constraint, it is not possible to sample units whose permits are issued less than 4 months in advance of the survey.

It is estimated that the 1981 AHS sample missed about 0.8 percent (i.e., about 157,000 units) of conventional housing units built after April 1970 because the permits for these units, which were built before October 1981, were issued less than 4 months in advance of the survey. The second stage of ratio estimation procedure was employed to reduce the effect of this deficiency although some bias in the AHS estimates of conventional new construction probably still exists. Review of the second-stage ratio estimation procedures indicates that we have consistently overcompensated for this deficiency every year since 1975 by ratio adjusting the new construction to counts of new construction for the end of the interview period, which has been December or January, instead of October. This overcompensation may inflate the new construction counts by 100,000 to 300,000 units. Changing this procedure to correct for this overcompensation would reduce the estimate of housing units built since the last survey since it would be based on a 10-month time period rather than 12 months. Thus, it was decided not to change this procedure so that the estimate of housing units built since the last survey would be consistent with previous years' estimates, which are generally based on a 12-month time period.

In addition, the 1976 Coverage Improvement Program also had certain deficiencies. First, when the canvassing was done to identify mobile homes and trailer parks that were not in the same sample frame or not on the commercial lists, only 92 percent of the census address frame ED's were represented. Second, it appears that the listing procedure (used to find mobile homes and trailers placed outside parks, units converted from nonresidential to residential, and houses that had been moved onto their present site) was not very efficient for finding nonresidential conversions (which might be primarily in business districts), since the listing procedure started from a residential unit. (The sample estimate of this component was approximately 16,000 housing units with a standard error of 12,000.)

Finally, it is felt that deficiencies also exist in ED's where area sampling methods are used. As before, it had been assumed that all housing units located inside these ED's would be represented in the sample. However, it has been estimated that the 1981 AHS sample missed as much as 2 percent (i.e., as much as 400,000 units) of all housing units in ED's where area sampling

methods are used because these units are not listed during the canvassing.

The third stage of ratio estimation corrects for these deficiencies as far as the total count of housing is concerned, i.e., it adjusts the estimate of the total housing inventory to the best available estimate. However, biases of subtotals would still remain.

Effects of ratio estimation on estimates of change—As stated previously, 1981 was the first AHS for which 1980 census-based estimates, rather than 1970 census-based estimates, were used in the third-stage ratio estimation procedure. Since the 1980-based estimates are about 2 percent higher than the corresponding 1970-based estimates, this will cause estimates of change between the 1981 AHS data and data for 1980 or earlier years to be overstated. This overstatement should be taken into consideration during the analysis of estimates of change.

Possible effects of telephone interviewing on the data—A preliminary analysis of the 1981 telephone interviewing experiment has been completed. It has been concluded that data collected, using the telephone interviewing procedures, were not sufficiently different from data collected by regular personal interviews to preclude basing the 1981 AHS national sample published data. However, since the percentage of the comparison tests that were significant was slightly above what would be expected, there was some evidence that telephone interviewing may have had an effect on the data. Thus, some caution should be exercised in making comparisons between 1981 AHS data and data from preceding years.

Rounding errors—In errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of this survey. Also, since medians in this report were computed using unrounded data, instead of the published rounded data, they can differ from medians calculated directly from the published data.

TABLE I. Standard Errors of Estimated Numbers of Housing Units: 1981 (Excluding Estimates of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water-Individual Well, Lacking Complete Plumbing, Mobile Homes, and Housing Units With Householder of Spanish Origin)

(68 chances out of 100)

Size of estimate (000)	Standard error		Size of estimate (000)	Standard error	
	Total or White (000)	Black (000)		Total or White (000)	Black (000)
0	2	2	2,500	71	61
5	3	3	5,000	98	68
10	5	5	7,500	119	51
25	7	7	10,000 . . .	135	–
50	10	10	25,000 . . .	193	–
100.	14	14	50,000 . . .	216	–
250.	23	22	75,000 . . .	167	–
500.	32	31	85,000 . . .	112	–
1,000 . . .	45	43			

TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water-Individual Well, Lacking Complete Plumbing, Mobile Homes, and Housing Units With Householder of Spanish Origin: 1981

(68 chances out of 100)

Size of estimate (000)	Standard error		Size of estimate (000)	Standard error	
	Total or White (000)	Black (000)		Total or White (000)	Black (000)
0	3	3	1,000	53	50
5	4	4	2,500	83	71
10	5	5	5,000	116	79
25	8	8	7,500	139	60
50	12	12	10,000 . . .	159	–
100.	17	17	25,000 . . .	227	–
250.	27	26	50,000 . . .	253	–
500.	37	37			

TABLE III. Standard Errors of Estimated Numbers of Lost Housing Units: 1973-1981 (Excluding Estimates of Lost Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Some or All Plumbing, Mobile Homes, and Other Vacants)

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	250	21
5	3	500	31
10	4	1,000	47
25	7	2,500	93
50	10	5,000	150
100	14		

TABLE IV. Standard Errors of Estimated Numbers of Lost Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Some or All Plumbing, Mobile Homes, and Other Vacants: 1973-1981

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	3	100	18
5	4	250	29
10	6	500	41
25	9	1,000	60
50	13		

TABLE Va. Standard Errors of Estimated Numbers of Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1981 (Excluding Estimates of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions and Excluding Source of Water-Individual Well, and Mobile Homes for Each of the Regions)

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	500	33
5	3	1,000	47
10	5	2,500	74
25	7	5,000	105
50	10	7,500	129
100	15	10,000	148
250	23	25,000	235

TABLE Vb. Standard Errors of Estimated Numbers of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions and to Source of Water-Individual Well, and Mobile Homes for the Northeast and North Central Regions: 1981

(68 chances out of 100. For estimates pertaining to source of water-individual well and mobile homes for the West Region, apply a factor of 1.66 to the standard errors)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	3	500	39
5	4	1,000	56
10	6	2,500	87
25	9	5,000	122
50	13	7,500	148
100	18	10,000	168
250	28	25,000	243

TABLE Vc. Standard Errors of Estimated Numbers of Housing Units Pertaining to Source of Water-Individual Well, and Mobile Homes for the South Region: 1981

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	5	500	50
5	5	1,000	70
10	7	2,500	109
25	11	5,000	150
50	16	7,500	179
100	22	10,000	201
250	35	25,000	259

TABLE Vd. Standard Errors of Estimated Numbers of Lost Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1973-1981 (Excluding Estimates of Lost Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Some or All Plumbing, Mobile Homes, and Other Vacants)

(68 chances out of 100)

Size of estimate (000)	Standard error	
	Northeast or North Central	South or West
0	2	2
5	3	3
10	4	5
25	7	8
50	10	11
100	14	15
250	24	25
500	38	36
1,000	62	54
2,500	130	98

Note: For standard errors of regional estimates of lost housing units (1973-1981) pertaining to new construction, lacking complete kitchen facilities, no bedrooms, no bathrooms, lacking some or all plumbing, mobile homes, and other vacants for the Northeast, North Central, South, and West Regions, use the national standard errors presented in table IV.

TABLE VI. Standard Errors of Estimated Percentages of Housing Units: 1981 (Excluding Estimated Percentages of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water-Individual Well, Lacking Complete Plumbing, Mobile Homes, and Housing Units With Householder of Spanish Origin)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	29.1	29.1	29.1	29.1	29.1	29.1	29.1	32.0
10	17.0	17.0	17.0	17.0	17.0	17.0	19.6	22.7
25	7.6	7.6	7.6	7.6	8.6	10.2	12.4	14.3
50	3.9	3.9	3.9	4.4	6.1	7.2	8.8	10.1
100	2.0	2.0	2.0	3.1	4.3	5.1	6.2	7.2
250	0.8	0.9	1.3	2.0	2.7	3.2	3.9	4.5
500	0.4	0.6	0.9	1.4	1.9	2.3	2.8	3.2
1,000	0.2	0.5	0.6	1.0	1.4	1.6	2.0	2.3
2,500	0.08	0.3	0.4	0.6	0.9	1.0	1.2	1.4
5,000	0.04	0.2	0.3	0.4	0.6	0.7	0.9	1.0
7,500	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.8
10,000	0.02	0.14	0.2	0.3	0.4	0.5	0.6	0.7
25,000	0.01	0.09	0.13	0.2	0.3	0.3	0.4	0.5
50,000	–	0.06	0.09	0.14	0.2	0.2	0.3	0.3
75,000	–	0.05	0.07	0.11	0.2	0.2	0.2	0.3
85,000	–	0.05	0.07	0.11	0.15	0.2	0.2	0.2

TABLE VII. Standard Errors of Estimated Percentages of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water-Individual Well, Lacking Complete Plumbing, Mobile Homes, and Housing Units With Householder of Spanish Origin: 1981

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	36.1	36.1	36.1	36.1	36.1	36.1	36.1	37.6
10	22.0	22.0	22.0	22.0	22.0	22.0	23.0	26.6
25	10.2	10.2	10.2	10.2	10.2	12.0	14.6	16.8
50	5.3	5.3	5.3	5.3	7.1	8.5	10.3	11.9
100	2.7	2.7	2.7	3.7	5.0	6.0	7.3	8.4
250	1.1	1.1	1.5	2.3	3.2	3.8	4.6	5.3
500	0.6	0.7	1.1	1.6	2.3	2.7	3.3	3.8
1,000	0.3	0.5	0.7	1.2	1.6	1.9	2.3	2.7
2,500	0.11	0.3	0.5	0.7	1.0	1.2	1.5	1.7
5,000	0.06	0.2	0.3	0.5	0.7	0.8	1.0	1.2
7,500	0.04	0.2	0.3	0.4	0.6	0.7	0.8	1.0
10,000	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.8
25,000	0.01	0.11	0.15	0.2	0.3	0.4	0.5	0.5
50,000	0.01	0.07	0.11	0.2	0.2	0.3	0.3	0.4

TABLE VIII. Standard Errors of Estimated Percentages of Lost Housing Units: 1973-1981 (Excluding Estimated Percentages of Lost Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Some or All Plumbing, Mobile Homes, and Other Vacants)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	27.7	27.7	27.7	27.7	27.7	27.7	27.7	30.9
10	16.1	16.1	16.1	16.1	16.1	16.1	18.9	21.9
25	7.1	7.1	7.1	7.1	8.3	9.9	12.0	13.8
50	3.7	3.7	3.7	4.3	5.9	7.0	8.5	9.8
100	1.9	1.9	1.9	3.0	4.1	4.9	6.0	6.9
250	0.8	0.9	1.2	1.9	2.6	3.1	3.8	4.4
500	0.4	0.6	0.9	1.3	1.9	2.2	2.7	3.1
1,000	0.2	0.4	0.6	1.0	1.3	1.6	1.9	2.2
2,500	0.08	0.3	0.4	0.6	0.8	1.0	1.2	1.4
5,000	0.04	0.2	0.3	0.4	0.6	0.7	0.8	1.0

TABLE IX. Standard Errors of Estimated Percentages of Lost Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Some or All Plumbing, Mobile Homes, and Other Vacants: 1973-1981

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	39.2	39.2	39.2	39.2	39.2	39.2	39.2	40.1
10	24.3	24.3	24.3	24.3	24.3	24.3	24.6	28.4
25	11.4	11.4	11.4	11.4	11.4	12.8	15.5	17.9
50	6.0	6.0	6.0	6.0	7.6	9.1	11.0	12.7
100	3.1	3.1	3.1	3.9	5.4	6.4	7.8	9.0
250	1.3	1.3	1.6	2.5	3.4	4.1	4.9	5.7
500	0.6	0.8	1.1	1.7	2.4	2.9	3.5	4.0
1,000	0.3	0.6	0.8	1.2	1.7	2.0	2.5	2.8
2,500	0.13	0.4	0.5	0.8	1.1	1.3	1.6	1.8
5,000	0.06	0.3	0.4	0.6	0.8	0.9	1.1	1.3

TABLE Xa. Standard Errors of Estimated Percentages of Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1981 (Excluding Estimated Percentages Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing Facilities for the Northeast, North Central, and West Regions and Exlcuding Source of Water-Individual Well and Mobile Homes for Each of the Regions)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	30.6	30.6	30.6	30.6	30.6	30.6	30.6	33.2
10	18.1	18.1	18.1	18.1	18.1	18.1	20.3	23.5
25	8.1	8.1	8.1	8.1	8.9	10.6	12.9	14.8
50	4.2	4.2	4.2	4.6	6.3	7.5	9.1	10.5
100	2.2	2.2	2.2	3.2	4.5	5.3	6.4	7.4
250	0.9	0.9	1.3	2.0	2.8	3.4	4.1	4.7
500	0.4	0.7	0.9	1.4	2.0	2.4	2.9	3.3
1,000	0.2	0.5	0.7	1.0	1.4	1.7	2.0	2.3
2,500	0.09	0.3	0.4	0.6	0.9	1.1	1.3	1.5
5,000	0.04	0.2	0.3	0.5	0.6	0.7	0.9	1.0
7,500	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.9
10,000	0.02	0.15	0.2	0.3	0.4	0.5	0.6	0.7
25,000	0.01	0.09	0.13	0.2	0.3	0.3	0.4	0.5

TABLE Xb. Standard Errors of Estimated Percentages of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing Facilities for the Northeast, North Central, and West Regions and to Source of Water-Individual Well, and Mobile Homes for the Northeast and North Central Regions: 1981

(68 chances out of 100. For estimated percentages pertaining to source of water-individual well and mobile homes for the West Region, apply a factor of 1.66 to the standard errors)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	38.5	38.5	38.5	38.5	38.5	38.5	38.5	39.6
10	23.9	23.9	23.9	23.9	23.9	23.9	24.2	28.0
25	11.1	11.1	11.1	11.1	11.1	12.6	15.3	17.7
50	5.9	5.9	5.9	5.9	7.5	8.9	10.8	12.5
100	3.0	3.0	3.0	3.9	5.3	6.3	7.7	8.8
250	1.2	1.2	1.6	2.4	3.4	4.0	4.8	5.6
500	0.6	0.8	1.1	1.7	2.4	2.8	3.4	4.0
1,000	0.3	0.6	0.8	1.2	1.7	2.0	2.4	2.8
2,500	0.13	0.4	0.5	0.8	1.1	1.3	1.5	1.8
5,000	0.06	0.2	0.4	0.5	0.8	0.9	1.1	1.3
7,500	0.04	0.2	0.3	0.4	0.6	0.7	0.9	1.0
10,000	0.03	0.2	0.2	0.4	0.5	0.6	0.8	0.9
25,000	0.01	0.11	0.2	0.2	0.3	0.4	0.5	0.6

TABLE Xc. Standard Errors of Estimated Percentages of Housing Units Pertaining to Source of Water-Individual Well and Mobile Homes for the South Region: 1981

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	49.7	49.7	49.7	49.7	49.7	49.7	49.7	49.7
10	33.1	33.1	33.1	33.1	33.1	33.1	33.1	35.2
25	16.5	16.5	16.5	16.5	16.5	16.5	19.3	22.2
50	9.0	9.0	9.0	9.0	9.4	11.2	13.6	15.7
100	4.7	4.7	4.7	4.8	6.7	7.9	9.6	11.1
250	1.9	1.9	2.0	3.1	4.2	5.0	6.1	7.0
500	1.0	1.0	1.4	2.2	3.0	3.6	4.3	5.0
1,000	0.5	0.7	1.0	1.5	2.1	2.5	3.0	3.5
2,500	0.2	0.4	0.6	1.0	1.3	1.6	1.9	2.2
5,000	0.10	0.3	0.4	0.7	0.9	1.1	1.4	1.6
7,500	0.07	0.3	0.4	0.6	0.8	0.9	1.1	1.3
10,000	0.05	0.2	0.3	0.5	0.7	0.8	1.0	1.1
25,000	0.02	0.14	0.2	0.3	0.4	0.5	0.6	0.7

TABLE Xd. Standard Errors of Estimated Percentages of Lost Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1973-1981 (Excluding Estimates of Lost Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Some or All Plumbing, Mobile Homes, and Other Vacants)

(68 chances out of 100)

Size of estimate (000)	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
	Estimated percentages for the Northeast or North Central							
5	27.1	27.1	27.1	27.1	27.1	27.1	27.1	30.5
10	15.6	15.6	15.6	15.6	15.6	15.6	18.6	21.5
25	6.9	6.9	6.9	6.9	8.2	9.7	11.8	13.6
50	3.6	3.6	3.6	4.2	5.8	6.9	8.3	9.6
100	1.8	1.8	1.9	3.0	4.1	4.9	5.9	6.8
250	0.7	0.9	1.2	1.9	2.6	3.1	3.7	4.3
500	0.4	0.6	0.9	1.3	1.8	2.2	2.6	3.0
1,000	0.2	0.4	0.6	0.9	1.3	1.5	1.9	2.2
2,500	0.07	0.3	0.4	0.6	0.8	1.0	1.2	1.4
5,000	0.04	0.2	0.3	0.4	0.6	0.7	0.8	1.0
	Estimated percentages for the South or West							
5	31.5	31.5	31.5	31.5	31.5	31.5	31.5	33.9
10	18.7	18.7	18.7	18.7	18.7	18.7	20.8	24.0
25	8.4	8.4	8.4	8.4	9.1	10.8	13.1	15.2
50	4.4	4.4	4.4	4.7	6.4	7.7	9.3	10.7
100	2.2	2.2	2.2	3.3	4.5	5.4	6.6	7.6
250	0.9	1.0	1.3	2.1	2.9	3.4	4.2	4.8
500	0.5	0.7	0.9	1.5	2.0	2.4	2.9	3.4
1,000	0.2	0.5	0.7	1.0	1.4	1.7	2.1	2.4
2,500	0.09	0.3	0.4	0.7	0.9	1.1	1.3	1.5
5,000	0.05	0.2	0.3	0.5	0.6	0.8	0.9	1.1

Note: For standard errors of regional estimated percentages of lost housing units (1973-1981) pertaining to new construction, lacking complete kitchen facilities, no bedrooms, no bathrooms, lacking some or all plumbing, mobile homes, and other vacants, use the national standard errors in table IX.

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-45
Selection of sample areas. . . . App-45
Designation of sample housing for the 1981 survey . . . App-45
Selection of the 1973 sample housing units. . . App-46
Splitting of the sample . . . App-46
CEN-SUP sample . . . App-46
Selection of supplemental sample housing units in rural areas . . . App-46
Selection of sample housing units for the 1976 Coverage Improvement Program . . . App-46
1977 and 1981 sample reductions . App-47
Supplemental sample from nonresidential conversions . . . App-47
1981 telephone interviewing experiment . . . App-47
ESTIMATION . . . App-47
RELIABILITY OF THE ESTIMATES App-49
Sampling errors . . . App-49
Standard errors of estimates of levels . . . App-49
Standard errors of estimates of percentages. . . App-49
Standard errors of ratios . . . App-49
Illustration of the use of the standard error tables. Illustration I . . . App-50
Illustration II . . . App-50
Differences. . . App-51
Illustration of the computation of the standard error of a difference . . . App-51
Medians . . . App-51
Illustration of the computation of the 95-percent confidence interval for a median. App-51
Nonsampling errors . . . App-52
Reinterview program . . . App-52
Coverage errors . . . App-52
Effects of ratio estimation on the estimates of change . . . App-53
Possible effects of telephone interviewing on the data . . . App-53
Rounding errors. . . App-53
Standard error tables . . . App-54

SAMPLE DESIGN

The 1981 estimates are based on data collected from September 1981 through December 1981 for the Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. The sample for this survey was spread over 461 sample areas (called primary sampling units), comprising 923 counties and independent cities with coverage in each of the 50 States and the District of Columbia.

Approximately 56,800 sample housing units (both occupied and vacant) were eligible for interview in the 1981 Annual Housing Survey. Of this number, 3,200 interviews were classified as "noninterview" for various reasons. Occupied housing units were classified as "noninterview" mainly because the occupants refused to be interviewed after repeated calls. For vacant housing units, interviews were not obtained because an informed respondent was not found after repeated visits. In addition to the 56,800 eligible housing units, there were also 3,600 sample units which were visited but were ineligible for interview for the AHS in terms of collecting information relevant to the 1981 housing inventory.

Selection of sample areas—The United States was divided into areas made up of counties and independent cities referred to as primary sampling units (PSU's). These PSU's were then grouped into 376 strata, 156 of which consisted of only one PSU which was in sample with certainty. These 156 strata were mostly the larger SMSA's and were called self-representing (SR) since the sample from the sample area represented just that PSU. Each one of the other 220 strata consisted of a group of PSU's and were referred to as non-self-representing (NSR), since the sample of housing units from the sample PSU in a stratum represented the other PSU's in the stratus as well.

One PSU was selected from each NSR stratum with probability proportionate to the 1970 census population of the PSU. (This resulted in 220 NSR sample PSU's.) In addition, the NSR strata were grouped into 110 pairs and one stratum was picked at random from each pair. From this stratum, an additional PSU was selected independently of the other PSU selected from this stratum. Since the two PSU's were independently selected, it was possible for the same PSU to be selected twice. This occurred in 25 instances, producing an additional 85 NSR sample PSU's, thus giving a grand total of 461 PSU's.

Designation of sample housing units for the 1981 survey—The sample housing units designated to be interviewed in the 1981 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1980 survey (which included all sample housing units that were selected as part of the 1976 Coverage Improvement Program) and were not part of the 1981 sample reduction.

2. All sample housing units that were either type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of survey but which could become eligible in the future) in the 1980 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1981 AHS questionnaire, page App-21).
3. All sample housing units that were selected from the list of building permits issued since the 1980 survey. (This sample represented the housing units built in permit-issuing areas, since the 1980 survey.)
4. Housing units added as the result of the updated listings in selected areas which do not issue building permits.

Selection of the 1973 sample housing units—The overall sampling rate used to select the sample for the 1973 AHS was about 1 in 1,366. The within-PSU sampling rate for the AHS was determined so that the overall probability of selection for each sample housing unit was the same (e.g., if the probability of selecting a NSR PSU was 1 in 10, then the within-PSU sampling rate would be 1 in 136.6).

Within the sample PSU's, a sample of the housing units enumerated in the 1970 Census of Population and Housing was selected for the AHS. In addition, a sample of new construction building permits was also selected to represent the units constructed since the 1970 census. These samples were selected at about twice the rate mentioned previously (i.e., at 2 in 1,366), thereby producing a sample twice as large as needed. This sample was split into two equal-sized samples—one to be used for the AHS, and one to be held in reserve for possible future use for the AHS. The procedure used to split this sample into equal-sized samples is described in the next section.

The sample of 1970 census units was selected in several stages. Within the sample PSU's, the first step was the selection of a sample of census enumeration districts (ED's), administrative units used in the 1970 census. The probability of selection for an ED was proportional to the following 1970 census counts of housing units (HU's) and persons in group quarters, combined in the following formula:

$$\frac{\text{Number of HU's in the ED} + \dfrac{\text{Number of group quarters persons in the ED}}{3}}{4}$$

The next step was to select an expected cluster of about four neighboring housing units within each sample ED. For most of the ED's, the selection was accomplished using the list of addresses for the ED as compiled in the 1970 census. (These ED's are referred to as address ED's.) However, in those ED's where addresses were incomplete or inadequate (mostly rural areas), the selection process was accomplished using area sampling methods. These ED's were divided into segments (i.e., small land areas with well-defined boundaries, having an expected size of four, or a multiple of four, housing units) and a segment was selected. Those selected segments with an expected size which was a multiple of four were further subsampled at the time of interview so that an expected four housing units were chosen for interview.

The sample of new construction housing units was selected from building permits issued since January 1970. Within each sample PSU, the building permits were chronologically ordered by month issued, and compact clusters of approximately four housing units were created. These clusters were then sampled at the rate of 2 in 1,366. Housing units constructed since the 1970 census in areas which do not issue building permits were brought into the sample as a result of the area sample described.

Splitting of the sample—The described sample selection procedure produced clusters (or segments) of size-four housing units for the sample taken from the census address frame, the new construction frame, and the area sampling frame (mainly rural areas). Clusters of this size should result in a minimum loss in precision for estimates of housing characteristics in rural areas because of the heterogeneity of neighboring units. However, clusters of size-two housing units were considered to be more optimum within those areas where the housing characteristics of neighboring units tend to be very similar (i.e., urban areas and new construction units). A splitting operation was then carried out for clusters selected from the census address and the new construction frames. This consisted of halving each sample cluster from these frames. Thus, two housing units from each of these clusters were included in the survey and two housing units were held in reserve. No splitting operation was carried out within the clusters selected from the area sampling frame; every other area sample cluster of four housing units was used for the survey and the remaining clusters were assigned to the reserve sample.

CEN-SUP sample—Housing units at addresses missed in the census or inadequately described in the 1970 census address register did not have a chance of being selected for the AHS sample in address ED's. Consequently, a special operation was undertaken to develop a separate sample, called the census supplemental (CEN-SUP) sample to represent these units. Due to time constraints on this operation, only about 40 percent of the CEN-SUP units were ready to be interviewed in 1973. The rest were interviewed for the first time in the 1974 AHS.

Selection of supplemental sample housing units in rural areas—In 1974, it was decided to increase the reliability of the AHS estimates of rural housing characteristics by doubling the number of sample housing units from rural areas. This was accomplished by reactivating the reserve sample, selected in the original sampling operations in 1973, from rural areas only. For the reserve sample selected in census address and new construction frames, the other half of each rural cluster (an expected two housing units) was reactivated in 1974. Similarly, for the area sampling frame, the entire reserve cluster (an expected four housing units) was reactivated in 1974 if the cluster was rural. This supplementation increased the overall probability of selection for sample housing units in rural areas to about 2 in 1,366; whereas, the overall probability of selection for sample housing units in urban areas remained at 1 in 1,366.

Selection of sample housing units for the 1976 Coverage Improvement Program—The 1976 Coverage Improvement Pro-

gram was undertaken to correct certain deficiencies in the AHS national sample from the census address and new construction frames. The coverage deficiencies included the following types of housing units:

1. New construction housing units from building permits issued prior to January 1970 for which construction had not been completed at the time of the 1970 census.
2. Housing units converted to residential use in structures totally nonresidential at the time of the 1970 census.
3. Houses that have been moved onto their present site since the 1970 census.
4. Mobile homes and trailers placed in parks either missed in the 1970 census or established since the 1970 census.
5. Mobile homes and trailers placed outside parks since the 1970 census or vacant at the time of the 1970 census.

A sample of new construction housing units whose permits were issued before January 1970 was selected in two stages. First, units whose permits were issued before January 1970, but which were completed after the census, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. Second, these units were then sampled so that the overall probability of selection was about 1 in 1,320.

A sample of mobile homes and trailers placed in parks missed by the census or established after the census was also selected in two stages. During the first stage, a list of mobile home parks was obtained from commercial listings. This list was then supplemented by additional parks identified by a canvassing operation similar to that performed in ED's where area sampling methods are used. The second stage consisted of dividing the parks into clusters of an expected size of four sites. These clusters were then sampled so that the overall probability of selection was about 1 in 1,366.

For the remaining housing units (i.e., mobile homes and trailers placed outside parks since the 1970 census, mobile homes and trailers vacant at the time of the 1970 census, units converted from nonresidential to residential use since the 1970 census, and houses that had been moved onto their present site since the 1970 census), the sampling was done in three stages. First, a subsample of the regular AHS sample housing units from the census address frame was selected. Second, succeeding structures that had been eligible to be selected from the census address frame were then listed until eight structures (excluding mobile home parks) were found. Finally, the intervening structures that had been listed which did not have a chance of selection in the AHS were identified and the units within these structures were interviewed.

1977 and 1981 sample reductions—By 1977, the addition to the sample, primarily from new construction and the coverage improvements, had increased the total sample size (interviews plus noninterviews) to about 81,000 housing units. The sample was reduced by about 7 percent to approximately 75,000 in 1977. This reduction did not include any CEN-SUP units or units which were selected as part of the 1976 Coverage Improvement Program. Initially, the sample was further reduced in 1981 by about 5 percent. Subsequently, the rural sample was reduced by about 25 percent in 125 "large" SMSA's[1] and by about 50 percent in "small" SMSA's[1] and outside the SMSA's. These reductions brought the sample size down to approximately 60,000 housing units in 1981.

Supplemental sample from nonresidential conversions—The purpose of this sample was to improve the coverage in address ED's (i.e., ED's in which 1970 census address listings were used for selecting the AHS sample) of housing units created in structures that were completely nonresidential at the time of the 1970 census. This sample was derived from listings created for the Commercial Victimization Survey (CVS), a nationally representative area sample survey conducted by the Census Bureau for the Department of Justice. Each of the CVS area segments was initially canvassed either in 1972 or 1975 and eligible business establishments were listed. These listings were updated every 6 months with the list updating in the last half of 1977. Each basic address containing a business establishment listed for CVS was visited for the 1980 Components of Inventory Change Survey (CINCH), which was conducted in conjunction with the 1980 AHS, to identify HU's in structures which were completely nonresidential at the time of the 1970 census. These cases were later matched to the 1970 census address listings to identify those cases in address ED's and as an additional check to see if housing units existed in these structures at the time of the 1970 census. Due to the timing restraints associated with these operations, these cases were not included for the 1980 AHS but were included for the 1981 survey.

1981 telephone interviewing experiment—A large scale telephone interviewing experiment was conducted in conjunction with the 1981 AHS national sample in order to provide more definitive information about the effect of telephone interviewing on the AHS data and cost. A random sample of about 12,500 AHS national sample units was selected for the experiment. Among the cases assigned for telephone interviewing, only those sample housing units that had been interviewed in 1980 and had a telephone number available were eligible to be interviewed by telephone. Since a large portion of the total AHS national sample was assigned for telephone interviewing, its possible effect on the data is mentioned in the section "Reliability of the Estimates" of this appendix.

ESTIMATION

In 1981, the AHS estimates employed a three-stage ratio estimation procedure. However, prior to implementation of the procedure, the basic weight (i.e., the inverse of the probability of selection) was adjusted to account for the type A noninterview housing units encountered in the AHS. This noninterview

[1] A "large" SMSA is one with a 1970 population of 250,000 persons or more and a "small" SMSA is one with a 1970 population of less than 250,000 persons.

adjustment was done separately for occupied and vacant units. The noninterview adjustment was equal to the following ratio:

$$\frac{\text{Interviewed housing units + Noninterviewed housing units}}{\text{Interviewed housing units}}$$

The first-stage ratio estimation procedure was employed for sample housing units from non-self-representing (NSR) PSU's only. This procedure was designed to reduce the contribution to the variance arising from the sampling of PSU's. The first-stage ratio estimation procedure takes into account the differences that existed at the time of the 1970 census in the distribution by tenure and residence of the housing population estimated from the sample NSR PSU's and that of the NSR housing population in each of the four census regions of the country.

The first-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{The 1970 census housing population in the residence-tenure category for all NSR strata in a census region}}{\text{Estimate of the housing population category using 1970 census housing counts for sample NSR PSU's in a census region}}$$

The numerators of the ratios were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR stratum and summing these counts across the NSR strata in each census region. The denominators were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR sample PSU, weighting these counts by the inverse of the probability of selecting that PSU, and summing these weighted counts across the NSR sample PSU's in each census region. The computed first-stage ratio estimation factor was then applied to the existing weight for each NSR sample unit in each first-stage ratio estimation category.

The second-stage ratio estimation procedure was designed to adjust the AHS sample estimate of conventional new construction housing units, i.e., sample units built April 1, 1970, or later, to independently derived current estimates where a known deficiency in the AHS sample exists (see the section on nonsampling error) for each of the four regions. These estimates were considered to be the best estimates available for the number of conventional new construction units in these categories.

The second-stage ratio estimation factor was as follows:

$$\frac{\text{Current best estimate of new construction in the category}}{\text{AHS sample estimate of new construction units in the category}}$$

The numerators of the ratios were derived from data based on the Survey of Construction (SOC). The denominators of the ratios were obtained from the weighted estimates for the AHS sample units using the existing weight after the first-stage ratio estimates procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in each second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all the AHS sample units. This procedure was designed to adjust the AHS sample estimates of housing (i.e., the estimates employing the noninterview, first-stage, and second-stage adjustments) to current vacant housing estimates for 4 categories of vacant housing units and to independently derived current housing estimates for 24 categories of occupied housing units. Each of these categories is a combination of the characteristics of residence, tenure, race of householder, and sex of householder.

The third-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{Current independent estimate of housing units in the category}}{\text{AHS sample estimate of housing units in the category}}$$

The numerators of the ratios for occupied housing units were derived from data based on the Current Population Survey (CPS), a sample household survey conducted monthly by the Bureau of the Census. This is the first time that CPS household estimates based on the 1980 census were used as the independent controls in this adjustment rather than the 1970 census-based controls. The 1970-based estimates were about 2 percent smaller than the 1980-based estimates. This difference should be taken into consideration when comparing the 1981 AHS estimates to the AHS estimates for 1980 or earlier years. The numerators of the ratios for vacant housing units were derived from data based on the Housing Vacancy Survey (HVS), a quarterly vacancy survey also conducted by the Bureau of the Census. The denominators of the ratios were obtained from the weighted estimates for the AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computer third-stage ratio estimation factor was then applied to the existing weight for each sample unit in each third-stage ratio estimation category.

The second-stage and the third-stage ratio estimation procedures were iterated in order to bring the AHS estimates into close agreement with both sets of independent estimates. The second stage was modified so that the estimates for all 27 categories of new construction would be identical to the estimates before the third stage. Hence, the repeated second stage had the effect of controlling the AHS sample estimates of new construction units to the unbiased sample estimates for 9 categories of new construction mobile homes and trailers for each of the four regions and of adjusting the AHS sample estimate of 18 categories of conventional new construction units to the independently derived current estimates.

The numerators were either the unbiased weighted estimates for the AHS sample housing units, using the existing weight after the first-stage ratio estimation procedure (i.e., the estimates employing the noninterview and first-stage adjustments) or the independent estimate derived from data based on the Survey of Construction (SOC).

The denominators of the ratios in this iterative process were obtained from the weighted estimates for the AHS sample units after the previous stage of ratio estimation. The factors resulting from this iterative process were then applied to the existing weight on the appropriate records, and the resulting product was used as the final weight for tabulation.

The third-stage ratio estimation procedure, as well as the overall estimation procedure, reduced the sampling error for most statistics below what would have been obtained by simply

weighting the results of the sample by the inverse of the probability of selection. The distribution of the housing population selected for the sample differed somewhat, by chance, from that of the Nation as a whole in such basic housing characteristics as tenure, vacancy status, residence, race of householder, and sex of householder. These characteristics are probably closely correlated with other housing characteristics measured for the AHS. Therefore, through the use of the three-stage ratio estimation procedure one can expect the sample estimate to be improved substantially.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS national sample.

Sampling errors—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same questionnaires, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The variability between estimates from all possible samples is defined as sampling error. One common measure of sampling error is the standard error which measures the precision with which an estimate from a sample approximates the average result of all possible samples. In addition, the standard error, as calculated for this report, also partially reflects the variation in the estimates due to some nonsampling errors, but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, and biases and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates so that the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, and each of these samples was surveyed under essentially the same general conditions and an estimate and its estimated standard error were calculated for each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples either is or is not contained in any particular computed interval. However, for a particular sample one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables (pages App-54 to App-56) of this appendix are approximations to the standard errors of various estimates shown in this report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than the precise standard error for any specific item.

Standard errors of estimates of levels—Tables I and II present the standard errors applicable to the 1981 national housing inventory estimates in this report. Tables III and IV present the standard errors applicable to estimates for the Northeast, North Central, and West Regions, and table III presents the standard errors applicable to estimates for the South Region. Linear interpolation should be used to determine standard errors for levels of estimates not specifically shown in tables I through IV.

Standard errors of estimates of percentages—The reliability of an estimated percentage, computed by using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more.

Tables V through VIII present the standard errors of estimated percentages. Table V shows the approximate standard errors of all national estimated percentages of housing units except those pertaining to the specified items in table II. The standard errors shown in table VI should be used for those specified items. Table VII shows the approximate standard errors of all regional estimated percentages of housing units except those pertaining to the specified items in table IV. The standard errors shown in table VIII should be used for those specified items. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables V through VIII.

Included in tables I through VIII are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered to be overestimates of the true standard errors and should be used primarily for construction of confidence intervals for characteristics when an estimate of zero is obtained.

Standard errors of ratios—For ratios of the form (100) (x/y), where x is not a subclass of y, tables V through VIII underestimate the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be obtained by letting the standard error of the ratio be approximately equal to:

$$(100)\left(\frac{x}{y}\right)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

where: x = the numerator of the ratio
y = the denominator of the ratio
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

Illustration of the use of the standard error tables. Illustration I—Tables A-2 (section 1 of this report shows that in the United States there were 14,678,000 renter-occupied housing units with common stairways in 1981. Interpolation in standard error table 1 shows that the standard error of an estimate of this size is approximately 153,000. The following interpolating procedure was used.

The information presented in the following table was extracted from standard error table I. The entry for "x" is the one sought.

Size of estimate (000)	Standard error (000)
10,000	135
14,678	x
25,000	193

By vertically interpolating between 135 and 193 the entry for "x" is determined to be 153

$$14{,}678-10{,}000 = 4{,}678$$
$$25{,}000-10{,}000 = 15{,}000$$
$$135 + \frac{4{,}678}{15{,}000}(193-135) = 153$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 14,525,000 to 14,831,000 housing units. Therefore, a conclusion that the average estimate of 1981 housing units of this type lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all possible samples, lies within the interval from 14,433,000 to 14,923,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 14,372,000 to 14,984,000 housing units with 95 percent confidence.

Table A-2 (section 1) also shows that of the 14,678,000 renter-occupied housing units with common stairways, 13,045,000, or 88.9 percent, were located inside SMSA's. Interpolation in standard error table V (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the above percentage is 0.4 percentage points. The following interpolating procedure was used.

The information presented in the following table was extracted from standard error table V. The entry for "p" is the one sought.

Base of percentage (000)	Estimated percentage		
	85	88.9	90
10,000	0.5	a	0.4
14,678		p	
25,000	0.3	b	0.3

1. By horizontal interpolation between 0.5 and 0.4 the entry for cell "a" is determined to be 0.4.

$$88.9-85.0 = 3.9$$
$$90.0-85.0 = 5.0$$
$$0.5 + \frac{3.9}{5.0}(0.4-0.5) = 0.4$$

2. Horizontal interpolation between 0.3 and 0.3 is not necessary.

3. By vertical interpolation between 0.4 and 0.3 the entry for "p" is determined to be 0.4.

$$14{,}678-10{,}000 = 4{,}678$$
$$25{,}000-10{,}000 = 15{,}000$$
$$0.4 + \frac{4{,}678}{15{,}000}(0.3-0.4) = 0.4$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 88.5 to 89.3 percent; the 90-percent confidence interval is from 88.3 to 89.5 percent; and the 95-percent confidence interval is from 88.1 to 89.7 percent.

Illustration II—Table A-2 (section 1) of this report shows that in the United States in 1981 there were 6,020,000 owner-occupied housing units which had blown fuses or tripped breaker switches. Interpolation in standard error table 1 of this appendix shows that the standard error of an estimate of this size is approximately 107,000. Consequently, the 68-percent confidence interval is from 5,913,000 to 6,127,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1981 owner-occupied housing units which had blown fuses or tripped breaker switches, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all possible samples, lies within the interval from 5,849,000 to 6,191,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 5,806,000 to 6,234,000 housing units with 95 percent confidence.

Table A-2 (section 1) also shows that of the 6,020,000 owner-occupied housing units in 1981 which had blown fuses or tripped breaker switches, 1,256,000, or 20.9 percent, had blown fuses or tripped breaker switches three times or more. Interpolation in standard error table V (i.e., interpolation on both the base and the percent) shows that the standard error of the above

percentage is 0.8 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 20.1 to 21.7 percent; the 90-percent confidence interval is from 19.6 to 22.2 percent; and the 95-percent confidence interval is from 19.3 to 22.5 percent.

Differences—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristics in two different areas or the difference between separate and uncorrelated characteristics in the same area. If there is a high positive correlation between the two characteristics, the formula will overestimate the true error; if there is a high negative correlation between the two characteristics, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference—Table A-2 (section 2) of this report shows that in the United States in 1981 there were 3,422,000 owner-occupied housing units, which had exactly one blown fuse or tripped breaker switch. Table A-2 (section 2) also shows that in the United States in 1981 there were 1,256,000 owner-occupied housing units which had blown fuses or tripped breaker switches three times or more. Thus, the apparent difference between the number of 1981 owner-occupied housing units that had blown fuses or tripped breaker switches three times or more and the number that had breakdowns just one time, is 2,166,000. Interpolation in standard error table 1 shows that the standard error on an estimate of 3,422,000 to be approximately 81,000 and the standard error on an estimate of 1,256,000 to be approximately 49,000. Therefore, the standard error of the estimated difference of 2,166,000 is about 95,000.

$$95{,}000 = \sqrt{(81{,}000)^2 + (49{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 2,166,000 difference is from 2,071,000 to 2,261,000 housing units. Therefore, a conclusion that the average estimate of this difference, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 2,014,000 to 2,318,000 housing units, and the 95-percent confidence interval is from 1,976,000 to 2,356,000. Thus, we can conclude with 95 percent confidence that the number of 1981 owner-occupied housing units, which had three or more blown fuses or tripped breaker switches, is different than the number that had exactly one blown fuse or tripped breaker switch since the 95-percent confidence interval of this difference does not include zero or negative values.

Medians—For medians, the sampling error depends on the size of the base and the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From the appropriate standard error tables, determine the standard error of a 50-percent characteristic on the base of the median.
2. Add to and subtract from 50 percent the standard error determined in step 1. This will give you a lower percentage limit (50 percent minus standard error of 50 percent) and an upper percentage limit (50 percent plus standard error of 50 percent).
3. Using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of a 95-percent confidence interval for a median—Table A-5 (section 2) of this report shows the median value of specified owner-occupied housing units with one bathroom was $40,700 in 1981. The base of the distribution from which this median was determined is 19,342,000 housing units.

1. From standard error table V, the standard error of a 50-percent characteristic on the base of 19,342,000 is approximately 0.6 percentage points.
2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 48.8 and 51.2.
3. From table A-5 (section 2), it can be seen by cumulating the frequencies for the first four categories that 9,403,000 specified owner-occupied housing units with one bathroom, or 48.6 percent, had a value less than $40,000 and that an additional 3,607,000 specified owner-occupied housing units with one bathroom, or 18.6 percent, had a value between $40,000 and $49,999. By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$\$40{,}000 + (\$50{,}000 - \$40{,}000)\ \frac{(48.8 - 48.6)}{18.6} = \$40{,}108$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$\$40{,}000 + (\$50{,}000 - \$40{,}000)\ \frac{(51.2 - 48.6)}{18.6} = \$41{,}398$$

Thus, the 95-percent confidence interval ranges from $40,108 to $41,398.

Nonsampling errors—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases; definitional difficulties; differences in the interpretation of questions; inability or unwillingness to provide correct information on the part of respondents; mistakes in recording or coding the data; and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for the 1981 AHS national sample.

Reinterview program—For the AHS national sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a subsample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and thus were the basis for the measurement of the "content" error of these AHS estimates.

As part of the reinterview, an additional check was carried out for interviewer evaluation and quality control. This check was made at each of these households to determine if the following were done during the original interview.

1. The correct housing unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

The results of the 1981 and 1980 reinterview studies were not available at the time of publication. In addition, the 1979 reinterview was carried out as an interviewer quality control only. Thus, the results of the 1977 and 1978 reinterview studies which were presented in the Census Bureau memoranda, "Reinterview Results for the Annual Housing Survey—National Sample 1977" and "Reinterview Results for the Annual Housing Survey—National Sample 1978" are presented here.

In 1978, a substantial portion of the reinterview questionnaire was devoted to testing the new questions 1 through 7 (parts a and b). The questions (part a), which were asked only at housing units interviewed in the previous year, determined whether there had been a change since last year in selected nonattitudinal items. If a change had been recorded or the respondent did not know if a change had occurred, part b of the question, which collects the value of the item, was asked. The reinterviewers asked these items using the questions as formatted in 1977. Comparing the responses from the differently formatted questions, the 1978 reinterview found that 80 percent of the questions showed low levels of inconsistency with the remainder showing moderate levels.

The 1977 reinterview program showed moderate to high levels of inconsistency with about 21 percent of the nonattitudinal and 56 percent of the attitudinal items showing high levels of inconsistency. A large portion (43 percent) of the nonattitudinal items showed a low level of inconsistency. Moderate levels indicate that there are some problems with inconsistent reporting and high levels indicate that improvements are needed in the data collection methods or that the category concepts themselves are ambiguous.

Cross-tabulations involving those items, which are subject to substantial levels of inconsistency may be subject to a large distortion as a consequence of the associated high level of response variance, and thus, are considered to be less reliable than comparable cross-tabulations which do not involve these data. As a consequence, the tables in this report, which contain such cross-tabulations, have been footnoted with a cautionary statement.

The 1970 census reinterview results provide illustrations of possible nonsampling errors for some of the items which also appear in the AHS. For example, median value of homes was consistently underestimated by about 5 percent, and the average monthly cost of electricity and utility gas were consistently overestimated although the net effect on average gross rent was fairly small.

A possible explanation for the results of the AHS and census reinterview studies, as well as the surveys themselves, is that the respondents may lack precise information. Also, because the results of the reinterview studies are derived from sample surveys, there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of these studies.

Coverage errors—A deficiency in the representation of conventional new construction (non-mobile home or trailer) for the AHS new construction sample (mentioned previously in the section on estimation) is an example of coverage errors. During the sampling of building permits, only those issued more than 4 months before the survey began were eligible to be selected to represent conventional new construction. Due to time constraint it is not possible to sample units whose permits are issued less than 4 months in advance of the survey.

It is estimated that the 1981 AHS sample missed about 0.8 percent (i.e., about 157,000 units) of conventional housing units built after April 1970 because the permits for these units, which were built before October 1981, were issued less than 4 months in advance of the survey. The second-stage ratio estimation procedure was employed to reduce the effect of this deficiency although some bias in the AHS estimates of conventional new construction probably still exists. Review of the

second-stage ratio estimation procedure indicates that we have consistently overcompensated for this deficiency every year since 1975 by ratio adjusting the new construction to counts of new construction for the end of the interview period, which has been December or January, instead of October. This overcompensation may inflate the new construction counts by 100,000 to 300,000 units.

Changing this procedure to correct for this overcompensation would reduce the estimate of housing units built since the last survey since it would be based on a 10-month time period rather than 12 months. Thus, it was decided not to change this procedure so that the estimate of housing units built since the last survey would be consistent with previous years' estimates, which are generally based on a 12-month time period.

In addition, the 1976 Coverage Improvement Program had certain deficiencies. First, when the canvassing was done to identify mobile homes and trailer parks that were not in the same sample frame or not on the commercial lists, only 92 percent of the census address frame ED's were represented. Second, it appears that the listing procedure (used to find mobile homes and trailers placed outside parks, units converted from nonresidential to residential, and houses that had been moved onto their present site) was not very efficient for finding nonresidential conversions (which might be primarily in business districts), since the listing procedure started from a residential unit. (The sample estimate of this component was approximately 16,000 housing units with a standard error of 12,000.)

Finally, it is felt that deficiencies also exist in ED's where area sampling methods are used. As before, it had been assumed that all housing units located inside these ED's would be represented in the sample. However, it has been estimated that the 1981 AHS sample missed as much as 2 percent (i.e., as much as 400,000 units) of all housing units in ED's where area sampling methods are used because these units are not listed during the canvassing.

The third-stage of ratio estimation corrects for these deficiencies as far as the total count of housing is concerned, i.e., it adjusts the estimate of the total housing inventory to the best available estimate. However, biases of subtotals would still remain.

Effects of ratio estimation on the estimates of change—As stated previously, 1981 was the first AHS for which 1980 census-based estimates, rather than 1970 census-based estimates, were used in the third-stage ratio estimation procedure. Since the 1980-based estimates are about 2 percent higher than the corresponding 1970-based estimates, this will cause estimates of change between the 1981 AHS data and data for 1980 or earlier years to be overstated. This overstatement should be taken into consideration during the analysis of estimates of change.

Possible effects of telephone interviewing on the data—A preliminary analysis of the 1981 telephone interviewing experiment has been completed. It has been concluded that data collected, using the telephone interviewing procedures, were not sufficiently different from data collected by regular personal interviews, to preclude basing the 1981 AHS national sample published data. However, since the percentage of the comparison tests that were significant was slightly above what could be expected by chance, there was some evidence that telephone interviewing may have had an effect on the data. Thus, some caution should be exercised in making comparisons between 1981 AHS data and data from preceding years.

Rounding errors—In errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of this survey. Also, since medians in this report were computed using unrounded data, instead of the published rounded data, they can differ from medians calculated directly from the published data.

TABLE I. Standard Errors of Estimated Numbers of Housing Units: 1981 (Excluding Estimates of Housing Units Pertaining to Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Complete Plumbing Facilities, and Housing Units With Householder of Spanish Origin)

(68 chances out of 100)

Size of estimate (000)	Standard error		Size of estimate (000)	Standard error	
	Total or White (000)	Black (000)		Total or White (000)	Black (000)
0	2	2	2,500	71	61
5	3	3	5,000	98	68
10	5	5	7,500	119	51
25	7	7	10,000	135	–
50	10	10	25,000	193	–
100	14	14	50,000	216	–
250	23	22	75,000	167	–
500	32	31	85,000	112	–
1,000	45	43			

TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Complete Plumbing Facilities, and Housing Units With Householder of Spanish Origin: 1981

(68 chances out of 100)

Size of estimate (000)	Standard error		Size of estimate (000)	Standard error	
	Total, White, or Spanish origin (000)	Black (000)		Total, White, or Spanish origin (000)	Black (000)
0	3	3	1,000	53	50
5	4	4	2,500	83	71
10	5	5	5,000	116	79
25	8	8	7,500	139	60
50	12	12	10,000	159	–
100	17	17	25,000	227	–
250	27	26	50,000	253	–
500	37	37			

TABLE III. Standard Errors of Estimated Numbers of Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1981 (Excluding Estimates of Housing Units Pertaining to Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing Facilities for the Northeast, North Central, and West Regions)

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	500	33
5	3	1,000	47
10	5	2,500	74
25	7	5,000	105
50	10	7,500	129
100	15	10,000	148
250	23	25,000	235

TABLE IV. Standard Errors of Estimated Numbers of Housing Units Pertaining to Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing Facilities for the Northeast, North Central, and West Regions: 1981

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	3	500	39
5	4	1,000	56
10	6	2,500	87
25	9	5,000	122
50	13	7,500	148
100	18	10,000	168
250	28	25,000	243

TABLE V. Standard Errors of Estimated Percentages of Housing Units: 1981 (Excluding Estimated Percentages of Housing Units Pertaining to Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Complete Plumbing Facilities, and Housing Units With Householder of Spanish Origin)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	29.1	29.1	29.1	29.1	29.1	29.1	29.1	32.0
10	17.0	17.0	17.0	17.0	17.0	17.0	19.6	22.7
25	7.6	7.6	7.6	7.6	8.6	10.2	12.4	14.3
50	3.9	3.9	3.9	4.4	6.1	7.2	8.8	10.1
100	2.0	2.0	2.0	3.1	4.3	5.1	6.2	7.2
250	0.8	0.9	1.3	2.0	2.7	3.2	3.9	4.5
500	0.4	0.6	0.9	1.4	1.9	2.3	2.8	3.2
1,000	0.2	0.5	0.6	1.0	1.4	1.6	2.0	2.3
2,500	0.08	0.3	0.4	0.6	0.9	1.0	1.2	1.4
5,000	0.04	0.2	0.3	0.4	0.6	0.7	0.9	1.0
7,500	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.8
10,000	0.02	0.14	0.2	0.3	0.4	0.5	0.6	0.7
25,000	0.01	0.09	0.13	0.2	0.3	0.3	0.4	0.5
50,000	—	0.06	0.09	0.14	0.2	0.2	0.3	0.3
75,000	—	0.05	0.07	0.11	0.2	0.2	0.2	0.3
85,000	—	0.05	0.07	0.11	0.15	0.2	0.2	0.2

TABLE VI. Standard Errors of Estimated Percentages of Housing Units Pertaining to Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Complete Plumbing Facilities, and Housing Units With Householder of Spanish Origin: 1981

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	36.1	36.1	36.1	36.1	36.1	36.1	36.1	37.6
10	22.0	22.0	22.0	22.0	22.0	22.0	23.0	26.6
25	10.2	10.2	10.2	10.2	10.2	12.0	14.6	16.8
50	5.3	5.3	5.3	5.3	7.1	8.5	10.3	11.9
100	2.7	2.7	2.7	3.7	5.0	6.0	7.3	8.4
250	1.1	1.1	1.5	2.3	3.2	3.8	4.6	5.3
500	0.6	0.7	1.1	1.6	2.3	2.7	3.3	3.8
1,000	0.3	0.5	0.7	1.2	1.6	1.9	2.3	2.7
2,500	0.11	0.3	0.5	0.7	1.0	1.2	1.5	1.7
5,000	0.06	0.2	0.3	0.5	0.7	0.8	1.0	1.2
7,500	0.04	0.2	0.3	0.4	0.6	0.7	0.8	1.0
10,000	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.8
25,000	0.01	0.11	0.15	0.2	0.3	0.4	0.5	0.5
50,000	0.01	0.07	0.11	0.2	0.2	0.3	0.3	0.4

TABLE VII. Standard Errors of Estimated Percentages of Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1981 (Excluding Estimated Percentages of Housing Units Pertaining to Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing Facilities for the Northeast, North Central, and West Regions)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	30.6	30.6	30.6	30.6	30.6	30.6	30.6	33.2
10	18.1	18.1	18.1	18.1	18.1	18.1	20.3	23.5
25	8.1	8.1	8.1	8.1	8.9	10.6	12.9	14.8
50	4.2	4.2	4.2	4.6	6.3	7.5	9.1	10.5
100	2.2	2.2	2.2	3.2	4.5	5.3	6.4	7.4
250	0.9	0.9	1.3	2.0	2.8	3.4	4.1	4.7
500	0.4	0.7	0.9	1.4	2.0	2.4	2.9	3.3
1,000	0.2	0.5	0.7	1.0	1.4	1.7	2.0	2.3
2,500	0.09	0.3	0.4	0.6	0.9	1.1	1.3	1.5
5,000	0.04	0.2	0.3	0.5	0.6	0.7	0.9	1.0
7,500	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.9
10,000	0.02	0.15	0.2	0.3	0.4	0.5	0.6	0.7
25,000	0.01	0.09	0.13	0.2	0.3	0.3	0.4	0.5

TABLE VIII. Standard Errors of Estimated Percentages of Housing Units Pertaining to Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing Facilities for the Northeast, North Central, and West Regions: 1981

(68 chances out of 100. For estimated percentages pertaining to source of water-individual well and mobile homes for the West Region, apply a factor of 1.66 to the standard error listed below)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	38.5	38.5	38.5	38.5	38.5	38.5	38.5	39.6
10	23.9	23.9	23.9	23.9	23.9	23.9	24.2	28.0
25	11.1	11.1	11.1	11.1	11.1	12.6	15.3	17.7
50	5.9	5.9	5.9	5.9	7.5	8.9	10.8	12.5
100	3.0	3.0	3.0	3.9	5.3	6.3	7.7	8.8
250	1.2	1.2	1.6	2.4	3.4	4.0	4.8	5.6
500	0.6	0.8	1.1	1.7	2.4	2.8	3.4	4.0
1,000	0.3	0.6	0.8	1.2	1.7	2.0	2.4	2.8
2,500	0.13	0.4	0.5	0.8	1.1	1.3	1.5	1.8
5,000	0.06	0.2	0.4	0.5	0.8	0.9	1.1	1.3
7,500	0.04	0.2	0.3	0.4	0.6	0.7	0.9	1.0
10,000	0.03	0.2	0.2	0.4	0.5	0.6	0.8	0.9
25,000	0.01	0.11	0.2	0.2	0.3	0.4	0.5	0.6

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-40
Selection of sample areas. App-40
Designation of sample housing units for the 1981 survey. App-40
Selection of the 1973 sample housing units. App-41
Splitting of the sample . App-41
CEN-SUP sample . App-41
Selection of supplemental sample housing units in rural areas. App-41
Selection of sample housing units for the 1976 Coverage Improvement Program . App-42
1977 and 1981 sample reductions . App-42
Supplemental sample from nonresidential conversions . App-42
1981 telephone interviewing experiment . App-42
ESTIMATION. App-42
RELIABILITY OF THE ESTIMATES. App-44
Sampling errors . App-44
Standard errors of estimates of levels . App-44
Standard errors of estimates of percentages. App-44
Standard errors of ratios . App-45
Illustration of the use of the standard error tables. Illustration I . App-45
Illustration II . App-45
Differences. App-46
Illustration of the computation of the standard error of a difference. App-46
Medians . App-46
Illustration of the computation of the 95-percent confidence interval for a median . App-46
Nonsampling errors . App-47
Reinterview program . App-47
Coverage errors . App-47
Effects of ratio estimation on the estimates of change . App-48
Possible effects of telephone interviewing on the data . App-48
Rounding errors. App-48

SAMPLE DESIGN

The 1981 estimates are based on data collected from September 1981 through December 1981 for the Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. The sample for this survey was spread over 461 sample areas (called primary sampling units), comprising 923 counties and independent cities with coverage in each of the 50 States and the District of Columbia.

Approximately 56,800 sample housing units (both occupied and vacant) were eligible for interview in the 1981 Annual Housing Survey. Of this number, 3,200 interviews were classified as "noninterview" for various reasons. Occupied housing units were classified as "noninterview" mainly, because the occupants refused to be interviewed after repeated calls. For vacant housing units, interviews were not obtained because an informed respondent was not found after repeated visits. In addition to the 56,800 eligible housing units, there were also 3,600 sample units which were visited but were ineligible for interview for the AHS in terms of collecting information relevant to the 1981 housing inventory.

Selection of sample areas—The United States was divided into areas made up of counties and independent cities referred to as primary sampling units (PSU's). These PSU's were then grouped into 376 strata, 156 of which consisted of only one PSU which was in sample with certainty. These 156 strata were mostly the larger SMSA's and were called self-representing (SR) since the sample from the sample area represented just that PSU. Each one of the other 220 strata consisted of a group of PSU's and were referred to as non-self-representing (NSR), since the sample of housing units from the sample PSU in a stratum represented the other PSU's in the stratum as well.

One PSU was selected from each NSR stratum with probability proportionate to the 1970 census population of the PSU. (This resulted in 220 NSR sample PSU's.) In addition, the NSR strata were grouped into 110 pairs and one stratum was picked at random from each pair. From this stratum, an additional PSU was selected independently of the other PSU selected from this stratum. Since the two PSU's were independently selected, it was possible for the same PSU to be selected twice. This occurred in 25 instances, producing an additional 85 NSR sample PSU's, thus giving a grand total of 461 PSU's.

Designation of sample housing units for the 1981 survey—The sample housing units designated to be interviewed in the 1981 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1980 survey (which included all sample housing units that were selected as part of the 1976 Coverage Improvement Program) and were not part of the 1981 sample reduction.
2. All sample housing units that were either type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of survey but which could become eligible in the future) in the 1980 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1981 AHS questionnaire, page App-16).
3. All sample housing units that were selected from the list of building permits issued since the 1980 survey. (This sample represented the housing units built in permit-issuing areas, since the 1980 survey.)
4. Housing units added as the result of the updated listings in selected areas which do not issue building permits.

Selection of the 1973 sample housing units—The overall sampling rate used to select the sample for the 1973 AHS was about 1 in 1,366. The within-PSU sampling rate for the AHS was determined so that the overall probability of selection for each sample housing unit was the same (e.g., if the probability of selecting a NSR PSU was 1 in 10, then the within-PSU sampling rate would be 1 in 136.6).

Within the sample PSU's, a sample of the housing units enumerated in the 1970 Census of Population and Housing was selected for the AHS. In addition, a sample of new construction building permits was also selected to represent the units constructed since the 1970 census. These samples were selected at about twice the rate mentioned previously (i.e., at 2 in 1,366), thereby producing a sample twice as large as needed. This sample was split into two equal-sized samples—one to be used for the AHS, and one to be held in reserve for possible future use for the AHS. The procedure used to split this sample into equal-sized samples is described in the next section.

The sample of 1970 census units was selected in several stages. Within the sample PSU's, the first step was the selection of a sample of census enumeration districts (ED's), administrative units used in the 1970 census. The probability of selection for an ED was proportional to the following 1970 census counts of housing units (HU's) and persons in group quarters, combined in the following formula:

$$\frac{\text{Number of HU's in the ED} + \dfrac{\text{Number of group quarters persons in the ED}}{3}}{4}$$

The next step was to select an expected cluster of about four neighboring housing units within each sample ED. For most of the ED's, the selection was accomplished using the list of addresses for the ED as compiled in the 1970 census (these ED's are referred to as address ED's). However, in those ED's where addresses were incomplete or inadequate (mostly rural areas), the selection process was accomplished using area sampling methods. These ED's were divided into segments (i.e., small land areas with well-defined boundaries, having an expected size of four, or a multiple of four, housing units) and a segment was selected. Those selected segments with an expected size which was a multiple of four were further subsampled at the time of interview so that an expected four housing units were chosen for interview.

The sample of new construction housing units was selected from building permits issued since January 1970. Within each sample PSU, the building permits were chronologically ordered by month issued, and compact clusters of approximately four housing units were created. These clusters were then sampled at the rate of 2 in 1,366. Housing units constructed since the 1970 census in areas which do not issue building permits were brought into the sample as a result of the area sample described.

Splitting of the sample—The described sample selection procedure produced clusters (or segments) of size-four housing units for the sample taken from the census address frame, the new construction frame, and the area sampling frame (mainly rural areas). Clusters of this size should result in a minimum loss in precision for estimates of housing characteristics in rural areas because of the heterogeneity of neighboring units. However, clusters of size-two housing units, were considered to be more optimum within those areas where the housing characteristics of neighboring units tend to be very similar (i.e., urban areas and new construction units). A splitting operation was then carried out for clusters selected from the census address and the new construction frames. This consisted of halving each sample cluster from these frames. Thus, two housing units from each of these clusters were included in the survey and two housing units were held in reserve. No splitting operation was carried out within the clusters selected from the area sampling frame; every other area sample cluster of four housing units was used for the survey and the remaining clusters were assigned to the reserve sample.

CEN-SUP sample—Housing units at addresses missed in the census or inadequately described in the 1970 census address register did not have a chance of being selected for the AHS sample in address ED's. Consequently, a special operation was undertaken to develop a separate sample, called the census supplemental (CEN-SUP) sample to represent these units. Due to time constraints on this operation, only about 40 percent of the CEN-SUP units were ready to be interviewed in 1973. The rest were interviewed for the first time in the 1974 AHS.

Selection of supplemental sample housing units in rural areas—In 1974, it was decided to increase the reliability of the AHS estimates of rural housing characteristics by doubling the number of sample housing units from rural areas. This was accomplished by reactivating the reserve sample, selected in the original sampling operations in 1973, from rural areas only. For the reserve sample selected in census address and new construction frames, the other half of each rural cluster (an expected two housing units) was reactivated in 1974. Similarly, for the area sampling frame, the entire reserve cluster (an expected four housing units) was reactivated in 1974 if the cluster was rural. This supplementation increased the overall probability of selection for sample housing units in rural areas to about 2 in

1,366; whereas, the overall probability of selection for sample housing units in urban areas remained at 1 in 1,366.

Selection of sample housing units for the 1976 Coverage Improvement Program—The 1976 Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS national sample from the census address and new construction frames. The coverage deficiencies included the following types of housing units:

1. New construction housing units, from building permits issued prior to January 1970 for which construction had not been completed at the time of the 1970 census.
2. Housing units converted to residential use in structures totally nonresidential at the time of the 1970 census.
3. Houses that have been moved onto their present site since the 1970 census.
4. Mobile homes and trailers placed in parks either missed in the 1970 census or established since the 1970 census.
5. Mobile homes and trailers placed outside parks since the 1970 census or vacant at the time of the 1970 census.

A sample of new construction housing units whose permits were issued before January 1970 was selected in two stages. In the first stage, the units whose permits were issued before January 1970, but which were completed after the census, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. In the second stage, these units were then sampled so that the overall probability of selection was about 1 in 1,320.

A sample of mobile homes and trailers placed in a park that was missed by the census or established after the census was also selected in two stages. During the first stage, a list of mobile home parks was obtained from commercial listings. This list was then supplemented by additional parks identified by a canvassing operation similar to that performed in ED's where area sampling methods are used. The second stage consisted of dividing the parks into clusters of an expected size of four sites. These clusters were then sampled so that the overall probability of selection was about 1 in 1,366.

For the remaining housing units (i.e., mobile homes and trailers placed outside parks since the 1970 census, mobile homes and trailers vacant at the time of the 1970 census, units converted from nonresidential to residential use since the 1970 census, and houses that had been moved onto their present site since the 1970 census), the sampling was done in three stages. First, a subsample of the regular AHS sample housing units from the census address frame was selected. Second, succeeding structures that had been eligible to be selected from the census address frame were then listed until eight structures (excluding mobile home parks) were found. Finally, the intervening structures that had been listed which did not have a chance of selection in the AHS were identified and the units within these structures were interviewed.

1977 and 1981 sample reductions—By 1977, the addition to the sample, primarily from new construction and the coverage improvements, had increased the total sample size (interviews plus noninterviews) to about 81,000 housing units. The sample was reduced by about 7 percent to approximately 75,000 in 1977. This reduction did not include any CEN-SUP units or units which were selected as part of the 1976 Coverage Improvement Program. Initially, the overall sample was further reduced in 1981 by about 5 percent. Subsequently, the rural sample was reduced by about 25 percent in 125 large SMSA's[1] and by about 50 percent in small SMSA's[1] and outside SMSA's. These reductions brought the sample size down to approximately 60,000 housing units in 1981.

Supplemental sample from nonresidential conversions—The purpose of this sample was to improve the coverage in address ED's (i.e., ED's in which 1970 census address listings were used for selecting the AHS sample) of housing units created in structures that were completely nonresidential at the time of the 1970 census. This sample was derived from listings created for the Commercial Victimization Survey (CVS), a nationally representative area sample survey conducted by the Census Bureau for the Department of Justice. Each of the CVS area segments was initially canvassed either in 1972 or 1975 and eligible business establishments were listed. These listings were updated every 6 months with the last updating in the last half of 1977. Each basic address containing a business establishment listed for CVS was visited for the 1980 Components of Inventory Change Survey (CINCH), which was conducted in conjunction with the 1980 AHS, to identify HU's in structures which were completely nonresidential at the time of the 1970 census. These cases were later matched to the 1970 census address listings to identify those cases in address ED's and as an additional check to see if housing units existed in these structures at the time of the 1970 census. Due to the timing restraints associated with these operations, these cases were not included for the 1980 AHS but were included for the 1981 survey.

1981 telephone interviewing experiment—A large scale telephone interviewing experiment was conducted in conjunction with the 1981 AHS national sample in order to provide more definitive information about the effect of telephone interviewing on the AHS data and cost. A random sample of about 12,500 AHS national sample units was selected for the experiment. Among the cases assigned for telephone interviewing, only those sample housing units that had been interviewed in 1980 and had a telephone number available were eligible to be interviewed by telephone. Since a large portion of the total AHS national sample was assigned for telephone interviewing, its possible effect on the data is mentioned in the section on Reliability of the Estimates of this appendix.

ESTIMATION

In 1981, the AHS estimates employed a three-stage ratio estimation procedure. However, prior to implementation of the

[1] A "large" SMSA is one with a 1970 population of 250,000 persons or more and a "small" SMSA is one with a 1970 population of less than 250,000 persons.

procedure, the basic weight (i.e., the inverse of the probability of selection) was adjusted to account for the type A noninterview housing units encountered in the AHS. This noninterview adjustment was done separately for occupied and vacant units. The noninterview adjustment was equal to the following ratio:

$$\frac{\text{Interviewed housing units + Noninterviewed housing units}}{\text{Interviewed housing units}}$$

The first-stage ratio estimation procedure was employed for sample housing units from non-self-representing (NSR) PSU's only. This procedure was designed to reduce the contribution to the variance arising from the sampling of PSU's. The first-stage ratio estimation procedure takes into account the differences that existed at the time of the 1970 census in the distribution by tenure and residence of the housing population estimated from the sample NSR PSU's and that of the NSR housing population in each of the four census regions of the country.

The first-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{The 1970 census housing population in the residence-tenure category for all NSR strata in a census region}}{\text{Estimates of the housing population category using 1970 census housing counts for sample NSR PSU's in a census region}}$$

The numerators of the ratios were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR stratum and summing these counts across the NSR strata in each census region. The denominators were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR sample PSU, weighting these counts by the inverse of the probability of selecting that PSU, and summing these weighted counts across the NSR sample PSU's in each census region. The computed first-stage ratio estimation factor was then applied to the existing weight for each NSR sample unit in each first-stage ratio estimation category.

The second-stage ratio estimation procedure was designed to adjust the AHS sample estimate of conventional new construction housing units, i.e., sample units built April 1, 1970, or later, to independently derived current estimates where a known deficiency in the AHS sample exists (see the section on nonsampling error) for each of the four regions. These estimates were considered to be the best estimates available for the number of conventional new construction units in these categories.

The second-stage ratio estimation factor was as follows:

$$\frac{\text{Current best estimate of new construction in the category}}{\text{AHS sample estimate of new construction units in the category}}$$

The numerators of the ratios were derived from data based on the Survey of Construction (SOC). The denominators of the ratios were obtained from the weighted estimates for the AHS sample units using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in each second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all the AHS sample units. This procedure was designed to adjust the AHS sample estimates of housing (i.e., the estimates employing the noninterview, first-stage, and second-stage adjustments) to current vacant housing estimates for four categories of vacant housing units and to independently derived current housing estimates for 24 categories of occupied housing units. Each of these categories is a combination of the characteristics of residence, tenure, race of householder, and sex of householder.

The third-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{Current independent estimate of housing units in the category}}{\text{AHS sample estimate of housing units in the category}}$$

The numerators of the ratios for occupied housing units were derived from data based on the Current Population Survey (CPS), a sample household survey conducted monthly by the Bureau of the Census. This is the first time that CPS household estimates based on the 1980 census were used as the independent controls in this adjustment rather than the 1970 census-based controls. The 1970-based estimates were about 2 percent smaller than the 1980-based estimates. This difference should be taken into consideration when comparing the 1981 AHS estimates to the AHS estimates for 1980 or earlier years. The numerators of the ratios for vacant housing units were derived from data based on the Housing Vacancy Survey (HVS), a quarterly vacancy survey also conducted by the Bureau of the Census. The denominators of the ratios were obtained from the weighted estimates for the AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit in each third-stage ratio estimation category.

The second-stage and the third-stage ratio estimation procedures were iterated in order to bring the AHS estimates into close agreement with both sets of independent estimates. The second-stage was modified so that the estimates for all 27 categories of new construction would be identical to the estimates before the third stage. Hence, the repeated second stage had the effect of controlling the AHS sample estimates of new construction units to the unbiased sample estimates for 9 categories of new construction mobile homes and trailers for each of the four regions and of adjusting the AHS sample estimate of 18 categories of conventional new construction units to the independently derived current estimates.

The numerators were either the unbiased weighted estimates for the AHS sample housing units, using the existing weight after the first-stage ratio estimation procedure (i.e., the estimates employing the noninterview and first-stage adjustments) or the independent estimate derived from data based on the Survey of Construction (SOC).

The denominators of the ratios in this iterative process were obtained from the weighted estimates for the AHS sample units after the previous stage of ratio estimation. The factors resulting from this iterative process were then applied to the

existing weight on the appropriate records, and the resulting product was used as the final weight for tabulation.

The third-stage ratio estimation procedure, as well as the overall estimation procedure, reduced the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. The distribution of the housing population selected for the sample differed somewhat, by chance, from that of the Nation as a whole in such basic housing characteristics as tenure, vacancy status, residence, race of householder, and sex of householder. These characteristics are probably closely correlated with other housing characteristics measured for the AHS. Therefore, through the use of the three-stage ratio estimation procedure one can expect the sample estimate to be improved substantially.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys; sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS national sample.

Sampling errors—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same questionnaires, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The variability between estimates from all possible samples is defined as sampling error. One common measure of sampling error is the standard error which measures the precision with which an estimate from a sample approximates the average result of all possible samples. In addition, the standard error, as calculated for this report, also partially reflects the variation in the estimates due to some nonsampling errors, but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, and biases and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates so that the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, and each of these samples was surveyed under essentially the same general conditions and an estimate and its estimated standard error were calculated for each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate, to one standard error above the estimate, would include the average result of all possible samples;
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples;
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples either is or is not contained in any particular computed interval. However, for a particular sample one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables (pages App-48 to App-52) of this appendix are approximations to the standard errors of various estimates shown in this report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than the precise standard error for any specific item.

Standard errors of estimates of levels—Tables I and II present the standard errors applicable to the 1981 national housing inventory estimates in this report. Tables III and IV present the standard errors applicable to estimates for the Northeast, North Central, and West Regions, and tables III and V present the standard errors applicable to estimates for the South Region. Linear interpolation should be used to determine standard errors for levels of estimates not specifically shown in tables I through V.

Standard errors of estimates of percentages—The reliability of an estimated percentage, computed by using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more.

Tables VI through X present the standard errors of estimated percentages. Table VI shows the approximate standard errors of all national estimated percentages of housing units except those pertaining to the specified items in table II. The standard errors shown in table VII should be used for those specified items. Table VIII shows the approximate standard errors of the estimated percentages of housing units for the Northeast, North Central, South, and West Regions except for those percentages pertaining to the specified items. Table IX should be used for those specified items for the Northeast, North Central, and West Regions and table X for the South Region. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables VI through X.

Included in tables I through X are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered to be overestimates of the true standard errors and should be used primarily for construction of confidence intervals for characteristics when an estimate of zero is obtained.

Standard errors of ratios—For ratios of the form (100) (x/y), where x is not a subclass of y, tables VI through X underestimate the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be obtained by letting the standard error of the ratio be approximately equal to:

$$(100)\left(\frac{x}{y}\right)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

where: x = the numerator of the ratio
y = the denominator of the ratio
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

Illustration of the use of the standard error tables. Illustration I—Table A-2 of this report shows that in the United States there were 13,667,000 specified owner-occupied housing units with two persons in 1981. Interpolation using standard errors in table I (page App-48) shows the standard error of an estimate of this size is approximately 149,000. The following procedure was used in interpolating.

The information presented in the following table was extracted from standard error table I. The entry for "x" is the one sought.

Size of estimate (000)	Standard error (000)
10,000	135
13,667	x
25,000	193

By vertically interpolating between 135 and 193, the entry for "x" is determined to be:

$$13{,}667-10{,}000 = 3{,}667$$
$$25{,}000-10{,}000 = 15{,}000$$

$$135 + \frac{3{,}667}{15{,}000}(193-135) = 149$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 13,518,000 to 13,816,000 housing units. Therefore, a conclusion that the average estimate of 1981 housing units of this type lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all possible samples, lies within the interval from 13,429,000 to 13,905,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 13,369,000 to 13,965,000 housing units with 95 percent confidence.

Table A-2 also shows that of the 9,777,000 specified owner-occupied housing units with two bedrooms in 1981, 1,542,000, or 15.8 percent, were valued between $20,000 and $29,999. Interpolation using standard error table VI (i.e., interpolation on both the base and percent) (page App-50) shows that the standard error of the above percentage is 0.5 percentage points. The following procedure was used in interpolating.

The information presented in the following table was extracted from table VI. The entry for "p" is the one sought.

Base of percentage (000)	Estimated percentage		
	15	15.8	25
7,500	0.6	a	0.7
9,777		p	
10,000	0.5	b	0.6

1. By horizontal interpolation between 0.6 and 0.7, the entry for cell "a" is determined to be 0.6.

$$15.8-15.0 = 0.8$$
$$25.0-15.0 = 10.0$$

$$0.6 + \frac{0.8}{10.0}(0.7-0.6) = 0.6$$

2. By horizontal interpolation between 0.5 and 0.6, the entry for cell "b" is determined to be 0.5.

$$15.8-15.0 = 0.8$$
$$25.0-15.0 = 10.0$$

$$0.5 + \frac{0.8}{10.0}(0.6-0.5) = 0.5$$

3. By vertical interpolation between 0.6 and 0.5, the entry for "p" is determined to be 0.5.

$$9{,}777-7{,}500 = 2{,}277$$
$$10{,}000-7{,}500 = 2{,}500$$

$$0.6 + \frac{2{,}277}{2{,}500}(0.5-0.6) = 0.5$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 15.3 to 16.3 percent; the 90-percent confidence interval is from 15.0 to 16.6 percent; and the 95-percent confidence interval is from 14.8 to 16.8 percent.

Illustration II—Table A-2 of this report shows that in the United States in 1981 there were 6,294,000 specified owner-occupied housing units whose source of water was an individual well. Interpolation of the data in table II (page App-49) shows that the standard error of an estimate of this size is aprpoximately 128,000. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1981 specified owner-occupied housing units whose source of water was an individual well, lies within the interval from 6,166,000 to 6,422,000 is correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all

possible samples, lies within the interval from 6,089,000 to 6,499,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 6,038,000 to 6,550,000 housing units with 95 percent confidence.

Table A-2 also shows that of the 6,294,000 specified owner-occupied housing units in 1981 whose source of water was an individual well, 157,000, or 2.5 percent, were valued at less than $10,000. Interpolation in table VII (i.e., interpolation on both the base and the percent) (page App-50) shows that the standard error of the above percentage is 0.3 percentage points. Consequently, the 68-percent confidence interval, as shown by these data is from 2.2 to 2.8 percent; the 90-percent confidence interval is from 2.0 to 3.0 percent; and the 95-percent confidence interval is from 1.9 to 3.1 percent.

Differences—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristics in two different areas or the difference between separate and uncorrelated characteristics in the same area. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error. However, if there is a high negative correlation between the two characteristics, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference—Table A-2 shows that in the United States in 1981 there were 1,542,000 specified owner-occupied housing units with two bedrooms valued between $20,000 and $29,999. It also shows that in the United States in 1981 there were 1,813,000 specified owner-occupied housing units with two bedrooms valued between $30,000 and $39,999. Thus, the apparent difference between the number of 1981 specified owner-occupied housing units with two bedrooms valued between $20,000 and $29,999 and those valued between $30,000 and $39,999 is 271,000. Interpolation using standard error table I shows the standard error on an estimate of 1,542,000 to be approximately 54,000 and the standard error on an estimate of 1,813,000 to be approximately 59,000. Therefore, the standard error of the estimated difference of 271,000 is about 80,000.

$$80{,}000 = \sqrt{(54{,}000)^2 + (59{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 271,000 difference is from 191,000 to 351,000 housing units. Therefore, a conclusion that the average estimate of this difference, derived from all possible samples, lies within a range computed in this way and would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 143,000 to 399,000 housing units, and the 95-percent confidence interval is from 111,000 to 431,000. Thus, we can conclude with 95 percent confidence that the number of 1981 specified owner-occupied housing units with two bedrooms valued between $30,000 and $39,999 is greater than the number valued between $20,000 and $29,999 since the 95-percent confidence interval of this difference does not include zero or negative values.

Medians—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From the appropriate standard error tables, determine the standard error of a 50-percent characteristic on the base of the median;
2. Add to and subtract from 50 percent the standard error determined in step 1; and
3. Using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval for a median—Table A-2 of this report shows the median value of specified owner-occupied housing units with two bedrooms in the United States was $41,500 in 1981. The base of the distribution, from which this median was determined is 9,777,000 housing units.

1. From table VI of this appendix, the standard error of a 50-percent characteristic on the base of 9,777,000 is 0.8 percentage points.
2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent, twice the standard error determined in step 1. This yields percentage limits of 48.4 and 51.6.
3. From table A-2, it can be seen by cumulating the frequencies for the first four categories that 4,647,000 owner-occupied housing units with two bedrooms, or 47.5 percent, had a value of less than $40,000 and an additional 1,569,000 owner-occupied housing units with two bedrooms, or 16.0 percent, had a value between $40,000 and $50,000. By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$\$40{,}000 + (\$50{,}000 - \$40{,}000)\,\frac{(48.4 - 47.5)}{16.0} = \$40{,}600$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$\$40{,}000 + (\$50{,}000 - \$40{,}000)\ \frac{(51.6 - 47.5)}{16.0} = \$42{,}600$$

Thus, the 95-percent confidence interval ranges from $40,600 to $42,600.

Nonsampling errors—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases; definitional difficulties; differences in the interpretation of questions; inability or unwillingness to provide correct information on the part of respondents; mistakes in recording or coding the data; and other errors of collection, response, processing, coverage, and estimation for missing data. As can be seen from the list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for the 1981 AHS national sample.

Reinterview program—For the AHS national sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a subsample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and thus were the basis for the measurement of the "content" error of these AHS estimates.

As part of the reinterview, an additional check was carried out for interviewer evaluation and quality control. This check was made at each of these households to determine if the following was done during the original interview.

1. The correct housing unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

The results of the 1981 and 1980 reinterview studies were not available at the time of publication. In addition, the 1979 reinterview was carried out as an interviewer quality control only. Thus, the results of the 1977 and 1978 reinterview studies which were presented in the Census Bureau memoranda, "Reinterview Results for the Annual Housing Survey—National Sample 1977" and "Reinterview Results for the Annual Housing Survey—National Sample 1978" are presented here.

In 1978, a substantial portion of the reinterview questionnaire was devoted to testing the new questions 1 through 7 (parts a and b). The questions (part a), which were asked only at housing units interviewed in the previous year, determined whether there had been a change since last year in selected nonattitudinal items. If a change had been recorded or the respondent did not know if a change had occurred, part b of the question, which collects the value of the item, was asked. The reinterviewers asked these items using the questions as formatted in 1977. Comparing the responses from the differently formatted questions, the 1978 reinterview found that 80 percent of the questions showed low levels of inconsistency with the remainder showing moderate levels.

The 1977 reinterview program showed moderate to high levels of inconsistency with about 21 percent of the nonattitudinal and 56 percent of the attitudinal items showing high levels of inconsistency. But a large proportion (43 percent) of the nonadditudinal items showed a low level of inconsistency. Moderate levels indicate that there are some problems with inconsistent reporting and high levels indicate that improvements are needed in the data collection methods or that the category concepts themselves are ambiguous.

Cross-tabulations involving those items, which are subject to substantial levels of inconsistency may be subject to a large distortion as a consequence of the associated high level of response variance, and thus, are considered to be less reliable than comparable cross-tabulations which do not involve these data. As a consequence, the tables in this report, which contain such cross-tabulations, have been footnoted with a cautionary statement.

The 1970 census reinterview results provide illustrations of possible nonsampling errors for some of the items which also appear in the AHS. For example, median value of homes was consistently underestimated by about 5 percent, and the average monthly cost of electricity and utility gas were consistently overestimated although the net effect on average gross rent was fairly small.

A possible explanation for the results of the AHS and census reinterview studies, as well as the surveys themselves, is that the respondents may lack precise information. Also, because the results of the reinterview studies are derived from sample surveys, there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of these studies.

Coverage errors—A deficiency in the representation of conventional new construction (non-mobile home or trailer) for the AHS new construction sample (mentioned previously in the section on estimation) is an example of coverage errors. During the sampling of building permits, only those issued more than 4 months before the survey began were eligible to be selected to represent conventional new construction. Due to time con-

straint, it is not possible to sample units whose permits are issued less than 4 months in advance of the survey.

It is estimated that the 1981 AHS sample missed about 0.8 percent (i.e., about 157,000 units) of conventional housing units built after April 1970 because the permits for these units, which were built before October 1981, were issued less than 4 months in advance of the survey. The second stage of ratio estimation procedure was employed to reduce the effect of this deficiency although some bias in the AHS estimates of conventional new construction probably still exists. Review of the second-stage ratio estimation procedures indicates that we have consistently overcompensated for this deficiency every year since 1975 by ratio adjusting the new construction to counts of new construction for the end of interview period, which has been December or January, instead of October. This overcompensation may inflate the new construction counts by 100,000 to 300,000 units. Changing this procedure to correct for this overcompensation would reduce the estimate of housing units built since the last survey since it would be based on a 10-month time period rather than 12 months. Thus, it was decided not to change this procedure so that the estimate of housing units built since the last survey would be consistent with previous years' estimates, which are generally based on a 12-month time period.

In addition, the 1976 Coverage Improvement Program had certain deficiencies. First, when the canvassing was done to identify mobile homes and trailer parks that were not in the same sample frame or not on the commercial lists, only 92 percent of the census address frame ED's were represented. Second, it appears that the listing procedure (used to find mobile homes and trailers placed outside parks, units converted from nonresidential to residential, and houses that had been moved onto their present site) was not very efficient for finding nonresidential conversions (which might be primarily in business districts), since the listing procedure started from a residential unit. (The sample estimate of this component was approximately 16,000 housing units with a standard error of 12,000.)

Finally, it is felt that deficiencies also exist in ED's where area sampling methods are used. As before, it had been assumed that all housing units located inside these ED's would be represented in the sample. However, it has been estimated that the 1981 AHS sample missed as much as 2 percent (i.e., as much as 400,000 units) of all housing units in ED's where area sampling methods are used because these units are not listed during the canvassing.

The third stage of ratio estimation corrects for these deficiencies as far as the total count of housing is concerned, i.e., it adjusts the estimate of the total housing inventory to the best available estimate. However, biases of subtotals would still remain.

Effects of ratio estimation on the estimates of change—As stated previously, 1981 was the first AHS for which 1980 census-based estimates, rather than 1970 census-based estimates, were used in the third-stage ratio estimation procedure. Since the 1980-based estimates are about 2 percent higher than the corresponding 1970-based estimates, this will cause estimates of change between the 1981 AHS data and data for 1980 or earlier year's to be overstated. This overstatement should be taken into consideration during the analysis of estimates of change.

Possible effects of telephone interviewing on the data—A preliminary analysis of the 1981 telephone interviewing experiment has been completed. It has been concluded that data collected, using the telephone interviewing procedures, were not sufficiently different from data collected by regular personal interviews, to preclude basing the 1981 AHS national sample published data. However, since the percentage of the comparison tests that were significant was slightly above what could be expected, there was some evidence that telephone interviewing may have had an effect on the data. Thus, some caution should be exercised in making comparisons between 1981 AHS data and data from preceding years.

Rounding errors—In errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of this survey. Also, since medians in this report were computed using unrounded data, instead of the published rounded data, they can differ from medians calculated directly from the published data.

TABLE I. Standard Errors of Estimated Numbers of Housing Units: 1981 (Excluding Estimates of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water-Individual Well, Lacking Complete Plumbing, Mobile Homes, and Housing Units With Householder of Spanish-Origin)

(68 chances out of 100)

Size of estimate (000)	Standard error		Size of estimate (000)	Standard error	
	Total or White (000)	Black (000)		Total or White (000)	Black (000)
0	2	2	2,500 . . .	71	61
5	3	3	5,000 . . .	98	68
10	5	5	7,500 . . .	119	51
25	7	7	10,000. . .	135	–
50	10	10	25,000. . .	193	–
100.	14	14	50,000. . .	216	–
250.	23	22	75,000. . .	167	–
500.	32	31	85,000. . .	112	–
1,000 . . .	45	43			

TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water-Individual Well, Lacking Complete Plumbing, Mobile Homes, and Housing Units With Householder of Spanish-Origin: 1981

(68 chances out of 100)

Size of estimate (000)	Standard error		Size of estimate (000)	Standard error	
	Total or White (000)	Black (000)		Total or White (000)	Black (000)
0	3	3	1,000	53	50
5	4	4	2,500	83	71
10	5	5	5,000	116	79
25	8	8	7,500	139	60
50	12	12	10,000	159	–
100	17	17	25,000	227	–
250	27	26	50,000	253	–
500	37	37			

TABLE III. Standard Errors of Estimated Numbers of Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1981 (Excluding Estimates of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions and Excluding Source of Water-Individual Well, and Mobile Homes for Each of the Regions)

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	500	33
5	3	1,000	47
10	5	2,500	74
25	7	5,000	105
50	10	7,500	129
100	15	10,000	148
250	23	25,000	235

TABLE IV. Standard Errors of Estimated Numbers of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions and to Source of Water-Individual Well, and Mobile Homes for the Northeast and North Central Regions: 1981

(68 chances out of 100. For estimates pertaining to source of water-individual well and mobile homes for the West Region, apply a factor of 1.66 to the standard errors listed below)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	3	500	39
5	4	1,000	56
10	6	2,500	87
25	9	5,000	122
50	13	7,500	148
100	18	10,000	168
250	28	25,000	243

TABLE V. Standard Errors of Estimated Numbers of Housing Units Pertaining to Source of Water-Individual Well and Mobile Homes for the South Region: 1981

(68 chance out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	5	500	50
5	5	1,000	70
10	7	2,500	109
25	11	5,000	150
50	16	7,500	179
100	22	10,000	201
250	35	25,000	259

TABLE VI. Standard Errors of Estimated Percentages of Housing Units: 1981 (Excluding Estimated Percentages of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water-Individual Well, Lacking Complete Plumbing, Mobile Homes, and Housing Units With Householder of Spanish-Origin)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	29.1	29.1	29.1	29.1	29.1	29.1	29.1	32.0
10	17.0	17.0	17.0	17.0	17.0	17.0	19.6	22.7
25	7.6	7.6	7.6	7.6	8.6	10.2	12.4	14.3
50	3.9	3.9	3.9	4.4	6.1	7.2	8.8	10.1
100	2.0	2.0	2.0	3.1	4.3	5.1	6.2	7.2
250	0.8	0.9	1.3	2.0	2.7	3.2	3.9	4.5
500	0.4	0.6	0.9	1.4	1.9	2.3	2.8	3.2
1,000	0.2	0.5	0.6	1.0	1.4	1.6	2.0	2.3
2,500	0.08	0.3	0.4	0.6	0.9	1.0	1.2	1.4
5,000	0.04	0.2	0.3	0.4	0.6	0.7	0.9	1.0
7,500	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.8
10,000	0.02	0.14	0.2	0.3	0.4	0.5	0.6	0.7
25,000	0.01	0.09	0.13	0.2	0.3	0.3	0.4	0.5
50,000	—	0.06	0.09	0.14	0.2	0.2	0.3	0.3
75,000	—	0.05	0.07	0.11	0.2	0.2	0.2	0.3
85,000	—	0.05	0.07	0.11	0.15	0.2	0.2	0.2

TABLE VII. Standard Errors of Estimated Percentages of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water-Individual Well, Lacking Complete Plumbing, Mobile Homes, and Housing Units With Householder of Spanish-Origin: 1981

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	36.1	36.1	36.1	36.1	36.1	36.1	36.1	37.6
10	22.0	22.0	22.0	22.0	22.0	22.0	23.0	26.6
25	10.2	10.2	10.2	10.2	10.2	12.0	14.6	16.8
50	5.3	5.3	5.3	5.3	7.1	8.5	10.3	11.9
100	2.7	2.7	2.7	3.7	5.0	6.0	7.3	8.4
250	1.1	1.1	1.5	2.3	3.2	3.8	4.6	5.3
500	0.6	0.7	1.1	1.6	2.3	2.7	3.3	3.8
1,000	0.3	0.5	0.7	1.2	1.6	1.9	2.3	2.7
2,500	0.11	0.3	0.5	0.7	1.0	1.2	1.5	1.7
5,000	0.06	0.2	0.3	0.5	0.7	0.8	1.0	1.2
7,500	0.04	0.2	0.3	0.4	0.6	0.7	0.8	1.0
10,000	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.8
25,000	0.01	0.11	0.15	0.2	0.3	0.4	0.5	0.5
50,000	0.01	0.07	0.11	0.2	0.2	0.3	0.3	0.4

TABLE VIII. Standard Errors of Estimated Percentages of Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1981 (Excluding Estimated Percentages Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions and Excluding Source of Water-Individual Well, and Mobile Homes for Each of the Regions)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	30.6	30.6	30.6	30.6	30.6	30.6	30.6	33.2
10	18.1	18.1	18.1	18.1	18.1	18.1	20.3	23.5
25	8.1	8.1	8.1	8.1	8.9	10.6	12.9	14.8
50	4.2	4.2	4.2	4.6	6.3	7.5	9.1	10.5
100	2.2	2.2	2.2	3.2	4.5	5.3	6.4	7.4
250	0.9	0.9	1.3	2.0	2.8	3.4	4.1	4.7
500	0.4	0.7	0.9	1.4	2.0	2.4	2.9	3.3
1,000	0.2	0.5	0.7	1.0	1.4	1.7	2.0	2.3
2,500	0.09	0.3	0.4	0.6	0.9	1.1	1.3	1.5
5,000	0.04	0.2	0.3	0.5	0.6	0.7	0.9	1.0
7,500	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.9
10,000	0.02	0.15	0.2	0.3	0.4	0.5	0.6	0.7
25,000	0.01	0.09	0.13	0.2	0.3	0.3	0.4	0.5

TABLE IX. Standard Errors of Estimated Percentages of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions and to Source of Water-Individual Well, and Mobile Homes for the Northeast and North Central Regions: 1981

(68 chances out of 100. For estimated percentages pertaining to source of water-individual well and mobile homes for the West Region, apply a factor of 1.66 to the standard errors lited below)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	38.5	38.5	38.5	38.5	38.5	38.5	38.5	39.6
10	23.9	23.9	23.9	23.9	23.9	23.9	24.2	28.0
25	11.1	11.1	11.1	11.1	11.1	12.6	15.3	17.7
50	5.9	5.9	5.9	5.9	7.5	8.9	10.8	12.5
100	3.0	3.0	3.0	3.9	5.3	6.3	7.7	8.8
250	1.2	1.2	1.6	2.4	3.4	4.0	4.8	5.6
500	0.6	0.8	1.1	1.7	2.4	2.8	3.4	4.0
1,000	0.3	0.6	0.8	1.2	1.7	2.0	2.4	2.8
2,500	0.13	0.4	0.5	0.8	1.1	1.3	1.5	1.8
5,000	0.06	0.2	0.4	0.5	0.8	0.9	1.1	1.3
7,500	0.04	0.2	0.3	0.4	0.6	0.7	0.9	1.0
10,000	0.03	0.2	0.2	0.4	0.5	0.6	0.8	0.9
25,000	0.01	0.11	0.2	0.2	0.3	0.4	0.5	0.6

TABLE X. Standard Errors of Estimated Percentages of Housing Units Pertaining to Source of Water-Individual Well and Mobile Homes for the South Region: 1981

(68 chances out of 100)

Base of percentage (000)	Estimated percentages							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	49.7	49.7	49.7	49.7	49.7	49.7	49.7	49.7
10	33.1	33.1	33.1	33.1	33.1	33.1	33.1	35.2
25	16.5	16.5	16.5	16.5	16.5	16.5	19.3	22.2
50	9.0	9.0	9.0	9.0	9.4	11.2	13.6	15.7
100	4.7	4.7	4.7	4.8	6.7	7.9	9.6	11.1
250	1.9	1.9	2.0	3.1	4.2	5.0	6.1	7.0
500	1.0	1.0	1.4	2.2	3.0	3.6	4.3	5.0
1,000	0.5	0.7	1.0	1.5	2.1	2.5	3.0	3.5
2,500	0.2	0.4	0.6	1.0	1.3	1.6	1.9	2.2
5,000	0.10	0.3	0.4	0.7	0.9	1.1	1.4	1.6
7,500	0.07	0.3	0.4	0.6	0.8	0.9	1.1	1.3
10,000	0.05	0.2	0.3	0.5	0.7	0.8	1.0	1.1
25,000	0.02	0.14	0.2	0.3	0.4	0.5	0.6	0.7

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-45
Selection of sample areas. App-45
Designation of sample housing units for the 1981 survey. App-45
Selection of the 1973 sample housing units. App-46
Splitting of the sample . App-46
CEN-SUP sample . App-46
Selection of supplemental sample housing units in rural areas . App-46
Selection of sample housing units for the 1976 Coverage Improvement Program . App-46
1977 and 1981 sample reductions . App-47
Supplemental sample from nonresidential conversions . App-47
1981 telephone interviewing experiment. App-47
ESTIMATION. App-47
RELIABILITY OF THE ESTIMATES. App-49
Sampling errors . App-49
Standard errors of estimates of levels . App-49
Standard errors of estimates of percentages. App-49
Standard errors of ratios . App-49
Illustration of the use of the standard error tables
Illustration I . App-50
Illustration II . App-50
Differences. App-50
Illustration of the computation of the standard error of a difference. App-51
Medians . App-51
Illustration of the computation of the 95-percent confidence interval for a median. App-51
Nonsampling errors App-52
Rounding errors. App-52
Reinterview program App-52
Coverage errors App-52
Effects of ratio estimation on estimates of change App-53
Possible effects of telephone interviewing on the data App-53
Standard error tables App-53

SAMPLE DESIGN

The 1981 estimates are based on data collected from September 1981 through December 1981 for the Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. The sample for this survey was spread over 461 sample areas (called primary sampling units), made up of 923 counties and independent cities with coverage in each of the 50 States and the District of Columbia.

Approximately 56,800 sample housing units (both occupied and vacant) were eligible for interview in the 1981 Annual Housing Survey. Of this number, 3,200 interviews were classified as "noninterview" for various reasons. Occupied housing units were classified as "noninterview" mainly, because the occupants refused to be interviewed after repeated calls. For vacant housing units, interviews were not obtained because an informed respondent was not found after repeated visits. In addition to the 56,800 eligible housing units, there were also 3,600 sample units which were visited but were ineligible for interview for the AHS in terms of collecting information relevant to the 1981 housing inventory.

Selection of sample areas—The United States was divided into areas made up of counties and independent cities referred to as primary sampling units (PSU's). These PSU's were then grouped into 376 strata, 156 of which consisted of only one PSU which was in sample. These 156 strata were mostly the larger SMSA's and were called self-representing (SR) since the sample from the sample area represented just that PSU. Each one of the other 220 strata consisted of a group of PSU's and were referred to as non-self-representing (NSR), since the sample of housing units from the sample PSU in a stratum represented the other PSU's in the stratum as well.

One PSU was selected from each NSR stratum with probability proportionate to the 1970 census population of the PSU. (This resulted in 220 NSR sample PSU's.) In addition, the NSR strata were grouped into 110 pairs and one stratum was picked at random from each pair. From this stratum, an additional PSU was selected independently of the other PSU selected from this stratum. Since the two PSU's were independently selected, it was possible for the same PSU to be selected twice. This occurred in 25 instances, producing an additional 85 NSR sample PSU's, thus giving a grand total of 461 PSU's.

Designation of sample housing units for the 1981 survey—The sample housing units designated to be interviewed in the 1981 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1980 survey (which included all sample housing units that were selected as part of the 1976 Coverage Improvement Program) and were not part of the 1981 sample reduction.

2. All sample housing units that were either type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of survey but which could become eligible in the future) in the 1980 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1981 AHS questionnaire, page (App-21).
3. All sample housing units that were selected from the list of building permits issued since the 1980 survey. (This sample represented the housing units built in permit-issuing areas, since the 1980 survey.)
4. Housing units added as the result of the updated listings in selected areas which do not issue building permits.

Selection of the 1973 sample housing units—The overall sampling rate used to select the sample for the 1973 AHS was about 1 in 1,366. The within-PSU sampling rate for the AHS was determined so that the overall probability of selection for each sample housing unit was the same (e.g., if the probability of selecting an NSR PSU was 1 in 10, then the within-PSU sampling rate would be 1 in 136.6).

Within the sample PSU's, a sample of the housing units enumerated in the 1970 Census of Population and Housing was selected for the AHS. In addtion, a sample of new construction building permits was also selected to represent the units constructed since the 1970 census. These samples were selected at about twice the rate mentioned previously (i.e., at 2 in 1,366), thereby producing a sample twice as large as needed. This sample was split into two equal-sized samples—one to be used for the AHS, and one to be held in reserve for possible future use for the AHS. The procedure used to split this sample into equal-sized samples is described in the next section.

The sample of 1970 census units was selected in several stages. Within the sample PSU's, the first step was the selection of a sample of census enumeration districts (ED's), administrative units used in the 1970 census. The probability of selection for an ED was proportional to the following 1970 census counts of housing units (HU's) and persons in group quarters, combined in the following formula:

$$\frac{\text{Number of HU's in the ED} + \dfrac{\text{Number of group quarters persons in the ED}}{3}}{4}$$

The next step was to select an expected cluster of about four neighboring housing units within each sample ED. For most of the ED's, the selection was accomplished using the list of addresses for the ED as compiled in the 1970 census. (These ED's are referred to as address ED's.) However, in those ED's where addresses were incomplete or inadequate (mostly rural areas), the selection process was accomplished using area sampling methods. These ED's were divided into segments (i.e., small land areas with well-defined boundaries, having an expected size of four, or a multiple of four, housing units) and a segment was selected. Those selected segments with an expected size which was a multiple of four were further subsampled at the time of interview so that an expected four housing units were chosen for interview.

The sample of new construction housing units was selected from building permits issued since January 1970. Within each sample PSU, the building permits were chronologically ordered by month issued, and compact clusters of approximately four housing units were created. These clusters were then sampled at the rate of 2 in 1,366. Housing units constructed since the 1970 census in areas which do not issue building permits were brought into the sample as a result of the area sample described above.

Splitting of the sample—The described sample selection procedure produced clusters (or segments) of size-four housing units for the sample taken from the census address frame, the new construction frame, and the area sampling frame (mainly rural areas). Clusters of this size should result in a minimum loss in precision for estimates of housing characteristics in rural areas because of the heterogeneity of neighboring units. However, clusters of size-two housing units, were considered to be more optimum within those areas where the housing characteristics of neighboring units tend to be very similar (i.e., urban areas and new construction units). A splitting operation was then carried out for clusters selected from the census address and the new construction frames. This consisted of halving each sample cluster from these frames. Thus, two housing units from each of these clusters were included in the survey and two housing units were held in reserve. No splitting operation was carried out within the clusters selected from the area sampling frame; every other area sample cluster of four housing units was used for the survey and the remaining clusters were assigned to the reserve sample.

CEN-SUP sample—Housing units at addresses missed in the census or inadequately described in the 1970 census address register did not have a chance of being selected for the AHS sample in address ED's. Consequently, a special operation was undertaken to develop a separate sample, called the census supplemental (CEN-SUP) sample to represent these units. Due to time constraints on this operation, only about 40 percent of the CEN-SUP units were ready to be interviewed in 1973. The rest were interviewed for the first time in the 1974 AHS.

Selection of supplemental sample housing units in rural areas—In 1974, it was decided to increase the reliability of the AHS estimates of rural housing characteristics by doubling the number of sample housing units from rural areas. This was accomplished by reactivating the reserve sample, selected in the original sampling operations in 1973, from rural areas only. For the reserve sample selected in census address and new construction frames, this meant the other half of each rural cluster (an expected two housing units) was reactivated in 1974. Similarly, for the area sampling frame, the entire reserve cluster (an expected four housing units) was reactivated in 1974 if the cluster was rural. This supplementation increased the overall probability of selection for sample housing units in rural areas to about 2 in 1,366; whereas, the overall probability of selection for sample housing units in urban areas remained at 1 in 1,366.

Selection of sample housing units for the 1976 Coverage Improvement Program—The 1976 Coverage Improvement Program was

undertaken to correct certain deficiencies in the AHS national sample from the census address and new construction frames. The coverage deficiences included the following types of housing units:

1. New construction housing units from building permits issued prior to January 1970 for which construction had not been completed at the time of the 1970 census.
2. Housing units converted to residential use in structures totally nonresidential at the time of the 1970 census.
3. Houses that have been moved onto their present site since the 1970 census.
4. Mobile homes and trailers placed in parks either missed in the 1970 census or established since the 1970 census.
5. Mobile homes and trailers placed outside parks since the 1970 census or vacant at the time of the 1970 census.

A sample of new construction housing units whose permits were issued before January 1970 was selected in two stages. In the first stage, units whose permits were issued before January 1970, but which were completed after the census, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. In the second stage, these units were then sampled so that the overall probability of selection was about 1 in 1,320.

A sample of mobile homes and trailers placed in parks missed by the census or established after the census was also selected in two stages. During the first stage, a list of mobile home parks was obtained from commercial listings. This list was then supplemented by additional parks identified by a canvassing operation similar to that performed in ED's where area sampling methods are used. The second stage consisted of dividing the parks into clusters of an expected size of four sites. These clusters were then sampled so that the overall probability of selection was about 1 in 1,366.

For the remaining housing units (i.e., mobile homes and trailers placed outside parks since the 1970 census, mobile homes and trailers vacant at the time of the 1970 census, units converted from nonresidential to residential use since the 1970 census, and houses that had been moved onto their present site since the 1970 census), the sampling was done in three stages. First, a subsample of the regular AHS sample housing units from the census address frame was selected. Second, succeeding structures that had been eligible to be selected from the census address frame were then listed until eight structures (excluding mobile home parks) were found. Finally, the intervening structures that had been listed which did not have a chance of selection in the AHS were identified and the units within these structures were interviewed.

1977 and 1981 sample reductions—By 1977, the addition to the sample, primarily from new construction and the coverage improvements, had increased the total sample size (interviews plus noninterviews) to about 81,000 housing units. The sample was reduced by about 7 percent to approximately 75,000 in 1977. This reduction did not include any CEN-SUP units or units which were selected as part of the 1976 Coverage Improvement Program. The sample was further reduced in 1981, first by about 5 percent overall and then the rural sample was reduced by about 25 percent in 125 "large" SMSA's[1] and by about 50 percent in "small" SMSA's[1] and outside of SMSA's. These reductions brought the sample size down to approximately 60,000 housing units in 1981.

Supplemental sample from nonresidential conversions—The purpose of this sample was to improve the coverage in address ED's (i.e., ED's in which 1970 census address listings were used for selecting the AHS sample) of housing units created in structures that were completely nonresidential at the time of the 1970 census. This sample was derived from listings created for the Commercial Victimization Survey (CVS), a nationally representative area sample survey conducted by the Census Bureau for the Department of Justice. Each of the CVS area segments was initially canvassed either in 1972 or 1975 and eligible business establishments were listed. These listings were updated every 6 months with the last updating having taken place in the last half of 1977. Each basic address containing a business establishment listed for CVS was visited for the 1980 Components of Inventory Change Survey (CINCH), which was conducted in conjunction with the 1980 AHS, to identify HU's in structures which were completely nonresidential at the time of the 1970 census. These cases were later matched to the 1970 census address listings to identify those cases in address ED's and as an additional check to see if housing units existed in these structures at the time of the 1970 census. Due to the timing restraints associated with these operations, these cases were not included for the 1980 AHS but were included for the 1981 survey.

1981 telephone interviewing experiment—A large scale telephone interviewing experiment was conducted in conjunction with the 1981 AHS national sample in order to provide more definitive information about the effect of telephone interviewing on the AHS data and cost. A random sample of about 12,500 AHS national sample units was selected for the experiment. Among the cases assigned for telephone interviewing, only those sample housing units that had been interviewed in 1980 and had a telephone number available were eligible to be interviewed by telephone. Since a large portion of the total AHS national sample was assigned for telephone interviewing, its possible effect on the data is mentioned in the section "Reliability of the Estimates" of this appendix.

ESTIMATION

In 1981, the AHS estimates employed a three-stage ratio estimation procedure. However, prior to implementation of the procedure, the basic weight (i.e., the inverse of the probability of selection) was adjusted to account for the type A noninterview housing units encountered in the AHS. This noninterview adjustment was done separately for occupied and

[1] A "large" SMSA is one with a 1970 population of 250,000 persons or more and a "small" SMSA is one with a 1970 population of less than 250,000 persons.

vacant units. The noninterview adjustment was equal to the following ratio:

$$\frac{\text{Interviewed housing units + Noninterviewed housing units}}{\text{Interviewed housing units}}$$

The first-stage ratio estimation procedure was employed for sample housing units from non-self-representing (NSR) PSU's only. This procedure was designed to reduce the contribution to the variance arising from the sampling of PSU's. The first-stage ratio estimation procedure takes into account the differences that existed at the time of the 1970 census in the distribution by tenure and residence of the housing population estimated from the sample NSR PSU's and that of the NSR housing population in each of the four census regions of the country.

The first-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{The 1970 census housing population in the residence-tenure category for all NSR strata in a census region}}{\text{Estimates of the housing population category using 1970 census housing counts for sample NSR PSU's in a census region}}$$

The numerators of the ratios were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR stratum and summing these counts across the NSR strata in each census region. The denominators were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR sample PSU, weighting these counts by the inverse of the probability of selecting that PSU, and summing these weighted counts across the NSR sample PSU's in each census region. The computed first-stage ratio estimation factor was then applied to the existing weight for each NSR sample unit in each first-stage ratio estimation category.

The second-stage ratio estimation procedure was designed to adjust the AHS sample estimate of conventional new construction housing units, i.e., sample units built April 1, 1970, or later, to independently derived current estimates where a known deficiency in the AHS sample exists (see the section on nonsampling error) for each of the four regions. These estimates were considered to be the best estimates available for the number of conventional new construction units in these categories.

The second-stage ratio estimation factor was as follows:

$$\frac{\text{Current best estimate of new construction in the category}}{\text{AHS sample estimate of new construction units in the category}}$$

The numerators of the ratios were derived from data based on the Survey of Construction (SOC). The denominators of the ratios were obtained from the weighted estimates for the AHS sample units using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in each second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all the AHS sample units. This procedure was designed to adjust the AHS sample estimates of housing (i.e., the estimates employing the noninterview, first-stage, and second-stage adjustments) to current vacant housing estimates for four categories of vacant housing units and to independently derived current housing estimates for 24 categories of occupied housing units. Each of these categories is a combination of the characteristics of residence, tenure, race and sex of householder.

The third-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{Current independent estimate of housing units in the category}}{\text{AHS sample estimate of housing units in the category}}$$

The numerators of the ratios for occupied housing units were derived from data based on the Current Population Survey (CPS), a sample household survey conducted monthly by the Bureau of the Census. This is the first time that CPS household estimates based on the 1980 census were used as the independent controls in this adjustment rather than the 1970 census-based controls. The 1970 based estimates were about 2 percent smaller than the 1980 based estimates. This difference should be taken into consideration when comparing the 1981 AHS estimates to the AHS estimates for 1980 or earlier years. The numerators of the ratios for vacant housing units were derived from data based on the Housing Vacancy Survey (HVS), a quarterly vacancy survey also conducted by the Bureau of the Census. The denominators of the ratios were obtained from the weighted estimates for the AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit in each third-stage ratio estimation category.

The second- and the third-stage ratio estimation procedures were iterated in order to bring the AHS estimates into close agreement with both sets of "independent" estimates. The second-stage was modified so that the estimates for all 27 categories of new construction would be identical to the estimates before the third-stage. Hence, the repeated second-stage had the effect of controlling the AHS sample estimates of new construction units to the "unbiased" sample estimates for 9 categories of new construction mobile homes and trailers for each of the four regions and of adjusting the AHS sample estimate of 18 categories of conventional new construction units to the independently derived current estimates.

The numerators were either the unbiased weighted estimates for the AHS sample housing units, using the existing weight after the first-stage ratio estimation procedure (i.e., the estimates employing the noninterview and first-stage adjustments) or the independent estimate derived from data based on the Survey of Construction (SOC).

The denominators of the ratios in this iterative process were obtained from the weighted estimates for the AHS sample units after the previous stage of ratio estimation. The factors resulting from this iterative process were than applied to the existing weight on the appropriate records, and the resulting product was used as the final weight for tabulation.

The third-stage ratio estimation procedure, as well as the overall estimation procedure, reduced the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. The distribution of the housing population

selected for the sample differed somewhat, by chance, from that of the Nation as a whole in such basic housing characteristics as tenure, vacancy status, residence, race and sex of householder. These characteristics are probably closely correlated with other housing characteristics measured for the AHS. Therefore, through the use of the three-stage ratio estimation procedure one can expect the sample estimate to be improved substantially.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS national sample.

Sampling errors—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same questionnaires, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The variability between estimates from all possible samples is defined as sampling error. One common measure of sampling error is the standard error which measures the precision with which an estimate from a sample approximates the average result of all possible samples. In addition, the standard error, as calculated for this report, also partially reflects the variation in the estimates due to some nonsampling errors, but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, and biases and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates so that the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, and each of these samples was surveyed under essentially the same general conditions and an estimate and its estimated standard error were calculated for each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples either is or is not contained in any particular computed interval. However, for a particular sample one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables (pages App-53 to App-57) of this appendix are approximations to the standard errors of various estimates shown in this report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than the precise standard error for any specific item.

Standard errors of estimates of levels—Tables I and II present the standard errors applicable to the 1981 national housing inventory estimates in this report. Tables III, IV, and V present the standard errors applicable to estimates for the Northeast, North Central, South, and West Regions. Linear interpolation should be used to determine standard errors for levels of estimates not specifically shown in tables I through V.

Standard errors of estimates of percentages—The reliability of an estimated percentage, computed by using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more.

Tables VI through X present the standard errors of estimated percentages. Table VI shows the approximate standard errors of all national estimated percentages of housing units except those pertaining to the specified items in table II. The standard errors shown in table VII should be used for specified items. Table VIII shows the approximate standard error of all regional estimated percentages of housing units except those pertaining to the specified items in tables IV and V. The standard errors shown in tables IX and X should be used for those specified items. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables VI through X.

Included in tables I through X are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered to be overestimates of the true standard errors and should be used primarily for construction of confidence intervals for characteristics when an estimate of zero is obtained.

Standard errors of ratios—For ratios of the form (100) (x/y), where x is not a subclass of y, tables VI through X underestimate the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be obtained by letting the standard error of the ratio be approximately equal to:

$$(100)\left(\frac{x}{y}\right)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

where: x = the numerator of the ratio
y = the denominator of the ratio
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

Illustration of the use of the standard error tables. Illustration I—Table A-1 of this report shows that inside the United States there were 3,775,000 owner-occupied housing units occupied by recent movers in 1981. Interpolation in standard error table I shows that the standard error of an estimate of this size is approximately 85,000. The following procedure was used in interpolating.

The information presented in the following table was extracted from standard error table I. The entry for "x" is the one sought.

Size of estimate (000)	Standard error (000)
2,500	71
3,775	x
5,000	98

By vertically interpolating between 71 and 98, the entry for "x" is determined to be 85.

$$3{,}775-2{,}500 = 1{,}275$$
$$5{,}000-2{,}500 = 2{,}500$$

$$71 + \frac{1{,}275}{2{,}500}(98-71) = 85$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 3,690,000 to 3,860,000 housing units. Therefore, a conclusion that the average estimate of 1981 housing units of this type lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples lies within the interval from 3,639,000 to 3,911,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 3,605,000 to 3,945,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 3,775,000 owner-occupied housing units occupied by recent movers in 1981, 167,000, or 4.4 percent, had six persons or more. Interpolation in standard error table VI (i.e., interpolation on both the base and percent) shows that the standard error of the above percentage is 0.5 percentage points. The following procedure was used in interpolating.

The information presented in the following table was extracted from table VI. The entry for "p" is the one sought.

Base of percentage (000)	Estimated percentage		
	2	4.4	5
2,500	0.4	a	0.6
3,775		p	
5,000	0.3	b	0.4

1. By horizontal interpolation between 0.4 and 0.6, the entry for cell "a" is determined to be 0.6.

$$4.4-2.0 = 2.4$$
$$5.0-2.0 = 3.0$$

$$0.4 + \frac{2.4}{3.0}(0.6-0.4) = 0.6$$

2. By horizontal interpolation between 0.3 and 0.4, the entry for cell "b" is determined to be 0.4.

$$4.4-2.0 = 2.4$$
$$5.0-2.0 = 3.0$$

$$0.3 + \frac{2.4}{3.0}(0.4-0.3) = 0.4$$

3. By vertical interpolation between 0.6 and 0.4, the entry for "p" is determined to be 0.5.

$$3{,}775-2{,}500 = 1{,}275$$
$$5{,}000-2{,}500 = 2{,}500$$

$$0.6 + \frac{1{,}275}{2{,}500}(0.4-0.6) = 0.5$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 3.9 to 4.9 percent; the 90-percent confidence interval is from 3.6 to 5.2 percent; and the 95-percent confidence interval is from 3.4 to 5.4 percent.

Illustration II—Table A-21 of this report shows that in the United States in 1981 there were 178,000 owner-occupied housing units having a recent mover householder of Spanish origin. Interpolation in standard error table II shows that the standard error of an estimate of this size is approximately 22,000. Consequently, the 68-percent confidence interval is from 156,000 to 200,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1981 owner-occupied housing units having a recent mover householder of Spanish origin lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval from 143,000 to 213,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 134,000 to 222,000 housing units with 95 percent confidence.

In 1981, table A-21 also shows that of the 178,000 owner-occupied housing units having a recent mover householder of Spanish origin, 88,000, or 49.4 percent, had three bedrooms. Interpolation in standard error table VII (i.e., interpolation on both the base and the percent) shows that the standard error of the above percentage is 6.8 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 42.6 to 56.2 percent; the 90-percent confidence interval is from 38.5 to 60.3 percent; and the 95-percent confidence interval is from 35.8 to 63.0 percent.

Differences—The standard errors shown are not directly applicable to differences between two sample estimates. The stand-

ard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristics in two different areas or the difference between separate and uncorrelated characteristics in the same area. If there is a high positive correlation between the two characteristics the formula will overestimate the true error; if there is a high negative correlation between the two characteristics, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference—Table A-1 of this report shows that in the United States there were 330,000 owner-occupied housing units occupied by recent movers with five persons. Table A-1 also shows that in the United States in 1981 there were 167,000 owner-occupied housing units occupied by recent movers with six persons or more. Thus, the apparent difference between the number of 1981 owner-occupied housing units occupied by recent movers with five persons and the number with six persons or more is 163,000. Interpolation in standard error table I shows that the standard error on an estimate of 330,000 to be approximately 26,000 and the standard error on an estimate of 167,000 to be approximately 18,000. Therefore, the standard error of the estimated difference of 163,000 is about 32,000.

$$32{,}000 = \sqrt{(26{,}000)^2 + (18{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 163,000 difference is from 131,000 to 195,000 housing units. Therefore, a conclusion that the average estimate of this difference, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 112,000 to 214,000 housing units, and the 95-percent confidence interval is from 99,000 to 227,000. Thus, we can conclude with 95 percent confidence that the number of 1981 owner-occupied housing units occupied by recent movers with five persons is different than the number of owner-occupied housing units occupied by recent movers with six persons or more since the 95-percent confidence interval of this difference does not include zero or negative values.

Medians—For medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From standard error tables VI through X, determine the standard error of a 50-percent characteristic on the base of the median.
2. Add to and subtract from 50 percent the standard error determined in step 1.
3. Using the distribution of the characteristics, determine the confidence interval corresponding to the two points established in step 2. To find the lower endpoint of the confidence interval, it is necessary to know into which interval of the distribution the lower percentage limit falls. Similarly, to find the upper endpoint of the confidence interval, it is necessary to know into which interval of the distribution the upper percentage limit falls. These two distribution intervals could be different, although this will not happen very often.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval for a median—Table A-1 of this report shows the median number of persons in owner-occupied housing units occupied by recent movers in the United States was 2.7 in 1981. The base of the distribution from which this median was determined is 3,775,000 housing units.

1. Interpolation using standard error table VI shows that the standard error of 50 percent on a base of 3,775,000 is approximately 1.2 percentage points.
2. To obtain a 95-percent confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 47.6 and 52.4.
3. From the distribution for "persons" in table A-1, the interval for owner-occupied housing units occupied by recent movers with three persons (for purposes of calculating the median, the category of three persons is considered to be from 2.5 to 3.5 persons) corresponds to the 47.6 percent derived in step 2. About 1,753,000 housing units, or 46.4 percent, fall below this interval, and 789,000 housing units, or 20.9 percent, fall within this interval. By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\,\frac{(47.6-46.4)}{20.9} = 2.6$$

Similarly, the interval for owner-occupied housing units occupied by recent movers with three persons corresponds to the 52.4 percent derived in step 2. About 1,753,000 housing units, or 46.4 percent, fall below this interval, and 789,000 housing units, or 20.9 percent, fall within this interval. The upper limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\,\frac{(52.4-46.4)}{20.9} = 2.8$$

Thus, the 95-percent confidence interval ranges from 2.6 to 2.8 persons.

Nonsampling errors—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases; definitional difficulties; differences in the interpretation of questions; inability or unwillingness to provide correct information on the part of respondents; mistakes in recording or coding the data; and other errors of collection, response, processing, coverage, and estimation for missing data. Nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for the 1981 AHS national sample.

Rounding errors—In errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of this survey. Also, since medians in this report were computed using unrounded data, they can differ from medians calculated directly from the published data.

Reinterview program—For the AHS national sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a subsample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and thus were the basis for the measurement of the "content" error of these AHS estimates.

As part of the reinterview, an additional check was carried out for interviewer evaluation and quality control. This check was made at each of these households to determine if the following were done during the original interview.

1. The correct housing unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

The results of the 1981 and 1980 reinterview studies were not available at the time of publication. In addition, the 1979 reinterview was carried out as an interviewer quality control only. Thus, the results of the 1977 and 1978 reinterview studies which were presented in the Census Bureau memoranda, "Reinterview Results for the Annual Housing Survey—National Sample 1977" and "Reinterview Results for the Annual Housing Survey—National Sample 1978" are presented below.

In 1978, a substantial portion of the reinterview questionnaire was devoted to testing the new questions 1 through 7 (parts a and b). The questions (part a), which were asked only at housing units interviewed in the previous year, determined whether there had been a change since last year in selected nonattitudinal items. If a change had been recorded or the respondent did not know if a change had occurred, part b of the question, which collects the value of the item, was asked. The reinterviewers asked these items using the questions as formatted in 1977. Comparing the responses from the differently formatted questions, the 1978 reinterview found that 80 percent of the questions showed low levels of inconsistency with the remainder showing moderate levels.

The 1977 reinterview program showed moderate to high levels of inconsistency with about 21 percent of the nonattitudinal and 56 percent of the attitudinal items showing high levels of inconsistency. A large proportion (43 percent) of the nonattitudinal items showed a low level of inconsistency. Moderate levels indicate that there are some problems with inconsistent reporting and high levels indicate that improvements are needed in the data collection methods or that the category concepts themselves are ambiguous.

Cross-tabulations involving those items which are subject to substantial levels of inconsistency may be subject to a large distortion as a consequence of the associated high level of response variance, and thus, are considered to be less reliable than comparable cross-tabulations which do not involve these data. As a consequence, the tables in this report which contain such cross-tabulations have been footnoted with a cautionary statement.

The 1970 census reinterview results provide illustrations of possible nonsampling errors for some of the items which also appear in the AHS. For example, median value of homes was consistently underestimated by about 5 percent, and the average monthly cost of electricity and utility gas were consistently overestimated although the net effect on average gross rent was fairly small.

A possible explanation for the results of the AHS and census reinterview studies, as well as the surveys themselves, is that the respondents may lack precise information. Also, because the results of the reinterview studies are derived from sample surveys, there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of these studies.

Coverage errors—Deficiencies in the representation of conventional new construction (non-mobile home or trailer) for the AHS new construction sample (mentioned previously in the

section on estimation) is an example of coverage errors. During the sampling of building permits, only those issued more than 4 months before the survey began were eligible to be selected to represent conventional new construction. Due to time constraints, it is not possible to sample units whose permits are issued less than 4 months in advance of the survey.

It is estimated that the 1981 AHS sample missed about 0.8 percent (i.e., about 157,000 units) of conventional housing units built after April 1970 because the permits for these units, which were built before October 1981, were issued less than 4 months in advance of the survey. The second-stage ratio estimation procedure was employed to reduce the effect of this deficiency although some bias in the AHS estimates of conventional new construction probably still exists. Review of the second-stage ratio estimation procedure indicates that we have consistently overcompensated for this deficiency every year since 1975 by ratio adjusting the new construction to counts of new construction for the end of the interview period, which has been December or January instead of October. This overcompensation may inflate the new construction counts by 100,000 to 300,000 units.

Changing this procedure to correct for this overcompensation would reduce the estimate of housing units built since the last survey since it would be based on a 10-month time period rather than 12 months. Thus, it was decided not to change this procedure so that the estimate of housing units built since the last survey would be consistent with previous years' estimates, which are generally based on a 12-month time period.

In addition, the 1976 Coverage Improvement Program had certain deficiencies. First, when the canvassing was done to identify mobile homes and trailer parks that were not in the same sample frame or not on the commercial lists, only 92 percent of the census address frame ED's were represented. Second, it appears that the listing procedure (used to find mobile homes and trailers placed outside parks, units converted from nonresidential to residential, and houses that had been moved onto their present site) was not very efficient for finding nonresidential conversions (which might be primarily in business districts), since the listing procedure started from a residential unit. (The sample estimate of this component was approximately 16,000 housing units with a standard error of 12,000.)

Finally, it is felt that deficiencies also exist in ED's where area sampling methods are used. As before, it had been assumed that all housing units located inside these ED's would be represented in the sample. However, it has been estimated that the 1981 AHS sample missed as much as 2 percent (i.e., as much as 400,000 units) of all housing units in ED's where area sampling methods are used because these units are not listed during the canvassing.

The third stage of ratio estimation corrects for these deficiencies as far as the total count of housing is concerned, i.e., it adjusts the estimate of the total housing inventory to the best available estimate. However, biases of subtotals would still remain.

Effects of ratio estimation on the estimates of change—As stated previously, 1981 was the first AHS for which 1980 census-based estimates, rather than 1970 census-based estimates, were used in the third-stage ratio estimation procedure. Since the 1980-based estimates are about 2 percent higher than the corresponding 1970-based estimates, this will cause estimates of change between 1981 AHS data and data for 1980 or earlier years to be overstated. This overstatement should be taken into consideration during the analysis of estimates of change.

Possible effects of telephone interviewing on the data—A preliminary analysis of the 1981 Telephone Interviewing Experiment has been completed. It has been concluded that data collected, using the telephone interviewing procedures, were not sufficiently different from data collected by regular personal interviews, to preclude basing the 1981 AHS national sample published data. However, since the percentage of the comparison tests that were significant was slightly above what could be expected, by chance, there was some evidence that telephone interviewing may have had an effect on the data. Thus, some caustion should be exercised, in making comparisons between the 1981 AHS data and data from preceding years.

TABLE I. Standard Errors of Estimated Numbers of Housing Units: 1981 (Excluding Estimates of Housing Units Pertaining to New Construction, No Bedrooms, Lacking Complete Plumbing, Mobile Homes, and Housing Units With Spanish-Origin Householder)

(68 chances out of 100)

Size of estimate (000)	Standard error		Size of estimate (000)	Standard error	
	Total or White (000)	Black (000)		Total or White (000)	Black (000)
0	2	2	2,500 . . .	71	61
5	3	3	5,000 . . .	98	68
10	5	5	7,500 . . .	119	51
25	7	7	10,000. . .	135	–
50	10	10	25,000. . .	193	–
100.	14	14	50,000. . .	216	–
250.	23	22	75,000. . .	167	–
500.	32	31	85,000. . .	112	–
1,000 . . .	45	43			

TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to New Construction, No Bedrooms, Lacking Complete Plumbing, Mobile Homes, and Housing Units With Spanish-Origin Householder: 1981

(68 chances out of 100)

Size of estimate (000)	Standard error: Total, White, or Spanish origin (000)	Standard error: Black (000)	Size of estimate (000)	Standard error: Total, White, or Spanish origin (000)	Standard error: Black (000)
0	3	3	1,000	53	50
5	4	4	2,500	83	71
10	5	5	5,000	116	79
25	8	8	7,500	139	60
50	12	12	10,000	159	–
100	17	17	25,000	227	–
250	27	26	50,000	253	–
500	37	37			

TABLE III. Standard Errors of Estimated Numbers of Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1981 (Excluding Estimates of Housing Units Pertaining to New Construction, No Bedrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions, and Excluding Mobile Homes for Each of the Regions)

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	500	33
5	3	1,000	47
10	5	2,500	74
25	7	5,000	105
50	10	7,500	129
100	15	10,000	148
250	23	25,000	235

TABLE IV. Standard Errors of Estimated Numbers of Housing Units Pertaining to New Construction, No Bedrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions and to Mobile Homes for the Northeast and North Central Regions: 1981

(68 chances out of 100. For estimates pertaining to mobile homes for the West Region, apply a factor of 1.66 to the standard errors)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	3	500	39
5	4	1,000	56
10	6	2,500	87
25	9	5,000	122
50	13	7,500	148
100	18	10,000	168
250	28	25,000	243

TABLE V. Standard Errors of Estimated Numbers of Housing Units Pertaining to Mobile Homes for the South Region: 1981

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	5	500	50
5	5	1,000	70
10	7	2,500	109
25	11	5,000	150
50	16	7,500	179
100	22	10,000	201
250	35	25,000	259

TABLE VI. Standard Errors of Estimated Percentages of Housing Units: 1981 (Excluding Estimated Percentages of Housing Units Pertaining to New Construction, No Bedrooms, Lacking Complete Plumbing, Mobile Homes, and Housing Units With Spanish-Origin Householder)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	29.1	29.1	29.1	29.1	29.1	29.1	29.1	32.0
10	17.0	17.0	17.0	17.0	17.0	17.0	19.6	22.7
25	7.6	7.6	7.6	7.6	8.6	10.2	12.4	14.3
50	3.9	3.9	3.9	4.4	6.1	7.2	8.8	10.1
100	2.0	2.0	2.0	3.1	4.3	5.1	6.2	7.2
250	0.8	0.9	1.3	2.0	2.7	3.2	3.9	4.5
500	0.4	0.6	0.9	1.4	1.9	2.3	2.8	3.2
1,000	0.2	0.5	0.6	1.0	1.4	1.6	2.0	2.3
2,500	0.08	0.3	0.4	0.6	0.9	1.0	1.2	1.4
5,000	0.04	0.2	0.3	0.4	0.6	0.7	0.9	1.0
7,500	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.8
10,000	0.02	0.14	0.2	0.3	0.4	0.5	0.6	0.7
25,000	0.01	0.09	0.13	0.2	0.3	0.3	0.4	0.5
50,000	–	0.06	0.09	0.14	0.2	0.2	0.3	0.3
75,000	–	0.05	0.07	0.11	0.2	0.2	0.2	0.3
85,000	–	0.05	0.07	0.11	0.15	0.2	0.2	0.2

TABLE VII. Standard Errors of Estimated Percentages of Housing Units Pertaining to New Construction, No Bedrooms, Lacking Complete Plumbing, Mobile Homes, and Housing Units With Spanish-Origin Householder: 1981

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	36.1	36.1	36.1	36.1	36.1	36.1	36.1	37.6
10	22.0	22.0	22.0	22.0	22.0	22.0	23.0	26.6
25	10.2	10.2	10.2	10.2	10.2	12.0	14.6	16.8
50	5.3	5.3	5.3	5.3	7.1	8.5	10.3	11.9
100	2.7	2.7	2.7	3.7	5.0	6.0	7.3	8.4
250	1.1	1.1	1.5	2.3	3.2	3.8	4.6	5.3
500	0.6	0.7	1.1	1.6	2.3	2.7	3.3	3.8
1,000	0.3	0.5	0.7	1.2	1.6	1.9	2.3	2.7
2,500	0.11	0.3	0.5	0.7	1.0	1.2	1.5	1.7
5,000	0.06	0.2	0.3	0.5	0.7	0.8	1.0	1.2
7,500	0.04	0.2	0.3	0.4	0.6	0.7	0.8	1.0
10,000	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.8
25,000	0.01	0.11	0.15	0.2	0.3	0.4	0.5	0.5
50,000	0.01	0.07	0.11	0.2	0.2	0.3	0.3	0.4

TABLE VIII. Standard Errors of Estimated Percentages of Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1981 (Excluding Estimated Percentages of Housing Units Pertaining to New Construction, No Bedrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions and Excluding Mobile Homes for Each of the Regions)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	30.6	30.6	30.6	30.6	30.6	30.6	30.6	33.2
10	18.1	18.1	18.1	18.1	18.1	18.1	20.3	23.5
25	8.1	8.1	8.1	8.1	8.9	10.6	12.9	14.8
50	4.2	4.2	4.2	4.6	6.3	7.5	9.1	10.5
100	2.2	2.2	2.2	3.2	4.5	5.3	6.4	7.4
250	0.9	0.9	1.3	2.0	2.8	3.4	4.1	4.7
500	0.4	0.7	0.9	1.4	2.0	2.4	2.9	3.3
1,000	0.2	0.5	0.7	1.0	1.4	1.7	2.0	2.3
2,500	0.09	0.3	0.4	0.6	0.9	1.1	1.3	1.5
5,000	0.04	0.2	0.3	0.5	0.6	0.7	0.9	1.0
7,500	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.9
10,000	0.02	0.15	0.2	0.3	0.4	0.5	0.6	0.7
25,000	0.01	0.09	0.13	0.2	0.3	0.3	0.4	0.5

TABLE IX. Standard Errors of Estimated Percentages of Housing Units Pertaining to New Construction, No Bedrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions, and to Mobile Homes for the Northeast and North Central Regions: 1981

(68 chances out of 100. For estimated percentages pertaining to mobile homes for the West Region, apply a factor of 1.66 to the standard errors)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	38.5	38.5	38.5	38.5	38.5	38.5	38.5	39.6
10	23.9	23.9	23.9	23.9	23.9	23.9	24.2	28.0
25	11.1	11.1	11.1	11.1	11.1	12.6	15.3	17.7
50	5.9	5.9	5.9	5.9	7.5	8.9	10.8	12.5
100	3.0	3.0	3.0	3.9	5.3	6.3	7.7	8.8
250	1.2	1.2	1.6	2.4	3.4	4.0	4.8	5.6
500	0.6	0.8	1.1	1.7	2.4	2.8	3.4	4.0
1,000	0.3	0.6	0.8	1.2	1.7	2.0	2.4	2.8
2,500	0.13	0.4	0.5	0.8	1.1	1.3	1.5	1.8
5,000	0.06	0.2	0.4	0.5	0.8	0.9	1.1	1.3
7,500	0.04	0.2	0.3	0.4	0.6	0.7	0.9	1.0
10,000	0.03	0.2	0.2	0.4	0.5	0.6	0.8	0.9
25,000	0.01	0.11	0.2	0.2	0.3	0.4	0.5	0.6

TABLE X. Standard Errors of Estimated Percentages of Housing Units Pertaining to Mobile Homes for the South Region: 1981

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	49.7	49.7	49.7	49.7	49.7	49.7	49.7	49.7
10	33.1	33.1	33.1	33.1	33.1	33.1	33.1	35.2
25	16.5	16.5	16.5	16.5	16.5	16.5	19.3	22.2
50	9.0	9.0	9.0	9.0	9.4	11.2	13.6	15.7
100	4.7	4.7	4.7	4.8	6.7	7.9	9.6	11.1
250	1.9	1.9	2.0	3.1	4.2	5.0	6.1	7.0
500	1.0	1.0	1.4	2.2	3.0	3.6	4.3	5.0
1,000	0.5	0.7	1.0	1.5	2.1	2.5	3.0	3.5
2,500	0.2	0.4	0.6	1.0	1.3	1.6	1.9	2.2
5,000	0.10	0.3	0.4	0.7	0.9	1.1	1.4	1.6
7,500	0.07	0.3	0.4	0.6	0.8	0.9	1.1	1.3
10,000	0.05	0.2	0.3	0.5	0.7	0.8	1.0	1.1
25,000	0.02	0.14	0.2	0.3	0.4	0.5	0.6	0.7

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-45
Selection of sample areas. App-45
Designation of sample housing for the 1981 survey App-45
Selection of the 1973 sample housing units. App-46
Splitting of the sample App-46
CEN-SUP sample App-46
Selection of supplemental sample housing units in rural areas App-46
Selection of sample housing units for the 1976 Coverage Improvement Program App-46
1977 and 1981 sample reductions . App-47
Supplemental sample from nonresidential conversions App-47
1981 telephone interviewing experiment App-47
ESTIMATION. App-47
RELIABILITY OF THE ESTIMATES App-49
Sampling errors App-49
Standard errors of estimates of levels App-49
Standard errors of estimates of percentages. App-49
Standard errors of ratios App-49
Illustration of the use of the standard error tables. Illustration I App-50
Illustration II App-50
Differences. App-50
Illustration of the computation of the standard error of a difference App-51
Medians App-51
Illustration of the computation of the 95-percent confidence interval for a median. App-51
Nonsampling errors App-51
Reinterview program App-51
Coverage errors App-52
Effects of ratio estimation on the estimates of change App-53
Possible effects of telephone interviewing on the data App-53
Rounding errors. App-53
Standard error tables App-53

SAMPLE DESIGN

The 1981 estimates are based on data collected from September 1981 through December 1981 for the Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. The sample for this survey was spread over 461 sample areas (called primary sampling units), made up of 923 counties and independent cities with coverage in each of the 50 States and the District of Columbia.

Approximately 56,800 sample housing units (both occupied and vacant) were eligible for interview in the 1981 Annual Housing Survey. Of this number, 3,200 interviews were classified as "noninterview" for various reasons. Occupied housing units were classified as "noninterview" mainly because the occupants refused to be interviewed after repeated calls. For vacant housing units, interviews were not obtained because an informed respondent was not found after repeated visits. In addition to the 56,800 eligible housing units, there were also 3,600 sample units which were visited but were ineligible for interview for the AHS in terms of collecting information relevant to the 1981 housing inventory.

Selection of sample areas—The United States was divided into areas made up of counties and independent cities referred to as primary sampling units (PSU's). These PSU's were then grouped into 376 strata, 156 of which consisted of only one PSU which was in sample with certainty. These 156 strata were mostly the larger SMSA's and were called self-representing (SR) since the sample from the sample area represented just that PSU. Each one of the other 220 strata consisted of a group of PSU's and were referred to as non-self-representing (NSR), since the sample of housing units from the sample PSU in a stratum represented the other PSU's in the stratum as well.

One PSU was selected from each NSR stratum with probability proportionate to the 1970 census population of the PSU. (This resulted in 220 NSR sample PSU's.) In addition, the NSR strata were grouped into 110 pairs and one stratum was picked at random from each pair. From this stratum, an additional PSU was selected independently of the other PSU selected from this stratum. Since the two PSU's were independently selected, it was possible for the same PSU to be selected twice. This occurred in 25 instances, producing an additional 85 NSR sample PSU's, thus giving a grand total of 461 PSU's.

Designation of sample housing units for the 1981 survey—The sample housing units designated to be interviewed in the 1981 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1980 survey (which included all sample housing units that were selected as part of the 1976 Coverage Improvement Program) and were not part of the 1981 sample reduction.

2. All sample housing units that were either type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of survey but which could become eligible in the future) in the 1980 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1981 AHS questionnaire, page App-21).
3. All sample housing units that were selected from the list of building permits issued since the 1980 survey. (This sample represented the housing units built in permit-issuing areas, since the 1980 survey.)
4. Housing units added as the result of the updated listings in selected areas which do not issue building permits.

Selection of the 1973 sample housing units—The overall sampling rate used to select the sample for the 1973 AHS was about 1 in 1,366. The within-PSU sampling rate for the AHS was determined so that the overall probability of selection for each sample housing unit was the same (e.g., if the probability of selecting a NSR PSU was 1 in 10, then the within-PSU sampling rate would be 1 in 136.6).

Within the sample PSU's, a sample of the housing units enumerated in the 1970 Census of Population and Housing was selected for the AHS. In addition, a sample of new construction building permits was also selected to represent the units constructed since the 1970 census. These samples were selected at about twice the rate mentioned previously (i.e., at 2 in 1,366), thereby producing a sample twice as large as needed. This sample was split into two equal-sized samples—one to be used for the AHS, and one to be held in reserve for possible future use for the AHS. The procedure used to split this sample into equal-sized samples is described in the next section.

The sample of 1970 census units was selected in several stages. Within the sample PSU's, the first step was the selection of a sample of census enumeration districts (ED's), administrative units used in the 1970 census. The probability of selection for an ED was proportional to the following 1970 census counts of housing units (HU's) and persons in group quarters, combined in the following formula:

$$\frac{\text{Number of HU's in the ED} + \dfrac{\text{Number of group quarters persons in the ED}}{3}}{4}$$

The next step was to select an expected cluster of about four neighboring housing units within each sample ED. For most of the ED's, the selection was accomplished using the list of addresses for the ED as compiled in the 1970 census. (These ED's are referred to as address ED's.) However, in those ED's where addresses were incomplete or inadequate (mostly rural areas), the selection process was accomplished using area sampling methods. These ED's were divided into segments (i.e., small land areas with well-defined boundaries, having an expected size of four, or a multiple of four, housing units) and a segment was selected. Those selected segments with an expected size which was a multiple of four were further subsampled at the time of interview so that an expected four housing units were chosen for interview.

The sample of new construction housing units was selected from building permits issued since January 1970. Within each sample PSU, the building permits were chronologically ordered by month issued, and compact clusters of approximately four housing units were created. These clusters were then sampled at the rate of 2 in 1,366. Housing units constructed since the 1970 census in areas which do not issue building permits were brought into the sample as a result of the area sample described.

Splitting of the sample—The described sample selection procedure produced clusters (or segments) of size-four housing units for the sample taken from the census address frame, the new construction frame, and the area sampling frame (mainly rural areas). Clusters of this size should result in a minimum loss in precision for estimates of housing characteristics in rural areas because of the heterogeneity of neighboring units. However, clusters of size-two housing units were considered to be more optimum within those areas where the housing characteristics of neighboring units tend to be very similar (i.e., urban areas and new construction units). A splitting operation was then carried out for clusters selected from the census address and the new construction frames. This consisted of halving each sample cluster from these frames. Thus, two housing units from each of these clusters were included in the survey and two housing units were held in reserve. No splitting operation was carried out within the clusters selected from the area sampling frame; every other area sample cluster of four housing units was used for the survey and the remaining clusters were assigned to the reserve sample.

CEN-SUP sample—Housing units at addresses missed in the census or inadequately described in the 1970 census address register did not have a chance of being selected for the AHS sample in address ED's. Consequently, a special operation was undertaken to develop a separate sample, called the census supplemental (CEN-SUP) sample to represent these units. Due to time constraints on this operation, only about 40 percent of the CEN-SUP units were ready to be interviewed in 1973. The rest were interviewed for the first time in the 1974 AHS.

Selection of supplemental sample housing units in rural areas—In 1974, it was decided to increase the reliability of the AHS estimates of rural housing characteristics by doubling the number of sample housing units from rural areas. This was accomplished by reactivating the reserve sample, selected in the original sampling operations in 1973, from rural areas only. For the reserve sample selected in census address and new construction frames, the other half of each rural cluster (an expected two housing units) was reactivated in 1974. Similarly, for the area sampling frame, the entire reserve cluster (an expected four housing units) was reactivated in 1974 if the cluster was rural. This supplementation increased the overall probability of selection for sample housing units in rural areas to about 2 in 1,366; whereas, the overall probability of selection for sample housing units in urban areas remained at 1 in 1,366.

Selection of sample housing units for the 1976 Coverage Improvement Program—The 1976 Coverage Improvement Pro-

gram was undertaken to correct certain deficiencies in the AHS national sample from the census address and new construction frames. The coverage deficiencies included the following types of housing units:

1. New construction housing units from building permits issued prior to January 1970 for which construction had not been completed at the time of the 1970 census.
2. Housing units converted to residential use in structures totally nonresidential at the time of the 1970 census.
3. Houses that have been moved onto their present site since the 1970 census.
4. Mobile homes and trailers placed in parks either missed in the 1970 census or established since the 1970 census.
5. Mobile homes and trailers placed outside parks since the 1970 census or vacant at the time of the 1970 census.

A sample of new construction housing units whose permits were issued before January 1970 was selected in two stages. First, units whose permits were issued before January 1970, but which were completed after the census, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. Second, these units were then sampled so that the overall probability of selection was about 1 in 1,320.

A sample of mobile homes and trailers placed in parks missed by the census or established after the census was also selected in two stages. During the first stage, a list of mobile home parks was obtained from commercial listings. This list was then supplemented by additional parks identified by a canvassing operation similar to that performed in ED's where area sampling methods are used. The second stage consisted of dividing the parks into clusters of an expected size of four sites. These clusters were then sampled so that the overall probability of selection was about 1 in 1,366.

For the remaining housing units (i.e., mobile homes and trailers placed outside parks since the 1970 census, mobile homes and trailers vacant at the time of the 1970 census, units converted from nonresidential to residential use since the 1970 census, and houses that had been moved onto their present site since the 1970 census), the sampling was done in three stages. First, a subsample of the regular AHS sample housing units from the census address frame was selected. Second, succeeding structures that had been eligible to be selected from the census address frame were then listed until eight structures (excluding mobile home parks) were found. Finally, the intervening structures that had been listed which did not have a chance of selection in the AHS were identified and the units within these structures were interviewed.

1977 and 1981 sample reductions—By 1977, the addition to the sample, primarily from new construction and the coverage improvements, had increased the total sample size (interviews plus noninterviews) to about 81,000 housing units. The sample was reduced by about 7 percent to approximately 75,000 in 1977. This reduction did not include any CEN-SUP units or units which were selected as part of the 1976 Coverage Improvement Program. Initially, the sample was further reduced in 1981 by about 5 percent. Subsequently, the rural sample was reduced by about 25 percent in 125 "large" SMSA's[1] and by about 50 percent in "small" SMSA's[1] and outside the SMSA's. These reductions brought the sample size down to approximately 60,000 housing units in 1981.

Supplemental sample from nonresidential conversions—The purpose of this sample was to improve the coverage in address ED's (i.e., ED's in which 1970 census address listings were used for selecting the AHS sample) of housing units created in structures that were completely nonresidential at the time of the 1970 census. This sample was derived from listings created for the Commercial Victimization Survey (CVS), a nationally representative area sample survey conducted by the Census Bureau for the Department of Justice. Each of the CVS area segments was initially canvassed either in 1972 or 1975 and eligible business establishments were listed. These listings were updated every 6 months with the list updating in the last half of 1977. Each basic address containing a business establishment listed for CVS was visted for the 1980 Components of Inventory Change Survey (CINCH), which was conducted in conjunction with the 1980 AHS, to identify HU's in structures which were completely nonresidential at the time of the 1970 census. These cases were later matched to the 1970 census address listings to identify those cases in address ED's and as an additional check to see if housing units existed in these structures at the time of the 1970 census. Due to the timing restraints associated with these operations, these cases were not included for the 1980 AHS but were included for the 1981 survey.

1981 telephone interviewing experiment—A large scale telephone interviewing experiment was conducted in conjunction with the 1981 AHS national sample in order to provide more definitive information about the effect of telephone interviewing on the AHS data and cost. A random sample of about 12,500 AHS national sample units was selected for the experiment. Among the cases assigned for telephone interviewing, only those sample housing units that had been interviewed in 1980 and had a telephone number available were eligible to be interviewed by telephone. Since a large portion of the total AHS national sample was assigned for telephone interviewing, its possible effect on the data is mentioned in the section "Reliability of the Estimates" of this appendix.

ESTIMATION

In 1981, the AHS estimates employed a three-stage ratio estimation procedure. However, prior to implementation of the procedure, the basic weight (i.e., the inverse of the probability of selection) was adjusted to account for the type A noninterview housing units encountered in the AHS. This noninterview

[1] A "large" SMSA is one with a 1970 population of 250,000 persons or more and a "small" SMSA is one with a 1970 population of less than 250,000 persons.

adjustment was done separately for occupied and vacant units. The noninterview adjustment was equal to the following ratio:

$$\frac{\text{Interviewed housing units + Noninterviewed housing units}}{\text{Interviewed housing units}}$$

The first-stage ratio estimation procedure was employed for sample housing units from non-self-representing (NSR) PSU's only. This procedure was designed to reduce the contribution to the variance arising from the sampling of PSU's. The first-stage ratio estimation procedure takes into account the differences that existed at the time of the 1970 census in the distribution by tenure and residence of the housing population estimated from the sample NSR PSU's and that of the NSR housing population in each of the four census regions of the country.

The first-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{The 1970 census housing population in the residence-tenure category for all NSR strata in a census region}}{\text{Estimate of the housing population category using 1970 census housing counts for sample NSR PSU's in a census region}}$$

The numerators of the ratios were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR stratum and summing these counts across the NSR strata in each census region. The denominators were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR sample PSU, weighting these counts by the inverse of the probability of selecting that PSU, and summing these weighted counts across the NSR sample PSU's in each census region. The computed first-stage ratio estimation factor was then applied to the existing weight for each NSR sample unit in each first-stage ratio estimation category.

The second-stage ratio estimation procedure was designed to adjust the AHS sample estimate of conventional new construction housing units; i.e., sample units built April 1, 1970, or later, to independently derived current estimates where a known deficiency in the AHS sample exists (see the section on nonsampling error) for each of the four regions. These estimates were considered to be the best estimates available for the number of conventional new construction units in these categories.

The second-stage ratio estimation factor was as follows:

$$\frac{\text{Current best estimate of new construction in the category}}{\text{AHS sample estimate of new construction units in the category}}$$

The numerators of the ratios were derived from data based on the Survey of Construction (SOC). The denominators of the ratios were obtained from the weighted estimates for the AHS sample units using the existing weight after the first-stage ratio estimates procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in each second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all the AHS sample units. This procedure was designed to adjust the AHS sample estimates of housing (i.e., the estimates employing the noninterview, first-stage, and second-stage adjustments) to current vacant housing estimates for 4 categories of vacant housing units and to independently derived current housing estimates for 24 categories of occupied housing units. Each of these categories is a combination of the characteristics of residence, tenure, race of householder, and sex of householder.

The third-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{Current independent estimate of housing units in the category}}{\text{AHS sample estimate of housing units in the category}}$$

The numerators of the ratios for occupied housing units were derived from data based on the Current Population Survey (CPS), a sample household survey conducted monthly by the Bureau of the Census. This is the first time that CPS household estimates based on the 1980 census were used as the independent controls in this adjustment rather than the 1970 census-based controls. The 1970-based estimates were about 2 percent smaller than the 1980-based estimates. This difference should be taken into consideration when comparing the 1981 AHS estimates to the AHS estimates for 1980 or earlier years. The numerators of the ratios for vacant housing units were derived from data based on the Housing Vacancy Survey (HVS), a quarterly vacancy survey also conducted by the Bureau of the Census. The denominators of the ratios were obtained from the weighted estimates for the AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computer third-stage ratio estimation factor was then applied to the existing weight for each sample unit in each third-stage ratio estimation category.

The second-stage and the third-stage ratio estimation procedures were iterated in order to bring the AHS estimates into close agreement with both sets of independent estimates. The second stage was modified so that the estimates for all 27 categories of new construction would be identical to the estimates before the third stage. Hence, the repeated second stage had the effect of controlling the AHS sample estimates of new construction units to the unbiased sample estimates for 9 categories of new construction mobile homes and trailers for each of the four regions and of adjusting the AHS sample estimate of 18 categories of conventional new construction units to the independently derived current estimates.

The numerators were either the unbiased weighted estimates for the AHS sample housing units, using the existing weight after the first-stage ratio estimation procedure (i.e., the estimates employing the noninterview and first-stage adjustments) or the independent estimate derived from data based on the Survey of Construction (SOC).

The denominators of the ratios in this iterative process were obtained from the weighted estimates for the AHS sample units after the previous stage of ratio estimation. The factors resulting from this iterative process were then applied to the existing weight on the appropriate records, and the resulting product was used as the final weight for tabulation.

The third-stage ratio estimation procedure, as well as the overall estimation procedure, reduced the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-45
Selection of sample areas. . . . App-45
Designation of sample housing for the 1981 survey . . . App-45
Selection of the 1973 sample housing units. . . App-46
Splitting of the sample . . . App-46
CEN-SUP sample . . . App-46
Selection of supplemental sample housing units in rural areas . . . App-46
Selection of sample housing units for the 1976 Coverage Improvement Program . . . App-46
1977 and 1981 sample reductions . App-47
Supplemental sample from nonresidential conversions . . . App-47
1981 telephone interviewing experiment . . . App-47
ESTIMATION. . . App-47
RELIABILITY OF THE ESTIMATES . . . App-49
Sampling errors . . . App-49
Standard errors of estimates of levels . . . App-49
Standard errors of estimates of percentages. . . App-49
Standard errors of ratios . . . App-49
Illustration of the use of the standard error tables. Illustration I . . . App-50
Illustration II . . . App-50
Differences. . . App-50
Illustration of the computation of the standard error of a difference . . . App-51
Medians . . . App-51
Illustration of the computation of the 95-percent confidence interval for a median. App-51
Nonsampling errors . . . App-51
Reinterview program . . . App-51
Coverage errors . . . App-52
Effects of ratio estimation on the estimates of change . . . App-53
Possible effects of telephone interviewing on the data . . . App-53
Rounding errors. . . App-53
Standard error tables . . . App-53

SAMPLE DESIGN

The 1981 estimates are based on data collected from September 1981 through December 1981 for the Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. The sample for this survey was spread over 461 sample areas (called primary sampling units), made up of 923 counties and independent cities with coverage in each of the 50 States and the District of Columbia.

Approximately 56,800 sample housing units (both occupied and vacant) were eligible for interview in the 1981 Annual Housing Survey. Of this number, 3,200 interviews were classified as "noninterview" for various reasons. Occupied housing units were classified as "noninterview" mainly because the occupants refused to be interviewed after repeated calls. For vacant housing units, interviews were not obtained because an informed respondent was not found after repeated visits. In addition to the 56,800 eligible housing units, there were also 3,600 sample units which were visited but were ineligible for interview for the AHS in terms of collecting information relevant to the 1981 housing inventory.

Selection of sample areas—The United States was divided into areas made up of counties and independent cities referred to as primary sampling units (PSU's). These PSU's were then grouped into 376 strata, 156 of which consisted of only one PSU which was in sample with certainty. These 156 strata were mostly the larger SMSA's and were called self-representing (SR) since the sample from the sample area represented just that PSU. Each one of the other 220 strata consisted of a group of PSU's and were referred to as non-self-representing (NSR), since the sample of housing units from the sample PSU in a stratum represented the other PSU's in the stratum as well.

One PSU was selected from each NSR stratum with probability proportionate to the 1970 census population of the PSU. (This resulted in 220 NSR sample PSU's.) In addition, the NSR strata were grouped into 110 pairs and one stratum was picked at random from each pair. From this stratum, an additional PSU was selected independently of the other PSU selected from this stratum. Since the two PSU's were independently selected, it was possible for the same PSU to be selected twice. This occurred in 25 instances, producing an additional 85 NSR sample PSU's, thus giving a grand total of 461 PSU's.

Designation of sample housing units for the 1981 survey—The sample housing units designated to be interviewed in the 1981 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1980 survey (which included all sample housing units that were selected as part of the 1976 Coverage Improvement Program) and were not part of the 1981 sample reduction.

2. All sample housing units that were either type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of survey but which could become eligible in the future) in the 1980 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1981 AHS questionnaire, page App-21).
3. All sample housing units that were selected from the list of building permits issued since the 1980 survey. (This sample represented the housing units built in permit-issuing areas, since the 1980 survey.)
4. Housing units added as the result of the updated listings in selected areas which do not issue building permits.

Selection of the 1973 sample housing units—The overall sampling rate used to select the sample for the 1973 AHS was about 1 in 1,366. The within-PSU sampling rate for the AHS was determined so that the overall probability of selection for each sample housing unit was the same (e.g., if the probability of selecting a NSR PSU was 1 in 10, then the within-PSU sampling rate would be 1 in 136.6).

Within the sample PSU's, a sample of the housing units enumerated in the 1970 Census of Population and Housing was selected for the AHS. In addition, a sample of new construction building permits was also selected to represent the units constructed since the 1970 census. These samples were selected at about twice the rate mentioned previously (i.e., at 2 in 1,366), thereby producing a sample twice as large as needed. This sample was split into two equal-sized samples—one to be used for the AHS, and one to be held in reserve for possible future use for the AHS. The procedure used to split this sample into equal-sized samples is described in the next section.

The sample of 1970 census units was selected in several stages. Within the sample PSU's, the first step was the selection of a sample of census enumeration districts (ED's), administrative units used in the 1970 census. The probability of selection for an ED was proportional to the following 1970 census counts of housing units (HU's) and persons in group quarters, combined in the following formula:

$$\frac{\text{Number of HU's in the ED} + \dfrac{\text{Number of group quarters persons in the ED}}{3}}{4}$$

The next step was to select an expected cluster of about four neighboring housing units within each sample ED. For most of the ED's, the selection was accomplished using the list of addresses for the ED as compiled in the 1970 census. (These ED's are referred to as address ED's.) However, in those ED's where addresses were incomplete or inadequate (mostly rural areas), the selection process was accomplished using area sampling methods. These ED's were divided into segments (i.e., small land areas with well-defined boundaries, having an expected size of four, or a multiple of four, housing units) and a segment was selected. Those selected segments with an expected size which was a multiple of four were further subsampled at the time of interview so that an expected four housing units were chosen for interview.

The sample of new construction housing units was selected from building permits issued since January 1970. Within each sample PSU, the building permits were chronologically ordered by month issued, and compact clusters of approximately four housing units were created. These clusters were then sampled at the rate of 2 in 1,366. Housing units constructed since the 1970 census in areas which do not issue building permits were brought into the sample as a result of the area sample described.

Splitting of the sample—The described sample selection procedure produced clusters (or segments) of size-four housing units for the sample taken from the census address frame, the new construction frame, and the area sampling frame (mainly rural areas). Clusters of this size should result in a minimum loss in precision for estimates of housing characteristics in rural areas because of the heterogeneity of neighboring units. However, clusters of size-two housing units were considered to be more optimum within those areas where the housing characteristics of neighboring units tend to be very similar (i.e., urban areas and new construction units). A splitting operation was then carried out for clusters selected from the census address and the new construction frames. This consisted of halving each sample cluster from these frames. Thus, two housing units from each of these clusters were included in the survey and two housing units were held in reserve. No splitting operation was carried out within the clusters selected from the area sampling frame; every other area sample cluster of four housing units was used for the survey and the remaining clusters were assigned to the reserve sample.

CEN-SUP sample—Housing units at addresses missed in the census or inadequately described in the 1970 census address register did not have a chance of being selected for the AHS sample in address ED's. Consequently, a special operation was undertaken to develop a separate sample, called the census supplemental (CEN-SUP) sample to represent these units. Due to time constraints on this operation, only about 40 percent of the CEN-SUP units were ready to be interviewed in 1973. The rest were interviewed for the first time in the 1974 AHS.

Selection of supplemental sample housing units in rural areas—In 1974, it was decided to increase the reliability of the AHS estimates of rural housing characteristics by doubling the number of sample housing units from rural areas. This was accomplished by reactivating the reserve sample, selected in the original sampling operations in 1973, from rural areas only. For the reserve sample selected in census address and new construction frames, the other half of each rural cluster (an expected two housing units) was reactivated in 1974. Similarly, for the area sampling frame, the entire reserve cluster (an expected four housing units) was reactivated in 1974 if the cluster was rural. This supplementation increased the overall probability of selection for sample housing units in rural areas to about 2 in 1,366; whereas, the overall probability of selection for sample housing units in urban areas remained at 1 in 1,366.

Selection of sample housing units for the 1976 Coverage Improvement Program—The 1976 Coverage Improvement Pro-

gram was undertaken to correct certain deficiencies in the AHS national sample from the census address and new construction frames. The coverage deficiencies included the following types of housing units:

1. New construction housing units from building permits issued prior to January 1970 for which construction had not been completed at the time of the 1970 census.
2. Housing units converted to residential use in structures totally nonresidential at the time of the 1970 census.
3. Houses that have been moved onto their present site since the 1970 census.
4. Mobile homes and trailers placed in parks either missed in the 1970 census or established since the 1970 census.
5. Mobile homes and trailers placed outside parks since the 1970 census or vacant at the time of the 1970 census.

A sample of new construction housing units whose permits were issued before January 1970 was selected in two stages. First, units whose permits were issued before January 1970, but which were completed after the census, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. Second, these units were then sampled so that the overall probability of selection was about 1 in 1,320.

A sample of mobile homes and trailers placed in parks missed by the census or established after the census was also selected in two stages. During the first stage, a list of mobile home parks was obtained from commercial listings. This list was then supplemented by additional parks identified by a canvassing operation similar to that performed in ED's where area sampling methods are used. The second stage consisted of dividing the parks into clusters of an expected size of four sites. These clusters were then sampled so that the overall probability of selection was about 1 in 1,366.

For the remaining housing units (i.e., mobile homes and trailers placed outside parks since the 1970 census, mobile homes and trailers vacant at the time of the 1970 census, units converted from nonresidential to residential use since the 1970 census, and houses that had been moved onto their present site since the 1970 census), the sampling was done in three stages. First, a subsample of the regular AHS sample housing units from the census address frame was selected. Second, succeeding structures that had been eligible to be selected from the census address frame were then listed until eight structures (excluding mobile home parks) were found. Finally, the intervening structures that had been listed which did not have a chance of selection in the AHS were identified and the units within these structures were interviewed.

1977 and 1981 sample reductions—By 1977, the addition to the sample, primarily from new construction and the coverage improvements, had increased the total sample size (interviews plus noninterviews) to about 81,000 housing units. The sample was reduced by about 7 percent to approximately 75,000 in 1977. This reduction did not include any CEN-SUP units or units which were selected as part of the 1976 Coverage Improvement Program. Initially, the sample was further reduced in 1981 by about 5 percent. Subsequently, the rural sample was reduced by about 25 percent in 125 "large" SMSA's[1] and by about 50 percent in "small" SMSA's[1] and outside the SMSA's. These reductions brought the sample size down to approximately 60,000 housing units in 1981.

Supplemental sample from nonresidential conversions—The purpose of this sample was to improve the coverage in address ED's (i.e., ED's in which 1970 census address listings were used for selecting the AHS sample) of housing units created in structures that were completely nonresidential at the time of the 1970 census. This sample was derived from listings created for the Commercial Victimization Survey (CVS), a nationally representative area sample survey conducted by the Census Bureau for the Department of Justice. Each of the CVS area segments was initially canvassed either in 1972 or 1975 and eligible business establishments were listed. These listings were updated every 6 months with the list updating in the last half of 1977. Each basic address containing a business establishment listed for CVS was visted for the 1980 Components of Inventory Change Survey (CINCH), which was conducted in conjunction with the 1980 AHS, to identify HU's in structures which were completely nonresidential at the time of the 1970 census. These cases were later matched to the 1970 census address listings to identify those cases in address ED's and as an additional check to see if housing units existed in these structures at the time of the 1970 census. Due to the timing restraints associated with these operations, these cases were not included for the 1980 AHS but were included for the 1981 survey.

1981 telephone interviewing experiment—A large scale telephone interviewing experiment was conducted in conjunction with the 1981 AHS national sample in order to provide more definitive information about the effect of telephone interviewing on the AHS data and cost. A random sample of about 12,500 AHS national sample units was selected for the experiment. Among the cases assigned for telephone interviewing, only those sample housing units that had been interviewed in 1980 and had a telephone number available were eligible to be interviewed by telephone. Since a large portion of the total AHS national sample was assigned for telephone interviewing, its possible effect on the data is mentioned in the section "Reliability of the Estimates" of this appendix.

ESTIMATION

In 1981, the AHS estimates employed a three-stage ratio estimation procedure. However, prior to implementation of the procedure, the basic weight (i.e., the inverse of the probability of selection) was adjusted to account for the type A noninterview housing units encountered in the AHS. This noninterview

[1] A "large" SMSA is one with a 1970 population of 250,000 persons or more and a "small" SMSA is one with a 1970 population of less than 250,000 persons.

adjustment was done separately for occupied and vacant units. The noninterview adjustment was equal to the following ratio:

$$\frac{\text{Interviewed housing units + Noninterviewed housing units}}{\text{Interviewed housing units}}$$

The first-stage ratio estimation procedure was employed for sample housing units from non-self-representing (NSR) PSU's only. This procedure was designed to reduce the contribution to the variance arising from the sampling of PSU's. The first-stage ratio estimation procedure takes into account the differences that existed at the time of the 1970 census in the distribution by tenure and residence of the housing population estimated from the sample NSR PSU's and that of the NSR housing population in each of the four census regions of the country.

The first-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{The 1970 census housing population in the residence-tenure category for all NSR strata in a census region}}{\text{Estimate of the housing population category using 1970 census housing counts for sample NSR PSU's in a census region}}$$

The numerators of the ratios were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR stratum and summing these counts across the NSR strata in each census region. The denominators were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR sample PSU, weighting these counts by the inverse of the probability of selecting that PSU, and summing these weighted counts across the NSR sample PSU's in each census region. The computed first-stage ratio estimation factor was then applied to the existing weight for each NSR sample unit in each first-stage ratio estimation category.

The second-stage ratio estimation procedure was designed to adjust the AHS sample estimate of conventional new construction housing units; i.e., sample units built April 1, 1970, or later, to independently derived current estimates where a known deficiency in the AHS sample exists (see the section on nonsampling error) for each of the four regions. These estimates were considered to be the best estimates available for the number of conventional new construction units in these categories.

The second-stage ratio estimation factor was as follows:

$$\frac{\text{Current best estimate of new construction in the category}}{\text{AHS sample estimate of new construction units in the category}}$$

The numerators of the ratios were derived from data based on the Survey of Construction (SOC). The denominators of the ratios were obtained from the weighted estimates for the AHS sample units using the existing weight after the first-stage ratio estimates procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in each second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all the AHS sample units. This procedure was designed to adjust the AHS sample estimates of housing (i.e., the estimates employing the noninterview, first-stage, and second-stage adjustments) to current vacant housing estimates for 4 categories of vacant housing units and to independently derived current housing estimates for 24 categories of occupied housing units. Each of these categories is a combination of the characteristics of residence, tenure, race of householder, and sex of householder.

The third-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{Current independent estimate of housing units in the category}}{\text{AHS sample estimate of housing units in the category}}$$

The numerators of the ratios for occupied housing units were derived from data based on the Current Population Survey (CPS), a sample household survey conducted monthly by the Bureau of the Census. This is the first time that CPS household estimates based on the 1980 census were used as the independent controls in this adjustment rather than the 1970 census-based controls. The 1970-based estimates were about 2 percent smaller than the 1980-based estimates. This difference should be taken into consideration when comparing the 1981 AHS estimates to the AHS estimates for 1980 or earlier years. The numerators of the ratios for vacant housing units were derived from data based on the Housing Vacancy Survey (HVS), a quarterly vacancy survey also conducted by the Bureau of the Census. The denominators of the ratios were obtained from the weighted estimates for the AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computer third-stage ratio estimation factor was then applied to the existing weight for each sample unit in each third-stage ratio estimation category.

The second-stage and the third-stage ratio estimation procedures were iterated in order to bring the AHS estimates into close agreement with both sets of independent estimates. The second stage was modified so that the estimates for all 27 categories of new construction would be identical to the estimates before the third stage. Hence, the repeated second stage had the effect of controlling the AHS sample estimates of new construction units to the unbiased sample estimates for 9 categories of new construction mobile homes and trailers for each of the four regions and of adjusting the AHS sample estimate of 18 categories of conventional new construction units to the independently derived current estimates.

The numerators were either the unbiased weighted estimates for the AHS sample housing units, using the existing weight after the first-stage ratio estimation procedure (i.e., the estimates employing the noninterview and first-stage adjustments) or the independent estimate derived from data based on the Survey of Construction (SOC).

The denominators of the ratios in this iterative process were obtained from the weighted estimates for the AHS sample units after the previous stage of ratio estimation. The factors resulting from this iterative process were then applied to the existing weight on the appropriate records, and the resulting product was used as the final weight for tabulation.

The third-stage ratio estimation procedure, as well as the overall estimation procedure, reduced the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the

probability of selection. The distribution of the housing population selected for the sample differed somewhat, by chance, from that of the Nation as a whole in such basic housing characteristics as tenure, vacancy status, residence, race of householder, and sex of householder. These characteristics are probably closely correlated with other housing characteristics measured for the AHS. Therefore, through the use of the three-stage ratio estimation procedure one can expect the sample estimate to be improved substantially.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS national sample.

Sampling errors—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same questionnaires, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The variability between estimates from all possible samples is defined as sampling error. One common measure of sampling error is the standard error which measures the precision with which an estimate from a sample approximates the average result of all possible samples. In addition, the standard error, as calculated for this report, also partially reflects the variation in the estimates due to some nonsampling errors, but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, and biases and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates so that the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, and each of these samples was surveyed under essentially the same general conditions and an estimate and its estimated standard error were calculated for each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples either is or is not contained in any particular computed interval. However, for a particular sample one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables (pages App-53 to App-59) of this appendix are approximations to the standard errors of various estimates shown in this report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than the precise standard error for any specific item.

Standard errors of estimates of levels—Tables I, II, and III present the standard errors applicable to the 1981 national housing inventory estimates in this report, and tables IV and V present the standard errors applicable to 1973-1981 lost housing unit estimates in this report. Table VI presents the standard errors applicable for the Northeast, North Central, South, and West Regions. Linear interpolation should be used to determine standard errors for levels of estimates not specifically shown in tables I through VI.

Standard errors of estimates of percentages—The reliability of an estimated percentage, computed by using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more.

Tables VII through XII present the standard errors of estimated percentages. Tables VII and VIII show the approximate standard errors of all national estimated percentages of housing units. Tables IX and X show the approximate standard errors of the estimated percentages of 1973-1981 lost housing units. Table XI shows the approximate standard error of all regional estimated percentages of housing units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables VII through XI.

Included in tables I through XI are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered to be overestimates of the true standard errors and should be used primarily for construction of confidence intervals for characteristics when an estimate of zero is obtained.

Standard errors of ratios—For ratios of the form (100) (x/y), where x is not a subclass of y, tables VII through XI underestimate the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be obtained by letting the standard error of the ratio be approximately equal to:

$$(100)\left(\frac{x}{y}\right)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

where: x = the numerator of the ratio
y = the denominator of the ratio
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

Illustration of the use of the standard error tables. Illustration I—Table A-1 of this report shows that in urban areas of the United States there were 23,637,000 renter-occupied housing units in 1981. Interpolation in standard error table III shows that the standard error of an estimate of this size is approximately 180,000. The following interpolating procedure was used.

The information presented in the following table was extracted from standard error table III. The entry for "x" is the one sought.

Size of estimate (000)	Standard error (000)
10,000	130
23,637	x
25,000	185

By vertically interpolating between 130 and 185, the entry for "x" is determined to be 180.

$$23{,}637-10{,}000 = 13{,}637$$
$$25{,}000-10{,}000 = 15{,}000$$

$$130 + \frac{13{,}637}{15{,}000}(185-130) = 180$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 23,457,000 to 23,817,000 housing units. Therefore, a conclusion that the average estimate of 1981 housing units of this type lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all possible samples lies within the interval from 23,349,000 to 23,925,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 23,277,000 to 23,997,000 housing units with 95 percent confidence.

Table A-1 also shows that of the 23,637,000 renter-occupied housing units in urban areas, 6,909,000, or 29.2 percent, were occupied by two persons. Interpolation in standard error table VII (i.e., interpolation on both the base and percent) of this appendix shows that the standard error of the above percentage is 0.4. The following interpolating procedure was used.

The information presented in the table below was extracted from standard error table VII. The entry for "p" is the one sought.

Base of percentage (000)	Estimated percentage		
	25	29.2	50
10,000	0.6	a	0.7
23,637		p	
25,000	0.4	b	0.5

1. By horizontal interpolation between 0.6 and 0.7 the entry for cell "a" is determined to be 0.6.

$$29.2-25.0 = 4.2$$
$$50.0-25.0 = 25.0$$

$$0.6 + \frac{4.2}{25.0}(0.7-0.6) = 0.6$$

2. By horizontal interpolation between 0.4 and 0.5 the entry for cell "b" is determined to be 0.4.

$$29.2-25.0 = 4.2$$
$$50.0-25.0 = 25.0$$

$$0.4 + \frac{4.2}{25.0}(0.5-0.4) = 0.4$$

3. By vertical interpolation between 0.6 and 0.4 the entry for "p" is determined to be 0.4.

$$23{,}637-10{,}000 = 13{,}637$$
$$25{,}000-10{,}000 = 15{,}000$$

$$0.6 + \frac{13{,}637}{15{,}000}(0.4-0.6) = 0.4$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 28.8 to 29.6 percent; the 90-percent confidence interval is from 28.6 to 29.8 percent; and the 95-percent confidence interval is from 28.4 to 30.0 percent.

Illustration II—Table A-2 of this report shows that in the rural areas of the United States in 1981 there were 13,543,000 specified owner-occupied housing units. Interpolation in standard error table III of this appendix shows that the standard error of an estimate of this size is approximately 217,000. Consequently, the 68-percent confidence interval is from 13,326,000 to 13,760,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1981 specified owner-occupied housing units lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all possible samples lies within the interval from 13,196,000 to 13,890,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 13,109,000 to 13,977,000 housing units with 95 percent confidence.

Table A-2 also shows that of the 13,543,000 specified owner-occupied housing units in rural areas, 5,335,000, or 39.4 percent, had no mortgage. Interpolation in standard error table VII (i.e., interpolation on both the base and the percent) shows that the standard error of the above percentage is 0.7 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 38.7 to 40.1 percent; the 90-percent confidence interval is from 38.3 to 40.5 percent; and the 95-percent confidence interval is from 38.0 to 40.8 percent.

Differences—The standard errors shown are not directly applicable to differences between two sample estimates. The stand-

ard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristics in two different areas or the difference between separate and uncorrelated characteristics in the same area. If there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; if there is a high negative correlation between the two characteristics, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference—Table A-1 of this report shows that in urban areas of the United States there were 3,614,000 renter-occupied housing units with three persons in 1981. Table A-1 also shows that in urban areas of the United States there were 6,909,000 renter-occupied housing units with two persons in 1981. Thus, the apparent difference between the number of 1981 renter-occupied housing units in urban areas with two persons and those with three persons is 3,295,000. Interpolation in standard error table III shows the standard error of 3,614,000 is approximately 80,000 and the standard error on an estimate of 6,909,000 is approximately 108,000. Therefore, the standard error of the estimated difference of 3,295,000 is about 134,000.

$$134{,}000 = \sqrt{(108{,}000)^2 + (80{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 3,295,000 difference is from 3,161,000 to 3,429,000 housing units. Therefore, a conclusion that the average estimate of this difference, derived from all possible samples lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 3,081,000 to 3,509,000 housing units, and the 95-percent confidence interval is from 3,027,000 to 3,563,000. Thus, we can conclude with 95 percent confidence that the number of 1981 renter-occupied housing units in urban areas with two persons is greater than the number with three persons.

Medians—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From the appropriate standard error table determine the standard error of a 50-percent characteristic on the base of the median.
2. Add to and subtract from 50 percent the standard error determined in step 1.
3. Using the distribution of the characteristic, determine the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples; the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval for a median—Table A-1 of this report shows the median number of persons in owner-occupied housing units in urban areas was 2.5 in 1981. The base of the distribution, from which this median was determined, is 34,389,000 housing units.

1. From standard error table VII, the standard error of a 50-percent characteristic on the base of 34,389,000 is 0.4 percentage points.
2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 49.2 and 50.8.
3. From the distribution for "persons" in table A-1 of part A, the interval for owner-occupied housing units with two persons (for purposes of calculating the median, the category of two persons is considered to be from 1.5 to 2.5 persons) contains the 49.2 percent derived in step 2. About 5,724,000 housing units or 16.6 percent fall below this interval, and 11,367,000 housing units or 33.1 percent fall within this interval. By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$1.5 + (2.5-1.5)\,\frac{(49.2-16.6)}{33.1} = 2.5$$

Similarly, the interval for owner-occupied housing units with three persons contains the 50.8 percent derived in step 2. About 17,091,000 housing units or 49.7 percent fall below this interval, and 6,148,000 housing units or 17.9 percent fall within this interval. The upper limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\,\frac{(50.8-49.7)}{17.9} = 2.6$$

Thus, the 95-percent confidence interval ranges from 2.5 to 2.6 persons.

Nonsampling errors—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases; definitional difficulties; differences in the interpretation of questions; inability or unwillingness to provide correct information on the part of respondents; mistakes in recording or coding the data; and other errors of collection, response, processing, coverage, and estimation for missing data. Nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Reinterview program—For the AHS national sample, a study was conducted to obtain a measurement of some of the com-

ponents of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a subsample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and thus were the basis for the measurement of the "content" error of these AHS estimates.

As part of the reinterview, an additional check was carried out for interviewer evaluation and quality control. This check was made at each of these households to determine if the following were done during the original interview.

1. The correct housing unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

The results of the 1981 and 1980 reinterview studies were not available at the time of publication. In addition, the 1979 reinterview was carried out as an interviewer quality control only. Thus, the results of the 1977 and 1978 reinterview studies which were presented in the Census Bureau memoranda, "Reinterview Results for the Annual Housing Survey—National Sample 1977" and "Reinterview Results for the Annual Housing Survey—National Sample 1978" are presented here.

In 1978, a substantial portion of the reinterview questionnaire was devoted to testing the new questions 1 through 7 (parts a and b). The questions (part a), which were asked only at housing units interviewed in the previous year, determined whether there had been a change since last year in selected nonattitudinal items. If a change had been recorded or the respondent did not know if a change had occurred, part b of the question, which collects the value of the item, was asked. The reinterviewers asked these items using the questions as formatted in 1977. Comparing the responses from the differently formatted questions, the 1978 reinterview found that 80 percent of the questions showed low levels of inconsistency with the remainder showing moderate levels.

The 1977 reinterview program showed moderate to high levels of inconsistency with about 21 percent of the nonattitudinal and 56 percent of the attitudinal items showing high levels of inconsistency. A large portion (43 percent) of the nonattitudinal items showed a low level of inconsistency. Moderate levels indicate that there are some problems with inconsistent reporting and high levels indicate that improvements are needed in the data collection methods or that the category concepts themselves are ambiguous.

Cross-tabulations involving those items, which are subject to substantial levels of inconsistency may be subject to a large distortion as a consequence of the associated high level of response variance, and thus, are considered to be less reliable than comparable cross-tabulations which do not involve these data. As a consequence, the tables in this report, which contain such cross-tabulations, have been footnoted with a cautionary statement.

The 1970 census reinterview results provide illustrations of possible nonsampling errors for some of the items which also appear in the AHS. For example, median value of homes was consistently underestimated by about 5 percent, and the average monthly cost of electricity and utility gas were consistently overestimated although the net effect on average gross rent was fairly small.

A possible explanation for the results of the AHS and census reinterview studies, as well as the surveys themselves, is that the respondents may lack precise information. Also, because the results of the reinterview studies are derived from sample surveys, there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of these studies.

Coverage errors—Deficiencies in the representation of conventional new construction (non-mobile home or trailer) for the AHS new construction sample (mentioned previously in the section on estimation) is an example of coverage errors. During the sampling of building permits, only those issued more than 4 months before the survey began were eligible to be selected to represent conventional new construction. Due to time constraints it is not possible to sample units whose permits are issued less than 4 months in advance of the survey.

It is estimated that the 1981 AHS sample missed about 0.8 percent (i.e., about 157,000 units) of conventional housing units built after April 1970 because the permits for these units, which were built before October 1981, were issued less than 4 months in advance of the survey. The second-stage ratio estimation procedure was employed to reduce the effect of this deficiency although some bias in the AHS estimates of conventional new construction probably still exists. Review of the second-stage ratio estimation procedure indicates that we have consistently overcompensated for this deficiency every year since 1975 by ratio adjusting the new construction to counts of new construction for the end of the interview period, which has been December or January, instead of October. This overcompensation may inflate the new construction counts by 100,000 to 300,000 units.

Changing this procedure to correct for this overcompensation would reduce the estimate of housing units built since the last survey since it would be based on a 10-month time period rather than 12 months. Thus, it was decided not to change this procedure so that the estimate of housing units built since the last survey would be consistent with previous years' estimates, which are generally based on a 12-month time period.

In addition, the 1976 Coverage Improvement Program had certain deficiencies. First, when the canvassing was done to identify mobile homes and trailer parks that were not in the same sample frame or not on the commercial lists, only 92 percent of the census address frame ED's were represented.

Second, it appears that the listing procedure (used to find mobile homes and trailers placed outside parks, units converted from nonresidential to residential, and houses that had been moved onto their present site) was not very efficient for finding nonresidential conversions (which might be primarily in business districts), since the listing procedure started from a residential unit. (The sample estimate of this component was approximately 16,000 housing units with a standard error of 12,000.)

Finally, it is felt that deficiencies also exist in ED's where area sampling methods are used. As before, it had been assumed that all housing units located inside these ED's would be represented in the sample. However, it has been estimated that the 1981 AHS sample missed as much as 2 percent (i.e., as much as 400,000 units) of all housing units in ED's where area sampling methods are used because these units are not listed during the canvassing.

The third-stage of ratio estimation corrects for these deficiencies as far as the total count of housing is concerned, i.e., it adjusts the estimate of the total housing inventory to the best available estimate. However, biases of subtotals would still remain.

Effects of ratio estimation on the estimates of change—As stated previously, 1981 was the first AHS for which 1980 census-based estimates, rather than 1970 census-based estimates, were used in the third-stage ratio estimation procedure. Since the 1980-based estimates are about 2 percent higher than the corresponding 1970-based estimates, this will cause estimates of change between the 1981 AHS data and data for 1980 or earlier years to be overstated. This overstatement should be taken into consideration during the analysis of estimates of change.

Possible effects of telephone interviewing on the data—A preliminary analysis of the 1981 telephone interviewing experiment has been completed. It has been concluded that data collected, using the telephone interviewing procedures, were not sufficiently different from data collected by regular personal interviews, to preclude basing the 1981 AHS national sample published data. However, since the percentage of the comparison tests that were significant was slightly above what could be expected by chance, there was some evidence that telephone interviewing may have had an effect on the data. Thus, some caution should be exercised in making comparisons between 1981 AHS data and data from preceding years.

Rounding errors—In errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of this survey. Also, since medians in this report were computed using unrounded data, instead of the published rounded data, they can differ from medians calculated directly from the published data.

TABLE I. Standard Errors of Estimated Numbers of Housing Units: 1981 (Excluding Estimates of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water-Individual Well, Cooking Fuel, Lacking Complete Plumbing, Mobile Homes, and Housing Units With Householder of Spanish Origin)

(68 chances out of 100)

Size of estimate (000)	Standard error: Total or White (000)	Standard error: Black (000)	Size of estimate (000)	Standard error: Total or White (000)	Standard error: Black (000)
0	2	2	2,500 . . .	71	61
5	3	3	5,000 . . .	98	68
10	5	5	7,500 . . .	119	51
25	7	7	10,000. . .	135	–
50	10	10	25,000. . .	193	–
100.	14	14	50,000. . .	216	–
250.	23	22	75,000. . .	167	–
500.	32	31	85,000. . .	112	–
1,000 . . .	45	43			

TABLE II. Standard Errors of Estimated Numbers of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water-Individual Well, Cooking Fuel, Lacking Complete Plumbing, Mobile Homes, and Housing Units With Householder of Spanish Origin: 1981

(68 chances out of 100)

Size of estimate (000)	Standard error: Total, White, or Spanish origin (000)	Standard error: Black (000)	Size of estimate (000)	Standard error: Total, White, or Spanish origin (000)	Standard error: Black (000)
0	3	3	1,000 . . .	53	50
5	4	4	2,500 . . .	83	71
10	5	5	5,000 . . .	116	79
25	8	8	7,500 . . .	139	60
50	12	12	10,000. . .	159	–
100.	17	17	25,000. . .	227	–
250.	27	26	50,000. . .	253	–
500.	37	37			

TABLE III. Standard Errors of Estimated Numbers of Urban or Rural Housing Units: 1981

(68 chances out of 100)

Size of estimate (000)	Rural housing units (except those in the next column) (000)	Rural housing units pertaining to new construction, no bedrooms, source of water-individual well, cooking fuel, lacking some or all plumbing, and mobile homes (000)	Urban housing units (except those in the next column) (000)	Urban housing units pertaining to new construction, lacking complete kitchen facilities, no bedrooms, no bathrooms, source of water-individual well, cooking fuel, lacking some or all plumbing, mobile homes, and housing units with householder of Spanish origin (000)
0	2	3	2	2
5	3	4	3	3
10	5	6	4	5
25	7	9	7	8
50	11	13	10	11
100	15	18	14	15
250	24	30	22	25
500	34	44	31	36
1,000	48	68	43	53
2,500	79	129	68	94
5,000	117	224	95	152
10,000	179	412	130	263
25,000	341	–	185	587
50,000	596	–	206	–
60,000	–	–	195	–

TABLE IV. Standard Errors of Estimated Numbers of Lost Housing Units and of Urban and Rural Lost Housing Units: 1973-1981 (Excluding Estimates of Lost Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Some or All Plumbing, Mobile Homes, Other Vacants, and Rural Vacants for Rent)

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	250	23
5	3	500	33
10	4	1,000	50
25	7	2,500	93
50	10	5,000	158
100	14		

TABLE V. Standard Errors of Estimated Numbers of Lost Housing Units and of Urban and Rural Lost Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Some or All Plumbing, Mobile Homes, Other Vacants, and Rural Vacants for Rent: 1973-1981

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	3	100	18
5	4	250	29
10	6	500	41
25	9	1,000	60
50	13		

TABLE VIa. Standard Errors of Estimated Numbers of Housing Units and of Urban Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1981 (Excluding Estimates of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions and Excluding Source of Water-Individual Well, Cooking Fuel, and Mobile Homes for Each of the Regions)

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	2	500	33
5	3	1,000	47
10	5	2,500	74
25	7	5,000	105
50	10	10,000	148
100	15	25,000	235
250	23		

TABLE VIb. Standard Errors of Estimated Numbers of Housing Units and of Urban Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing for the Northeast, North Central, and West Regions and to Source of Water-Individual Well, Cooking Fuel, and Mobile Homes for the Northeast and North Central Regions: 1981

(68 chances out of 100. For estimates pertaining to source of water-individual well, cooking fuel, and mobile homes for the West Region, apply a factor of 1.66 to the standard errors)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	3	500	39
5	4	1,000	56
10	6	2,500	87
25	9	5,000	122
50	13	10,000	168
100	18	25,000	243
250	28		

TABLE VIc. Standard Errors of Estimated Numbers of Housing Units Pertaining to Source of Water-Individual Well, Cooking Fuel, and Mobile Homes for the South Region: 1981

(68 chances out of 100. For estimates of urban housing units pertaining to source of water-individual well, cooking fuel, and mobile homes for the South Region, apply a factor of 0.91 to the standard errors)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	5	500	49
5	5	1,000	70
10	7	2,500	109
25	11	5,000	150
50	16	10,000	201
100	22	25,000	259
250	35		

TABLE VId. Standard Errors of Estimated Numbers of Rural Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1981

(68 chances out of 100)

Size of estimate (000)	Rural housing units (except in the following columns) for the Northeast, North Central, South, and West Regions (000)	Rural housing units pertaining to new construction, no bedrooms, lacking complete plumbing for the Northeast, North Central, and West Regions and to source of water-individual well, cooking fuel, and mobile homes for the Northeast and North Central Regions (000)	Rural housing units pertaining to source of water-individual well, cooking fuel, and mobile homes for the West Region (000)	Rural housing units pertaining to source of water-individual well, cooking fuel, and mobile homes for the South Region (000)
0	2	4	10	6
5	4	4	10	6
10	5	6	10	7
25	8	9	16	12
50	11	13	22	17
100	16	19	31	24
250	25	30	49	37
500	35	42	69	53
1,000	50	59	96	74
2,500	79	93	144	115
5,000	112	130	184	159
10,000	158	179	192	214

TABLE VII. Standard Errors of Estimated Percentages of Housing Units and of Urban and Rural Housing Units: 1981 (Excluding Estimated Percentages of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water-Individual Well, Cooking Fuel, Lacking Complete Plumbing, Mobile Homes, and Housing Units With Householder of Spanish Origin)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	29.1	29.1	29.1	29.1	29.1	29.1	29.1	32.0
10	17.0	17.0	17.0	17.0	17.0	17.0	19.6	22.7
25	7.6	7.6	7.6	7.6	8.6	10.2	12.4	14.3
50	3.9	3.9	3.9	4.4	6.1	7.2	8.8	10.1
100	2.0	2.0	2.0	3.1	4.3	5.1	6.2	7.2
250	0.8	0.9	1.3	2.0	2.7	3.2	3.9	4.5
500	0.4	0.6	0.9	1.4	1.9	2.3	2.8	3.2
1,000	0.2	0.5	0.6	1.0	1.4	1.6	2.0	2.3
2,500	0.08	0.3	0.4	0.6	0.9	1.0	1.2	1.4
5,000	0.04	0.2	0.3	0.4	0.6	0.7	0.9	1.0
7,500	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.8
10,000	0.02	0.14	0.2	0.3	0.4	0.5	0.6	0.7
25,000	0.01	0.09	0.13	0.2	0.3	0.3	0.4	0.5
50,000	–	0.06	0.09	0.14	0.2	0.2	0.3	0.3
75,000	–	0.05	0.07	0.11	0.2	0.2	0.2	0.3
85,000	–	0.05	0.07	0.11	0.15	0.2	0.2	0.2

TABLE VIII. Standard Errors of Estimated Percentages of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Source of Water-Individual Well, Cooking Fuel, Lacking Complete Plumbing, Mobile Homes, and Housing Units With Householder of Spanish Origin: 1981

(68 chances out of 100. For estimates pertaining to urban housing untis, apply a factor of 0.91 to the standard errors. For standard errors of rural housing units pertaining to new construction, no bedrooms, source of water-individual well, cooking fuel, lacking some or all plumbing, and mobile homes, use the standard errors in table XIb)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	36.1	36.1	36.1	36.1	36.1	36.1	36.1	37.6
10	22.0	22.0	22.0	22.0	22.0	22.0	23.0	26.6
25	10.2	10.2	10.2	10.2	10.2	12.0	14.6	16.8
50	5.3	5.3	5.3	5.3	7.1	8.5	10.3	11.9
100	2.7	2.7	2.7	3.7	5.0	6.0	7.3	8.4
250	1.1	1.1	1.5	2.3	3.2	3.8	4.6	5.3
500	0.6	0.7	1.1	1.6	2.3	2.7	3.3	3.8
1,000	0.3	0.5	0.7	1.2	1.6	1.9	2.3	2.7
2,500	0.1	0.3	0.5	0.7	1.0	1.2	1.5	1.7
5,000	0.06	0.2	0.3	0.5	0.7	0.8	1.0	1.2
10,000	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.8
25,000	0.01	0.1	0.15	0.2	0.3	0.4	0.5	0.5
50,000	0.01	0.07	0.11	0.2	0.2	0.3	0.3	0.4
75,000	–	0.06	0.09	0.13	0.2	0.2	0.3	0.3

TABLE IX. Standard Errors of Estimated Percentages of Lost Housing Units and of Urban and Rural Lost Housing Units: 1973-1981 (Excluding Estimated Percentages of Lost Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Some or All Plumbing, Mobile Homes, Other Vacants, and Rural Vacants for Rent)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	27.7	27.7	27.7	27.7	27.7	27.7	27.7	30.9
10	16.1	16.1	16.1	16.1	16.1	16.1	18.9	21.9
25	7.1	7.1	7.1	7.1	8.3	9.9	12.0	13.8
50	3.7	3.7	3.7	4.3	5.9	7.0	8.5	9.8
100	1.9	1.9	1.9	3.0	4.1	4.9	6.0	6.9
250	0.8	0.9	1.2	1.9	2.6	3.1	3.8	4.4
500	0.4	0.6	0.9	1.3	1.9	2.2	2.7	3.1
1,000	0.2	0.4	0.6	1.0	1.3	1.6	1.9	2.2
2,500	0.08	0.3	0.4	0.6	0.8	1.0	1.2	1.4
5,000	0.04	0.2	0.3	0.4	0.6	0.7	0.8	1.0

TABLE X. Standard Errors of Estimated Percentages of Lost Housing Units and of Urban and Rural Lost Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, Lacking Some or All Plumbing, Mobile Homes, Other Vacants, and Rural Vacants for Rent: 1973-1981

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	39.2	39.2	39.2	39.2	39.2	39.2	39.2	40.1
10	24.3	24.3	24.3	24.3	24.3	24.3	24.6	28.4
25	11.4	11.4	11.4	11.4	11.4	12.8	15.5	17.9
50	6.0	6.0	6.0	6.0	7.6	9.1	11.0	12.7
100	3.1	3.1	3.1	3.9	5.4	6.4	7.8	9.0
250	1.3	1.3	1.6	2.5	3.4	4.1	4.9	5.7
500	0.6	0.8	1.1	1.7	2.4	2.9	3.5	4.0
1,000	0.3	0.6	0.8	1.2	1.7	2.0	2.5	2.8
2,500	0.13	0.4	0.5	0.8	1.1	1.3	1.6	1.8
5,000	0.06	0.3	0.4	0.6	0.8	0.9	1.1	1.3

TABLE XIa. Standard Errors of Estimated Percentages of Housing Units and of Urban Housing Units Pertaining to the Northeast, North Central, South, and West Regions: 1981 (Excluding Estimated Percentages Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing Facilities for the Northeast, North Central, and West Regions and Excluding Source of Water-Individual Well, Cooking Fuel, and Mobile Homes for Each of the Regions)

(68 chances out of 100. For standard errors of estimated percentages of rural housing units pertaining to the Northeast, North Central, South, and West Regions, excluding estimates of rural housing units pertaining to new construction, no bedrooms, and lacking complete plumbing for the Northeast, North Central, and West Regions, and excluding source of water-individual well, cooking fuel, and mobile homes for each of the regions, multiply the standard errors by 1.04)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	30.6	30.6	30.6	30.6	30.6	30.6	30.6	33.2
10	18.1	18.1	18.1	18.1	18.1	18.1	20.3	23.5
25	8.1	8.1	8.1	8.1	8.9	10.6	12.9	14.8
50	4.2	4.2	4.2	4.6	6.3	7.5	9.1	10.5
100	2.2	2.2	2.2	3.2	4.5	5.3	6.4	7.4
250	0.9	0.9	1.3	2.0	2.8	3.4	4.1	4.7
500	0.4	0.7	0.9	1.4	2.0	2.4	2.9	3.3
1,000	0.2	0.5	0.7	1.0	1.4	1.7	2.0	2.3
2,500	0.09	0.3	0.4	0.6	0.9	1.1	1.3	1.5
5,000	0.04	0.2	0.3	0.5	0.6	0.7	0.9	1.0
7,500	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.9
10,000	0.02	0.15	0.2	0.3	0.4	0.5	0.6	0.7
25,000	0.01	0.09	0.13	0.2	0.3	0.3	0.4	0.5

TABLE XIb. Standard Errors of Estimated Percentages of Housing Units Pertaining to New Construction, Lacking Complete Kitchen Facilities, No Bedrooms, No Bathrooms, and Lacking Complete Plumbing Facilities for the Northeast, North Central, and West Regions and to Source of Water-Individual Well, Cooking Fuel, and Mobile Homes for the Northeast and North Central Regions: 1981

(68 chances out of 100. For estimated percentages of urban housing units pertaining to new construction, lacking complete kitchen facilities, no bedrooms, no bathrooms, and lacking complete plumbing facilities for the Northeast, North Central, and West Regions and to source of water-individual well, cooking fuel, and mobile homes for the Northeast and North Central Regions, apply a factor of 0.91 to the standard errors. For estimated percentages pertaining to source of water-individual well, cooking fuel, and mobile homes for the West Region, apply a factor of 1.66 to the standard errors. For standard errors of regional rural estimates pertaining to new construction, no bedrooms, lacking complete plumbing, source of water-individual well, cooking fuel, and mobile homes for the Northeast, North Central, and West Regions, multiply the standard errors by 1.04, except for estimates for the West Region pertaining to source of water-individual well, cooking fuel, and mobile homes, multiply the standard errors by 1.77)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	38.5	38.5	38.5	38.5	38.5	38.5	38.5	39.6
10	23.9	23.9	23.9	23.9	23.9	23.9	24.2	28.0
25	11.1	11.1	11.1	11.1	11.1	12.6	15.3	17.7
50	5.9	5.9	5.9	5.9	7.5	8.9	10.8	12.5
100	3.0	3.0	3.0	3.9	5.3	6.3	7.7	8.8
250	1.2	1.2	1.6	2.4	3.4	4.0	4.8	5.6
500	0.6	0.8	1.1	1.7	2.4	2.8	3.4	4.0
1,000	0.3	0.6	0.8	1.2	1.7	2.0	2.4	2.8
2,500	0.13	0.4	0.5	0.8	1.1	1.3	1.5	1.8
5,000	0.06	0.2	0.4	0.5	0.8	0.9	1.1	1.3
7,500	0.04	0.2	0.3	0.4	0.6	0.7	0.9	1.0
10,000	0.03	0.2	0.2	0.4	0.5	0.6	0.8	0.9
25,000	0.01	0.11	0.2	0.2	0.3	0.4	0.5	0.6

TABLE XIc. Standard Errors of Estimated Percentages of Housing Units Pertaining to Source of Water-Individual Well, Cooking Fuel, and Mobile Homes for the South Region: 1981

(68 chances out of 100. For estimated percentages of urban housing units pertaining to source of water-individual well, cooking fuel, and mobile homes for the South Region, apply a factor of 0.91 to the standard errors. For standard errors of estimated percentages of rural housing units pertaining to source of water-individual well, cooking fuel, and mobile homes for the South Region, multiply the standard errors by 1.06)

Base of percentage (000)	Estimated percentages							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	49.7	49.7	49.7	49.7	49.7	49.7	49.7	49.7
10	33.1	33.1	33.1	33.1	33.1	33.1	33.1	35.2
25	16.5	16.5	16.5	16.5	16.5	16.5	19.3	22.2
50	9.0	9.0	9.0	9.0	9.4	11.2	13.6	15.7
100	4.7	4.7	4.7	4.8	6.7	7.9	9.6	11.1
250	1.9	1.9	2.0	3.1	4.2	5.0	6.1	7.0
500	1.0	1.0	1.4	2.2	3.0	3.6	4.3	5.0
1,000	0.5	0.7	1.0	1.5	2.1	2.5	3.0	3.5
2,500	0.2	0.4	0.6	1.0	1.3	1.6	1.9	2.2
5,000	0.10	0.3	0.4	0.7	0.9	1.1	1.4	1.6
7,500	0.07	0.3	0.4	0.6	0.8	0.9	1.1	1.3
10,000	0.05	0.2	0.3	0.5	0.7	0.8	1.0	1.1
25,000	0.02	0.14	0.2	0.3	0.4	0.5	0.6	0.7

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-45
Selection of sample areas. App-45
Designation of sample housing for the 1981 survey App-45
Selection of the 1973 sample housing units. App-46
Splitting of the sample App-46
CEN-SUP sample App-46
Selection of supplemental sample housing units in rural areas App-46
Selection of sample housing units for the 1976 Coverage Improvement Program App-46
1977 and 1981 sample reductions . App-47
Supplemental sample from nonresidential conversions App-47
1981 telephone interviewing experiment App-47
ESTIMATION. App-47
RELIABILITY OF THE ESTIMATES. App-49
Sampling errors App-49
Standard errors of estimates of levels App-49
Standard errors of estimates of percentages. App-49
Standard errors of ratios App-49
Illustration of the use of the standard error tables. Illustration I App-49
Illustration II App-50
Differences. App-50
Illustration of the computation of the standard error of a difference App-50
Medians App-51
Illustration of the computation of the 95-percent confidence interval for a median. App-51
Nonsampling errors App-51
Reinterview program App-51
Coverage errors App-52
Effects of ratio estimation on the estimates of change App-53
Possible effects of telephone interviewing on the data App-53
Rounding errors. App-53
Standard error tables App-53

SAMPLE DESIGN

The 1981 estimates are based on data collected from September 1981 through December 1981 for the Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. The sample for this survey was spread over 461 sample areas (called primary sampling units), comprising 923 counties and independent cities with coverage in each of the 50 States and the District of Columbia.

Approximately 56,800 sample housing units (both occupied and vacant) were eligible for interview in the 1981 Annual Housing Survey. Of this number, 3,200 interviews were classified as "noninterview" for various reasons. Occupied housing units were classified as "noninterview" mainly because the occupants refused to be interviewed after repeated calls. For vacant housing units, interviews were not obtained because an informed respondent was not found after repeated visits. In addition to the 56,800 eligible housing units, there were also 3,600 sample units which were visited but were ineligible for interview for the AHS in terms of collecting information relevant to the 1981 housing inventory.

Selection of sample areas—The United States was divided into areas made up of counties and independent cities referred to as primary sampling units (PSU's). These PSU's were then grouped into 376 strata, 156 of which consisted of only one PSU which was in sample with certainty. These 156 strata were mostly the larger SMSA's and were called self-representing (SR) since the sample from the sample area represented just that PSU. Each one of the other 220 strata consisted of a group of PSU's and were referred to as non-self-representing (NSR), since the sample of housing units from the sample PSU in a stratum represented the other PSU's in the stratum as well.

One PSU was selected from each NSR stratum with probability proportionate to the 1970 census population of the PSU. (This resulted in 220 NSR sample PSU's.) In addition, the NSR strata were grouped into 110 pairs and one stratum was picked at random from each pair. From this stratum, an additional PSU was selected independently of the other PSU selected from this stratum. Since the two PSU's were independently selected, it was possible for the same PSU to be selected twice. This occurred in 25 instances, producing an additional 85 NSR sample PSU's, thus giving a grand total of 461 PSU's.

Designation of sample housing units for the 1981 survey—The sample housing units designated to be interviewed in the 1981 survey consisted of the following categories, which are described in detail in succeeding sections.

1. All sample housing units that were interviewed in the 1980 survey (which included all sample housing units that were selected as part of the 1976 Coverage Improvement Program) and were not part of the 1981 sample reduction.

2. All sample housing units that were either type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of survey but which could become eligible in the future) in the 1980 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1981 AHS questionnaire, page App-21).
3. All sample housing units that were selected from the list of building permits issued since the 1980 survey. (This sample represented the housing units built in permit-issuing areas, since the 1980 survey.)
4. Housing units added as the result of the updated listings in selected areas which do not issue building permits.

Selection of the 1973 sample housing units—The overall sampling rate used to select the sample for the 1973 AHS was about 1 in 1,366. The within-PSU sampling rate for the AHS was determined so that the overall probability of selection for each sample housing unit was the same (e.g., if the probability of selecting a NSR PSU was 1 in 10, then the within-PSU sampling rate would be 1 in 136.6).

Within the sample PSU's, a sample of the housing units enumerated in the 1970 Census of Population and Housing was selected for the AHS. In addition, a sample of new construction building permits was also selected to represent the units constructed since the 1970 census. These samples were selected at about twice the rate mentioned previously (i.e., at 2 in 1,366), thereby producing a sample twice as large as needed. This sample was split into two equal-sized samples—one to be used for the AHS, and one to be held in reserve for possible future use for the AHS. The procedure used to split this sample into equal-sized samples is described in the next section.

The sample of 1970 census units was selected in several stages. Within the sample PSU's, the first step was the selection of a sample of census enumeration districts (ED's), administrative units used in the 1970 census. The probability of selection for an ED was proportional to the following 1970 census counts of housing units (HU's) and persons in group quarters, combined in the following formula:

$$\frac{\text{Number of HU's in the ED} + \dfrac{\text{Number of group quarters persons in the ED}}{3}}{4}$$

The next step was to select an expected cluster of about four neighboring housing units within each sample ED. For most of the ED's, the selection was accomplished using the list of addresses for the ED as compiled in the 1970 census. (These ED's are referred to as address ED's.) However, in those ED's where addresses were incomplete or inadequate (mostly rural areas), the selection process was accomplished using area sampling methods. These ED's were divided into segments (i.e., small land areas with well-defined boundaries, having an expected size of four, or a multiple of four, housing units) and a segment was selected. Those selected segments with an expected size which was a multiple of four were further subsampled at the time of interview so that an expected four housing units were chosen for interview.

The sample of new construction housing units was selected from building permits issued since January 1970. Within each sample PSU, the building permits were chronologically ordered by month issued, and compact clusters of approximately four housing units were created. These clusters were then sampled at the rate of 2 in 1,366. Housing units constructed since the 1970 census in areas which do not issue building permits were brought into the sample as a result of the area sample described.

Splitting of the sample—The described sample selection procedure produced clusters (or segments) of size-four housing units for the sample taken from the census address frame, the new construction frame, and the area sampling frame (mainly rural areas). Clusters of this size should result in a minimum loss in precision for estimates of housing characteristics in rural areas because of the heterogeneity of neighboring units. However, clusters of size-two housing units were considered to be more optimum within those areas where the housing characteristics of neighboring units tend to be very similar (i.e., urban areas and new construction units). A splitting operation was then carried out for clusters selected from the census address and the new construction frames. This consisted of halving each sample cluster from these frames. Thus, two housing units from each of these clusters were included in the survey and two housing units were held in reserve. No splitting operation was carried out within the clusters selected from the area sampling frame; every other area sample cluster of four housing units was used for the survey and the remaining clusters were assigned to the reserve sample.

CEN-SUP sample—Housing units at addresses missed in the census or inadequately described in the 1970 census address register did not have a chance of being selected for the AHS sample in address ED's. Consequently, a special operation was undertaken to develop a separate sample, called the census supplemental (CEN-SUP) sample to represent these units. Due to time constraints on this operation, only about 40 percent of the CEN-SUP units were ready to be interviewed in 1973. The rest were interviewed for the first time in the 1974 AHS.

Selection of supplemental sample housing units in rural areas—In 1974, it was decided to increase the reliability of the AHS estimates of rural housing characteristics by doubling the number of sample housing units from rural areas. This was accomplished by reactivating the reserve sample, selected in the original sampling operations in 1973, from rural areas only. For the reserve sample selected in census address and new construction frames, the other half of each rural cluster (an expected two housing units) was reactivated in 1974. Similarly, for the area sampling frame, the entire reserve cluster (an expected four housing units) was reactivated in 1974 if the cluster was rural. This supplementation increased the overall probability of selection for sample housing units in rural areas to about 2 in 1,366; whereas, the overall probability of selection for sample housing units in urban areas remained at 1 in 1,366.

Selection of sample housing units for the 1976 Coverage Improvement Program—The 1976 Coverage Improvement Pro-

gram was undertaken to correct certain deficiencies in the AHS national sample from the census address and new construction frames. The coverage deficiencies included the following types of housing units:

1. New construction housing units from building permits issued prior to January 1970 for which construction had not been completed at the time of the 1970 census.
2. Housing units converted to residential use in structures totally nonresidential at the time of the 1970 census.
3. Houses that have been moved onto their present site since the 1970 census.
4. Mobile homes and trailers placed in parks either missed in the 1970 census or established since the 1970 census.
5. Mobile homes and trailers placed outside parks since the 1970 census or vacant at the time of the 1970 census.

A sample of new construction housing units whose permits were issued before January 1970 was selected in two stages. First, units whose permits were issued before January 1970, but which were completed after the census, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. Second, these units were then sampled so that the overall probability of selection was about 1 in 1,320.

A sample of mobile homes and trailers placed in parks missed by the census or established after the census was also selected in two stages. During the first stage, a list of mobile home parks was obtained from commercial listings. This list was then supplemented by additional parks identified by a canvassing operation similar to that performed in ED's where area sampling methods are used. The second stage consisted of dividing the parks into clusters of an expected size of four sites. These clusters were then sampled so that the overall probability of selection was about 1 in 1,366.

For the remaining housing units (i.e., mobile homes and trailers placed outside parks since the 1970 census, mobile homes and trailers vacant at the time of the 1970 census, units converted from nonresidential to residential use since the 1970 census, and houses that had been moved onto their present site since the 1970 census), the sampling was done in three stages. First, a subsample of the regular AHS sample housing units from the census address frame was selected. Second, succeeding structures that had been eligible to be selected from the census address frame were then listed until eight structures (excluding mobile home parks) were found. Finally, the intervening structures that had been listed which did not have a chance of selection in the AHS were identified and the units within these structures were interviewed.

1977 and 1981 sample reductions—By 1977, the addition to the sample, primarily from new construction and the coverage improvements, had increased the total sample size (interviews plus noninterviews) to about 81,000 housing units. The sample was reduced by about 7 percent to approximately 75,000 in 1977. This reduction did not include any CEN-SUP units or units which were selected as part of the 1976 Coverage Improvement Program. Initially, the sample was further reduced in 1981 by about 5 percent. Subsequently, the rural sample was reduced by about, 25 percent in 125 "large" SMSA's[1] and by about 50 percent in "small" SMSA's[1] and outside the SMSA's. These reductions brought the sample size down to approximately 60,000 housing units in 1981.

Supplemental sample from nonresidential conversions—The purpose of this sample was to improve the coverage in address ED's (i.e., ED's in which 1970 census address listings were used for selecting the AHS sample) of housing units created in structures that were completely nonresidential at the time of the 1970 census. This sample was derived from listings created for the Commercial Victimization Survey (CVS), a nationally representative area sample survey conducted by the Census Bureau for the Department of Justice. Each of the CVS area segments was initially canvassed either in 1972 or 1975 and eligible business establishments were listed. These listings were updated every 6 months with the last updating in the last half of 1977. Each basic address containing a business establishment listed for CVS was visted for the 1980 Components of Inventory Change Survey (CINCH), which was conducted in conjunction with the 1980 AHS, to identify HU's in structures which were completely nonresidential at the time of the 1970 census. These cases were later matched to the 1970 census address listings to identify those cases in address ED's and as an additional check to see if housing units existed in these structures at the time of the 1970 census. Due to the timing restraints associated with these operations, these cases were not included for the 1980 AHS but were included for the 1981 survey.

1981 telephone interviewing experiment—A large scale telephone interviewing experiment was conducted in conjunction with the 1981 AHS national sample in order to provide more definitive information about the effect of telephone interviewing on the AHS data and cost. A random sample of about 12,500 AHS national sample units was selected for the experiment. Among the cases assigned for telephone interviewing, only those sample housing units that had been interviewed in 1980 and had a telephone number available were eligible to be interviewed by telephone. Since a large portion of the total AHS national sample was assigned for telephone interviewing, its possible effect on the data is mentioned in the section "Reliability of the Estimates" of this appendix.

ESTIMATION

In 1981, the AHS estimates employed a three-stage ratio estimation procedure. However, prior to implementation of the procedure, the basic weight (i.e., the inverse of the probability of selection) was adjusted to account for the type A noninterview housing units encountered in the AHS. This noninterview adjustment was done separately for occupied and vacant units. The noninterview adjustment was equal to the following ratio:

$$\frac{\text{Interviewed housing units + Noninterviewed housing units}}{\text{Interviewed housing units}}$$

[1] A "large" SMSA is one with a 1970 population of 250,000 persons or more and a "small" SMSA is one with a 1970 population of less than 250,000 persons.

The first-stage ratio estimation procedure was employed for sample housing units from non-self-representing (NSR) PSU's only. This procedure was designed to reduce the contribution to the variance arising from the sampling of PSU's. The first-stage ratio estimation procedure takes into account the differences that existed at the time of the 1970 census in the distribution by tenure and residence of the housing population estimated from the sample NSR PSU's and that of the NSR housing population in each of the four census regions of the country.

The first-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{The 1970 census housing population in the residence-tenure category for all NSR strata in a census region}}{\text{Estimate of the housing population category using 1970 census housing counts for sample NSR PSU's in a census region}}$$

The numerators of the ratios were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR stratum and summing these counts across the NSR strata in each census region. The denominators were calculated by obtaining the 1970 census housing counts for each of the residence-tenure categories for each NSR sample PSU, weighting these counts by the inverse of the probability of selecting that PSU, and summing these weighted counts across the NSR sample PSU's in each census region. The computed first-stage ratio estimation factor was then applied to the existing weight for each NSR sample unit in each first-stage ratio estimation category.

The second-stage ratio estimation procedure was designed to adjust the AHS sample estimate of conventional new construction housing units, i.e., sample units built April 1, 1970, or later, to independently derived current estimates where a known deficiency in the AHS sample exists (see the section on nonsampling error) for each of the four regions. These estimates were considered to be the best estimates available for the number of conventional new construction units in these categories.

The second-stage ratio estimation factor was as follows:

$$\frac{\text{Current best estimate of new construction in the category}}{\text{AHS sample estimate of new construction units in the category}}$$

The numerators of the ratios were derived from data based on the Survey of Construction (SOC). The denominators of the ratios were obtained from the weighted estimates for the AHS sample units using the existing weight after the first-stage ratio estimates procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in each second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all the AHS sample units. This procedure was designed to adjust the AHS sample estimates of housing (i.e., the estimates employing the noninterview, first-stage, and second-stage adjustments) to current vacant housing estimates for 4 categories of vacant housing units and to independently derived current housing estimates for 24 categories of occupied housing units. Each of these categories is a combination of the characteristics of residence, tenure, race of householder, and sex of householder.

The third-stage ratio estimation factor for each specified category was as follows:

$$\frac{\text{Current independent estimate of housing units in the category}}{\text{AHS sample estimate of housing units in the category}}$$

The numerators of the ratios for occupied housing units were derived from data based on the Current Population Survey (CPS), a sample household survey conducted monthly by the Bureau of the Census. This is the first time that CPS household estimates based on the 1980 census were used as the independent controls in this adjustment rather than the 1970 census-based controls. The 1970-based estimates were about 2 percent smaller than the 1980-based estimates. This difference should be taken into consideration when comparing the 1981 AHS estimates to the AHS estimates for 1980 or earlier years. The numerators of the ratios for vacant housing units were derived from data based on the Housing Vacancy Survey (HVS), a quarterly vacancy survey also conducted by the Bureau of the Census. The denominators of the ratios were obtained from the weighted estimates for the AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computer third-stage ratio estimation factor was then applied to the existing weight for each sample unit in each third-stage ratio estimation category.

The second-stage and the third-stage ratio estimation procedures were iterated in order to bring the AHS estimates into close agreement with both sets of independent estimates. The second stage was modified so that the estimates for all 27 categories of new construction would be identical to the estimates before the third stage. Hence, the repeated second stage had the effect of controlling the AHS sample estimates of new construction units to the unbiased sample estimates for 9 categories of new construction mobile homes and trailers for each of the four regions and of adjusting the AHS sample estimate of 18 categories of conventional new construction units to the independently derived current estimates.

The numerators were either the unbiased weighted estimates for the AHS sample housing units, using the existing weight after the first-stage ratio estimation procedure (i.e., the estimates employing the noninterview and first-stage adjustments) or the independent estimate derived from data based on the Survey of Construction (SOC).

The denominators of the ratios in this iterative process were obtained from the weighted estimates for the AHS sample units after the previous stage of ratio estimation. The factors resulting from this iterative process were then applied to the existing weight on the appropriate records, and the resulting product was used as the final weight for tabulation.

The third-stage ratio estimation procedure, as well as the overall estimation procedure, reduced the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. The distribution of the housing population selected for the sample differed somewhat, by chance, from that of the Nation as a whole in such basic housing

characteristics as tenure, vacancy status, residence, race of householder, and sex of householder. These characteristics are probably closely correlated with other housing characteristics measured for the AHS. Therefore, through the use of the three-stage ratio estimation procedure one can expect the sample estimate to be improved substantially.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS national sample.

Sampling errors—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same questionnaires, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The variability between estimates from all possible samples is defined as sampling error. One common measure of sampling error is the standard error which measures the precision with which an estimate from a sample approximates the average result of all possible samples. In addition, the standard error, as calculated for this report, also partially reflects the variation in the estimates due to some nonsampling errors, but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, and biases and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates so that the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, and each of these samples was surveyed under essentially the same general conditions and an estimate and its estimated standard error were calculated for each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples either is or is not contained in any particular computed interval. However, for a particular sample one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables (pages App-53 and App-54) of this appendix are approximations to the standard errors of various estimates shown in this report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the tables of standard errors provide an indication of the order of magnitude of the standard errors rather than the precise standard error for any specific item.

Standard errors of estimates of levels—Tables I and II present the standard errors applicable to the 1981 national and regional housing inventory estimates in this report. Linear interpolation should be used to determine standard errors for levels of estimates not specifically shown in tables I and II.

Standard errors of estimates of percentages—The reliability of an estimated percentage, computed by using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more. Tables III and IV present the standard errors for estimated percentages.

Included in tables I through IV are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered to be overestimates of the true standard errors and should be used primarily for construction of confidence intervals for characteristics when an estimate of zero is obtained.

Standard errors of ratios—For ratios of the form (100) (x/y), where x is not a subclass of y, tables III and IV underestimate the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio, a better approximation of the standard error may be obtained by letting the standard error of the ratio be approximately equal to:

$$(100)\left(\frac{x}{y}\right)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

where: x = the numerator of the ratio
y = the denominator of the ratio
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

Illustration of the use of the standard error tables. Illustration I—Table A-1 of this report shows that in the United States there were 28,833,000 renter-occupied housing units in 1981. Interpolation in the standard error table I shows that the standard error of an estimate of this size is approximately 211,000. The following interpolating procedure was used.

The information presented in the following table was extracted from standard error table I. The entry for "x" is the one sought.

Size of estimate (000)	Standard error (000)
25,000	207
28,833	x
50,000	231

By vertically interpolating between 207 and 231, the entry for "x" is determined to be 211.

$$28{,}833-25{,}000 = 3{,}833$$
$$50{,}000-25{,}000 = 25{,}000$$

$$207 + \frac{3{,}833}{25{,}000}(231-207) = 211$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 28,622,000 to 29,044,000 housing units. Therefore, a conclusion that the average estimate of 1981 housing units of this type lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all possible samples, lies within the interval from 28,495,000 to 29,171,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 28,411,000 to 29,255,000 housing units with 95-percent confidence.

Table A-1 also shows that of the 28,833,000 renter-occupied housing units in 1981, 15,287,000, or 53.0 percent, heat with utility gas. Interpolation in standard error table III (i.e., interpolation on both the base and percent) shows that the standard error of the above percentage is 0.5 percentage points. The following interpolating procedure was used.

The information presented in the following table was extracted from standard error table III. The entry for "p" is the one sought.

Base of percentage (000)	Estimated percentage		
	75	53.0	50
25,000	0.4	a	0.5
28,833		p	
50,000	0.3	b	0.3

1. By horizontal interpolation between 0.5 and 0.4, the entry for cell "a" is determined to be 0.5.

$$53.0-50.0 = 3.0$$
$$75.0-50.0 = 25.0$$

$$0.5 + \frac{3.0}{25.0}(0.4-0.5) = 0.5$$

2. By horizontal interpolation between 0.3 and 0.3, using the same procedure as in step 1, the entry for cell "b" is determined to be 0.3.
3. By vertical interpolation between 0.5 and 0.3, using the same procedure as in step 1, the entry for "p" is determined to be 0.5.

Consequently, the 68-percent confidence interval, as shown by these data, is from 52.5 to 53.5 percent; the 90-percent confidence interval is from 52.2 to 53.8 percent; and the 95-percent confidence interval is from 52.0 to 54.0 percent.

Illustration II—Table E-1 of this report shows that in all occupied housing units of the West Region in 1981 there were 9,765,000 owner-occupied housing units. Interpolation in standard error table II shows that the standard error of an estimate of this size is approximately 229,000. Consequently, the 68-percent confidence interval is from 9,536,000 to 9,994,000 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1981 owner-occupied housing units in the West Region lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate, derived from all possible samples, lies with the interval from 9,399,000 to 10,131,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 9,307,000 to 10,223,000 housing units with 95 percent confidence.

Table E-I also shows that of the 9,765,000 owner-occupied housing units in the West Region, 2,540,000 or 26.0 percent, have a central air-conditioning system. Interpolation in standard error table IV (i.e., interpolation on both the base and the percent) shows that the standard error of the above percentage is 0.8 percentage points. Consequently, the 68-percent confidence interval, as shown by these data, is from 25.2 to 26.8 percent; the 90-percent confidence interval is from 24.7 to 27.3 percent; and the 95-percent confidence interval is from 24.4 to 27.6 percent.

Differences—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between separate and uncorrelated characteristics in the same area. If there is a high positive correlation between the two characteristics, the formula will overestimate the true error; if there is a high negative correlation, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference—Table A-1 of this report shows that of all renter-occupied housing units in the United States there were 6,354,000 units which heated with electricity. Thus, the apparent difference between the number of 1981 renter-occupied units which heated with utility gas and those which heated with electricity is 8,933,000. Interpolation in standard error table I shows that the standard error of 15,287,000 is approximately

167,000 and that the standard error on an estimate of 6,354,000 is approximately 117,000. Therefore, the standard error of the estimated difference of 8,933,000 is about:

$$204{,}000 = \sqrt{(167{,}000)^2 + (117{,}000)^2}$$

Consequently, the 68-percent confidence interval for the 8,933,000 difference is from 8,729,000 to 9,137,000 housing units. Therefore, a conclustion that the average estimate of this difference, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 8,607,000 to 9,259,000 housing units, and the 95-percent confidence interval is from 8,525,000 to 9,341,000. Thus, we can conclude with 95-percent confidence that the number of 1981 renter-occupied units which heated with utility gas is greater than the number which heated with electricity.

Medians—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median so that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

1. From the appropriate standard error table determine the standard error of a 50-percent characteristic on the base of the median.
2. Add to and subtract from 50 percent the standard error determined in step 1.
3. Using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 95-percent confidence interval for a median—Table A-2 of this report shows the median income of families and primary individuals in specified owner-occupied housing units was $23,200 in 1981. The base of the distribution, from which this median was determined is 43,293,000 housing units.

1. From standard error table III, the standard error of a 50-percent characteristic on the base of 43,293,000 is 0.4 percentage points.
2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 49.2 and 50.8.
3. From table A-2, it can be seen by cumulating the frequencies for the first 5 categories that 18,227,000 owner-occupied housing units, or 42.1 percent, had income less than $20,000 and that an additional 5,351,000 specified owner-occupied housing units, or 12.4 percent, had income from $20,000 to less than $25,000. By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$\$20{,}000 + (\$25{,}000 - \$20{,}000)\,\frac{(49.2 - 42.1)}{12.4} = \$22{,}900$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$\$20{,}000 + (\$25{,}000 - \$20{,}000)\,\frac{(50.8 - 42.1)}{12.4} = \$23{,}500$$

Thus, the 95-percent confidence interval ranges from $22,900 to $23,500.

Nonsampling errors—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases; definitional difficulties; differences in the interpretation of questions; inability or unwillingness to provide correct information on the part of respondents; mistakes in recording or coding the data; and other errors of collection, response, processing, coverage, and estimation for missing data. Nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for the 1981 AHS national sample.

Reinterview program—For the AHS national sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a subsample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and thus were the basis for the measurement of the "content" error of these AHS estimates.

As part of the reinterview, an additional check was carried out for interviewer evaluation and quality control. This check was made at each of these households to determine if the following were done during the original interview.

1. The correct housing unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.

6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

The results of the 1981 and 1980 reinterview studies were not available at the time of publication. In addition, the 1979 reinterview was carried out as an interviewer quality control only. Thus, the results of the 1977 and 1978 reinterview studies which were presented in the Census Bureau memoranda, "Reinterview Results for the Annual Housing Survey—National Sample 1977" and "Reinterview Results for the Annual Housing Survey—National Sample 1978" are presented here.

In 1978, a substantial portion of the reinterview questionnaire was devoted to testing the new questions 1 through 7 (parts a and b). The questions (part a), which were asked only at housing units interviewed in the previous year, determined whether there had been a change since last year in selected nonattitudinal items. If a change had been recorded or the respondent did not know if a change had occurred, part b of the question, which collects the value of the item, was asked. The reinterviewers asked these items using the questions as formatted in 1977. Comparing the responses from the differently formatted questions, the 1978 reinterview found that 80 percent of the questions showed low levels of inconsistency with the remainder showing moderate levels.

The 1977 reinterview program showed moderate to high levels of inconsistency with about 21 percent of the nonattitudinal and 56 percent of the attitudinal items showing high levels of inconsistency. A large portion (43 percent) of the nonadditudinal items showed a low level of inconsistency. Moderate levels indicate that there are some problems with inconsistent reporting and high levels indicate that improvements are needed in the data collection methods or that the category concepts themselves are ambiguous.

Cross-tabulations involving those items, which are subject to substantial levels of inconsistency may be subject to a large distortion as a consequence of the associated high level of response variance, and thus, are considered to be less reliable than comparable cross-tabulations which do not involve these data. As a consequence, the tables in this report, which contain such cross-tabulations, have been footnoted with a cautionary statement.

The 1970 census reinterview results provide illustrations of possible nonsampling errors for some of the items which also appear in the AHS. For example, median value of homes was consistently underestimated by about 5 percent, and the average monthly cost of electricity and utility gas were consistently overestimated although the net effect on average gross rent was fairly small.

A possible explanation for the results of the AHS and census reinterview studies, as well as the surveys themselves, is that the respondents may lack precise information. Also, because the results of the reinterview studies are derived from sample surveys, there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results of these studies.

Coverage errors—A deficiency in the representation of conventional new construction (non-mobile home or trailer) for the AHS new construction sample (mentioned previously in the section on estimation) is an example of coverage errors. During the sampling of building permits, only those issued more than 4 months before the survey began were eligible to be selected to represent conventional new construction. Due to time constraint it is not possible to sample units whose permits are issued less than 4 months in advance of the survey.

It is estimated that the 1981 AHS sample missed about 0.8 percent (i.e., about 157,000 units) of conventional housing units built after April 1970 because the permits for these units, which were built before October 1981, were issued less than 4 months in advance of the survey. The second-stage ratio estimation procedure was employed to reduce the effect of this deficiency although some bias in the AHS estimates of conventional new construction probably still exists. Review of the second-stage ratio estimation procedure indicates that we have consistently overcompensated for this deficiency every year since 1975 by ratio adjusting the new construction to counts of new construction for the end of the interview period, which has been December or January, instead of October. This overcompensation may inflate the new construction counts by 100,000 to 300,000 units.

Changing this procedure to correct for this overcompensation would reduce the estimate of housing units built since the last survey since it would be based on a 10-month time period rather than 12 months. Thus, it was decided not to change this procedure so that the estimate of housing units built since the last survey would be consistent with previous years' estimates, which are generally based on a 12-month time period.

In addition, the 1976 Coverage Improvement Program had certain deficiencies. First, when the canvassing was done to identify mobile homes and trailer parks that were not in the same sample frame or not on the commercial lists, only 92 percent of the census address frame ED's were represented. Second, it appears that the listing procedure (used to find mobile homes and trailers placed outside parks, units converted from nonresidential to residential, and houses that had been moved onto their present site) was not very efficient for finding nonresidential conversions (which might be primarily in business districts), since the listing procedure started from a residential unit. (The sample estimate of this component was approximately 16,000 housing units with a standard error of 12,000.)

Finally, it is felt that deficiencies also exist in ED's where area sampling methods are used. As before, it had been assumed that all housing units located inside these ED's would be represented in the sample. However, it has been estimated that the 1981 AHS sample missed as much as 2 percent (i.e., as much as 400,000 units) of all housing units in ED's where area sampling methods are used because these units are not listed during the canvassing.

The third-stage of ratio estimation corrects for these deficiencies as far as the total count of housing is concerned, i.e., it adjusts the estimate of the total housing inventory to the best available estimate. However, biases of subtotals would still remain.

Effects of ratio estimation on the estimates of change—As stated previously, 1981 was the first AHS for which 1980 census-based estimates, rather than 1970 census-based estimates, were used in the third-stage ratio estimation procedure. Since the 1980-based estimates are about 2 percent higher than the corresponding 1970-based estimates, this will cause estimates of change between the 1981 AHS data and data for 1980 or earlier years to be overstated. This overstatement should be taken into consideration during the analysis of estimates of change.

Possible effects of telephone interviewing on the data—A preliminary analysis of the 1981 telephone interviewing experiment has been completed. It has been concluded that data collected, using the telephone interviewing procedures, were not sufficiently different from data collected by regular personal interviews, to preclude basing the 1981 AHS-national sample published data. However, since the percentage of the comparison tests that were significant was slightly above what could be expected by chance, there was some evidence that telephone interviewing may have had an effect on the data. Thus, some caution should be exercised in making comparisons between 1981 AHS data and data from preceding years.

Rounding errors—In errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of this survey. Also, since medians in this report were computed using unrounded data, instead of the published rounded data, they can differ from medians calculated directly from the published data.

TABLE I. Standard Errors of Estimated Numbers of National and Regional Housing Units: 1981 (Excluding Estimates of Housing Units for the West Region and of Housing Units With Spanish-Origin Householder)

(68 chances out of 100)

Size of estimate (000)	Standard error (000)		Size of estimate (000)	Standard error (000)	
	National	Regional		National	Regional
0	2	2	1,000	48	47
5	3	3	2,500	76	76
10	5	5	5,000	106	110
25	8	7	10,000	145	164
50	11	10	25,000	207	295
100	15	15	50,000	231	488
250	24	23	75,000	179	—
500	34	33			

TABLE II. Standard Errors of Estimated Numbers of Housing Units for the West Region and of Housing Units With Spanish-Origin Householder: 1981

(68 chances out of 100)

Size of estimate (000)	Standard error (000)	Size of estimate (000)	Standard error (000)
0	3	500	38
5	4	1,000	55
10	5	2,500	93
25	8	5,000	144
50	12	10,000	234
100	17	25,000	488
250	27		

TABLE III. Standard Errors of Estimated Percentages of Housing Units: 1981 (Excluding Estimated Percentages of Housing Units for the West Region and of Housing Units With Spanish-Origin Householder)

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	32.0	32.0	32.0	32.0	32.0	32.0	32.0	34.3
10	19.1	19.1	19.1	19.1	19.1	19.1	21.0	24.3
25	8.6	8.6	8.6	8.6	9.2	11.0	13.3	15.3
50	4.5	4.5	4.5	4.7	6.5	7.7	9.4	10.9
100	2.3	2.3	2.3	3.3	4.6	5.5	6.6	7.7
250	0.9	1.0	1.4	2.1	2.9	3.5	4.2	4.9
500	0.5	0.7	1.0	1.5	2.1	2.5	3.0	3.4
1,000	0.2	0.5	0.7	1.1	1.5	1.7	2.1	2.4
2,500	0.09	0.3	0.4	0.7	0.9	1.1	1.3	1.5
5,000	0.05	0.2	0.3	0.5	0.7	0.8	0.9	1.1
10,000	0.02	0.15	0.2	0.3	0.5	0.5	0.7	0.8
25,000	0.01	0.10	0.14	0.2	0.3	0.3	0.4	0.5
50,000	–	0.07	0.10	0.15	0.2	0.2	0.3	0.3
75,000	–	0.06	0.08	0.12	0.2	0.2	0.2	0.3

TABLE IV. Standard Errors of Estimated Percentages of Housing Units for the West Region and of Housing Units With Spanish-Origin Householder: 1981

(68 chances out of 100)

Base of percentage (000)	Estimated percentage							
	0 or 100	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	50
5	35.8	35.8	35.8	35.8	35.8	35.8	35.8	37.3
10	21.8	21.8	21.8	21.8	21.8	21.8	22.8	26.4
25	10.0	10.0	10.0	10.0	10.0	11.9	14.5	16.7
50	5.3	5.3	5.3	5.3	7.1	8.4	10.2	11.8
100	2.7	2.7	2.7	3.6	5.0	6.0	7.2	8.3
250	1.1	1.1	1.5	2.3	3.2	3.8	4.6	5.3
500	0.6	0.7	1.0	1.6	2.2	2.7	3.2	3.7
1,000	0.3	0.5	0.7	1.2	1.6	1.9	2.3	2.6
2,500	0.11	0.3	0.5	0.7	1.0	1.2	1.4	1.7
5,000	0.06	0.2	0.3	0.5	0.7	0.8	1.0	1.2
10,000	0.03	0.2	0.2	0.4	0.5	0.6	0.7	0.8
25,000	0.01	0.11	0.15	0.2	0.3	0.4	0.5	0.5