1976 Metropolitan Appendix B: Source and Accuracy

Table of Contents

Allentown-Bethlehem-Easton 3

Baltimore 12

Birmingham 21

Buffalo 30

Cleveland 39

Denver 49

Grand Rapids 58

Honolulu 67

Houston 76

Indianapolis 86

Las Vegas 95

Louisville 104

New York 113

Oklahoma City 123

Omaha 132

Providence-Pawtucket-Warwick 141

Raleigh 150

Sacramento 159

St. Louis 168

Seattle-Everett 178

Allentown-Bethlehem-Easton 1976

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Annual Housing Survey App-42
Selection of the sample App-42
Coverage improvement sample selection App-43
Coverage improvement for deficiency 1 ..... App-43
Coverage improvement for deficiency 2 ..... App-43
Coverage improvement for deficiencies 3-6 ..... App-44
Building loss sample selection ... App-44
1970 Census of Population and Housing App-44
ESTIMATION App-44
1976 housing inventory App-44
1970-1976 lost units App-45
Ratio estimation procedure of the 1970 Census of Population and Housing App-45
RELIABILITY OF THE ESTIMATES App-45
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-46
Coverage errors App-46
Rounding errors App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences App-48
Illustration of the computation of the standard error of a difference App-49
Medians App-49
Illustration of the computation of the 95-percent confidence interval of a median .. App-49

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 20 SMSA's are based on data collected from the 1976 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. In each of the 20 SMSA's, the data were collected for the 12-month period from April 1976 through March 1977 with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and data for a second group of 21 SMSA's were collected from April 1975 through March 1976. The sample housing units for each group of the AHS SMSA's are interviewed on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the third group (1976-77) are: Houston, Tex., New York, N.Y., St. Louis, Mo.-Ill., and Seattle-Everett, Wash.

The remaining SMSA's in the third group are: Allentown-Bethlehem-Easton, Pa.-N.J., Baltimore, Md., Birmingham, Ala., Buffalo, N.Y., Cleveland, Ohio, Denver, Colo., Grand Rapids, Mich., Honolulu, Hawaii, Indianapolis, Ind., Las Vegas, Nev., Louisville, Ky.-Ind., Oklahoma City, Okla., Omaha, Nebr.-Iowa, Providence-Pawtucket-Warwick, R.I.-Mass., Raleigh, N.C., and Sacramento, Calif.

In this SMSA, 4,686 units were eligible for interview. Of these sample units, 138 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to the units eligible for interview, 330 units were visited but were not eligible for interview, because they were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit-issuing (Honolulu, Las Vegas, New York, and Sacramento) was selected from two sample frames—units enumerated in the 1970 Census of Population and Housing in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for the 16 SMSA's which are *not* 100-percent permit-issuing[1] included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe).

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of the sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and balance of the SMSA, since the sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total housing units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant housing unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit

records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to 1 of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—Family size	Renter—Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000-$5,999. . .		
$6,000-$9,999. . .		
$10,000-$14,999 .		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to 1 of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to 1 of the 4 vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. However, the housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of the interview, the units at each of the sample special places were listed and subsampled at a rate to produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since January 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit-issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then 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. At the time of the survey, those segments that did not have an expected size of four were further subdivided to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Coverage improvement sample selection.—The Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS-SMSA sample from the permit-issuing and new construction universes. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970, but completed after April 1, 1970.
2. Mobile homes placed in parks either missed during the 1970 census or established since the census.
3. Units missed in the 1970 census.
4. Units converted to residential use that were nonresidential at the time of the 1970 census.
5. Houses that have been moved onto their present site since the 1970 census.
6. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

Coverage improvement for deficiency 1.—A sample of new construction units whose permits were issued before January 1970, but completed after April 1970, was selected independently for each SMSA. The sampling was carried out in two stages for one-and-two-unit structures and in three stages for three-or-more-unit structures. These units were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units).

The first stage was a sample of permit offices and the second stage a sample of the 1969 permits within each of the selected permit offices. In the New York, N.Y., SMSA, Honolulu, Hawaii, SMSA, and Grand Rapids, Mich., SMSA, an additional sample of 1968 permits for three-or-more-unit structures was included in the second stage sampling. For the third stage, structures of size three or more were divided into clusters of an expected size of four units and a sample of clusters was selected. This procedure added an estimated 625 new construction units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiency 2.—In permit-issuing areas, a sample of mobile homes placed in a park missed by the census or established after the census was selected in two stages. First, for each 1976-77 SMSA, a sample of tracts was selected and canvassed. All parks were listed and then matched back to the 1970 census to identify parks missed by the census and parks established after the

census. Second, the parks were divided into clusters of an expected size of four sites and a sample of clusters was selected and interviewed. Each of the sample units represented the same number of units that the regular AHS sample units represented. This procedure added an estimated 395 units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiencies 3-6.—The remaining missed units were sampled by one of two procedures. The first procedure was designed to represent units from the following types of missed structures (structures that had no chance of selection for the AHS):

1. Structures missed in the 1970 census.
2. Structures that were completely nonresidential in the 1970 census but contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and had a utility hookup, or were on the site but not occupied on April 1, 1970, or had no utility hookup but were occupied by persons with no usual residence elsewhere.
4. Houses that had been moved onto their present site since the 1970 census.

Initially, a subsample of regular AHS units was selected at a rate of 1 in 22.2152. Then succeeding structures, in a defined path of travel to the right of the structure containing the sample unit, were listed until eight structures (excluding the sample unit structure) were found that had been eligible to be selected for the AHS. Finally, the intervening structures that did not have a chance of selection in the AHS were identified and units within these structures were interviewed. In cases where the interviewer workload would have been too great, a representative subsample of units within these structures was selected. This procedure added an estimated 4,516 units to the coverage of the housing inventory of this SMSA.

The second procedure was designed to represent missed units from structures represented in the AHS. These missed units were:

1. Units missed in the 1970 census.
2. Units converted to residential use since the 1970 census in structures that contained some residential units in 1970.

First, a subsample of regular AHS units in multi-unit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multi-unit structures selected, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 711 units to the coverage of the housing inventory of this SMSA.

Building loss sample selection.—Some tables in this report show estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS interview). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample previously described. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

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 employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at the time of the interview (i.e., the 1976 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1976 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1976 housing inventory.—The AHS estimates of characteristics of the 1976 housing inventory employed a one-stage ratio estimation procedure. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 138 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units} + \text{Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 50 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of 1 or more of the different strata used in the stratification of the universe as previously

illustrated). In addition, a noninterview factor was computed for one noninterview cell for new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes, and one noninterview cell for sample units from both the nonpermit universe (if applicable) and the coverage improvement universe (if units were not included above).

The following ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding ratio estimation category.

The effect of this ratio estimation procedure was to somewhat reduce the variance due to the variation in sample size of strata used in the sample selection of the permit-issuing universe, thereby reducing the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

1970-1976 lost units.—The AHS estimate of characteristics of the 1970-1976 lost units employed a one-stage ratio estimation procedure which is similar to the ratio estimation procedure described previously. The file of 1970-1976 lost units was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error 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 housing inventory from 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 this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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-SMSA 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 can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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 both the 1970 Census of Population and Housing and the 1976 AHS-SMSA 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 included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census,** and PHC(E)-10, **Accu-**

racy of Data for Selected Housing Characteristics as Measured by Reinterviews.

AHS-SMSA.—For the 1976 AHS-SMSA 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 sample 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 accuracy of the AHS data collected from interviewed households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct 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.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview check. The results of this study are available and are presented in the following Census Bureau memorandum, "Response Error in the Annual Housing Survey as Measured by Reinterviews—Year III (1976-1977) SMSA Sample."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year III SMSA's and not for any specific SMSA):

1. For attitudinal items which were not reconciled (i.e., after the question is answered in the reinterview, the enumerator does not present the previous response and then ask the respondent to decide upon the best answer), approximately 67 percent of the indices of inconsistency showed moderate levels of response disagreement while the remaining 33 percent showed high levels.
2. Some differences beyond those due to sampling error did occur. Most of the categories affected by bias were categories of attitudinal items.

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low; indices from 20—50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

The results of this study were based on sample data so 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 this study.

Coverage errors.—With respect to errors of coverage and estimation for missing data, the AHS new construction sample had deficiencies in the representation of conventional new construction. Due to time constraints, only those building permits issued more than 5 months before the survey ended were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. Also, permits were sampled on a monthly basis and were divided equally among the 12 panels in which interviews were conducted. Due to this procedure, some of the permits issued in November 1975 through October 1976 were not interviewed because they were assigned to panels in which the interviewing had already been completed. In this SMSA, 5.0 percent of the permits sampled were not interviewed because of this procedure. This percentage does not include permits issued during the last 5 months of the survey. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units since, due to the relatively short time-span involved, it is possible that construction of these units was not completed at the time the survey was conducted and they were not eligible for interview. In addition to these deficiencies, new construction in special places that do not require building permits, such as military bases, are also not adequately represented.

The Coverage Improvement Program also had certain deficiencies. It appears that the listing procedure used to find units missed in the 1970 census, units converted from nonresidential to residential, houses moved onto their present site, and mobile homes placed outside parks, was not very efficient for finding nonresidential conversions. Such conversions were primarily in business districts, whereas the listing procedure started from a residential unit.

Deficiencies also exist in ED's where area sampling methods are used. It had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing.

Rounding errors.—For 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 or small medians, when these figures are derived from relatively large bases (e.g., median number of persons per household). 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 the survey.

Sampling errors for the AHS-SMSA sample.—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 schedules, instructions, and interviewers were used, estimates from each of the different

samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variance in the estimates due to response and interviewer errors (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 error measured by the standard error, 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 in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from 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 following tables are approximations to the standard errors of various estimates shown in this report for this SMSA. 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 precise standard errors for any specific items.

Table I presents the standard errors applicable to estimates of characteristics of the 1976 housing inventory as well as estimates of characteristics of the 1970-1976 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in this table. The standard errors on the AHS estimates of the population in housing units shown in tables A-1, B-1, and C-1 of part A of this report are 6,420 for the total SMSA, 4,000 for the central city of the SMSA, and 5,610 for the balance of the SMSA.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1976 Housing Inventory and for Estimated Number of 1970-1976 Lost Units for the Allentown-Bethlehem-Easton, Pa.-N.J., SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central city	Not in central city
0	50	50	60
100	70	70	80
200	100	100	110
500	160	160	170
700	190	180	200
1,000 ...	230	220	240
2,500 ...	370	350	380
5,000 ...	510	480	530
10,000 ..	720	670	740
25,000 ..	1,110	970	1,130
50,000 ..	1,490	1,160	1,510
75,000 ..	1,730	1,090	1,720
100,000 .	1,870	—	1,840
150,000 .	1,960	—	—
200,000 .	1,790	—	—

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.

Table II presents the standard errors of estimated percentages of the 1976 housing inventory as well as estimated percentages of the 1970-1976 lost units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in table II.

Included in tables I and II are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered as 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.

For ratios, 100 (x/y), where x is not a subclass of y, table II underestimates 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.—Table A-1 of part A of this

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the Allentown-Bethlehem-Easton, Pa.-N.J., SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100	35.1	35.1	35.1	35.1	35.1	36.7
200	21.3	21.3	21.3	21.3	22.5	26.0
500	9.7	9.7	9.7	9.9	14.2	16.4
700	7.2	7.2	7.2	8.3	12.0	13.9
1,000	5.1	5.1	5.1	7.0	10.1	11.6
2,500	2.1	2.1	3.2	4.4	6.4	7.3
5,000	1.1	1.1	2.3	3.1	4.5	5.2
10,000	.5	.7	1.6	2.2	3.2	3.7
25,000	.2	.5	1.0	1.4	2.0	2.3
50,000	.11	.3	.7	1.0	1.4	1.6
75,000	.07	.3	.6	.8	1.2	1.3
100,000	.05	.2	.5	.7	1.0	1.2
150,000	.04	.2	.4	.6	.8	.9
200,000	.03	.2	.4	.5	.7	.8

[1] Standard errors are presented to nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to nearest one-hundredth of one percentage point.

report shows that in 1976 there were 137,900 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 1,940. The following interpolation procedure was used:

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

Size of estimate	Standard error
100,000	1,870
137,900	x
150,000	1,960

The entry for x is determined by vertically interpolating between 1,870 and 1,960.

$$137{,}900-100{,}000 = 37{,}900$$
$$150{,}000-100{,}000 = 50{,}000$$

$$1{,}870 + \frac{37{,}900}{50{,}000}(1{,}960-1{,}870) = 1{,}940$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 135,960 to 139,840 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 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 134,800 to 141,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 134,020 to 141,780 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 137,900 owner-occupied housing units, 29,100, or 21.1 percent, had two bedrooms. Interpolation in table II of the appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 21.1 percent is approximately 0.8 percentage points. The following interpolation procedure was used:

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

Base of percentage	Estimated percentage		
	10 or 90	21.1	25 or 75
100,000	.7	a	1.0
137,900		p	
150,000	.6	b	.8

1. The entry for cell "a" is determined by horizontal interpolation between 0.7 and 1.0.

$$21.1-10.0 = 11.1$$
$$25.0-10.0 = 15.0$$

$$0.7 + \frac{11.1}{15.0}(1.0-0.7) = .92$$

2. The entry for cell "b" is determined by horizontal interpolation between .6 and .8.

$$21.1-10.0 = 11.1$$
$$25.0-10.0 = 15.0$$

$$0.6 + \frac{11.1}{15.0}(0.8-0.6) = .75$$

3. The entry for "p" was then determined by vertical interpolation between .75 and .92.

$$137{,}900-100{,}000 = 37{,}900$$
$$150{,}000-100{,}000 = 50{,}000$$

$$.92-\frac{37{,}900}{50{,}000}(.92-.75) = 0.8$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 20.3 to 21.9 percent; the 90-percent confidence interval is from 19.8 to 22.4 percent; and the 95-percent confidence interval is from 19.5 to 22.7 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 error of each estimate

considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; but, 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 part A of this report shows that in 1976 there were 74,900 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 45,800. Table I shows the standard error of 29,100 is approximately 1,170 and the standard error of 74,900 is approximately 1,730. Therefore, the standard error of the estimated difference of 45,800 is about:

$$2{,}090 = \sqrt{(1{,}170)^2 + (1{,}730)^2}$$

Consequently, the 68-percent confidence interval for the 45,800 difference is from 43,710 to 47,890 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 42,460 to 49,140 housing units, and the 95-percent confidence interval is from 41,620 to 49,980. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval 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 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 table II 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 of a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 2.8. The base of the distribution from which this median was determined is 137,900 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 137,900 is approximately 1.0 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 48.0 and 52.0.
3. From table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 62,200 owner-occupied housing units, or 45.1 percent, had one or two persons (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 26,400 owner-occupied housing units, or 19.1 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:

$$2.5 + (3.5-2.5)\left(\frac{48.0-45.1}{19.1}\right) = 2.7$$

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

$$2.5 + (3.5-2.5)\left(\frac{52.0-45.1}{19.1}\right) = 2.9$$

Thus, the 95-percent confidence interval ranges from 2.7 to 2.9 persons.

Baltimore 1976

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Annual Housing Survey App-42
Selection of the sample App-42
Coverage improvement sample selection App-43
Coverage improvement for deficiency 1 . . . . . App-43
Coverage improvement for deficiency 2 . . . . . App-43
Coverage improvement for deficiencies 3-6 . . . . . App-44
Building loss sample selection . . . App-44
1970 Census of Population and Housing App-44
ESTIMATION App-44
1976 housing inventory App-44
1970-1976 lost units App-45
Ratio estimation procedure of the 1970 Census of Population and Housing App-45
RELIABILITY OF THE ESTIMATES App-45
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-45
Coverage errors App-46
Rounding errors App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences App-49
Illustration of the computation of the standard error of a difference App-49
Medians App-49
Illustration of the computation of the 95-percent confidence interval of a median . . App-49

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 20 SMSA's are based on data collected from the 1976 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. In each of the 20 SMSA's, the data were collected for the 12-month period from April 1976 through March 1977 with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and data for a second group of 21 SMSA's were collected from April 1975 through March 1976. The sample housing units for each group of the AHS SMSA's are interviewed on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the third group (1976-77) are: Houston, Tex., New York, N.Y., St. Louis, Mo.-Ill., and Seattle-Everett, Wash.

The remaining SMSA's in the third group are: Allentown-Bethlehem-Easton, Pa.-N.J., Baltimore, Md., Birmingham, Ala., Buffalo, N.Y., Cleveland, Ohio, Denver, Colo., Grand Rapids, Mich., Honolulu, Hawaii, Indianapolis, Ind., Las Vegas, Nev., Louisville, Ky.-Ind., Oklahoma City, Okla., Omaha, Nebr.-Iowa, Providence-Pawtucket-Warwick, R.I.-Mass., Raleigh, N.C., and Sacramento, Calif.

In this SMSA, 4,813 units were eligible for interview. Of these sample units, 196 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to the units eligible for interview, 452 units were visited but were not eligible for interview, because they were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit-issuing (Honolulu, Las Vegas, New York, and Sacramento) was selected from two sample frames—units enumerated in the 1970 Census of Population and Housing in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for the 16 SMSA's which are *not* 100-percent permit-issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe).

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of the sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA, since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and balance of the SMSA, since the sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total housing units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant housing unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit

records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to 1 of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—Family size	Renter—Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000-$5,999. . .		
$6,000-$9,999. . .		
$10,000-$14,999 .		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to 1 of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to 1 of the 4 vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. However, the housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of the interview, the units at each of the sample special places were listed and subsampled at a rate to produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since January 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit-issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then 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. At the time of the survey, those segments that did not have an expected size of four were further subdivided to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Coverage improvement sample selection.—The Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS-SMSA sample from the permit-issuing and new construction universes. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970, but completed after April 1, 1970.
2. Mobile homes placed in parks either missed during the 1970 census or established since the census.
3. Units missed in the 1970 census.
4. Units converted to residential use that were nonresidential at the time of the 1970 census.
5. Houses that have been moved onto their present site since the 1970 census.
6. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

Coverage improvement for deficiency 1.—A sample of new construction units whose permits were issued before January 1970, but completed after April 1970, was selected independently for each SMSA. The sampling was carried out in two stages for one-and-two-unit structures and in three stages for three-or-more-unit structures. These units were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units).

The first stage was a sample of permit offices and the second stage a sample of the 1969 permits within each of the selected permit offices. In the New York, N.Y., SMSA, Honolulu, Hawaii, SMSA, and Grand Rapids, Mich., SMSA, an additional sample of 1968 permits for three-or-more-unit structures was included in the second stage sampling. For the third stage, structures of size three or more were divided into clusters of an expected size of four units and a sample of clusters was selected. This procedure added an estimated 1,767 new construction units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiency 2.—In permit-issuing areas, a sample of mobile homes placed in a park missed by the census or established after the census was selected in two stages. First, for each 1976-77 SMSA, a sample of tracts was selected and canvassed. All parks were listed and then matched back to the 1970 census to identify parks missed by the census and parks established after the census. Second, the parks were divided into clusters of an expected size of four

sites and a sample of clusters was selected and interviewed. Each of the sample units represented tha same number of units that the regular AHS sample units represented. This procedure added an estimated 1,960 units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiencies 3-6.—The remaining missed units were sampled by one of two procedures. The first procedure was designed to represent units from the following types of missed structures (structures that had no chance of selection for the AHS):

1. Structures missed in the 1970 census.
2. Structures that were completely nonresidential in the 1970 census but contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and had a utility hookup, or were on the site but not occupied on April 1, 1970, or had no utility hookup but were occupied by persons with no usual residence elsewhere.
4. Houses that had been moved onto their present site since the 1970 census.

Initially, a subsample of regular AHS units was selected at a rate of 1 in 22.2152. Then succeeding structures, in a defined path of travel to the right of the structure containing the sample unit, were listed until eight structures (excluding the sample unit structure) were found that had been eligible to be selected for the AHS. Finally, the intervening structures that did not have a chance of selection in the AHS were identified and units within these structures were interviewed. In cases where the interviewer workload would have been too great, a representative subsample of units within these structures was selected. This procedure added an estimated 9,244 units to the coverage of the housing inventory of this SMSA.

The second procedure was designed to represent missed units from structures represented in the AHS. These missed units were:

1. Units missed in the 1970 census.
2. Units converted to residential use since the 1970 census in structures that contained some residential units in 1970.

First, a subsample of regular AHS units in multi-unit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multi-unit structures selected, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 2,060 units to the coverage of the housing inventory of this SMSA.

Building loss sample selection.—Some tables in this report show estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS interview). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample previously described. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

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 employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at the time of the interview (i.e., the 1976 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1976 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1976 housing inventory.—The AHS estimates of characteristics of the 1976 housing inventory employed a one-stage ratio estimation procedure. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 196 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units} + \text{Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 50 noninterview cells for sample housing units from the permit-issuing universe, (where the cells consisted of 1 or more of the different strata used in the stratification of the universe as previously illustrated). In addition, a noninterview factor was computed for one noninterview cell for new construction sample housing units from both the permit-issuing universe and the coverage

improvement universe, one noninterview cell for mobile homes, and one noninterview cell for sample units from both the nonpermit universe (if applicable) and the coverage improvement universe (if units were not included above).

The following ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding ratio estimation category.

The effect of this ratio estimation procedure was to somewhat reduce the variance due to the variation in sample size of strata used in the sample selection of the permit-issuing universe, thereby reducing the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

1970-1976 lost units.—The AHS estimate of characteristics of the 1970-1976 lost units employed a one-stage ratio estimation procedure which is similar to the ratio estimation procedure described previously. The file of 1970-1976 lost units was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error 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 housing inventory from 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 this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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-SMSA 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 can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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 both the 1970 Census of Population and Housing and the 1976 AHS-SMSA 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 included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series 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.**

AHS-SMSA.—For the 1976 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A rein-

terview program was conducted for a sample 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 accuracy of the AHS data collected from interviewed households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct 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.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview check. The results of this study are available and are presented in the following Census Bureau memorandum, "Response Error in the Annual Housing Survey as Measured by Reinterviews—Year III (1976-1977) SMSA Sample."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year III SMSA's and not for any specific SMSA):

1. For attitudinal items which were not reconciled (i.e., after the question is answered in the reinterview, the enumerator does not present the previous response and then ask the respondent to decide upon the best answer), approximately 67 percent of the indices of inconsistency showed moderate levels of response disagreement while the remaining 33 percent showed high levels.
2. Some differences beyond those due to sampling error did occur. Most of the categories affected by bias were categories of attitudinal items.

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low, indices from 20—50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

The results of this study were based on sample data so 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 this study.

Coverage errors.—With respect to errors of coverage and estimation for missing data, the AHS new construction sample had deficiencies in the representation of conventional new construction. Due to time constraints, only those building permits issued more than five months before the survey ended were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. Also, permits were sampled on a monthly basis and were divided equally among the 12 panels in which interviews were conducted. Due to this procedure, some of the permits issued in November 1975 through October 1976 were not interviewed because they were assigned to panels in which the interviewing had already been completed. In this SMSA, 6.2 percent of the permits sampled were not interviewed because of this procedure. This percentage does not include permits issued during the last 5 months of the survey. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units since, due to the relatively short time span involved, it is possible that construction of these units was not completed at the time the survey was conducted and they were not eligible for interview. In addition to these deficiencies, new construction in special places that do not require building permits, such as military bases, are also not adequately represented.

The Coverage Improvement Program also had certain deficiencies. It appears that the listing procedure used to find units missed in the 1970 census, units converted from nonresidential to residential, houses moved onto their present site, and mobile homes placed outside parks, was not very efficient for finding nonresidential conversions. Such conversions were primarily in business districts, whereas the listing procedure started from a residential unit.

Deficiencies also exist in ED's where area sampling methods are used. It had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing.

Rounding errors.—For 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 or small medians, when these figures are derived from relatively large bases (e.g., median number of persons per household). 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 the survey.

Sampling errors for the AHS-SMSA sample.—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 schedules, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and thus, is a measure of the precision

with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variance in the estimates due to response and interviewer errors (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 error measured by the standard error, 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 in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from 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 following tables are approximations to the standard errors of various estimates shown in this report for this SMSA. 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 precise standard errors for any specific items.

Table I presents the standard errors applicable to estimates of characteristics of the 1976 housing inventory as well as estimates of characteristics of the 1970-1976 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in this table. The standard errors on the AHS estimates of the population in housing units shown in tables A-1, B-1, and C-1 of part A of this report are 21,480 for the total SMSA, 14,200 for the central city of the SMSA, and 16,810 for the balance of the SMSA.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1976 Housing Inventory and for Estimated Number of 1970-1976 Lost Units for the Baltimore, Md., SMSA, for the Central City, and for the Balance of the SMSA.

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central city	Not in central city
0	190	170	190
100	190	170	190
200	190	180	190
500	310	290	310
700	360	350	360
1,000	430	410	430
2,500	690	650	680
5,000	970	920	960
10,000	1,370	1,290	1,350
25,000	2,140	1,990	2,120
50,000	2,990	2,700	2,930
75,000	3,610	3,170	3,520
100,000	4,100	3,490	3,980
150,000	4,870	3,830	4,650
200,000	5,440	3,840	5,100
250,000	5,870	3,510	5,380
300,000	6,190	—	5,520
400,000	6,550	—	5,420
500,000	6,590	—	—
600,000	6,310	—	—
700,000	5,670	—	—

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.

Table II presents the standard errors of estimated percentages of the 1976 housing inventory as well as estimated percentages of the 1970-1976 lost units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in table II.

Included in tables I and II are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered as 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.

For ratios, 100 (x/y), where x is not a subclass of y, table II underestimates 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.—Table A-1 of part A of this report shows that in 1976 there were 413,200 owner-occupied housing units in

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the Baltimore, Md., SMSA, for the Central City, and for the Balance of the SMSA.

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100	65.4	65.4	65.4	65.4	65.4	68.7
200	48.6	48.6	48.6	48.6	48.6	48.6
500	27.4	27.4	27.4	27.4	27.4	30.7
700	21.2	21.2	21.2	21.2	22.5	26.0
1,000	15.9	15.9	15.9	15.9	18.8	21.7
2,500	7.0	7.0	7.0	8.2	11.9	13.7
5,000	3.6	3.6	4.2	5.8	8.4	9.7
10,000	1.9	1.9	3.0	4.1	5.9	6.9
25,000	.7	.9	1.9	2.6	3.8	4.3
50,000	.4	.6	1.3	1.8	2.7	3.1
75,000	.3	.5	1.1	1.5	2.2	2.5
100,000	.2	.4	.9	1.3	1.9	2.2
150,000	.13	.4	.8	1.1	1.5	1.8
200,000	.09	.3	.7	.9	1.3	1.5
250,000	.08	.3	.6	.8	1.2	1.4
300,000	.06	.2	.5	.8	1.1	1.3
400,000	.05	.2	.5	.7	.9	1.1
500,000	.04	.2	.4	.6	.8	1.0
600,000	.03	.2	.4	.5	.8	.9
700,000	.03	.2	.4	.5	.7	.8

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point.

this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 6,560. The following interpolation procedure was used:

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

Size of estimate	Standard error
400,000	6,550
413,200	x
500,000	6,590

The entry for x is determined by vertically interpolating between 6,550 and 6,590.

$$413{,}200-400{,}000 = 13{,}200$$
$$500{,}000-400{,}000 = 100{,}000$$

$$6{,}550 + \frac{13{,}200}{100{,}000}(6{,}590-6{,}550) = 6{,}560$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 406,640 to 419,760 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 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 402,700 to 423,700 housing units with 90 percent confidence; and that the average estimate lies within the interval from 400,080 to 426,320 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 413,200 owner-occupied housing units, 88,400, or 21.4 percent, had two bedrooms. Interpolation in table II of the appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 21.4 percent is approximately 0.8 percentage points. The following interpolation procedure was used:

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

Base of percentage	Estimated percentage		
	10 or 90	21.4	25 or 75
400,000 .	.7	a	.9
413,200 .		p	
500,000 .	.6	b	.8

1. The entry for cell "a" is determined by horizontal interpolation between 0.7 and 0.9.

$$21.4-10.0 = 11.4$$
$$25.0-10.0 = 15.0$$

$$0.7 + \frac{11.4}{15.0}(0.9-0.7) = .85$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.6 and 0.8.

$$21.4-10.0 = 11.4$$
$$25.0-10.0 = 15.0$$

$$0.6 + \frac{11.4}{15.0}(0.8-0.6) = .75$$

3. The entry for "p" was then determined by vertical interpolation between .75 and .85.

$$413{,}200-400{,}000 = 13{,}200$$
$$500{,}000-400{,}000 = 100{,}000$$

$$.85 - \frac{13{,}200}{100{,}000}(.85-.75) = 0.8$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 20.6 to 22.2 percent; the 90-percent confidence interval is from 20.1 to 22.7 percent; and the 95-percent confidence interval is from 19.8 to 23.0 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 error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; but, 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 part A of this report shows that in 1976 there were 232,500 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 144,100. Table I shows the standard error of 88,400 is approximately 3,870 and the standard error of 232,500 is approximately 5,720. Therefore, the standard error of the estimated difference of 144,100 is about:

$$6{,}910 = \sqrt{(3{,}870)^2 + (5{,}720)^2}$$

Consequently, the 68-percent confidence interval for the 144,100 difference is from 137,190 to 151,010 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 133,040 to 155,160 housing units, and the 95-percent confidence interval is from 130,280 to 157,920. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval 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 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 table II 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 of a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 3.0. The base of the distribution from which this median was determined is 413,200 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 413,200 is approximately 1.1 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 47.8 and 52.2.
3. From table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 166,000 owner-occupied housing units, or 40.2 percent, had one or two persons (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 83,600 owner-occupied housing units, or 20.2 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:

$$2.5 + (3.5-2.5)\left(\frac{47.8-40.2}{20.2}\right) = 2.9$$

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

$$2.5 + (3.5-2.5)\left(\frac{52.2-40.2}{20.2}\right) = 3.1$$

Thus, the 95-percent confidence interval ranges from 2.9 to 3.1 persons.

Birmingham 1976

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Annual Housing Survey App-42
Selection of the sample App-42
Coverage improvement sample selection App-43
Coverage improvement for deficiency 1 ..... App-43
Coverage improvement for deficiency 2 ..... App-43
Coverage improvement for deficiencies 3-6 ..... App-44
Building loss sample selection ... App-44
1970 Census of Population and Housing App-44
ESTIMATION App-44
1976 housing inventory App-44
1970-1976 lost units App-45
Ratio estimation procedure of the 1970 Census of Population and Housing App-45
RELIABILITY OF THE ESTIMATES App-45
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-46
Coverage errors App-46
Rounding errors App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-48
Differences App-49
Illustration of the computation of the standard error of a difference App-49
Medians App-49
Illustration of the computation of the 95-percent confidence interval of a median .. App-49

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 20 SMSA's are based on data collected from the 1976 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. In each of the 20 SMSA's, the data were collected for the 12-month period from April 1976 through March 1977 with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and data for a second group of 21 SMSA's were collected from April 1975 through March 1976. The sample housing units for each group of the AHS SMSA's are interviewed on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the third group (1976-77) are: Houston, Tex., New York, N.Y., St. Louis, Mo.-Ill., and Seattle-Everett, Wash.

The remaining SMSA's in the third group are: Allentown-Bethlehem-Easton, Pa.-N.J., Baltimore, Md., Birmingham, Ala., Buffalo, N.Y., Cleveland, Ohio, Denver, Colo., Grand Rapids, Mich., Honolulu, Hawaii, Indianapolis, Ind., Las Vegas, Nev., Louisville, Ky.-Ind., Oklahoma City, Okla., Omaha, Nebr.-Iowa, Providence-Pawtucket-Warwick, R.I.-Mass., Raleigh, N.C., and Sacramento, Calif.

In this SMSA, 4,697 units were eligible for interview. Of these sample units, 161 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to the units eligible for interview, 526 units were visited but were not eligible for interview, because they were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit-issuing (Honolulu, Las Vegas, New York, and Sacramento) was selected from two sample frames—units enumerated in the 1970 Census of Population and Housing in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for the 16 SMSA's which are *not* 100-percent permit-issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe).

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of the sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA, since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and balance of the SMSA, since the sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total housing units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant housing unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit

records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to 1 of 50 strata according to its tenure; family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—Family size	Renter—Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000-$5,999. . .		
$6,000-$9,999. . .		
$10,000-$14,999 .		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to 1 of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to 1 of the 4 vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. However, the housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of the interview, the units at each of the sample special places were listed and subsampled at a rate to produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since January 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit-issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then 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. At the time of the survey, those segments that did not have an expected size of four were further subdivided to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Coverage improvement sample selection.—The Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS-SMSA sample from the permit-issuing and new construction universes. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970, but completed after April 1, 1970.
2. Mobile homes placed in parks either missed during the 1970 census or established since the census.
3. Units missed in the 1970 census.
4. Units converted to residential use that were nonresidential at the time of the 1970 census.
5. Houses that have been moved onto their present site since the 1970 census.
6. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

Coverage improvement for deficiency 1.—A sample of new construction units whose permits were issued before January 1970, but completed after April 1970, was selected independently for each SMSA. The sampling was carried out in two stages for one-and-two-unit structures and in three stages for three-or-more-unit structures. These units were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units).

The first stage was a sample of permit offices and the second stage a sample of the 1969 permits within each of the selected permit offices. In the New York, N.Y., SMSA, Honolulu, Hawaii, SMSA, and Grand Rapids, Mich., SMSA, an additional sample of 1968 permits for three-or-more-unit structures was included in the second stage sampling. For the third stage, structures of size three or more were divided into clusters of an expected size of four units and a sample of clusters was selected. This procedure yielded no additional units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiency 2.—In permit-issuing areas, a sample of mobile homes placed in a park missed by the census or established after the census was selected in two stages. First, for each 1976-77 SMSA, a sample of tracts was selected and canvassed. All parks were listed and then matched back to the 1970 census to identify parks missed by the census and parks established after the census. Second, the parks were divided

into clusters of an expected size of four sites and a sample of clusters was selected and interviewed. Each of the sample units represented the same number of units that the regular AHS sample units represented. This procedure added an estimated 438 units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiencies 3-6.—The remaining missed units were sampled by one of two procedures. The first procedure was designed to represent units from the following types of missed structures (structures that had no chance of selection for the AHS):

1. Structures missed in the 1970 census.
2. Structures that were completely nonresidential in the 1970 census but contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and had a utility hookup, or were on the site but not occupied on April 1, 1970, or had no utility hookup but were occupied by persons with no usual residence elsewhere.
4. Houses that had been moved onto their present site since the 1970 census.

Initially, a subsample of regular AHS units was selected at a rate of 1 in 22.2152. Then succeeding structures, in a defined path of travel to the right of the structure containing the sample unit, were listed until eight structures (excluding the sample unit structure) were found that had been eligible to be selected for the AHS. Finally, the intervening structures that did not have a chance of selection in the AHS were identified and units within these structures were interviewed. In cases where the interviewer workload would have been too great, a representative subsample of units within these structures was selected. This procedure added an estimated 7,990 units to the coverage of the housing inventory of this SMSA.

The second procedure was designed to represent missed units from structures represented in the AHS. These missed units were:

1. Units missed in the 1970 census.
2. Units converted to residential use since the 1970 census in structures that contained some residential units in 1970.

First, a subsample of regular AHS units in multi-unit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multi-unit structures selected, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 658 units to the coverage of the housing inventory of this SMSA.

Building loss sample selection.—Some tables in this report show estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS interview). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample previously described. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

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 employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at the time of the interview (i.e., the 1976 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1976 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1976 housing inventory.—The AHS estimates of characteristics of the 1976 housing inventory employed a one-stage ratio estimation procedure. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 161 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units} + \text{Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 50 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of 1 or more of the different strata used

in the stratification of the universe as previously illustrated). In addition, a noninterview factor was computed for one noninterview cell for new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes, and one noninterview cell for sample units from both the nonpermit universe (if applicable) and the coverage improvement universe (if units were not included above).

The following ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding ratio estimation category.

The effect of this ratio estimation procedure was to somewhat reduce the variance due to the variation in sample size of strata used in the sample selection of the permit-issuing universe, thereby reducing the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units, already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

1970-1976 lost units.—The AHS estimate of characteristics of the 1970-1976 lost units employed a one-stage ratio estimation procedure which is similar to the ratio estimation procedure described previously. The file of 1970-1976 lost units was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error 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 housing inventory from 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 this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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-SMSA 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 can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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 both the 1970 Census of Population and Housing and the 1976 AHS-SMSA 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 included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and

Research Program Reports, Series 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.**

AHS-SMSA.—For the 1976 AHS-SMSA 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 sample 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 accuracy of the AHS data collected from interviewed households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct 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.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview check. The results of this study are available and are presented in the following Census Bureau memorandum, "Response Error in the Annual Housing Survey as Measured by Reinterviews—Year III (1976-1977) SMSA Sample."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year III SMSA's and not for any specific SMSA):

1. For attitudinal items which were not reconciled (i.e., after the question is answered in the reinterview, the enumerator does not present the previous response and then ask the respondent to decide upon the best answer), approximately 67 percent of the indices of inconsistency showed moderate levels of response disagreement while the remaining 33 percent showed high levels.
2. Some differences beyond those due to sampling error did occur. Most of the categories affected by bias were categories of attitudinal items.

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low; indices from 20—50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

The results of this study were based on sample data so 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 this study.

Coverage errors.—With respect to errors of coverage and estimation for missing data, the AHS new construction sample had deficiencies in the representation of conventional new construction. Due to time constraints, only those building permits issued more than 5 months before the survey ended were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. Also, permits were sampled on a monthly basis and were divided equally among the 12 panels in which interviews were conducted. Due to this procedure, some of the permits issued in November 1975 through October 1976 were not interviewed because they were assigned to panels in which the interviewing had already been completed. In this SMSA, 5.0 percent of the permits sampled were not interviewed because of this procedure. This percentage does not include permits issued during the last 5 months of the survey. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units since, due to the relatively short time-span involved, it is possible that construction of these units was not completed at the time the survey was conducted and they were not eligible for interview. In addition to these deficiencies, new construction in special places that do not require building permits, such as military bases, are also not adequately represented.

The Coverage Improvement Program also had certain deficiencies. It appears that the listing procedure used to find units missed in the 1970 census, units converted from nonresidential to residential, houses moved onto their present site, and mobile homes placed outside parks, was not very efficient for finding nonresidential conversions. Such conversions were primarily in business districts, whereas the listing procedure started from a residential unit.

Deficiencies also exist in ED's where area sampling methods are used. It had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing.

Rounding errors.—For 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 or small medians, when these figures are derived from relatively large bases (e.g., median number of persons per household). 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 the survey.

Sampling errors for the AHS-SMSA sample.—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 schedules, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variance in the estimates due to response and interviewer errors (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 error measured by the standard error, 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 in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from 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 following tables are approximations to the standard errors of various estimates shown in this report for this SMSA. 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 precise standard errors for any specific items.

Table I presents the standard errors applicable to estimates of characteristics of the 1976 housing inventory as well as estimates of characteristics of the 1970-1976 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in this table. The standard errors on the AHS estimates of the population in housing units shown in tables A-1, B-1, and C-1 of part A of this report are 9,470 for the total SMSA, 8,870 for the central city of the SMSA, and 6,690 for the balance of the SMSA.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1976 Housing Inventory and for Estimated Number of 1970-1976 Lost Units for the Birmingham, Ala., SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central city	Not in central city
0	70	70	70
100.	80	80	80
200.	120	110	110
500.	190	180	180
700.	220	210	210
1,000 . . .	260	260	260
2,500 . . .	420	400	400
5,000 . . .	590	560	570
10,000 . .	830	790	810
25,000 . .	1,290	1,190	1,270
50,000 . .	1,770	1,540	1,780
75,000 . .	2,110	1,690	2,170
100,000. .	2,360	1,710	2,480
150,000. .	2,710	–	3,000
200,000. .	2,900	–	–
250,000. .	2,970	–	–

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.

Table II presents the standard errors of estimated percentages of the 1976 housing inventory as well as estimated percentages of the 1970-1976 lost units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in table II.

Included in tables I and II are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered as 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.

For ratios, 100 (x/y), where x is not a subclass of y, table II underestimates 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

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the Birmingham, Ala., SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100.	41.0	41.0	41.0	41.0	41.0	41.7
200.	25.8	25.8	25.8	25.8	25.8	29.5
500.	12.2	12.2	12.2	12.2	16.2	18.6
700.	9.0	9.0	9.0	9.5	13.7	15.8
1,000	6.5	6.5	6.5	7.9	11.4	13.2
2,500	2.7	2.7	3.6	5.0	7.2	8.3
5,000	1.4	1.4	2.6	3.5	5.1	5.9
10,000	0.7	0.8	1.8	2.5	3.6	4.2
25,000	0.3	0.5	1.1	1.6	2.3	2.6
50,000	0.14	0.4	0.8	1.1	1.6	1.9
75,000	0.09	0.3	0.7	0.9	1.3	1.5
100,000. . . .	0.07	0.3	0.6	0.8	1.1	1.3
150,000. . . .	0.05	0.2	0.5	0.6	0.9	1.1
200,000. . . .	0.03	0.2	0.4	0.6	0.8	0.9
250,000. . . .	0.03	0.2	0.4	0.5	0.7	0.8

[1] Standard errors are presented to nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to nearest one-hundredth of one percentage point.

Illustration of the use of the standard error tables.—Table A-1 of part A of this report shows that in 1976 there were 173,400 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 2,800. The following interpolation procedure was used:

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

Size of estimate	Standard error
150,000.	2,710
173,400.	x
200,000.	2,900

The entry for x is determined by vertically interpolating between 2,710 and 2,900.

$$173{,}400-150{,}000 = 23{,}400$$
$$200{,}000-150{,}000 = 50{,}000$$

$$2{,}710 + \frac{23{,}400}{50{,}000}(2{,}900-2{,}710) = 2{,}800$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 170,600 to 176,200 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 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 168,920 to 177,880 housing units with 90 percent confidence; and that the average estimate lies within the interval from 167,800 to 179,000 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 173,400 owner-occupied housing units, 54,900, or 31.7 percent, had two bedrooms. Interpolation in table II of the appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 31.7 percent is approximately 0.9 percentage points. The following interpolation procedure was used:

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

Base of percentage	Estimated percentage		
	25 or 75	31.7	50
150,000. . . .	0.9	a	1.1
173,400. . . .		p	
200,000. . . .	0.8	b	0.9

1. The entry for cell "a" is determined by horizontal interpolation between 0.9 and 1.1.

$$31.7-25.0 = 6.7$$
$$50.0-25.0 = 25.0$$

$$0.9 + \frac{6.7}{25.0}(1.1-0.9) = .95$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.8 and 0.9.

$$31.7-25.0 = 6.7$$
$$50.0-25.0 = 25.0$$

$$0.8 + \frac{6.7}{25.0}(0.9-0.8) = .83$$

3. The entry for "p" was then determined by vertical interpolation between .83 and .95.

$$173{,}400-150{,}000 = 23{,}400$$
$$200{,}000-150{,}000 = 50{,}000$$

$$.95 - \frac{23{,}400}{50{,}000}(.95-.83) = 0.9$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 30.8 to 32.6 percent; the 90-percent confidence interval is from 30.3 to 33.1 percent; and the 95-percent confidence interval is from 29.9 to 33.5 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 error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; but, 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 part A of this report shows that in 1976 there were 92,100 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 37,200. Table I shows the standard error of 54,900 is approximately 1,840 and the standard error of 92,100 is approximately 2,280. Therefore, the standard error of the estimated difference of 37,200 is about:

$$2{,}930 = \sqrt{(1{,}840)^2 + (2{,}280)^2}$$

Consequently, the 68-percent confidence interval for the 37,200 difference is from 34,270 to 40,130 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 32,510 to 41,890 housing units, and the 95-percent confidence interval is from 31,340 to 43,060. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval 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 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 table II 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 of a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 2.7. The base of the distribution from which this median was determined is 173,400 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 173,400 is approximately 1.0 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 48.0 and 52.0.
3. From table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 80,800 owner-occupied housing units, or 46.6 percent, had one or two persons (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 34,800 owner-occupied housing units, or 20.1 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:

$$2.5 + (3.5-2.5)\left(\frac{48.0-46.6}{20.1}\right) = 2.6$$

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

$$2.5 + (3.5-2.5)\left(\frac{52.0-46.6}{20.1}\right) = 2.8$$

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

Buffalo 1976

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Annual Housing Survey App-42
Selection of the sample App-42
Coverage improvement sample selection App-43
Coverage improvement for deficiency 1 ..... App-43
Coverage improvement for deficiency 2 ..... App-43
Coverage improvement for deficiencies 3-6 ..... App-44
Building loss sample selection ... App-44
1970 Census of Population and Housing App-44
ESTIMATION App-44
1976 housing inventory App-44
1970-1976 lost units App-45
Ratio estimation procedure of the 1970 Census of Population and Housing App-45
RELIABILITY OF THE ESTIMATES App-45
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-46
Coverage errors App-46
Rounding errors App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-48
Differences App-49
Illustration of the computation of the standard error of a difference App-49
Medians App-49
Illustration of the computation of the 95-percent confidence interval of a median .. App-49

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 20 SMSA's are based on data collected from the 1976 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. In each of the 20 SMSA's, the data were collected for the 12-month period from April 1976 through March 1977 with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and data for a second group of 21 SMSA's were collected from April 1975 through March 1976. The sample housing units for each group of the AHS SMSA's are interviewed on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the third group (1976-77) are: Houston, Tex., New York, N.Y., St. Louis, Mo.-Ill., and Seattle-Everett, Wash.

The remaining SMSA's in the third group are: Allentown-Bethlehem-Easton, Pa.-N.J., Baltimore, Md., Birmingham, Ala., Buffalo, N.Y., Cleveland, Ohio, Denver, Colo., Grand Rapids, Mich., Honolulu, Hawaii, Indianapolis, Ind., Las Vegas, Nev., Louisville, Ky.-Ind., Oklahoma City, Okla., Omaha, Nebr.-Iowa, Providence-Pawtucket-Warwick, R.I.-Mass., Raleigh, N.C., and Sacramento, Calif.

In this SMSA, 4,928 units were eligible for interview. Of these sample units, 287 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to the units eligible for interview, 354 units were visited but were not eligible for interview, because they were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit-issuing (Honolulu, Las Vegas, New York, and Sacramento) was selected from two sample frames—units enumerated in the 1970 Census of Population and Housing in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for the 16 SMSA's which are *not* 100-percent permit-issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe).

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of the sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA, since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and balance of the SMSA, since the sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total housing units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant housing unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four cate-

gories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to 1 of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—Family size	Renter—Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000-$5,999. . .		
$6,000-$9,999. . .		
$10,000-$14,999 .		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to 1 of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to 1 of the 4 vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. However, the housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of the interview, the units at each of the sample special places were listed and subsampled at a rate to produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since January 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit-issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then 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. At the time of the survey, those segments that did not have an expected size of four were further subdivided to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Coverage improvement sample selection.—The Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS-SMSA sample from the permit-issuing and new construction universes. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970, but completed after April 1, 1970.
2. Mobile homes placed in parks either missed during the 1970 census or established since the census.
3. Units missed in the 1970 census.
4. Units converted to residential use that were nonresidential at the time of the 1970 census.
5. Houses that have been moved onto their present site since the 1970 census.
6. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

Coverage improvement for deficiency 1.—A sample of new construction units whose permits were issued before January 1970, but completed after April 1970, was selected independently for each SMSA. The sampling was carried out in two stages for one-and-two-unit structures and in three stages for three-or-more-unit structures. These units were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units).

The first stage was a sample of permit offices and the second stage a sample of the 1969 permits within each of the selected permit offices. In the New York, N.Y., SMSA, Honolulu, Hawaii, SMSA, and Grand Rapids, Mich., SMSA, an additional sample of 1968 permits for three-or-more-unit structures was included in the second stage sampling. For the third stage, structures of size three or more were divided into clusters of an expected size of four units and a sample of clusters was selected. This procedure added an estimated 743 units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiency 2.—In permit-issuing areas, a sample of mobile homes placed in a park missed by the census or established after the census was selected in two stages. First, for each 1976-77 SMSA, a sample of tracts was selected and canvassed. All parks were listed and then matched back to the 1970 census to identify parks missed by the census and parks established after the census. Second, the parks were divided into clusters of an expected size of four sites and a sample of clusters was selected

and interviewed. Each of the sample units represented the same number of units that the regular AHS sample units represented. This procedure added an estimated 579 units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiencies 3-6.—The remaining missed units were sampled by one of two procedures. The first procedure was designed to represent units from the following types of missed structures (structures that had no chance of selection for the AHS):

1. Structures missed in the 1970 census.
2. Structures that were completely nonresidential in the 1970 census but contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and had a utility hookup, or were on the site but not occupied on April 1, 1970, or had no utility hookup but were occupied by persons with no usual residence elsewhere.
4. Houses that had been moved onto their present site since the 1970 census.

Initially, a subsample of regular AHS units was selected at a rate of 1 in 22.2152. Then, succeeding structures, in a defined path of travel to the right of the structure containing the sample unit, were listed until eight structures (excluding the sample unit structure) were found that had been eligible to be selected for the AHS. Finally, the intervening structures that did not have a chance of selection in the AHS were identified and units within these structures were interviewed. In cases where the interviewer workload would have been too great, a representative subsample of units within these structures was selected. This procedure added an estimated 5,167 units to the coverage of the housing inventory of this SMSA.

The second procedure was designed to represent missed units from structures represented in the AHS. These missed units were:

1. Units missed in the 1970 census.
2. Units converted to residential use since the 1970 census in structures that contained some residential units in 1970.

First, a subsample of regular AHS units in multi-unit structures of less than ten units was selected from the permit-issuing universe. Second, for the multi-unit structures selected, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 1,942 units to the coverage of the housing inventory of this SMSA.

Building loss sample selection.—Some tables in this report show estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS interview). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample previously described. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

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 employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at the time of the interview (i.e., the 1976 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1976 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1976 housing inventory.—The AHS estimates of characteristics of the 1976 housing inventory employed a one-stage ratio estimation procedure. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 287 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units} + \text{Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 50 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of 1 or more of the different strata used

in the stratification of the universe as previously illustrated). In addition, a noninterview factor was computed for one noninterview cell for new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes, and one noninterview cell for sample units from both the nonpermit universe (if applicable) and the coverage improvement universe (if units were not included above).

The following ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding ratio estimation category.

The effect of this ratio estimation procedure was to somewhat reduce the variance due to the variation in sample size of strata used in the sample selection of the permit-issuing universe, thereby reducing the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

1970-1976 lost units.—The AHS estimate of characteristics of the 1970-1976 lost units employed a one-stage ratio estimation procedure which is similar to the ratio estimation procedure described previously. The file of 1970-1976 lost units was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error 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 housing inventory from 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 this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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-SMSA 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 can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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 both the 1970 Census of Population and Housing and the 1976 AHS-SMSA 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 included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series 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.

AHS-SMSA.—For the 1976 AHS-SMSA 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 sample 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 accuracy of the AHS data collected from interviewed households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct 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.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview check. The results of this study are available and are presented in the following Census Bureau memorandum, "Response Error in the Annual Housing Survey as Measured by Reinterviews—Year III (1976-77) SMSA Sample."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year III SMSA's and not for any specific SMSA):

1. For attitudinal items which were not reconciled (i.e., after the question is answered in the reinterview, the enumerator does not present the previous response and then ask the respondent to decide upon the best answer), approximately 67 percent of the indices of inconsistency showed moderate levels of response disagreement while the remaining 33 percent showed high levels.
2. Some differences beyond those due to sampling error did occur. Most of the categories affected by bias were categories of attitudinal items.

The range for evaluating inconsistency is from 0–100. The rule of thumb is that indices below 20 are low, indices from 20–50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

The results of this study were based on sample data so 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 this study.

Coverage errors.—With respect to errors of coverage and estimation for missing data, the AHS new construction sample had deficiencies in the representation of conventional new construction. Due to time constraints, only those building permits issued more than 5 months before the survey ended were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. Also, permits were sampled on a monthly basis and were divided equally among the 12 panels in which interviews were conducted. Due to this procedure, some of the permits issued in November 1975 through October 1976 were not interviewed because they were assigned to panels in which the interviewing had already been completed. In this SMSA, 4.6 percent of the permits sampled were not interviewed because of this procedure. This percentage does not include permits issued during the last 5 months of the survey. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units, since, due to the relatively short time span involved, it is possible that construction of these units was not completed at the time the survey was conducted and they were not eligible for interview. In addition to these deficiencies, new construction in special places that do not require building permits, such as military bases, are also not adequately represented.

The Coverage Improvement Program also had certain deficiencies. It appears that the listing procedure used to find units missed in the 1970 census, units converted from nonresidential to residential, houses moved onto their present site, and mobile homes placed outside parks, was not very efficient for finding nonresidential conversions. Such conversions were primarily in business districts, whereas the listing procedure started from a residential unit.

Deficiencies also exist in ED's where area sampling methods are used. It had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing.

Rounding errors.—For 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 or small medians, when these figures are derived from relatively large bases (e.g., median number of persons per household). 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 the survey.

Sampling errors for the AHS-SMSA sample.—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 schedules, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from

the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variance in the estimates due to response and interviewer errors (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 error measured by the standard error, 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 in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from 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 following tables are approximations to the standard errors of various estimates shown in this report for this SMSA. 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 precise standard errors for any specific items.

Table I presents the standard errors applicable to estimates of characteristics of the 1976 housing inventory as well as estimates of characteristics of the 1970-1976 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in this table. The standard errors on the AHS estimates of the population in housing units shown in tables A-1, B-1, and C-1 of part A of this report are 12,600 for the total SMSA, 7,490 for the central city of the SMSA, and 11,140 for the balance of the SMSA.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1976 Housing Inventory and for Estimated Number of 1970-1976 Lost Units for the Buffalo, N.Y., SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central city	Not in central city
0	110	100	110
100.	110	100	110
200.	150	140	150
500.	230	220	240
700.	280	260	280
1,000 . . .	330	310	330
2,500 . . .	520	490	530
5,000 . . .	740	690	740
10,000 . .	1,040	970	1,040
25,000 . .	1,620	1,460	1,610
50,000 . .	2,230	1,880	2,200
75,000 . .	2,660	2,070	2,580
100,000. .	2,990	2,080	2,850
150,000. .	3,430	1,520	3,160
200,000. .	3,690	–	3,210
250,000. .	3,790	–	3,020
300,000. .	3,760	–	–
400,000. .	3,220	–	–

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.

Table II presents the standard errors of estimated percentages of the 1976 housing inventory as well as estimated percentages of the 1970-1976 lost units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in table II.

Included in tables I and II are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered as 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.

For ratios, 100 (x/y), where x is not a subclass of y, table II underestimates 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

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the Buffalo, N.Y., SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100.	52.4	52.4	52.4	52.4	52.4	52.5
200.	35.5	35.5	35.5	35.5	35.5	37.1
500.	18.1	18.1	18.1	18.1	20.3	23.5
700.	13.6	13.6	13.6	13.6	17.2	19.8
1,000	9.9	9.9	9.9	10.0	14.4	16.6
2,500	4.2	4.2	4.6	6.3	9.1	10.5
5,000	2.2	2.2	3.2	4.5	6.4	7.4
10,000	1.1	1.1	2.3	3.1	4.5	5.2
25,000	.4	.7	1.4	2.0	2.9	3.3
50,000	.2	.5	1.0	1.4	2.0	2.3
75,000	.15	.4	.8	1.1	1.7	1.9
100,000. . . .	.11	.3	.7	1.0	1.4	1.7
150,000. . . .	.07	.3	.6	.8	1.2	1.4
200,000. . . .	.06	.2	.5	.7	1.0	1.2
250,000. . . .	.04	.2	.5	.6	.9	1.0
300,000. . . .	.04	.2	.4	.6	.8	1.0
400,000. . . .	.03	.2	.4	.5	.7	.8

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point.

Illustration of the use of the standard error tables.—Table A-1 of part A of this report shows that in 1976 there were 275,900 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 3,770. The following interpolation procedure was used:

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

Size of estimate	Standard error
250,000.	3,790
275,900.	x
300,000.	3,760

The entry for x is determined by vertically interpolating between 3,790 and 3,760.

$$275{,}900-250{,}000 = 25{,}900$$
$$300{,}000-250{,}000 = 50{,}000$$

$$3{,}790-\frac{25{,}900}{50{,}000}(3{,}790-3{,}760) = 3{,}770$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 272,130 to 279,670 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 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 269,870 to 281,930 housing units with 90 percent confidence; and that the average estimate lies within the interval from 268,360 to 283,440 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 275,900 owner-occupied housing units, 56,800, or 20.6 percent, had two bedrooms. Interpolation in table II of the appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 20.6 percent is approximately 0.8 percentage points. The following interpolation procedure was used:

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

Base of percentage	Estimated percentage		
	10 or 90	20.6	25 or 75
250,000. . . .	.6	a	.9
275,900. . . .		p	
300,000. . . .	.6	b	.8

1. The entry for cell "a" is determined by horizontal interpolation between 0.6 and 0.9.

$$20.6-10.0 = 10.6$$
$$25.0-10.0 = 15.0$$

$$0.6+\frac{10.6}{15.0}(0.9-0.6) = .81$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.6 and 0.8.

$$20.6-10.0 = 10.6$$
$$25.0-10.0 = 15.0$$

$$0.6+\frac{10.6}{15.0}(0.8-0.6) = .74$$

3. The entry for "p" was then determined by vertical interpolation between .74 and .81.

$$275{,}900-250{,}000 = 25{,}900$$
$$300{,}000-250{,}000 = 50{,}000$$

$$.81-\frac{25{,}900}{50{,}000}(.81-.74) = 0.8$$

Consequently, the 68-percent confidence interval, as shown by these data, is from

19.8 to 21.4 percent; the 90-percent confidence interval is from 19.3 to 21.9 percent; and the 95-percent confidence interval is from 19.0 to 22.2 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 error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; but, 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 part A of this report shows that in 1976 there were 149,100 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 92,300. Table I shows the standard error of 56,800 is approximately 2,350 and the standard error of 149,100 is approximately 3,420. Therefore, the standard error of the estimated difference of 92,300 is about:

$$4{,}150 = \sqrt{(2{,}350)^2 + (3{,}420)^2}$$

Consequently, the 68-percent confidence interval for the 92,300 difference is from 88,150 to 96,450 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 85,660 to 98,940 housing units, and the 95-percent confidence interval is from 84,000 to 100,600. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval 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 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 table II 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 of a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 2.9. The base of the distribution from which this median was determined is 275,900 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 275,900 is approximately 1.0 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 48.0 and 52.0.
3. From table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 119,200 owner-occupied housing units, or 43.2 percent, had one or two persons (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 46,600 owner-occupied housing units, or 16.9 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:

$$2.5 + (3.5-2.5)\left(\frac{48.0-43.2}{16.9}\right) = 2.8$$

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

$$2.5 + (3.5-2.5)\left(\frac{52.0-43.2}{16.9}\right) = 3.0$$

Thus, the 95-percent confidence interval ranges from 2.8 to 3.0 persons.

Cleveland 1976

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Annual Housing Survey App-42
Selection of the sample App-42
Coverage improvement sample selection App-43
Coverage improvement for deficiency 1 ..... App-43
Coverage improvement for deficiency 2 ..... App-43
Coverage improvement for deficiencies 3-6 ..... App-44
Building loss sample selection ... App-44
1970 Census of Population and Housing App-44
ESTIMATION App-44
1976 housing inventory App-44
1970-1976 lost units App-45
Ratio estimation procedure of the 1970 Census of Population and Housing App-45
RELIABILITY OF THE ESTIMATES App-45
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-46
Coverage errors App-46
Rounding errors App-46
Sampling errors for the AHS-SMSA sample App-47
Illustration of the use of the standard error tables App-48
Differences App-49
Illustration of the computation of the standard error of a difference App-49
Medians App-49
Illustration of the computation of the 95-percent confidence interval of a median .. App-49

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 20 SMSA's are based on data collected from the 1976 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. In each of the 20 SMSA's, the data were collected for the 12-month period from April 1976 through March 1977 with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and data for a second group of 21 SMSA's were collected from April 1975 through March 1976. The sample housing units for each group of the AHS SMSA's are interviewed on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the third group (1976-77) are: Houston, Tex., New York, N.Y., St. Louis, Mo.-Ill., and Seattle-Everett, Wash.

The remaining SMSA's in the third group are: Allentown-Bethlehem-Easton, Pa.-N.J., Baltimore, Md., Birmingham, Ala., Buffalo, N.Y., Cleveland, Ohio, Denver, Colo., Grand Rapids, Mich., Honolulu, Hawaii, Indianapolis, Ind., Las Vegas, Nev., Louisville, Ky.-Ind., Oklahoma City, Okla., Omaha, Nebr.-Iowa, Providence-Pawtucket-Warwick, R.I.-Mass., Raleigh, N.C., and Sacramento, Calif.

In this SMSA, 4,908 units were eligible for interview. Of these sample units, 279 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to the units eligible for interview, 319 units were visited but were not eligible for interview, because they were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit-issuing (Honolulu, Las Vegas, New York, and Sacramento) was selected from two sample frames—units enumerated in the 1970 Census of Population and Housing in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for the 16 SMSA's which are *not* 100-percent permit-issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe).

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of the sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and balance of the SMSA, since the sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total housing units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant housing unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units.

The occupied housing unit records were further stratified so that each unit was assigned to 1 of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—Family size	Renter—Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000-$5,999. . .		
$6,000-$9,999. . .		
$10,000-$14,999 .		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to 1 of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to 1 of the 4 vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. However, the housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of the interview, the units at each of the sample special places were listed and subsampled at a rate to produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since January 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit-issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then 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. At the time of the survey, those segments that did not have an expected size of four were further subdivided to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Coverage improvement sample selection.—The Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS-SMSA sample from the permit-issuing and new construction universes. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970, but completed after April 1, 1970.
2. Mobile homes placed in parks either missed during the 1970 census or established since the census.
3. Units missed in the 1970 census.
4. Units converted to residential use that were nonresidential at the time of the 1970 census.
5. Houses that have been moved onto their present site since the 1970 census.
6. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

Coverage improvement for deficiency 1.—A sample of new construction units whose permits were issued before January 1970, but completed after April 1970, was selected independently for each SMSA. The sampling was carried out in two stages for one-and-two-unit structures and in three stages for three-or-more-unit structures. These units were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units).

The first stage was a sample of permit offices and the second stage a sample of the 1969 permits within each of the selected permit offices. In the New York, N.Y., SMSA, Honolulu, Hawaii, SMSA, and Grand Rapids, Mich., SMSA, an additional sample of 1968 permits for three-or-more-unit structures was included in the second stage sampling. For the third stage, structures of size three or more were divided into clusters of an expected size of four units and a sample of clusters was selected. This procedure added an estimated 5,329 new construction units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiency 2.—In permit-issuing areas, a sample of mobile homes placed in a park missed by the census or established after the census was selected in two stages. First, for each 1976-77 SMSA, a sample of tracts was selected and canvassed. All parks were listed and then matched back to the 1970 census to identify parks missed by the

census and parks established after the census. Second, the parks were divided into clusters of an expected size of four sites and a sample of clusters was selected and interviewed. Each of the sample units represented the same number of units that the regular AHS sample units represented. This procedure yielded no additional units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiencies 3-6.—The remaining missed units were sampled by one of two procedures. The first procedure was designed to represent units from the following types of missed structures (structures that had no chance of selection for the AHS):

1. Structures missed in the 1970 census.
2. Structures that were completely nonresidential in the 1970 census but contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and had a utility hookup, or were on the site but not occupied on April 1, 1970, or had no utility hookup but were occupied by persons with no usual residence elsewhere.
4. Houses that had been moved onto their present site since the 1970 census.

Initially, a subsample of regular AHS units was selected at a rate of 1 in 22.2152. Then succeeding structures, in a defined path of travel to the right of the structure containing the sample unit, were listed until eight structures (excluding the sample unit structure) were found that had been eligible to be selected for the AHS. Finally, the intervening structures that did not have a chance of selection in the AHS were identified and units within these structures were interviewed. In cases where the interviewer workload would have been too great, a representative subsample of units within these structures was selected. This procedure added an estimated 10,437 units to the coverage of the housing inventory of this SMSA.

The second procedure was designed to represent missed units from structures represented in the AHS. These missed units were:

1. Units missed in the 1970 census.
2. Units converted to residential use since the 1970 census in structures that contained some residential units in 1970.

First, a subsample of regular AHS units in multi-unit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multi-unit structures selected, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 2,787 units to the coverage of the housing inventory of this SMSA.

Building loss sample selection.—Some tables in this report show estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS interview). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample previously described. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

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 employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at the time of the interview (i.e., the 1976 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1976 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1976 housing inventory.—The AHS estimates of characteristics of the 1976 housing inventory employed a one-stage ratio estimation procedure. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 279 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units} + \text{Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 50 noninterview cells for

sample housing units from the permit-issuing universe (where the cells consisted of 1 or more of the different strata used in the stratification of the universe as previously illustrated). In addition, a noninterview factor was computed for one noninterview cell for new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes, and one noninterview cell for sample units from both the nonpermit universe (if applicable) and the coverage improvement universe (if units were not included above).

The following ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding ratio estimation category.

The effect of this ratio estimation procedure was to somewhat reduce the variance due to the variation in sample size of strata used in the sample selection of the permit-issuing universe, thereby reducing the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

1970-1976 lost units.—The AHS estimate of characteristics of the 1970-1976 lost units employed a one-stage ratio estimation procedure which is similar to the ratio estimation procedure described previously. The file of 1970-1976 lost units was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error 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 housing inventory from 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 this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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-SMSA 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 can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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 both the 1970 Census of Population and Housing and the 1976 AHS-SMSA 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 included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by rein-

terviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series 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.**

AHS-SMSA.—For the 1976 AHS-SMSA 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 sample 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 accuracy of the AHS data collected from interviewed households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct 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.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview check. The results of this study are available and are presented in the following Census Bureau memorandum, "Response Error in the Annual Housing Survey as Measured by Reinterviews—Year III (1976-1977) SMSA Sample."

Some of the results of this study are (note that these results are based on the reinterviews across all Year III SMSA's and not for any specific SMSA):

1. For attitudinal items which were not reconciled (i.e., after the question is answered in the reinterview, the enumerator does not present the previous response and then ask the respondent to decide upon the best answer), approximately 67 percent of the indices of inconsistency showed moderate levels of response disagreement while the remaining 33 percent showed high levels.

2. Some differences beyond those due to sampling error did occur. Most of the categories affected by bias were categories of attitudinal items.

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low; indices from 20—50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

The results of this study were based on sample data, so 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 this study.

Coverage errors.—With respect to errors of coverage and estimation for missing data, the AHS new construction sample had deficiencies in the representation of conventional new construction. Due to time constraints, only those building permits issued more than 5 months before the survey ended were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. Also, permits were sampled on a monthly basis and were divided equally among the 12 panels in which interviews were conducted. Due to this procedure, some of the permits issued in November 1975 through October 1976 were not interviewed because they were assigned to panels in which the interviewing had already been completed. In this SMSA, 7.1 percent of the permits sampled were not interviewed because of this procedure. This percentage does not include permits issued during the last 5 months of the survey. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units since, due to the relatively short time span involved, it is possible that construction of these units was not completed at the time the survey was conducted and they were not eligible for interview. In addition to these deficiencies, new construction in special places that do not require building permits, such as military bases, are also not adequately represented.

The Coverage Improvement Program also had certain deficiencies. It appears that the listing procedure used to find units missed in the 1970 census, units converted from nonresidential to residential, houses moved onto their present site, and mobile homes placed outside parks, was not very efficient for finding nonresidential conversions. Such conversions were primarily in business districts, whereas the listing procedure started from a residential unit.

Deficiencies also exist in ED's where area sampling methods are used. It had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing.

Rounding errors.—For 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 or small medians, when these figures are derived from relatively large bases (e.g., median number of persons per household). 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 the survey.

Sampling errors for the AHS-SMSA sample.—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 schedules, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variance in the estimates due to response and interviewer errors (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 error measured by the standard error, 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 in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from 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 following tables are approximations to the standard errors of various estimates shown in this report for this SMSA. 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 precise standard errors for any specific items.

Table I presents the standard errors applicable to estimates of characteristics of the 1976 housing inventory as well as estimates of characteristics of the 1970-1976 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in this table. The standard errors on the AHS estimates of the population in housing units shown in tables A-1, B-1, and C-1 of part A of this report are 22,000 for the total SMSA, 12,820 for the central city of the SMSA, and 17,470 for the balance of the SMSA.

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.

Table II presents the standard errors of estimated percentages of the 1976 housing inventory as well as estimated percentages of the 1970-1976 lost units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in table II.

Included in tables I and II are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered as 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.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1976 Housing Inventory and for Estimated Number of 1970-1976 Lost Units for the Cleveland, Ohio, SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard error[1]		
	SMSA	In central city	Not in central city
0	180	160	190
100	180	160	190
200	190	180	190
500	300	290	300
700	360	340	360
1,000 ...	430	400	430
2,500 ...	680	640	680
5,000 ...	950	900	960
10,000 ..	1,350	1,260	1,350
25,000 ..	2,120	1,960	2,110
50,000 ..	2,960	2,670	2,930
75,000 ..	3,570	3,150	3,510
100,000 .	4,080	3,490	3,960
150,000 .	4,860	3,890	4,630
200,000 .	5,460	4,000	5,070
250,000 .	5,920	3,830	5,340
300,000 .	6,290	–	5,470
400,000 .	6,770	–	5,340
500,000 .	6,990	–	–
600,000 .	6,950	–	–
700,000 .	6,660	–	–

[1] For estimates pertaining to new construction in the central city of the SMSA, the standard errors shown in the table should be multiplied by a factor of 1.5.

For ratios, 100 (x/y), where x is not a subclass of y, table II underestimates 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.—Table A-1 of part A of this report shows that in 1976 there were 428,700 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 6,830. The following interpolation procedure was used:

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

Size of estimates	Standard error
400,000	6,770
428,700	X
500,000	6,990

The entry for x is determined by vertically interpolating between 6,770 and 6,990.

$$428{,}700-400{,}000 = 28{,}700$$
$$500{,}000-400{,}000 = 100{,}000$$

$$6{,}770+\frac{28{,}700}{100{,}000}(6{,}990-6{,}770) = 6{,}830$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 421,870 to 435,530 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 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 417,770 to 439,630 housing units with 90 percent confidence; and that the average estimate lies within the interval from 415,040 to 442,360 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 428,700 owner-occupied housing units, 91,900, or 21.4 percent, had two bedrooms. Interpolation in table II of the appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 21.4 percent is approximately 0.8 percentage points. The following interpolation procedure was used:

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

Base of percentage	Estimated percentage		
	10 or 90	21.4	25 or 75
400,000	.6	a	.9
428,700		p	
500,000	.6	b	.8

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the Cleveland, Ohio, SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100	64.7	64.7	64.7	64.7	64.7	67.7
200	47.8	47.8	47.8	47.8	47.8	47.9
500	26.8	26.8	26.8	26.8	26.8	30.3
700	20.7	20.7	20.7	20.7	22.2	25.6
1,000	15.5	15.5	15.5	15.5	18.5	21.4
2,500	6.8	6.8	6.8	8.1	11.7	13.5
5,000	3.5	3.5	4.2	5.7	8.3	9.6
10,000	1.8	1.8	3.0	4.1	5.9	6.8
25,000	.7	.9	1.9	2.6	3.7	4.3
50,000	.4	.6	1.3	1.8	2.6	3.0
75,000	.2	.5	1.1	1.5	2.1	2.5
100,000	.2	.4	.9	1.3	1.9	2.1
150,000	.12	.3	.8	1.0	1.5	1.7
200,000	.09	.3	.7	.9	1.3	1.5
250,000	.07	.3	.6	.8	1.2	1.4
300,000	.06	.2	.5	.7	1.1	1.2
400,000	.05	.2	.5	.6	.9	1.1
500,000	.04	.2	.4	.6	.8	1.0
600,000	.03	.2	.4	.5	.8	.9
700,000	.03	.2	.4	.5	.7	.8

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point. For estimates pertaining to new construction in the central city of the SMSA, the standard errors shown in the table should be multiplied by a factor of 1.5.

1. The entry for cell "a" is determined by horizontal interpolation between 0.6 and 0.9.

$$21.4-10.0 = 11.4$$
$$25.0-10.0 = 15.0$$

$$0.6 + \frac{11.4}{15.0}(0.9-0.6) = .83$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.6 and 0.8.

$$21.4-10.0 = 11.4$$
$$25.0-10.0 = 15.0$$

$$0.6 + \frac{11.4}{15.0}(0.8-0.6) = .75$$

3. The entry for "p" was then determined by vertical interpolation between .75 and .83.

$$428{,}700-400{,}000 = 28{,}700$$
$$500{,}000-400{,}000 = 100{,}000$$

$$.83-\frac{28{,}700}{100{,}000}(.83-.75) = 0.8$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 20.6 to 22.2 percent; the 90-percent confidence interval is from 20.1 to 22.7 percent; and the 95-percent confidence interval is from 19.8 to 23.0 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 square of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; but, 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 part A of this report shows that in 1976 there were 233,800 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 141,900. Table I shows the standard error of 91,900 is approximately 3,910 and the standard error of 233,800 is approximately 5,770. Therefore, the standard error of the estimated difference of 141,900 is about:

$$6{,}970 = \sqrt{(3{,}910)^2 + (5{,}770)^2}$$

Consequently, the 68-percent confidence interval for the 141,900 difference is from 134,930 to 148,870 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 130,750 to 153,050 housing units, and the 95-percent confidence interval is from 127,960 to 155,840. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval 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 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 table II 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 of a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 3.0. The base of the distribution from which this median was determined is 428,700 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 428,700 is approximately 1.1 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 47.8 and 52.2.

3. From table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 175,800 owner-occupied housing units, or 41.0 percent, had one or two persons (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 80,300 owner-

occupied housing units, or 18.7 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:

$$2.5 + (3.5-2.5)\left(\frac{47.8-41.0}{18.7}\right) = 2.9$$

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

$$2.5 + (3.5-2.5)\left(\frac{52.2-41.0}{18.7}\right) = 3.1$$

Thus, the 95-percent confidence interval ranges from 2.9 to 3.1 persons.

Denver 1976

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Annual Housing Survey App-42
Selection of the sample App-42
Coverage improvement sample selection App-43
Coverage improvement for deficiency 1 ..... App-43
Coverage improvement for deficiency 2 ..... App-43
Coverage improvement for deficiencies 3-6 ..... App-44
Building loss sample selection ... App-44
1970 Census of Population and Housing App-44
ESTIMATION App-44
1976 housing inventory App-44
1970-1976 lost units App-45
Ratio estimation procedure of the 1970 Census of Population and Housing App-45
RELIABILITY OF THE ESTIMATES App-45
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-46
Coverage errors App-46
Rounding errors App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-48
Differences App-49
Illustration of the computation of the standard error of a difference App-49
Medians App-49
Illustration of the computation of the 95-percent confidence interval of a median .. App-49

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 20 SMSA's are based on data collected from the 1976 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. In each of the 20 SMSA's, the data were collected for the 12-month period from April 1976 through March 1977 with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and data for a second group of 21 SMSA's were collected from April 1975 through March 1976. The sample housing units for each group of the AHS SMSA's are interviewed on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the third group (1976-77) are: Houston, Tex., New York, N.Y., St. Louis, Mo.-Ill., and Seattle-Everett, Wash.

The remaining SMSA's in the third group are: Allentown-Bethlehem-Easton, Pa.-N.J., Baltimore, Md., Birmingham, Ala., Buffalo, N.Y., Cleveland, Ohio, Denver, Colo., Grand Rapids, Mich., Honolulu, Hawaii, Indianapolis, Ind., Las Vegas, Nev., Louisville, Ky.-Ind., Oklahoma City, Okla., Omaha, Nebr.-Iowa, Providence-Pawtucket-Warwick, R.I.-Mass., Raleigh, N.C., and Sacramento, Calif.

In this SMSA, 4,901 units were eligible for interview. Of these sample units, 187 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to the units eligible for interview, 414 units were visited but were not eligible for interview, because they were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit-issuing (Honolulu, Las Vegas, New York, and Sacramento) was selected from two sample frames—units enumerated in the 1970 Census of Population and Housing in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for the 16 SMSA's which are *not* 100-percent permit-issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe).

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of the sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA, since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and balance of the SMSA, since the sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total housing units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant housing unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four cate-

gories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to 1 of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—Family size	Renter—Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000-$5,999. . .		
$6,000-$9,999. . .		
$10,000-$14,999 .		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to 1 of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to 1 of the 4 vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. However, the housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of the interview, the units at each of the sample special places were listed and subsampled at a rate to produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since January 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit-issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then 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. At the time of the survey, those segments that did not have an expected size of four were further subdivided to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Coverage improvement sample selection.—The Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS-SMSA sample from the permit-issuing and new construction universes. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970, but completed after April 1, 1970.
2. Mobile homes placed in parks either missed during the 1970 census or established since the census.
3. Units missed in the 1970 census.
4. Units converted to residential use that were nonresidential at the time of the 1970 census.
5. Houses that have been moved onto their present site since the 1970 census.
6. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

Coverage improvement for deficiency 1.—A sample of new construction units whose permits were issued before January 1970, but completed after April 1970, was selected independently for each SMSA. The sampling was carried out in two stages for one-and-two-unit structures and in three stages for three-or-more-unit structures. These units were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units).

The first stage was a sample of permit offices and the second stage a sample of the 1969 permits within each of the selected permit offices. In the New York, N.Y., SMSA, Honolulu, Hawaii, SMSA, and Grand Rapids, Mich., SMSA, an additional sample of 1968 permits for three-or-more-unit structures was included in the second stage sampling. For the third stage, structures of size three or more were divided into clusters of an expected size of four units and a sample of clusters was selected. This procedure added an estimated 4,120 new construction units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiency 2.—In permit-issuing areas, a sample of mobile homes placed in a park missed by the census or established after the census was selected in two stages. First, for each 1976-77 SMSA, a sample of tracts was selected and canvassed. All parks were listed and then matched back to the 1970 census to identify parks missed by the

census and parks established after the census. Second, the parks were divided into clusters of an expected size of four sites and a sample of clusters was selected and interviewed. Each of the sample units represented the same number of units that the regular AHS sample units represented. This procedure added an estimated 6,125 units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiencies 3-6.—The remaining missed units were sampled by one of two procedures. The first procedure was designed to represent units from the following types of missed structures (structures that had no chance of selection for the AHS):

1. Structures missed in the 1970 census.
2. Structures that were completely nonresidential in the 1970 census but contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and had a utility hookup, or were on the site but not occupied on April 1, 1970, or had no utility hookup but were occupied by persons with no usual residence elsewhere.
4. Houses that had been moved onto their present site since the 1970 census.

Initially, a subsample of regular AHS units was selected at a rate of 1 in 22.2152. Then, succeeding structures, in a defined path of travel to the right of the structure containing the sample unit, were listed until eight structures (excluding the sample unit structure) were found that had been eligible to be selected for the AHS. Finally, the intervening structures that did not have a chance of selection in the AHS were identified and units within these structures were interviewed. In cases where the interviewer workload would have been too great, a representative subsample of units within these structures was selected. This procedure added an estimated 4,621 units to the coverage of the housing inventory of this SMSA.

The second procedure was designed to represent missed units from structures represented in the AHS. These missed units were:

1. Units missed in the 1970 census.
2. Units converted to residential use since the 1970 census in structures that contained some residential units in 1970.

First, a subsample of regular AHS units in multi-unit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multi-unit structures selected, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure yielded no additional units to the coverage of the housing inventory of this SMSA.

Building loss sample selection.—Some tables in this report show estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS interview). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample previously described. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

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 employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at the time of the interview (i.e., the 1976 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1976 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1976 housing inventory.—The AHS estimates of characteristics of the 1976 housing inventory employed a one-stage ratio estimation procedure. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 187 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units} + \text{Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 50 noninterview cells for sample housing units from the permit-issuing

universe (where the cells consisted of 1 or more of the different strata used in the stratification of the universe as previously illustrated). In addition, a noninterview factor was computed for one noninterview cell for new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes, and one noninterview cell for sample units from both the nonpermit universe (if applicable) and the coverage improvement universe (if units were not included above).

The following ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding ratio estimation category.

The effect of this ratio estimation procedure was to somewhat reduce the variance due to the variation in sample size of strata used in the sample selection of the permit-issuing universe, thereby reducing the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

1970-1976 lost units.—The AHS estimate of characteristics of the 1970-1976 lost units employed a one-stage ratio estimation procedure which is similar to the ratio estimation procedure described previously. The file of 1970-1976 lost units was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error 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 housing inventory from 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 this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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-SMSA 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 can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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 both the 1970 Census of Population and Housing and the 1976 AHS-SMSA 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 included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Popula-

tion and Housing Evaluation and Research Program Reports, Series 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.**

AHS-SMSA.—For the 1976 AHS-SMSA 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 sample 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 accuracy of the AHS data collected from interviewed households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct 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.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview check. The results of this study are available and are presented in the following Census Bureau memorandum, "Response Error in the Annual Housing Survey as Measured by Reinterviews—Year III (1976-1977) SMSA Sample."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year III SMSA's and not for any specific SMSA):

1. For attitudinal items which were not reconciled (i.e., after the question is answered in the reinterview, the enumerator does not present the previous response and then ask the respondent to decide upon the best answer), approximately 67 percent of the indices of inconsistency showed moderate levels of response disagreement while the remaining 33 percent showed high levels.

2. Some differences beyond those due to sampling error did occur. Most of the categories affected by bias were categories of attitudinal items.

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low; indices from 20—50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

The results of this study were based on sample data so 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 this study.

Coverage errors.—With respect to errors of coverage and estimation for missing data, the AHS new construction sample had deficiencies in the representation of conventional new construction. Due to time constraints, only those building permits issued more than 5 months before the survey ended were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. Also, permits were sampled on a monthly basis and were divided equally among the 12 panels in which interviews were conducted. Due to this procedure, some of the permits issued in November 1975 through October 1976 were not interviewed because they were assigned to panels in which the interviewing had already been completed. In this SMSA, 4.5 percent of the permits sampled were not interviewed because of this procedure. This percentage does not include permits issued during the last 5 months of the survey. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units, since, due to the relatively short time span involved, it is possible that construction of these units was not completed at the time the survey was conducted and they were not eligible for interview. In addition to these deficiencies, new construction in special places that do not require building permits, such as military bases, are also not adequately represented.

The Coverage Improvement Program also had certain deficiencies. It appears that the listing procedure used to find units missed in the 1970 census, units converted from nonresidential to residential, houses moved onto their present site, and mobile homes placed outside parks, was not very efficient for finding nonresidential conversions. Such conversions were primarily in business districts, whereas the listing procedure started from a residential unit.

Deficiencies also exist in ED's where area sampling methods are used. It had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing.

Rounding errors.—For 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 or small medians, when these figures are derived from relatively large bases (e.g., median number of persons per household). 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 the survey.

Sampling errors for the AHS-SMSA sample.—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 schedules, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variance in the estimates due to response and interviewer errors (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 error measured by the standard error, 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 in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from 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 following tables are approximations to the standard errors of various estimates shown in this report for this SMSA. 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 precise standard errors for any specific items.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1976 Housing Inventory and for Estimated Number of 1970-1976 Lost Units for the Denver, Colo., SMSA, for the Central City of the SMSA, and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central city	Not in central city
0	140	140	140
100	140	140	140
200	170	160	170
500	260	260	270
700	310	310	320
1,000	370	370	380
2,500	590	580	600
5,000	830	820	850
10,000	1,170	1,140	1,190
25,000	1,840	1,760	1,870
50,000	2,550	2,370	2,610
75,000	3,080	2,750	3,150
100,000	3,490	2,990	3,580
150,000	4,130	3,170	4,250
200,000	4,590	2,970	4,740
250,000	4,920	—	5,120
300,000	5,150	—	5,390
400,000	5,350	—	—
500,000	5,230	—	—

Table I presents the standard errors applicable to estimates of characteristics of the 1976 housing inventory as well as estimates of characteristics of the 1970-1976 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in this table. The standard errors on the AHS estimates of the population in housing units shown in tables A-1, B-1, and C-1 of part A of this report are 15,940 for the total SMSA, 9,920 for the central city of the SMSA, and 18,340 for the balance of the SMSA.

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.

Table II presents the standard errors of estimated percentages of the 1976 housing inventory as well as estimated percentages of the 1970-1976 lost units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in table II.

Included in tables I and II are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered as 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.

For ratios, 100 (x/y), where x is not a subclass of y, table II underestimates 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:

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

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the Denver, Colo., SMSA, for the Central City of the SMSA, and for the Balance of the SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100	58.1	58.1	58.1	58.1	58.1	58.9
200	41.0	41.0	41.0	41.0	41.0	41.7
500	21.7	21.7	21.7	21.7	22.8	26.4
700	16.6	16.6	16.6	16.6	19.3	22.3
1,000	12.2	12.2	12.2	12.2	16.1	18.6
2,500	5.3	5.3	5.3	7.1	10.2	11.8
5,000	2.7	2.7	3.6	5.0	7.2	8.3
10,000	1.4	1.4	2.6	3.5	5.1	5.9
25,000	.6	.7	1.6	2.2	3.2	3.7
50,000	.3	.5	1.1	1.6	2.3	2.6
75,000	.2	.4	.9	1.3	1.9	2.2
100,000	.14	.4	.8	1.1	1.6	1.9
150,000	.09	.3	.7	.9	1.3	1.5
200,000	.07	.3	.6	.8	1.1	1.3
250,000	.06	.2	.5	.7	1.0	1.2
300,000	.05	.2	.5	.6	.9	1.1
400,000	.03	.2	.4	.6	.8	.9
500,000	.03	.2	.4	.5	.7	.8

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point.

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 errors.—Table A-1 of part A of this report shows that in 1976 there were 311,200 owner-occupied housing units in the SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 5,170. The following interpolation procedure was used:

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

Size of estimate	Standard error
300,000	5,150
311,200	x
400,000	5,350

The entry for x is determined by vertically interpolating between 5,150 and 5,350.

$$311{,}200-300{,}000 = 11{,}200$$
$$400{,}000-300{,}000 = 100{,}000$$

$$5{,}150 + \frac{11{,}200}{100{,}000}(5{,}350-5{,}150) = 5{,}170$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 306,030 to 316,700 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 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 302,930 to 319,470 housing units with 90 percent confidence; and that the average estimate lies within the interval from 300,860 to 321,540 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 311,200 owner-occupied housing units, 83,800, or 26.9 percent, had two bedrooms. Interpolation in table II of the appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 26.9 percent is approximately 0.9 percentage points. The following interpolation procedure was used:

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

Base of percentage	Estimated percentage		
	25 or 75	26.9	50
300,000	.9	a	1.1
311,200		p	
400,000	.8	b	.9

1. The entry for cell "a" is determined by horizontal interpolation between 0.9 and 1.1

$$26.9-25.0 = 1.9$$
$$50.0-25.0 = 25.0$$

$$0.9 + \frac{1.9}{25.0}(1.1-0.9) = .92$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.8 and 0.9.

$$26.9-25.0 = 1.9$$
$$50.0-25.0 = 25.0$$

$$0.8 + \frac{1.9}{25.0}(0.9-0.8) = .81$$

3. The entry for "p" was then determined by vertical interpolation between .81 and .92.

$$311{,}200-300{,}000 = 11{,}200$$
$$400{,}000-300{,}000 = 100{,}000$$

$$.92-\frac{11{,}200}{100{,}000}(.92-.81) = 0.9$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 26.0 to 27.8 percent; the 90-percent confidence interval is from 25.5 to 28.3 percent; and the 95-percent confidence interval is from 25.1 to 28.7 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 error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; but, 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 part A of this report shows that in 1976 there were 122,300 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 38,500. Table I shows the standard error of 83,800 is approximately 3,220 and the standard error of 122,300 is approximately 3,780. Therefore, the standard error of the estimated difference of 38,500 is about:

$$4{,}970 = \sqrt{(3{,}220)^2 + (3{,}780)^2}$$

Consequently, the 68-percent confidence interval for the difference is from 33,530 to 43,470 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 30,550 to 46,450 housing units, and the 95-percent confidence interval is from 28,560 to 48,440. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval 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 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 table II 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 of a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 2.9. The base of the distribution from which this median was determined is 311,200 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 311,200 is approximately 1.8 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 46.4 and 53.6.
3. From table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 136,500 owner-occupied housing units, or 43.9 percent, had one or two persons (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 52,800 owner-occupied housing units, or 17.0 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:

$$2.5 + (3.5-2.5)\left(\frac{46.4-43.9}{17.0}\right) = 2.6$$

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

$$2.5 + (3.5-2.5)\left(\frac{53.6-43.9}{17.0}\right) = 3.1$$

Thus, the 95-percent confidence interval ranges from 2.6 to 3.1 persons.

Grand Rapids 1976

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Annual Housing Survey App-42
Selection of the sample App-42
Coverage improvement sample selection App-43
Coverage improvement for deficiency 1 ..... App-43
Coverage improvement for deficiency 2 ..... App-43
Coverage improvement for deficiencies 3-6 ..... App-44
Building loss sample selection ... App-44
1970 Census of Population and Housing App-44
ESTIMATION App-44
1976 housing inventory App-44
1970-1976 lost units App-45
Ratio estimation procedure of the 1970 Census of Population and Housing App-45
RELIABILITY OF THE ESTIMATES App-45
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-46
Coverage errors App-46
Rounding errors App-46
Sampling errors for the AHS-SMSA sample App-47
Illustration of the use of the standard error tables App-48
Differences App-49
Illustration of the computation of the standard error of a difference App-49
Medians App-49
Illustration of the computation of the 95-percent confidence interval of a median .. App-49

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 20 SMSA's are based on data collected from the 1976 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. In each of the 20 SMSA's, the data were collected for the 12-month period from April 1976 through March 1977 with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and data for a second group of 21 SMSA's were collected from April 1975 through March 1976. The sample housing units for each group of the AHS SMSA's are interviewed on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the third group (1976-77) are: Houston, Tex., New York, N.Y., St. Louis, Mo.-Ill., and Seattle-Everett, Wash.

The remaining SMSA's in the third group are: Allentown-Bethlehem-Easton, Pa.-N.J., Baltimore, Md., Birmingham, Ala., Buffalo, N.Y., Cleveland, Ohio, Denver, Colo., Grand Rapids, Mich., Honolulu, Hawaii, Indianapolis, Ind., Las Vegas, Nev., Louisville, Ky.-Ind., Oklahoma City, Okla., Omaha, Nebr.-Iowa, Providence-Pawtucket-Warwick, R.I.-Mass., Raleigh, N.C., and Sacramento, Calif.

In this SMSA, 4,730 units were eligible for interview. Of these sample units, 199 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to the units eligible for interview, 382 units were visited but were not eligible for interview, because they were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit-issuing (Honolulu, Las Vegas, New York, and Sacramento) was selected from two sample frames—units enumerated in the 1970 Census of Population and Housing in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for the 16 SMSA's which are *not* 100-percent permit-issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe).

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of the sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and balance of the SMSA, since the sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total housing units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant housing unit records, the occupied housing unit records were

stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to 1 of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—Family size	Renter—Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000-$5,999. . .		
$6,000-$9,999. . .		
$10,000-$14,999 .		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to 1 of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to 1 of the 4 vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. However, the housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of the interview, the units at each of the sample special places were listed and subsampled at a rate to produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since January 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit-issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then 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. At the time of the survey, those segments that did not have an expected size of four were further subdivided to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Coverage improvement sample selection.—The Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS-SMSA sample from the permit-issuing and new construction universes. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970, but completed after April 1, 1970.
2. Mobile homes placed in parks either missed during the 1970 census or established since the census.
3. Units missed in the 1970 census.
4. Units converted to residential use that were nonresidential at the time of the 1970 census.
5. Houses that have been moved onto their present site since the 1970 census.
6. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

Coverage improvement for deficiency 1.—A sample of new construction units whose permits were issued before January 1970, but completed after April 1970, was selected independently for each SMSA. The sampling was carried out in two stages for one-and-two-unit structures and in three stages for three-or-more-unit structures. These units were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units).

The first stage was a sample of permit offices and the second stage a sample of the 1969 permits within each of the selected permit offices. In the New York, N.Y., SMSA, Honolulu, Hawaii, SMSA, and Grand Rapids, Mich., SMSA, an additional sample of 1968 permits for three-or-more-unit structures was included in the second stage sampling. For the third stage, structures of size three or more were divided into clusters of an expected size of four units and a sample of clusters was selected. This procedure added an estimated 153 new construction units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiency 2.—In permit-issuing areas, a sample of mobile homes placed in a park missed by the census or established after the census was selected in two stages. First, for each

1976-77 SMSA, a sample of tracts was selected and canvassed. All parks were listed and then matched back to the 1970 census to identify parks missed by the census and parks established after the census. Second, the parks were divided into clusters of an expected size of four sites and a sample of clusters was selected and interviewed. Each of the sample units represented the same number of units that the regular AHS sample units represented. This procedure added an estimated 620 units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiencies 3-6.—The remaining missed units were sampled by one of two procedures. The first procedure was designed to represent units from the following types of missed structures (structures that had no chance of selection for the AHS):

1. Structures missed in the 1970 census.
2. Structures that were completely nonresidential in the 1970 census but contain units converted to residential use:
3. Mobile homes that had been placed outside parks since the 1970 census and had a utility hookup, or were on the site but not occupied on April 1, 1970, or had no utility hookup but were occupied by persons with no usual residence elsewhere.
4. Houses that had been moved onto their present site since the 1970 census.

Initially, a subsample of regular AHS units was selected at a rate of 1 in 22.2152. Then, succeeding structures, in a defined path of travel to the right of the structure containing the sample unit, were listed until eight structures (excluding the sample unit structure) were found that had been eligible to be selected for the AHS. Finally, the intervening structures that did not have a chance of selection in the AHS were identified and units within these structures were interviewed. In cases where the interviewer workload would have been too great, a representative subsample of units within these structures was selected. This procedure added an estimated 2,158 units to the coverage of the housing inventory of this SMSA.

The second procedure was designed to represent missed units from structures represented in the AHS. These missed units were:

1. Units missed in the 1970 census.
2. Units converted to residential use since the 1970 census in structures that contained some residential units in 1970.

First, a subsample of regular AHS units in multi-unit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multi-unit structures selected, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 171 units to the coverage of the housing inventory of this SMSA.

Building loss sample selection.—Some tables in this report show estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS interview). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample previously described. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

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 employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at the time of the interview (i.e., the 1976 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1976 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1976 housing inventory.—The AHS estimates of characteristics of the 1976 housing inventory employed a one-stage ratio estimation procedure. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 199 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units} + \text{Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 50 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of 1 or more of the different strata used in the stratification of the universe as previously illustrated). In addition, a noninterview factor was computed for one noninterview cell for new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes, and one noninterview cell for sample units from both the nonpermit universe (if applicable) and the coverage improvement universe (if units were not included above).

The following ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding ratio estimation category.

The effect of this ratio estimation procedure was to somewhat reduce the variance due to the variation in sample size of strata used in the sample selection of the permit-issuing universe, thereby reducing the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

1970-1976 lost units.—The AHS estimate of characteristics of the 1970-1976 lost units employed a one-stage ratio estimation procedure which is similar to the ratio estimation procedure described previously. The file of 1970-1976 lost units was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error 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 housing inventory from 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 this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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-SMSA 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 can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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 both the 1970 Census of Population and Housing and the 1976 AHS-SMSA 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 included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by

reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series 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.**

AHS-SMSA.—For the 1976 AHS-SMSA 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 sample 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 accuracy of the AHS data collected from interviewed households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct 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.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview check. The results of this study are available and are presented in the following Census Bureau memorandum, "Response Error in the Annual Housing Survey as Measured by Reinterviews—Year III (1976-1977) SMSA Sample."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year III SMSA's and not for any specific SMSA):

1. For attitudinal items which were not reconciled (i.e., after the question is answered in the reinterview, the enumerator does not present the previous response and then ask the respondent to decide upon the best answer), approximately 67 percent of the indices of inconsistency showed moderate levels of response disagreement while the remaining 33 percent showed high levels.
2. Some differences beyond those due to sampling error did occur. Most of the categories affected by bias were categories of attitudinal items.

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low; indices from 20—50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

The results of this study were based on sample data so 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 this study.

Coverage errors.—With respect to errors of coverage and estimation for missing data, the AHS new construction sample had deficiencies in the representation of conventional new construction. Due to time constraints, only those building permits issued more than 5 months before the survey ended were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. Also, permits were sampled on a monthly basis and were divided equally among the 12 panels in which interviews were conducted. Due to this procedure, some of the permits issued in November 1975 through October 1976 were not interviewed because they were assigned to panels in which the interviewing had already been completed. In this SMSA, 6.6 percent of the permits sampled were not interviewed because of this procedure. This percentage does not include permits issued during the last 5 months of the survey. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units since, due to the relatively short time-span involved, it is possible that construction of these units was not completed at the time the survey was conducted and they were not eligible for interview. In addition to these deficiencies, new construction in special places that do not require building permits, such as military bases, are also not adequately represented.

The Coverage Improvement Program also had certain deficiencies. It appears that the listing procedure used to find units missed in the 1970 census, units converted from nonresidential to residential, houses moved onto their present site, and mobile homes placed outside parks, was not very efficient for finding nonresidential conversions. Such conversions were primarily in business districts, whereas the listing procedure started from a residential unit.

Deficiencies also exist in ED's where area sampling methods are used. It had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing.

Rounding errors.—For 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 or small medians, when these figures are derived from relatively large bases (e.g., median number of persons per household). 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 the survey.

Sampling errors for the AHS-SMSA sample.—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 schedules, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variance in the estimates due to response and interviewer errors (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 error measured by the standard error, 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 in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from 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 following tables are approximations to the standard errors of various estimates shown in this report for this SMSA. 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 precise standard errors for any specific items.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1976 Housing Inventory and for Estimated Number of 1970-1976 Lost Units for the Grand Rapids, Mich., SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central city	Not in central city
0	50	40	50
100	70	70	70
200	100	90	100
500	150	150	160
700	180	180	190
1,000 . . .	220	210	220
2,500 . . .	340	330	350
5,000 . . .	480	460	490
10,000 . .	680	630	690
25,000 . .	1,040	880	1,050
50,000 . .	1,380	930	1,380
75,000 . .	1,580	–	1,560
100,000 .	1,690	–	1,630
150,000 .	1,680	–	–

Table I presents the standard errors applicable to estimates of characteristics of the 1976 housing inventory as well as estimates of characteristics of the 1970-1976 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in this table. The standard errors on the AHS estimates of the population in housing units shown in tables A-1, B-1, and C-1 of part A of this report are 5,500 for the total SMSA, 2,300 for the central city of the SMSA, and 4,570 for the balance of the SMSA.

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.

Table II presents the standard errors of estimated percentages of the 1976 housing inventory as well as estimated percentages of the 1970-1976 lost units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in table II.

Included in tables I and II are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered as 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.

For ratios, 100 (x/y), where x is not a subclass of y, table II underestimates 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:

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

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the Grand Rapids, Mich., SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100	33.4	33.4	33.4	33.4	33.4	35.4
200	20.0	20.0	20.0	20.0	21.7	25.0
500	9.1	9.1	9.1	9.5	13.7	15.8
700	6.7	6.7	6.7	8.0	11.6	13.4
1,000	4.8	4.8	4.9	6.7	9.7	11.2
2,500	2.0	2.0	3.1	4.2	6.1	7.1
5,000	1.0	1.0	2.2	3.0	4.3	5.0
10,000	0.5	0.7	1.5	2.1	3.1	3.5
25,000	0.2	0.4	1.0	1.3	1.9	2.2
50,000	0.1	0.3	0.7	1.0	1.4	1.6
75,000	0.07	0.3	0.6	0.8	1.1	1.3
100,000 . . .	0.05	0.2	0.5	0.7	1.0	1.1
150,000 . . .	0.03	0.2	0.4	0.5	0.8	0.9
200,000 . . .	0.03	0.2	0.3	0.5	0.7	0.8
250,000 . . .	0.02	0.14	0.3	0.4	0.6	0.7
300,000 . . .	0.02	0.13	0.3	0.4	0.6	0.6
400,000 . . .	0.01	0.11	0.2	0.3	0.5	0.6
500,000 . . .	0.01	0.10	0.2	0.3	0.4	0.5

[1] Standard errors are presented to nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to nearest one-hundredth of one percentage point.

Let: 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.—Table A-1 of part A of this report shows that in 1976 there were 136,100 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 1,680. The following interpolation procedure was used:

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

Size of estimate	Standard error
100,000	1,690
136,100	x
150,000	1,680

The entry for x is determined by vertically interpolating between 1,680 and 1,690

$$136{,}100-100{,}000 = 36{,}100$$
$$150{,}000-100{,}000 = 50{,}000$$

$$1{,}690-\frac{36{,}100}{50{,}000}(1{,}690-1{,}680) = 1{,}680$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 134,420 to 137,780 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 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 133,410 to 138,790 housing units with 90 percent confidence; and that the average estimates lies within the interval from 132,740 to 139,460 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 136,100 owner-occupied housing units, 31,500, or 23.1 percent, had two bedrooms. Interpolation in table II of the appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 23.1 percent is approximately 0.8 percentage points. The following interpolation procedure was used:

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

Base of percentage	Estimated percentage		
	10 or 90	23.1	25 or 75
100,000 .	0.7	a	1.0
136,100 .		p	
150,000 .	0.5	b	0.8

1. The entry for cell "a" is determined by horizontal interpolation between 0.7 and 1.0.

$$23.1-10.0 = 13.1$$
$$25.0-10.0 = 15.0$$

$$0.7 + \frac{13.1}{15.0}(1.0-0.7) = .96$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.5 and 0.8.

$$23.1-10.0 = 13.1$$
$$25.0-10.0 = 15.0$$

$$0.5 + \frac{13.1}{15.0}(0.8-0.5) = .76$$

3. The entry for "p" was then determined by vertical interpolation between .76 and .96.

$$136{,}100-100{,}000 = 36{,}100$$
$$150{,}000-100{,}000 = 50{,}000$$
$$.96-\frac{36{,}100}{50{,}000}(.96-.76) = 0.8$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 22.3 to 23.9 percent; the 90-percent confidence interval is from 21.8 to 24.4 percent; and the 95-percent confidence interval is from 21.5 to 24.7 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 error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; but, 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 part A of this report shows that in 1976 there were 71,300 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 39,800. Table I shows the standard error of 31,500 is approximately 1,130 and the standard error of 71,300 is approximately 1,550. Therefore, the standard error of the estimated difference of 39,800 is about:

$$1{,}920 = \sqrt{(1{,}130)^2 + (1{,}550)^2}$$

Consequently, the 68-percent confidence interval for the 39,800 difference is from 37,880 to 41,720 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 36,730 to 42,870 housing units, and the 95-percent confidence interval is from 35,960 to 43,640. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval 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 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 table II 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 of a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 3.0. The base of the distribution from which this median was determined is 136,100 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 136,100 is approximately 1.0 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 48.0 and 52.0.
3. From table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 56,400 owner-occupied housing units, or 41.4 percent, had one or two persons (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 23,300 owner-occupied housing units, or 17.1 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:

$$2.5 + (3.5-2.5)\left(\frac{48.0-41.4}{17.1}\right) = 2.9$$

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

$$2.5 + (3.5-2.5)\left(\frac{52.0-41.4}{17.1}\right) = 3.1$$

Thus, the 95-percent confidence interval ranges from 2.9 to 3.1 persons.

Honolulu 1976

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Annual Housing Survey App-42
Selection of the sample App-42
Coverage improvement sample selection App-43
Coverage improvement for deficiency 1 ..... App-43
Coverage improvement for deficiency 2 ..... App-43
Coverage improvement for deficiencies 3-6 ..... App-44
Building loss sample selection ... App-44
1970 Census of Population and Housing App-44
ESTIMATION App-44
1976 housing inventory App-44
1970-1976 lost units App-45
Ratio estimation procedure of the 1970 Census of Population and Housing App-45
RELIABILITY OF THE ESTIMATES App-45
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-46
Coverage errors App-46
Rounding errors App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-48
Differences App-49
Illustration of the computation of the standard error of a difference App-49
Medians App-49
Illustration of the computation of the 95-percent confidence interval of a median .. App-49

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 20 SMSA's are based on data collected from the 1976 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. In each of the 20 SMSA's, the data were collected for the 12-month period from April 1976 through March 1977 with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and data for a second group of 21 SMSA's were collected from April 1975 through March 1976. The sample housing units for each group of the AHS SMSA's are interviewed on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the third group (1976-77) are: Houston, Tex., New York, N.Y., St. Louis, Mo.-Ill., and Seattle-Everett, Wash.

The remaining SMSA's in the third group are: Allentown-Bethlehem-Easton, Pa.-N.J., Baltimore, Md., Birmingham, Ala., Buffalo, N.Y., Cleveland, Ohio, Denver, Colo., Grand Rapids, Mich., Honolulu, Hawaii, Indianapolis, Ind., Las Vegas, Nev., Louisville, Ky.-Ind., Oklahoma City, Okla., Omaha, Nebr.-Iowa, Providence-Pawtucket-Warwick, R.I.-Mass., Raleigh, N.C., and Sacramento, Calif.

In this SMSA, 4,492 units were eligible for interview. Of these sample units, 225 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to the units eligible for interview, 457 units were visited but were not eligible for interview, because they were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit-issuing (Honolulu, Las Vegas, New York, and Sacramento) was selected from two sample frames—units enumerated in the 1970 Census of Population and Housing in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for the 16 SMSA's which are *not* 100-percent permit-issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe).

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of the sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA, since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and balance of the SMSA, since the sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total housing units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant housing unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four cate-

gories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to 1 of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—Family size	Renter—Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000-$5,999. . .		
$6,000-$9,999. . .		
$10,000-$14,999 .		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to 1 of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to 1 of the 4 vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. However, the housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of the interview, the units at each of the sample special places were listed and subsampled at a rate to produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since January 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit-issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then 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. At the time of the survey, those segments that did not have an expected size of four were further subdivided to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Coverage improvement sample selection.—The Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS-SMSA sample from the permit-issuing and new construction universes. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970, but completed after April 1, 1970.
2. Mobile homes placed in parks either missed during the 1970 census or established since the census.
3. Units missed in the 1970 census.
4. Units converted to residential use that were nonresidential at the time of the 1970 census.
5. Houses that have been moved onto their present site since the 1970 census.
6. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

Coverage improvement for deficiency 1.—A sample of new construction units whose permits were issued before January 1970, but completed after April 1970, was selected independently for each SMSA. The sampling was carried out in two stages for one-and-two-unit structures and in three stages for three-or-more-unit structures. These units were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units).

The first stage was a sample of permit offices and the second stage a sample of the 1969 permits within each of the selected permit offices. In the New York, N.Y., SMSA, Honolulu, Hawaii, SMSA, and Grand Rapids, Mich., SMSA, an additional sample of 1968 permits for three-or-more-unit structures was included in the second stage sampling. For the third stage, structures of size three or more were divided into clusters of an expected size of four units and a sample of clusters was selected. This procedure added an estimated 1,398 new construction units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiency 2.—In permit-issuing areas, a sample of mobile homes placed in a park missed by the census or established after the census was selected in two stages. First, for each 1976-77 SMSA, a sample of tracts was selected and canvassed. All parks were listed and then matched back to the 1970 census to identify parks missed by the

census and parks established after the census. Second, the parks were divided into clusters of an expected size of four sites and a sample of clusters was selected and interviewed. Each of the sample units represented the same number of units that the regular AHS sample units represented. This procedure yielded no additional units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiencies 3-6.—The remaining missed units were sampled by one of two procedures. The first procedure was designed to represent units from the following types of missed structures (structures that had no chance of selection for the AHS):

1. Structures missed in the 1970 census.
2. Structures that were completely nonresidential in the 1970 census but contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and had a utility hookup, or were on the site but not occupied on April 1, 1970, or had no utility hookup but were occupied by persons with no usual residence elsewhere.
4. Houses that had been moved onto their present site since the 1970 census.

Initially, a subsample of regular AHS units was selected at a rate of 1 in 22.2152. Then, succeeding structures, in a defined path of travel to the right of the structure containing the sample unit, were listed until eight structures (excluding the sample unit structure) were found that had been eligible to be selected for the AHS. Finally, the intervening structures that did not have a chance of selection in the AHS were identified and units within these structures were interviewed. In cases where the interviewer workload would have been too great, a representative subsample of units within these structures was selected. This procedure added an estimated 1,639 units to the coverage of the housing inventory of this SMSA.

The second procedure was designed to represent missed units from structures represented in the AHS. These missed units were:

1. Units missed in the 1970 census.
2. Units converted to residential use since the 1970 census in structures that contained some residential units in 1970.

First, a subsample of regular AHS units in multi-unit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multi-unit structures selected, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure yielded no additional units to the coverage of the housing inventory of this SMSA.

Building loss sample selection.—Some tables in this report show estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS interview). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample previously described. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

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 employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at the time of the interview (i.e., the 1976 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1976 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1976 housing inventory.—The AHS estimates of characteristics of the 1976 housing inventory employed a one-stage ratio estimation procedure. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 225 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units} + \text{Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 50 noninterview cells for sample housing units from the permit-

issuing universe (where the cells consisted of 1 or more of the different strata used in the stratification of the universe as previously illustrated). In addition, a noninterview factor was computed for one noninterview cell for new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes, and one noninterview cell for sample units from both the nonpermit universe (if applicable) and the coverage improvement universe (if units were not included above).

The following ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding ratio estimation category.

The effect of this ratio estimation procedure was to somewhat reduce the variance due to the variation in sample size of strata used in the sample selection of the permit-issuing universe, thereby reducing the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

1970-1976 lost units.—The AHS estimate of characteristics of the 1970-1976 lost units employed a one-stage ratio estimation procedure which is similar to the ratio estimation procedure described previously. The file of 1970-1976 lost units was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error 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 housing inventory from 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 this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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-SMSA 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 can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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 both the 1970 Census of Population and Housing and the 1976 AHS-SMSA 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 included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and

Research Program Reports, Series 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.**

AHS-SMSA.—For the 1976 AHS-SMSA 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 sample 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 accuracy of the AHS data collected from interviewed households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct 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.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview check. The results of this study are available and are presented in the following Census Bureau memorandum, "Response Error in the Annual Housing Survey as Measured by Reinterviews—Year III (1976-1977) SMSA Sample."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year III SMSA's and not for any specific SMSA):

1. For attitudinal items which were not reconciled (i.e., after the question is answered in the reinterview, the enumerator does not present the previous response and then ask the respondent to decide upon the best answer), approximately 67 percent of the indices of inconsistency showed moderate levels of response disagreement while the remaining 33 percent showed high levels.

2. Some differences beyond those due to sampling error did occur. Most of the categories affected by bias were categories of attitudinal items.

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low; indices from 20—50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

The results of this study were based on sample data so 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 this study.

Coverage errors.—With respect to errors of coverage and estimation for missing data, the AHS new construction sample had deficiencies in the representation of conventional new construction. Due to time constraints, only those building permits issued more than 5 months before the survey ended were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. Also, permits were sampled on a monthly basis and were divided equally among the 12 panels in which interviews were conducted. Due to this procedure, some of the permits issued in November 1975 through October 1976 were not interviewed because they were assigned to panels in which the interviewing had already been completed. In this SMSA, 4.7 percent of the permits sampled were not interviewed because of this procedure. This percentage does not include permits issued during the last 5 months of the survey. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units since, due to the relatively short time-span involved, it is possible that construction of these units was not completed at the time the survey was conducted and they were not eligible for interview. In addition to these deficiencies, new construction in special places that do not require building permits, such as military bases, are also not adequately represented.

The Coverage Improvement Program also had certain deficiencies. It appears that the listing procedure used to find units missed in the 1970 census, units converted from nonresidential to residential, houses moved onto their present site, and mobile homes placed outside parks, was not very efficient for finding nonresidential conversions. Such conversions were primarily in business districts, whereas the listing procedure started from a residential unit.

Deficiencies also exist in ED's where area sampling methods are used. It had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing.

Rounding errors.—For 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 or small medians, when these figures are derived from relatively large bases (e.g., median number of persons per household). 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 the survey.

Sampling errors for the AHS-SMSA sample.—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 schedules, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variance in the estimates due to response and interviewer errors (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 error measured by the standard error, 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 in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from 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 following tables are approximations to the standard errors of various estimates shown in this report for this SMSA. 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 precise standard errors for any specific items.

Table I presents the standard errors applicable to estimates of characteristics of the 1976 housing inventory as well as estimates of characteristics of the 1970-1976 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in this table. The standard errors on the AHS estimates of the population in housing units shown in tables A-1, B-1, and C-1 of part A of this report are 6,850 for the total SMSA, 5,970 for the central city of the SMSA, and 4,730 for the balance of the SMSA.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1976 Housing Inventory and for Estimated Number of 1970-1976 Lost Units for the Honolulu, Hawaii, SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central city	Not in central city
0	50	60	50
100.	70	70	70
200.	100	100	100
500.	170	170	160
700.	200	200	190
1,000 . . .	230	230	230
2,500 . . .	370	370	360
5,000 . . .	520	520	510
10,000 . .	730	720	710
25,000 . .	1,110	1,100	1,090
50,000 . .	1,490	1,430	1,430
75,000 . .	1,720	1,600	1,620
100,000. .	1,850	1,650	1,700
150,000. .	1,900	—	—
200,000. .	1,660	—	—

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.

Table II presents the standard errors of estimated percentages of the 1976 housing inventory as well as estimated percentages of the 1970-1976 lost units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in table II.

Included in tables I and II are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered as 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.

For ratios, 100 (x/y), where x is not a subclass of y, table II underestimates 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

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the Honolulu, Hawaii, SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100.	35.6	35.6	35.6	35.6	35.6	37.1
200.	21.6	21.6	21.6	21.6	22.7	26.3
500.	9.9	9.9	9.9	10.0	14.4	16.6
700.	7.3	7.3	7.3	8.4	12.2	14.0
1,000	5.2	5.2	5.2	7.0	10.2	11.7
2,500	2.2	2.2	3.2	4.5	6.4	7.4
5,000	1.1	1.1	2.3	3.2	4.5	5.3
10,000	0.5	0.7	1.6	2.2	3.2	3.7
25,000	0.2	0.5	1.0	1.4	2.0	2.3
50,000	0.11	0.3	0.7	1.0	1.4	1.7
75,000	0.07	0.3	0.6	0.8	1.2	1.4
100,000. . . .	0.07	0.2	0.5	0.7	1.0	1.2

[1] Standard errors are presented to nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to nearest one-hundredth of one percentage point.

Illustration of the use of the standard error tables.—Table A-1 of part A of this report shows that in 1976 there were 96,100 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 1,830. The following interpolation procedure was used:

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

Size of estimate	Standard error
75,000	1,720
96,100	x
100,000.	1,850

The entry for x is determined by vertically interpolating between 1,720 and 1,850.

$$96{,}100-75{,}000 = 21{,}100$$
$$100{,}000-75{,}000 = 25{,}000$$

$$1{,}720 + \frac{21{,}100}{25{,}000}(1{,}850-1{,}720) = 1{,}830$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 94,270 to 97,930 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 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 93,170 to 99,030 housing units with 90 percent confidence; and that the average estimate lies within the interval from 92,440 to 99,760 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 96,100 owner-occupied housing units, 12,900, or 13.4 percent, had two bedrooms. Interpolation in table II of the appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 13.4 percent is approximately 0.8 percentage points. The following interpolation procedure was used:

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

Base of percentage	Estimated percentage		
	10 or 90	13.4	25 or 75
75,000	0.8	a	1.2
96,100		p	
100,000. . . .	0.7	b	1.0

1. The entry for cell "a" is determined by horizontal interpolation between 0.8 and 1.2.

$$13.4-10.0 = 3.4$$
$$25.0-10.0 = 15.0$$

$$0.8 + \frac{3.4}{15.0}(1.2-0.8) = .89$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.7 and 1.0.

$$13.4-10.0 = 3.4$$
$$25.0-10.0 = 15.0$$

$$0.7 + \frac{3.4}{15.0}(1.0-0.7) = .77$$

3. The entry for "p" was then determined by vertical interpolation between .77 and .89.

$$96{,}100-75{,}000 = 21{,}100$$
$$100{,}000-75{,}000 = 25{,}000$$

$$.89-\frac{21{,}100}{25{,}000}(.89-.77) = 0.8$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 12.6 to 14.2 percent; the 90-percent confidence interval is from 12.1 to 14.7 percent; and the 95-percent confidence interval is from 11.8 to 15.0 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 error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; but, 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 part A of this report shows that in 1976 there were 56,500 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 43,600. Table I shows the standard error of 12,900 is approximately 800 and the standard error of 56,500 is approximately 1,550. Therefore, the standard error of the estimated difference of 43,600 is about:

$$1{,}740 = \sqrt{(800)^2 + (1{,}550)^2}$$

Consequently, the 68-percent confidence interval for the 43,600 difference is from 41,860 to 45,340 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 40,820 to 46,380 housing units, and the 95-percent confidence interval is from 40,120 to 47,080. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval 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 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 table II 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 of a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 3.5. The base of the distribution from which this median was determined is 96,100 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 96,100 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 table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 31,700 owner-occupied housing units, or 33.0 percent, had one or two persons (for purposes of calculating the median, the category of two persons is considered to be from 1.5 to 2.5 persons); that an additional 17,000 owner-occupied housing units, or 17.7 percent, had three persons (i.e., 2.5 to 3.5 persons); and that an additional 20,700 owner-occupied housing units, or 21.5 percent, had four persons (i.e., 3.5 to 4.5 persons).

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$2.5 + (3.5-2.5)\left(\frac{47.6-33.0}{17.7}\right) = 3.3$$

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

$$3.5 + (4.5-3.5)\left(\frac{52.4-33.0-17.7}{21.5}\right) = 3.6$$

Thus, the 95-percent confidence interval ranges from 3.3 to 3.6 persons.

Houston 1976

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Annual Housing Survey App-42
Selection of the sample. App-42
Coverage improvement sample selection App-43
Coverage improvement for deficiency 1 . . . . . App-43
Coverage improvement for deficiency 2 . . . . . App-43
Coverage improvement for deficiencies 3-6 . . . . . App-44
Building loss sample selection . . . App-44
1970 Census of Population and Housing. App-44
ESTIMATION. App-44
1976 housing inventory App-44
1970-1976 lost units App-45
Ratio estimation procedure of the 1970 Census of Population and Housing. App-45
RELIABILITY OF THE ESTIMATES App-45
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-46
Coverage errors App-46
Rounding errors. App-47
Sampling errors for the AHS-SMSA sample App-47
Illustration of the use of the standard error tables App-48
Differences. App-49
Illustration of the computation of the standard error of a difference App-49
Medians App-49
Illustration of the computation of the 95-percent confidence interval of a median . . App-50

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 20 SMSA's are based on data collected from the 1976 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. In each of the 20 SMSA's, the data were collected for the 12-month period from April 1976 through March 1977 with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and data for a second group of 21 SMSA's were collected from April 1975 through March 1976. The sample housing units for each group of the AHS SMSA's are interviewed on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the third group (1976-77) are: Houston, Tex., New York, N.Y., St. Louis, Mo.-Ill., and Seattle-Everett, Wash.

The remaining SMSA's in the third group are: Allentown-Bethlehem-Easton, Pa.-N.J., Baltimore, Md., Birmingham, Ala., Buffalo, N.Y., Cleveland, Ohio, Denver, Colo., Grand Rapids, Mich., Honolulu, Hawaii, Indianapolis, Ind., Las Vegas, Nev., Louisville, Ky.-Ind., Oklahoma City, Okla., Omaha, Nebr.-Iowa, Providence-Pawtucket-Warwick, R.I.-Mass., Raleigh, N.C., and Sacramento, Calif.

In this SMSA, 16,025 units were eligible for interview. Of these sample units, 903 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to the units eligible for interview, 1,284 units were visited but were not eligible for interview, because they were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit-issuing (Honolulu, Las Vegas, New York, and Sacramento) was selected from two sample frames—units enumerated in the 1970 Census of Population and Housing in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for the 16 SMSA's which are *not* 100-percent permit-issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe).

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of the sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA, since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and balance of the SMSA, since the sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total housing units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit

records. Before the sample was selected from the occupied and vacant housing unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to 1 of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—Family size	Renter—Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000-$5,999. . .		
$6,000-$9,999. . .		
$10,000-$14,999 .		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to 1 of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to 1 of the 4 vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. However, the housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of the interview, the units at each of the sample special places were listed and subsampled at a rate to produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since January 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit-issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then 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. At the time of the survey, those segments that did not have an expected size of four were further subdivided to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Coverage improvement sample selection.—The Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS-SMSA sample from the permit-issuing and new construction universes. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970, but completed after April 1, 1970.
2. Mobile homes placed in parks either missed during the 1970 census or established since the census.
3. Units missed in the 1970 census.
4. Units converted to residential use that were nonresidential at the time of the 1970 census.
5. Houses that have been moved onto their present site since the 1970 census.
6. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

Coverage improvement for deficiency 1.—A sample of new construction units whose permits were issued before January 1970, but completed after April 1970, was selected independently for each SMSA. The sampling was carried out in two stages for one-and-two-unit structures and in three stages for three-or-more-unit structures. These units were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units).

The first stage was a sample of permit offices and the second stage a sample of the 1969 permits within each of the selected permit offices. In the New York, N.Y., SMSA, Honolulu, Hawaii, SMSA, and Grand Rapids, Mich., SMSA, an additional sample of 1968 permits for three-or-more unit structures was included in the second stage sampling. For the third stage, structures of size three or more were divided into clusters of an expected size of four units and a sample of clusters was selected. This procedure added an estimated 10,593 units to the housing inventory of this SMSA.

Coverage improvement for deficiency 2.—In permit-issuing areas, a sample of mobile homes placed in a park missed by the

census or established after the census was selected in two stages. First, for each 1976-77 SMSA, a sample of tracts was selected and canvassed. All parks were listed and then matched back to the 1970 census to identify parks missed by the census and parks established after the census. Second, the parks were divided into clusters of an expected size of four sites and a sample of clusters was selected and interviewed. Each of the sample units represented the same number of units that the regular AHS sample units represented. This procedure added an estimated 2,270 units to the housing inventory of this SMSA.

Coverage improvement for deficiencies 3-6.—The remaining missed units were sampled by one of two procedures. The first procedure was designed to represent units from the following types of missed structures (structures that had no chance of selection for the AHS):

1. Structures missed in the 1970 census.
2. Structures that were completely nonresidential in the 1970 census but contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and had a utility hookup, or were on the site but not occupied on April 1, 1970, or had no utility hookup but were occupied by persons with no usual residence elsewhere.
4. Houses that had been moved onto their present site since the 1970 census.

Initially, a subsample of regular AHS units was selected at a rate of 1 in 22.2152. Then, succeeding structures in a defined path of travel to the right of the structure containing the sample unit were listed until eight structures (excluding the sample unit structure) were found that had been eligible to be selected for the AHS. Finally, the intervening structures that did not have a chance of selection in the AHS were identified and units within these structures were interviewed. In cases where the interviewer workload would have been too great, a representative subsample of units within these structures was selected. This procedure added an estimated 6,815 units to the housing inventory of this SMSA.

The second procedure was designed to represent missed units from structures represented in AHS. These missed units were:

1. Units missed in the 1970 census.
2. Units converted to residential use since the 1970 census in structures that contained some residential units in 1970.

First, a subsample of regular AHS units in multi-unit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multi-unit structures selected, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 830 units to the housing inventory of this SMSA.

Building loss sample selection.—Some tables in this report show estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS interview). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample previously described. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a non-permit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

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 employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at the time of the interview (i.e., the 1976 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1976 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1976 housing inventory.—The AHS estimates of characteristics of the 1976 housing inventory employed a one-stage ratio estimation procedure. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 903 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units} + \text{Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 50 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of 1 or more of the different strata used in the stratification of the universe as previously illustrated). In addition, a noninterview factor was computed for one noninterview cell for new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes, and one noninterview cell for sample units from both the nonpermit universe (if applicable) and the coverage improvement universe (if units were not included above).

The following ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding ratio estimation category.

The effect of this ratio estimation procedure was to somewhat reduce the variance due to the variation in sample size of strata used in the sample selection of the permit-issuing universe, thereby reducing the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

1970-1976 lost units.—The AHS estimate of characteristics of the 1970-1976 lost units employed a one-stage ratio estimation procedure which is similar to the ratio estimation procedure described previously. The file of 1970-1976 lost units was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error 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 housing inventory from 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 this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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-SMSA 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 can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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 both the 1970 Census of Population and Housing and the 1976 AHS-SMSA 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 included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series 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.**

AHS-SMSA.—For the 1976 AHS-SMSA 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 sample 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 accuracy of the AHS data collected from interviewed households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct 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.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview check. The results of this study are not yet available; however, it is believed that they will be similar to the results of the Year II (1975-76) AHS SMSA reinterview study which were presented in the following Census Bureau memorandum, "Response Error in the Annual Housing Survey Data—Year II SMSA Sample."

Some of the results of this study are (note that these results are based on the reinterviews across all Year II SMSA's and not for any specific SMSA.):

1. The results indicate that the interviewers are doing a good job. From a total of 207 interviewers checked, 10 failed in classification checks, 7 in household composition, and 2 in content reinterview.

2. Of the nonattitudinal items which had enough data to compute reliable measures of response error, approximately 55 percent of these items had low levels of inconsistency in responses and about 45 percent had moderate levels of inconsistency in responses.

3. For the attitudinal items, approximately 75 percent of the indices showed low levels of inconsistency when reconciled; i.e., after the question is answered in the reinterview, the enumerator presents the previous response and then asks the respondent to decide upon the best answer. The remaining items showed moderate levels of inconsistent response.

4. When no reconciliation was done for these six attitudinal items, about half of the items showed moderate levels of inconsistency, 30 percent were in the low level and 20 percent were in the high inconsistency level.

The range for evaluating inconsistency is from 0-100. The rule of thumb is that indices below 20 are low; indices from 20-50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

The results of this study were based on sample data, so 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 this study.

Coverage errors.—With respect to errors of coverage and estimation for missing data, the AHS new construction sample had deficiencies in the representation of conventional new construction. Due to time constraints, only those building permits issued more than 5 months before the survey ended were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. Also, permits were sampled on a monthly basis and were divided equally among the 12 panels in which interviews were conducted. Due to this procedure, some of the permits issued in November 1975 through October 1976 were not interviewed because they were assigned to panels in which the interviewing had already been completed. In this SMSA, 9.0 percent of the permits sampled were not interviewed because of this procedure. This percentage does not include permits issued during the last 5 months of the survey. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units, since, due to the relatively short time span involved, it is possible that construction of these units was not completed at the time the survey was conducted and they were not eligible for interview. In addition to these deficiencies, new construction in special places that do not require building permits, such as military bases, are also not adequately represented.

The Coverage Improvement Program also had certain deficiencies. It appears that the listing procedure used to find units missed in the 1970 census, units converted from nonresidential to residential, houses moved onto their present site, and mobile homes placed outside parks was not very efficient for finding nonresidential conversions. Such conversions were primarily in business districts, whereas the listing procedure started from a residential unit.

Deficiencies also exist in ED's where area sampling methods are used. It had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing.

Rounding errors.—For 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 or small medians when these figures are derived from relatively large bases (e.g., median number of persons per household). 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 the survey.

Sampling errors for the AHS-SMSA sample.—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 schedules, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variance in the estimates due to response and interviewer errors (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 error measured by the standard error, 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 in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from 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 accompanying tables are approximations to the standard errors of various estimates shown in this report for this SMSA. 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 precise standard errors for any specific items.

Table I presents the standard errors applicable to estimates of characteristics of the 1976 housing inventory as well as estimates of characteristics of the 1970-1976 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in this table.

The standard errors on the AHS estimates of population in housing units shown in tables A-1, B-1, and C-1 of part A of this report are 23,980 for the SMSA, 16,900 for the central city of this SMSA, and 16,900 for the balance of this SMSA.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1976 Housing Inventory and for Estimated Number of 1970-1976 Lost Units for the Houston, Tex., SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard error[1]		
	SMSA	In central city	Not in central city
0	70	80	50
100.	80	90	70
200.	120	120	100
500.	190	200	160
700.	220	230	190
1,000 . . .	270	280	230
2,500 . . .	420	440	360
5,000 . . .	600	620	500
10,000 . .	840	880	710
25,000 . .	1,330	1,380	1,110
50,000 . .	1,860	1,930	1,540
75,000 . .	2,260	2,330	1,850
100,000. .	2,590	2,660	2,100
150,000. .	3,130	3,170	2,470
250,000. .	3,910	3,860	2,910
500,000. .	5,040	4,530	–

[1] For estimates pertaining to new construction, the standard errors shown in the table should be multiplied by a factor of 1.4 for the total SMSA and the In central city estimates, and 1.5 for the Not in central city estimates.

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 II through IV present the standard errors of estimated percentages of the 1976 housing inventory as well as estimated percentages of the 1970-1976 lost units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables II through IV.

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 as 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.

For ratios, 100 (x/y), where x is not a subclass of y, tables II through 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.—Table A-1 of part A of this report shows that in 1976 there were 429,500 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 4,720. The following interpolation procedure was used:

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

Size of estimate	Standard error
250,000.	3,910
429,500.	x
500,000.	5,040

The entry for x is determined by vertically interpolating between 3,910 and 5,040.

429,500−250,000 = 179,500
500,000−250,000 = 250,000

$$3{,}910+\frac{179{,}500}{250{,}000}(5{,}040-3{,}910)=4{,}720$$

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the Houston, Tex., SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100.	41.6	41.6	41.6	41.6	41.6	42.2
200.	26.3	26.3	26.3	26.3	26.3	29.9
500.	12.5	12.5	12.5	12.5	16.4	18.9
700.	9.2	9.2	9.2	9.6	13.8	16.0
1,000	6.7	6.7	6.7	8.0	11.6	13.4
2,500	2.8	2.8	3.7	5.1	7.3	8.4
5,000	1.4	1.4	2.6	3.6	5.2	6.0
10,000	.7	.8	1.8	2.5	3.7	4.2
25,000	.3	.5	1.2	1.6	2.3	2.7
50,000	.14	.4	.8	1.1	1.6	1.9
75,000	.09	.3	.7	.9	1.3	1.5
100,000. . . .	.07	.3	.6	.8	1.2	1.3
150,000. . . .	.05	.2	.5	.7	.9	1.1
250,000. . . .	.03	.2	.4	.5	.7	.8
500,000. . . .	.01	.12	.3	.4	.5	.6

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point. For estimated percentages pertaining to new construction, the standard errors shown in the table should be multiplied by a factor of 1.4.

TABLE III. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the Central City of the Houston, Tex., SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100.	43.8	43.8	43.8	43.8	43.8	44.2
200.	28.1	28.1	28.1	28.1	28.1	31.2
500.	13.5	13.5	13.5	13.5	17.1	19.7
700.	10.0	10.0	10.0	10.0	14.5	16.7
1,000	7.2	7.2	7.2	8.4	12.1	14.0
2,500	3.0	3.0	3.8	5.3	7.6	8.8
5,000	1.5	1.5	2.7	3.7	5.4	6:2
10,000	.8	.9	1.9	2.6	3.8	4.4
25,000	.3	.6	1.2	1.7	2.4	2.8
50,000	.2	.4	.9	1.2	1.7	2.0
75,000	.10	.3	.7	1.0	1.4	1.6
100,000. . . .	.08	.3	.6	.8	1.2	1.4
150,000. . . .	.05	.2	.5	.7	1.0	1.1
250,000. . . .	.03	.2	.4	.5	.8	.9
500,000. . . .	.02	.12	.3	.4	.5	.6

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point. For estimated percentages pertaining to new construction, the standard errors shown in the table should be multiplied by a factor of 1.4.

Consequently, the 68-percent confidence interval, as shown by these data, is from 424,780 to 434,220 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 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 421,950 to 437,050 housing units with 90 percent confidence; and the average estimate lies within the interval from 420,060 to 438,940 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 429,500 owner-occupied housing units, 104,400, or 24.3 percent, had two bedrooms. Interpolation in table II of the appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 24.3 percent is approximately 0.6 percentage points. The following interpolation procedure was used:

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

Base of percentage	Estimated percentage		
	10 or 90	24.3	25 or 75
250,000.	0.5	a	0.7
429,500.		p	
500,000.	0.4	b	0.5

1. The entry for cell "a" is determined by horizontal interpolation between 0.5 and 0.7.

$$24.3-10.0 = 14.3$$
$$25.0-10.0 = 15.0$$

$$0.5 + \frac{14.3}{15.0}(0.7-0.5) = 0.69$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.4 and 0.5.

$$24.3-10.0 = 14.3$$
$$25.0-10.0 = 15.0$$

$$0.4 + \frac{14.3}{15.0}(0.5-0.4) = 0.50$$

3. The entry for "p" was then determined by vertical interpolation between 0.50 and 0.69

$$429{,}500-250{,}000 = 179{,}500$$
$$500{,}000-250{,}000 = 250{,}000$$

$$0.69-\frac{179{,}500}{250{,}000}(0.69-0.50) = 0.6$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 23.7 to 24.9 percent; the 90-percent confidence interval is from 23.3 to 25.3 percent; and the 95-percent confidence interval is from 23.1 to 25.5 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 error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; but 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 part A of this report shows that in 1976 there were 239,300 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 134,900. Table I shows the standard error of 104,400 is approximately 2,640 and the standard error of 239,300 is approximately 3,830. Therefore, the standard error of the estimated difference of 134,900 is about

$$4{,}650 = \sqrt{(2{,}640)^2 + (3{,}830)^2}.$$

Consequently, the 68-percent confidence interval for the difference is from 130,250 to 139,550 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 127,460 to 142,340 housing units, and the 95-percent confidence interval is from 125,600 to 144,200. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval 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 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 tables II through IV, 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.

TABLE IV. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the Balance of the Houston, Tex., SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100.	33.7	33.7	33.7	33.7	33.7	35.7
200.	20.3	20.3	20.3	20.3	21.8	25.2
500.	9.2	9.2	9.2	9.6	13.8	16.0
700.	6.8	6.8	6.8	8.1	11.7	13.5
1,000	4.8	4.8	4.9	6.8	9.8	11.3
2,500	2.0	2.0	3.1	4.3	6.2	7.1
5,000	1.0	1.0	2.2	3.0	4.4	5.0
10,000	.5	.7	1.6	2.1	3.1	3.6
25,000	.2	.4	1.0	1.4	2.0	2.3
50,000	.10	.3	.7	1.0	1.4	1.6
75,000	.07	.3	.6	.8	1.1	1.3
100,000. . . .	.05	.2	.5	.7	1.0	1.1
150,000. . . .	.03	.2	.4	.6	.8	.9
250,000. . . .	.02	.14	.3	.4	.6	.7

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point. For estimated percentages pertaining to new construction, the standard errors shown in the table should be multiplied by a factor of 1.5.

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 of a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 3.0. The base of the distribution from which this median was determined is 429,500 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 429,500 is approximately 0.7 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 48.6 and 51.4.
3. From table A-1 of part A, it can be seen, by cumulating the frequencies for the first two categories, that 171,700 owner-occupied housing units, or 40.0 percent, had one or two persons (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 82,800 owner-occupied housing units, or 19.3 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:

$$2.5 + (3.5-2.5)\left(\frac{48.6-40.0}{19.3}\right) = 2.9$$

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

$$2.5 + (3.5-2.5)\left(\frac{51.4-40.0}{19.3}\right) = 3.1$$

Thus, the 95-percent confidence interval ranges from 2.9 to 3.1 persons.

Indianapolis 1976

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Annual Housing Survey App-42
Selection of the sample App-42
Coverage improvement sample selection App-43
Coverage improvement for deficiency 1 ..... App-43
Coverage improvement for deficiency 2 ..... App-43
Coverage improvement for deficiencies 3-6 ..... App-44
Building loss sample selection App-44
1970 Census of Population and Housing App-44
ESTIMATION App-44
1976 housing inventory App-44
1970-1976 lost units App-45
Ratio estimation procedure of the 1970 Census of Population and Housing App-45
RELIABILITY OF THE ESTIMATES App-45
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-46
Coverage errors App-46
Rounding errors App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-48
Differences App-49
Illustration of the computation of the standard error of a difference App-49
Medians App-49
Illustration of the computation of the 95-percent confidence interval of a median .. App-49

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 20 SMSA's are based on data collected from the 1976 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. In each of the 20 SMSA's, the data were collected for the 12-month period from April 1976 through March 1977 with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and data for a second group of 21 SMSA's were collected from April 1975 through March 1976. The sample housing units for each group of the AHS SMSA's are interviewed on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the third group (1976-77) are: Houston, Tex., New York, N.Y., St. Louis, Mo.-Ill., and Seattle-Everett, Wash.

The remaining SMSA's in the third group are: Allentown-Bethlehem-Easton, Pa.-N.J., Baltimore, Md., Birmingham, Ala., Buffalo, N.Y., Cleveland, Ohio, Denver, Colo., Grand Rapids, Mich., Honolulu, Hawaii, Indianapolis, Ind., Las Vegas, Nev., Louisville, Ky.-Ind., Oklahoma City, Okla., Omaha, Nebr.-Iowa, Providence-Pawtucket-Warwick, R.I.-Mass., Raleigh, N.C., and Sacramento, Calif.

In this SMSA, 4,796 units were eligible for interview. Of these sample units, 154 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to the units eligible for interview, 418 units were visited but were not eligible for interview, because they were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit-issuing (Honolulu, Las Vegas, New York, and Sacramento) was selected from two sample frames—units enumerated in the 1970 Census of Population and Housing in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for the 16 SMSA's which are *not* 100-percent permit-issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe).

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of the sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and balance of the SMSA, since the sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total housing units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant housing unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit

records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to 1 of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—Family size	Renter—Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000-$5,999. . .		
$6,000-$9,999. . .		
$10,000-$14,999 .		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to 1 of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to 1 of the 4 vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. However, the housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of the interview, the units at each of the sample special places were listed and subsampled at a rate to produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since January 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit-issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then 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. At the time of the survey, those segments that did not have an expected size of four were further subdivided to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Coverage improvement sample selection.—The Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS-SMSA sample from the permit-issuing and new construction universes. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970, but completed after April 1, 1970.
2. Mobile homes placed in parks either missed during the 1970 census or established since the census.
3. Units missed in the 1970 census.
4. Units converted to residential use that were nonresidential at the time of the 1970 census.
5. Houses that have been moved onto their present site since the 1970 census.
6. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

Coverage improvement for deficiency 1.—A sample of new construction units whose permits were issued before January 1970, but completed after April 1970, was selected independently for each SMSA. The sampling was carried out in two stages for one-and-two-unit structures and in three stages for three-or-more-unit structures. These units were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units).

The first stage was a sample of permit offices and the second stage a sample of the 1969 permits within each of the selected permit offices. In the New York, N.Y., SMSA, Honolulu, Hawaii, SMSA, and Grand Rapids, Mich., SMSA, an additional sample of 1968 permits for three-or-more-unit structures was included in the second stage sampling. For the third stage, structures of size three or more were divided into clusters of an expected size of four units and a sample of clusters was selected. This procedure added an estimated 1,674 new construction units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiency 2.—In permit-issuing areas, a sample of mobile homes placed in a park missed by the census or established after the census was selected in two stages. First, for each 1976-77 SMSA, a sample of tracts was selected and canvassed. All parks were listed and then matched back to the 1970 census to identify parks missed by the census and parks established after the

census. Second, the parks were divided into clusters of an expected size of four sites and a sample of clusters was selected and interviewed. Each of the sample units represented the same number of units that the regular AHS sample units represented. This procedure added an estimated 428 units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiencies 3-6.—The remaining missed units were sampled by one of two procedures. The first procedure was designed to represent units from the following types of missed structures (structures that had no chance of selection for the AHS):

1. Structures missed in the 1970 census.
2. Structures that were completely nonresidential in the 1970 census but contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and had a utility hookup, or were on the site but not occupied on April 1, 1970, or had no utility hookup but were occupied by persons with no usual residence elsewhere.
4. Houses that had been moved onto their present site since the 1970 census.

Initially, a subsample of regular AHS units was selected at a rate of 1 in 22.2152. Then succeeding structures, in a defined path of travel to the right of the structure containing the sample unit, were listed until eight structures (excluding the sample unit structure) were found that had been eligible to be selected for the AHS. Finally, the intervening structures that did not have a chance of selection in the AHS were identified and units within these structures were interviewed. In cases where the interviewer workload would have been too great, a representative subsample of units within these structures was selected. This procedure added an estimated 5,361 units to the coverage of the housing inventory of this SMSA.

The second procedure was designed to represent missed units from structures represented in the AHS. These missed units were:

1. Units missed in the 1970 census.
2. Units converted to residential use since the 1970 census in structures that contained some residential units in 1970.

First, a subsample of regular AHS units in multi-unit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multi-unit structures selected, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 2,286 units to the coverage of the housing inventory of this SMSA.

Building loss sample selection.—Some tables in this report show estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS interview). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample previously described. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

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 employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at the time of the interview (i.e., the 1976 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1976 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1976 housing inventory.—The AHS estimates of characteristics of the 1976 housing inventory employed a one-stage ratio estimation procedure. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 154 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units} + \text{Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 50 noninterview cells for sample housing units from the permit-

issuing universe (where the cells consisted of 1 or more of the different strata used in the stratification of the universe as previously illustrated). In addition, a noninterview factor was computed for one noninterview cell for new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes, and one noninterview cell for sample units from both the nonpermit universe (if applicable) and the coverage improvement universe (if units were not included above).

The following ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding ratio estimation category.

The effect of this ratio estimation procedure was to somewhat reduce the variance due to the variation in sample size of strata used in the sample selection of the permit-issuing universe, thereby reducing the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

1970-1976 lost units.—The AHS estimate of characteristics of the 1970-1976 lost units employed a one-stage ratio estimation procedure which is similar to the ratio estimation procedure described previously. The file of 1970-1976 lost units was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error 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 housing inventory from 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 this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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-SMSA 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 can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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 both the 1970 Census of Population and Housing and the 1976 AHS-SMSA 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 included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and

Research Program Reports, series 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.**

AHS-SMSA.—For the 1976 AHS-SMSA 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 sample 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 accuracy of the AHS data collected from interviewed households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct 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.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview check. The results of this study are available and are presented in the following Census Bureau memorandum, "Response Error in the Annual Housing Survey as Measured by Reinterviews—Year III (1976-1977) SMSA Sample."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year III SMSA's and not for any specific SMSA):

1. For attitudinal items which were not reconciled (i.e., after the question is answered in the reinterview, the enumerator does not present the previous response and then ask the respondent to decide upon the best answer), approximately 67 percent of the indices of inconsistency showed moderate levels of response disagreement while the remaining 33 percent showed high levels.
2. Some differences beyond those due to sampling error did occur. Most of the categories affected by bias were categories of attitudinal items.

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low, indices from 20—50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

The results of this study were based on sample data so 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 this study.

Coverage errors.—With respect to errors of coverage and estimation for missing data, the AHS new construction sample had deficiencies in the representation of conventional new construction. Due to time constraints, only those building permits issued more than five months before the survey ended were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. Also, permits were sampled on a monthly basis and were divided equally among the 12 panels in which interviews were conducted. Due to this procedure, some of the permits issued in November 1975 through October 1976 were not interviewed because they were assigned to panels in which the interviewing had already been completed. In this SMSA, 5.8 percent of the permits sampled were not interviewed because of this procedure. This percentage does not include permits issued during the last 5 months of the survey. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units since, due to the relatively short time span involved, it is possible that construction of these units was not completed at the time the survey was conducted and they were not eligible for interview. In addition to these deficiencies, new construction in special places that do not require building permits, such as military bases, are also not adequately represented.

The Coverage Improvement Program also had certain deficiencies. It appears that the listing procedure used to find units missed in the 1970 census, units converted from nonresidential to residential, houses moved onto their present site, and mobile homes placed outside parks, was not very efficient for finding nonresidential conversions. Such conversions were primarily in business districts, whereas the listing procedure started from a residential unit.

Deficiencies also exist in ED's where area sampling methods are used. It had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing.

Rounding errors.—For 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 or small medians, when these figures are derived from relatively large bases (e.g., median number of persons per household). 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 the survey.

Sampling errors for the AHS-SMSA sample.—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 schedules, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variance in the estimates due to response and interviewer errors (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 error measured by the standard error, 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 in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from 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 following tables are approximations to the standard errors of various estimates shown in this report for this SMSA. 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 precise standard errors for any specific items.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1976 Housing Inventory and for Estimated Number of 1970-1976 Lost Units for the Indianapolis, Ind., SMSA, for the Central City and for the Balance of the SMSA.

(68 chances out of 100)

Size of estimate	Standard error[1]		
	SMSA	In central city	Not in central city
0	110	110	90
100.	110	110	100
200.	140	150	140
500.	230	230	220
700.	270	280	260
1,000 . . .	320	330	310
2,500 . . .	510	520	490
5,000 . . .	720	730	690
10,000 . .	1,020	1,030	970
25,000 . .	1,590	1,600	1,530
50,000 . .	2,200	2,190	2,150
75,000 . .	2,640	2,590	2,610
100,000. .	2,980	2,880	2,990
150,000. .	3,490	3,230	–
200,000. .	3,820	3,360	–
250,000. .	4,030	3,280	–
300,000. .	4,130	–	–
400,000. .	4,040	–	–

[1] For estimates pertaining to new construction, the standard errors shown in the table should be multiplied by a factor of 1.4 for the total SMSA, the In central city, and the Not in central city estimates.

Table I presents the standard errors applicable to estimates of characteristics of the 1976 housing inventory as well as estimates of characteristics of the 1970-1976 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in this table.

The standard errors on the AHS estimates of population in housing units shown in tables A-1, B-1, and C-1 of this report are 12,210 for the SMSA, 9,870 for the central city of this SMSA, and 9,560 for the balance of this SMSA.

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.

Table II presents the standard errors of estimated percentages of the 1976 housing inventory as well as estimated percentages of the 1970-1976 lost units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in table II.

Included in tables I and II are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered as 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.

For ratios, 100 (x/y), where x is not a subclass of y, table II underestimates 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:

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

TABLE II. Standard errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the Indianapolis, Ind., SMSA, for the Central City of the SMSA, and for the Balance of the SMSA.

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100.	51.3	51.3	51.3	51.3	51.3	51.3
200.	34.5	34.5	34.5	34.5	34.5	36.3
500.	17.4	17.4	17.4	17.4	19.9	22.9
700.	13.1	13.1	13.1	13.1	16.8	19.4
1,000	9.5	9.5	9.5	9.7	14.0	16.2
2,500	4.0	4.0	4.5	6.2	8.9	10.3
5,000	2.1	2.1	3.2	4.4	6.3	7.3
10,000	1.0	1.0	2.2	3.1	4.4	5.1
25,000	.4	.6	1.4	1.9	2.8	3.2
50,000	.2	.5	1.0	1.4	2.0	2.3
75,000	.14	.4	.8	1.1	1.6	1.9
100,000. . . .	.11	.3	.7	1.0	1.4	1.6
150,000. . . .	.07	.3	.6	.8	1.1	1.3
200,000. . . .	.05	.2	.5	.7	1.0	1.1
250,000. . . .	.04	.2	.4	.6	.9	1.0
300,000. . . .	.04	.2	.4	.6	.8	.9
400,000. . . .	.03	.2	.4	.5	.7	.8

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point. For estimated percentages pertaining to new construction, the standard errors shown in the table should be multiplied by a factor of 1.4.

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.—Table A-1 of part A of this report shows that in 1976 there were 245,800 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 4,010. The following interpolation procedure was used:

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

Size of estimate	Standard error
200,000.	3,820
245,800.	x
250,000.	4,030

The entry for x is determined by vertically interpolating between 3,820 and 4,030.

$$245{,}800-200{,}000 = 45{,}800$$
$$250{,}000-200{,}000 = 50{,}000$$

$$3{,}820 + \frac{45{,}800}{50{,}000}(4{,}030-3{,}820) = 4{,}010$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 241,790 to 249,810 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 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 239,380 to 252,220 housing units with 90 percent confidence; and that the average estimate lies within the interval from 237,780 to 253,820 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 245,800 owner-occupied housing units, 74,100, or 30.1 percent, had two bedrooms. Interpolation in table II of the appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 30.1 percent is approximately 0.9 percentage points. The following interpolation procedure was used:

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

Base of percentage	Estimated percentage		
	25 or 75	30.1	50
200,000. . . .	1.0	a	1.1
245,800. . . .		p	
250,000. . . .	.9	b	1.0

1. The entry for cell "a" is determined by horizontal interpolation between 1.0 and 1.1.

$$30.1-25.0 = 5.1$$
$$50.0-25.0 = 25.0$$

$$1.0 + \frac{5.1}{25.0}(1.1-1.0) = 1.02$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.9 and 1.0.

$$30.1-25.0 = 5.1$$
$$50.0-25.0 = 25.0$$

$$0.9 + \frac{5.1}{25.0}(1.0-0.9) = .92$$

3. The entry for "p" was then determined by vertical interpolation between 0.92 and 1.02.

$$245{,}800-200{,}000 = 45{,}800$$
$$250{,}000-200{,}000 = 50{,}000$$
$$1.02-\frac{45{,}800}{50{,}000}(1.02-0.92) = 0.9$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 29.2 to 31.0 percent; the 90-percent confidence interval is from 28.7 to 31.5 percent; and the 95-percent confidence interval is from 28.3 to 31.9 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 error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; whereas, 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 part A of this report shows that in 1976 there were 124,800 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 50,700. Table I shows the standard error of 74,100 is approximately 2,620 and the standard error of 124,800 is approximately 3,230. Therefore, the standard error of the estimated difference of 50,700 is about

$$4{,}160 = \sqrt{(2{,}620)^2 + (3{,}230)^2}$$

Consequently, the 68-percent confidence interval for the difference is from 46,540 to 54,860 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 44,040 to 57,360 housing units, and the 95-percent confidence interval is from 42,380 to 59,020. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval 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 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 table II 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 of a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 2.8. The base of the distribution from which this median was determined is 245,800 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 245,800 is approximately 1.0 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 48.0 and 52.0.
3. From table A-I of part A, it can be seen by cumulating the frequencies for the first two categories that 108,600 owner-occupied housing units, or 44.2 percent, had one or two persons (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 43,000 owner-occupied housing units, or 17.5 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:

$$2.5 + (3.5-2.5)\left(\frac{48.0-44.2}{17.5}\right) = 2.7$$

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

$$2.5 + (3.5-2.5)\left(\frac{52.0-44.2}{17.5}\right) = 2.9$$

Thus, the 95-percent confidence interval ranges from 2.7 to 2.9 persons.

Las Vegas 1976

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Annual Housing Survey App-42
Selection of the sample App-42
Coverage improvement sample selection App-43
Coverage improvement for deficiency 1 . . . . . App-43
Coverage improvement for deficiency 2 . . . . . App-43
Coverage improvement for deficiencies 3-6 . . . . . App-44
Building loss sample selection . . . App-44
1970 Census of Population and Housing App-44
ESTIMATION App-44
1976 housing inventory App-44
1970-1976 lost units App-45
Ratio estimation procedure of the 1970 Census of Population and Housing App-45
RELIABILITY OF THE ESTIMATES App-45
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-46
Coverage errors App-46
Rounding errors App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences App-48
Illustration of the computation of the standard error of a difference App-49
Medians App-49
Illustration of the computation of the 95-percent confidence interval of a median . App-49

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 20 SMSA's are based on data collected from the 1976 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. In each of the 20 SMSA's, the data were collected for the 12-month period from April 1976 through March 1977 with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and data for a second group of 21 SMSA's were collected from April 1975 through March 1976. The sample housing units for each group of the AHS SMSA's are interviewed on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the third group (1976-77) are: Houston, Tex., New York, N.Y., St. Louis, Mo.-Ill., and Seattle-Everett, Wash.

The remaining SMSA's in the third group are: Allentown-Bethlehem-Easton, Pa.-N.J., Baltimore, Md., Birmingham, Ala., Buffalo, N.Y., Cleveland, Ohio, Denver, Colo., Grand Rapids, Mich., Honolulu, Hawaii, Indianapolis, Ind., Las Vegas, Nev., Louisville, Ky.-Ind., Oklahoma City, Okla., Omaha, Nebr.-Iowa, Providence-Pawtucket-Warwick, R.I.-Mass., Raleigh, N.C., and Sacramento, Calif.

In this SMSA, 4,965 units were eligible for interview. Of these sample units, 218 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to the units eligible for interview, 367 units were visited but were not eligible for interview, because they were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit-issuing (Honolulu, Las Vegas, New York, and Sacramento) was selected from two sample frames—units enumerated in the 1970 Census of Population and Housing in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for the 16 SMSA's which are *not* 100-percent permit-issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe).

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of the sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and balance of the SMSA, since the sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total housing units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant housing unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit

records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to 1 of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure: Owner—family size	Tenure: Renter—family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000		
$3,000-$5,999		
$6,000-$9,999		
$10,000-$14,999 ..		
$15,000 and over ..		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to 1 of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to 1 of the 4 vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected in such a way as to produce one-half of the desired sample size. However, the housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of the interview, the units at each of the sample special places were listed and subsampled at a rate to produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since January 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit-issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then 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. At the time of the survey, those segments that did not have an expected size of four were further subdivided to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Coverage improvement sample selection.—The Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS-SMSA sample from the permit-issuing and new construction universes. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970, but completed after April 1, 1970.
2. Mobile homes placed in parks either missed during the 1970 census or established since the census.
3. Units missed in the 1970 census.
4. Units converted to residential use that were nonresidential at the time of the 1970 census.
5. Houses that have been moved onto their present site since the 1970 census.
6. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

Coverage improvement for deficiency 1.—A sample of new construction units whose permits were issued before January 1970, but completed after April 1970, was selected independently for each SMSA. The sampling was carried out in two stages for one-and-two-unit structures and in three stages for three-or-more-unit structures. These units were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units).

The first stage was a sample of permit offices and the second stage a sample of the 1969 permits within each of the selected permit offices. In the New York, N.Y., SMSA, Honolulu, Hawaii, SMSA, and Grand Rapids, Mich., SMSA, an additional sample of 1968 permits for three-or-more-unit structures was included in the second stage sampling. For the third stage, structures of size three or more were divided into clusters of an expected size of four units and a sample of clusters was selected. This procedure added an estimated 912 new construction units to the housing inventory of this SMSA.

Coverage improvement for deficiency 2.—In permit-issuing areas, a sample of mobile homes placed in a park missed by the census or established after the census was selected in two stages. First, for each 1976-77 SMSA, a sample of tracts was selected and canvassed. All parks were

listed and then matched back to the 1970 census to identify parks missed by the census and parks established after the census. Second, the parks were divided into clusters of an expected size of four sites and a sample of clusters was selected and interviewed. Each of the sample units represented the same number of units that the regular AHS sample units represented. This procedure added an estimated 3,189 units to the housing inventory of this SMSA.

Coverage improvement for deficiencies 3-6.—The remaining missed units were sampled by one of two procedures. The first procedure was designed to represent units from the following types of missed structures (structures that had no chance of selection for AHS):

1. Structures missed in the 1970 census.
2. Structures that were completely nonresidential in the 1970 census but contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and had a utility hookup, or were on the site but not occupied on April 1, 1970, or had no utility hookup but were occupied by persons with no usual residence elsewhere.
4. Houses that had been moved onto their present site since the 1970 census.

Initially, a subsample of regular AHS units was selected at a rate of 1 in 22.2152. Then succeeding structures, in a defined path of travel to the right of the structure containing the sample unit, were listed until eight structures (excluding the sample unit structure) were found that had been eligible to be selected for the AHS. Finally, the intervening structures that did not have a chance of selection in the AHS were identified and units within these structures were interviewed. In cases where the interviewer workload would have been too great, a representative subsample of units within these structures was selected. This procedure added an estimated 181 units to the housing inventory of this SMSA.

The second procedure was designed to represent missed units from structures represented in the AHS. These missed units were:

1. Units missed in the 1970 census.
2. Units converted to residential use since the 1970 census in structures that contained some residential units in 1970.

First, a subsample of regular AHS units in multi-unit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multi-unit structures selected, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 50 units to the housing inventory of this SMSA.

Building loss sample selection.—Some tables in this report show estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS interview). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample previously described. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

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 employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at the time of the interview (i.e., the 1976 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1976 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1976 housing inventory.—The AHS estimates of characteristics of the 1976 housing inventory employed a one-stage ratio estimation procedure. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 218 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units} + \text{Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 50 noninterview cells for

sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated). In addition, a noninterview factor was computed for one noninterview cell for new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes, and one noninterview cell for sample units from both the nonpermit universe (if applicable) and the coverage improvement universe (if units were not included above).

The following ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding ratio estimation category.

The effect of this ratio estimation procedure was to somewhat reduce the variance due to the variation in sample size of strata used in the sample selection of the permit-issuing universe, thereby reducing the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

1970-1976 lost units.—The AHS estimate of characteristics of the 1970-1976 lost units employed a one-stage ratio estimation procedure which is similar to the ratio estimation procedure mentioned previously. The file of 1970-1976 lost units was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error 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 housing inventory from 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 this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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-SMSA 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 can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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 both the 1970 Census of Population and Housing and the 1976 AHS-SMSA 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 included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and

Research Program Reports, Series 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.**

AHS-SMSA.—For the 1976 AHS-SMSA 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 sample 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 accuracy of the AHS data collected from interviewed households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct 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.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview check. The results of this study are available and are presented in the following Census Bureau memorandum, "Response Error in the Annual Housing Survey as Measured by Reinterviews—Year III (1976-1977) SMSA Sample."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year III SMSA's and not for any specific SMSA):

1. For attitudinal items which were not reconciled (i.e., after the question is answered in the reinterview, the enumerator does not present the previous response and then ask the respondent to decide upon the best answer), approximately 67 percent of the indices of inconsistency showed moderate levels of response disagreement while the remaining 33 percent showed high levels.
2. Some differences beyond those due to sampling error did occur. Most of the categories affected by bias were categories of attitudinal items.

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low, indices from 20—50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

The results of this study were based on sample data so 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 this study.

Coverage errors.—With respect to errors of coverage and estimation for missing data, the AHS new construction sample had deficiencies in the representation of conventional new construction. Due to time constraints, only those building permits issued more than five months before the survey ended were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. Also, permits were sampled on a monthly basis and were divided equally among the 12 panels in which interviews were conducted. Due to this procedure, some of the permits issued in November 1975 through October 1976 were not interviewed because they were assigned to panels in which the interviewing had already been completed. In this SMSA, 6.3 percent of the permits sampled were not interviewed because of this procedure. This percentage does not include permits issued during the last 5 months of the survey. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units since, due to the relatively short time span involved, it is possible that construction of these units was not completed at the time the survey was conducted and they were not eligible for interview. In addition to these deficiencies, new construction in special places that do not require building permits, such as military bases, are also not adequately represented.

The Coverage Improvement Program also had certain deficiencies. It appears that the listing procedure used to find units missed in the 1970 census, units converted from nonresidential to residential, houses moved onto their present site, and mobile homes placed outside parks, was not very efficient for finding nonresidential conversions. Such conversions were primarily in business districts, whereas the listing procedure started from a residential unit.

It is also felt that deficiencies exist in ED's where area sampling methods are used. It had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing.

Rounding errors.—For 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 or small medians, when these figures are derived from relatively large bases (e.g., median number of persons per household). 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 the survey.

Sampling errors for the AHS-SMSA sample.—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 schedules, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variance in the estimates due to response and interviewer errors (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 error measured by the standard error, 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 in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from 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 following tables are approximations to the standard errors of various estimates shown in this report for this SMSA. 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 precise standard errors for any specific items.

Table I presents the standard errors applicable to estimates of characteristics of the 1976 housing inventory as well as estimates of characteristics of the 1970-1976 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in this table.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1976 Housing Inventory and for Estimated Number of 1970-1976 Lost Units for the Las Vegas, Nev., SMSA

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central city	Not in central city
0	30	30	30
100	60	60	50
200	80	80	80
500	120	130	120
700	150	150	140
1,000 ...	180	180	170
2,500 ...	280	280	270
5,000 ...	390	390	380
10,000 ..	540	540	530
25,000 ..	820	810	810
50,000 ..	1,060	1,000	1,070
75,000 ..	1,180	–	1,200
100,000 .	1,200	–	–

The standard errors on the AHS estimates of the population in housing units shown in tables A-1, B-1, and C-1 of part A of this report are 4,770 for the total SMSA, 3,310 for the central city of the SMSA, and 4,400 for the balance of the SMSA.

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.

Table II presents the standard errors of estimated percentages of the 1976 housing inventory as well as estimated percentages of the 1970-1976 lost units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in table II.

For ratios, 100 (x/y), where x is not a subclass of y, table II underestimates 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.—Table A-1 of part A of this report shows that in 1976 there were 75,100 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the Las Vegas, Nev., SMSA, for the Central City and for the Balance of the SMSA.

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100	24.1	24.1	24.1	24.1	24.4	28.2
200	13.7	13.7	13.7	13.7	17.3	19.9
500	6.0	6.0	6.0	7.6	10.9	12.6
700	4.3	4.3	4.6	6.4	9.2	10.7
1,000	3.1	3.1	3.9	5.3	7.7	8.9
2,500	1.3	1.3	2.5	3.4	4.9	5.6
5,000	.6	.8	1.7	2.4	3.5	4.0
10,000	.3	.6	1.2	1.7	2.4	2.8
25,000	.13	.4	.8	1.1	1.5	1.8
50,000	.06	.3	.5	.8	1.1	1.3
75,000	.04	.2	.4	.6	.9	1.0
100,000	.03	.2	.4	.5	.8	.9

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point.

1,180. The following interpolation procedure was used:

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

Size of estimate	Standard error
75,000	1,180
75,100	x
100,000	1,200

The entry for x is determined by vertically interpolating between 1,180 and 1,200.

$$75{,}100-75{,}000 = 100$$
$$100{,}000-75{,}000 = 25{,}000$$

$$1{,}180 + \frac{100}{25{,}000}(1{,}200-1{,}180) = 1{,}180$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 73,920 to 76,280 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 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 73,210 to 76,990 housing units with 90 percent confidence; and that the average estimate lies within the interval from 72,740 to 77,460 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 75,100 owner-occupied housing units, 15,900, or 21.2 percent, had two bedrooms. Interpolation in table II of the appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 21.2 percent is approximately 0.8 percentage points. The following interpolation procedure was used:

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

Base of percentage	Estimated percentage		
	10 or 90	21.2	25 or 75
75,000 ..	0.6	a	0.9
75,100 ..		p	
100,000 .	0.5	b	0.8

1. The entry for cell "a" is determined by horizontal interpolation between 0.6 and 0.9.

$$21.2-10.0 = 11.2$$
$$25.0-10.0 = 15.0$$

$$0.6 + \frac{11.2}{15.0}(0.9-0.6) = 0.82$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.5 and 0.8.

$$21.2-10.0 = 11.2$$
$$25.0-10.0 = 15.0$$

$$0.5 + \frac{11.2}{15.0}(0.8-0.5) = 0.72$$

3. The entry for "p" was then determined by vertical interpolation between 0.72 and 0.82.

$$75{,}100-75{,}000 = 100$$
$$100{,}000-75{,}000 = 25{,}000$$

$$0.82-\frac{100}{25{,}000}(0.82-0.72) = 0.8$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 20.4 to 22.0 percent; the 90-percent confidence interval is from 19.9 to 22.5 percent; and the 95-percent confidence interval is from 19.6 to 22.8 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 error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference

between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; but, 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 part A of this report shows that in 1976 there were 38,900 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 23,000. Table I shows the standard error of 15,900 is approximately 650 and the standard error of 38,900 is approximately 950. Therefore, the standard error of the estimated difference of 23,000 is about

$$1{,}150 = \sqrt{(650)^2 + (950)^2}$$

Consequently, the 68-percent confidence interval for the 23,000 difference is from 21,850 to 24,150 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 21,160 to 24,840 housing units, and the 95-percent confidence interval is from 20,700 to 25,300. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval 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 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 table II, 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 of a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 2.8. The base of the distribution from which this median was determined is 75,100 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 75,100 is approximately 1.0 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 48.0 and 52.0.
3. From table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 33,700 owner-occupied housing units, or 44.9 percent, had one or two persons (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 13,900 owner-occupied housing units, or 18.5 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:

$$2.5 + (3.5-2.5)\left(\frac{48.0-44.9}{18.5}\right) = 2.7$$

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

$$2.5 + (3.5-2.5)\left(\frac{52.0-44.9}{18.5}\right) = 2.9$$

Thus, the 95-percent confidence interval ranges from 2.7 to 2.9 persons.

Louisville 1976

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Annual Housing Survey App-42
Selection of the sample App-42
Coverage improvement sample selection App-43
Coverage improvement for deficiency 1 ..... App-43
Coverage improvement for deficiency 2 ..... App-43
Coverage improvement for deficiencies 3-6 ..... App-44
Building loss sample selection ... App-44
1970 Census of Population and Housing App-44
ESTIMATION App-44
1976 housing inventory App-44
1970-1976 lost units App-45
Ratio estimation procedure of the 1970 Census of Population and Housing App-45
RELIABILITY OF THE ESTIMATES App-45
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-46
Coverage errors App-46
Rounding errors App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences App-48
Illustration of the computation of the standard error of a difference App-49
Medians App-49
Illustration of the computation of the 95-percent confidence interval of a median .. App-49

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 20 SMSA's are based on data collected from the 1976 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. In each of the 20 SMSA's, the data were collected for the 12-month period from April 1976 through March 1977 with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and data for a second group of 21 SMSA's were collected from April 1975 through March 1976. The sample housing units for each group of the AHS SMSA's are interviewed on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the third group (1976-77) are: Houston, Tex., New York, N.Y., St. Louis, Mo.-Ill., and Seattle-Everett, Wash.

The remaining SMSA's in the third group are: Allentown-Bethlehem-Easton, Pa.-N.J., Baltimore, Md., Birmingham, Ala., Buffalo, N.Y., Cleveland, Ohio, Denver, Colo., Grand Rapids, Mich., Honolulu, Hawaii, Indianapolis, Ind., Las Vegas, Nev., Louisville, Ky.-Ind., Oklahoma City, Okla., Omaha, Nebr.-Iowa, Providence-Pawtucket-Warwick, R.I.-Mass., Raleigh, N.C., and Sacramento, Calif.

In this SMSA, 4,896 units were eligible for interview. Of these sample units, 288 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to the units eligible for interview, 363 units were visited but were not eligible for interview, because they were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit-issuing (Honolulu, Las Vegas, New York, and Sacramento) was selected from two sample frames—units enumerated in the 1970 Census of Population and Housing in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for the 16 SMSA's which are *not* 100-percent permit-issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe).

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of the sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and balance of the SMSA, since the sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total housing units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant housing unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent

associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to 1 of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—Family size	Renter—Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000-$5,999. . .		
$6,000-$9,999. . .		
$10,000-$14,999 .		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to 1 of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to 1 of the 4 vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. However, the housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of the interview, the units at each of the sample special places were listed and subsampled at a rate to produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since January 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit-issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then 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. At the time of the survey, those segments that did not have an expected size of four were further subdivided to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Coverage improvement sample selection.—The Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS-SMSA sample from the permit-issuing and new construction universes. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970, but completed after April 1, 1970.
2. Mobile homes placed in parks either missed during the 1970 census or established since the census.
3. Units missed in the 1970 census.
4. Units converted to residential use that were nonresidential at the time of the 1970 census.
5. Houses that have been moved onto their present site since the 1970 census.
6. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

Coverage improvement for deficiency 1.—A sample of new construction units whose permits were issued before January 1970, but completed after April 1970, was selected independently for each SMSA. The sampling was carried out in two stages for one-and-two-unit structures and in three stages for three-or-more-unit structures. These units were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units).

The first stage was a sample of permit offices and the second stage a sample of the 1969 permits within each of the selected permit offices. In the New York, N.Y., SMSA, Honolulu, Hawaii, SMSA, and Grand Rapids, Mich., SMSA, an additional sample of 1968 permits for three-or-more-unit structures was included in the second stage sampling. For the third stage, structures of size three or more were divided into clusters of an expected size of four units and a sample of clusters was selected. This procedure added an estimated 249 new construction units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiency 2.—In permit-issuing areas, a sample of mobile homes placed in a park missed by the census or established after the census was selected in two stages. First, for each 1976-77 SMSA, a sample of tracts was selected and canvassed. All parks were listed and then matched back to the 1970 census to identify parks missed by the census and parks established after the census. Second, the parks were divided into clusters of an expected size of four sites and a sample of clusters was selected

and interviewed. Each of the sample units represented the same number of units that the regular AHS sample units represented. This procedure added an estimated 525 units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiencies 3-6.—The remaining missed units were sampled by one of two procedures. The first procedure was designed to represent units from the following types of missed structures (structures that had no chance of selection for the AHS):

1. Structures missed in the 1970 census.
2. Structures that were completely nonresidential in the 1970 census but contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and had a utility hookup, or were on the site but not occupied on April 1, 1970, or had no utility hookup but were occupied by persons with no usual residence elsewhere.
4. Houses that had been moved onto their present site since the 1970 census.

Initially, a subsample of regular AHS units was selected at a rate of 1 in 22.2152. Then succeeding structures, in a defined path of travel to the right of the structure containing the sample unit, were listed until eight structures (excluding the sample unit structure) were found that had been eligible to be selected for the AHS. Finally, the intervening structures that did not have a chance of selection in the AHS were identified and units within these structures were interviewed. In cases where the interviewer workload would have been too great, a representative subsample of units within these structures was selected. This procedure added an estimated 2,629 units to the coverage of the housing inventory of this SMSA.

The second procedure was designed to represent missed units from structures represented in the AHS. These missed units were:

1. Units missed in the 1970 census.
2. Units converted to residential use since the 1970 census in structures that contained some residential units in 1970.

First, a subsample of regular AHS units in multi-unit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multi-unit structures selected, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 3,577 units to the coverage of the housing inventory of this SMSA.

Building loss sample selection.—Some tables in this report show estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS interview). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample previously described. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

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 employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at the time of the interview (i.e., the 1976 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1976 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1976 housing inventory.—The AHS estimates of characteristics of the 1976 housing inventory employed a one-stage ratio estimation procedure. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 288 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units} + \text{Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 50 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of 1 or more of the different strata used in the stratification of the universe as previously illustrated). In addition, a noninterview factor was computed for one noninterview cell for new construc-

tion sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes, and one noninterview cell for sample units from both the nonpermit universe (if applicable) and the coverage improvement universe (if units were not included above).

The following ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding ratio estimation category.

The effect of this ratio estimation procedure was to somewhat reduce the variance due to the variation in sample size of strata used in the sample selection of the permit-issuing universe, thereby reducing the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

1970-1976 lost units.—The AHS estimate of characteristics of the 1970-1976 lost units employed a one-stage ratio estimation procedure which is similar to the ratio estimation procedure described previously. The file of 1970-1976 lost units was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error 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 housing inventory from 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 this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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-SMSA 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 can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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 both the 1970 Census of Population and Housing and the 1976 AHS-SMSA 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 included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series 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.**

AHS-SMSA.—For the 1976 AHS-SMSA 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 sample 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 accuracy of the AHS data collected from interviewed households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct 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.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview check. The results of this study are available and are presented in the following Census Bureau memorandum, "Response Error in the Annual Housing Survey as Measured by Reinterviews—Year III (1976-1977) SMSA Sample."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year III SMSA's and not for any specific SMSA):

1. For attitudinal items which were not reconciled (i.e., after the question is answered in the reinterview, the enumerator does not present the previous response and then ask the respondent to decide upon the best answer), approximately 67 percent of the indices of inconsistency showed moderate levels of response disagreement while the remaining 33 percent showed high levels.
2. Some differences beyond those due to sampling error did occur. Most of the categories affected by bias were categories of attitudinal items.

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low; indices from 20—50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

The results of this study were based on sample data so 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 this study.

Coverage errors.—With respect to errors of coverage and estimation for missing data, the AHS new construction sample had deficiencies in the representation of conventional new construction. Due to time constraints, only those building permits issued more than 5 months before the survey ended were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. Also, permits were sampled on a monthly basis and were divided equally among the 12 panels in which interviews were conducted. Due to this procedure, some of the permits issued in November 1975 through October 1976 were not interviewed because they were assigned to panels in which the interviewing had already been completed. In this SMSA, 2.1 percent of the permits sampled were not interviewed because of this procedure. This percentage does not include permits issued during the last 5 months of the survey. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units since, due to the relatively short time-span involved, it is possible that construction of these units was not completed at the time the survey was conducted and they were not eligible for interview. In addition to these deficiencies, new construction in special places that do not require building permits, such as military bases, are also not adequately represented.

The Coverage Improvement Program also had certain deficiencies. It appears that the listing procedure used to find units missed in the 1970 census, units converted from nonresidential to residential, houses moved onto their present site, and mobile homes placed outside parks, was not very efficient for finding nonresidential conversions. Such conversions were primarily in business districts, whereas the listing procedure started from a residential unit.

Deficiencies also exist in ED's where area sampling methods are used. It had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing.

Rounding errors.—For 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 or small medians, when these figures are derived from relatively large bases (e.g., median number of persons per household). 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 the survey.

Sampling errors for the AHS-SMSA sample.—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 schedules, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The stand-

ard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variance in the estimates due to response and interviewer errors (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 error measured by the standard error, 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 in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from 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 following tables are approximations to the standard errors of various estimates shown in this report for this SMSA. 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 precise standard errors for any specific items.

Table I presents the standard errors applicable to estimates of characteristics of the 1976 housing inventory as well as estimates of characteristics of the 1970-1976 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in this table. The standard errors on the AHS estimates of the population in housing units shown in tables A-1, B-1, and C-1 of part A of this report are 9,730 for the total SMSA, 5,650 for the central city of the SMSA, and 7,470 for the balance of the SMSA.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1976 Housing Inventory and for Estimated Number of 1970-1976 Lost Units for the Louisville, Ky.-Ind., SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central city	Not in central city
0	70	70	70
100	90	80	90
200	120	120	120
500	190	190	190
700	230	220	230
1,000 . . .	270	260	270
2,500 . . .	430	420	430
5,000 . . .	610	580	610
10,000 . .	860	810	850
25,000 . .	1,330	1,230	1,330
50,000 . .	1,810	1,580	1,810
100,000 .	2,380	1,730	2,390
150,000 .	2,670	–	2,690
200,000 .	2,770	–	–
250,000 .	2,700	–	–
300,000 .	2,450	–	–

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.

Table II presents the standard errors of estimated percentages of the 1976 housing inventory as well as estimated percentages of the 1970-1976 lost units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in table II.

Included in tables I and II are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered as 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.

For ratios, 100 (x/y), where x is not a subclass of y, table II underestimates 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.—Table A-1 of part A of this report shows that in 1976 there were

189,600 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 2,750. The following interpolation procedure was used:

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

Size of estimate	Standard error
150,000	2,670
189,600	x
200,000	2,770

The entry for x is determined by vertically interpolating between 2,670 and 2,770.

$$189{,}600-150{,}000 = 39{,}600$$
$$200{,}000-150{,}000 = 50{,}000$$

$$2{,}670 + \frac{39{,}600}{50{,}000}(2{,}770-2{,}670) = 2{,}750$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 186,850 to 192,350 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 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 185,200 to 194,000 housing units with 90 percent confidence; and that the average estimate lies within the interval from 184,100 to 195,100 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 189,600 owner-occupied housing units, 50,200, or 26.5 percent, had two bedrooms. Interpolation in table II of the appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 26.5 percent is approximately 0.9 percentage points. The following interpolation procedure was used:

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

Base of percentage	Estimated percentage		
	25 or 75	26.5	50
150,000 ...	1.0	a	1.1
189,600 ...		p	
200,000 ...	.8	b	1.0

1. The entry for cell "a" is determined by horizontal interpolation between 1.0 and 1.1.

$$26.5-25.0 = 1.5$$
$$50.0-25.0 = 25.0$$

$$1.0 + \frac{1.5}{25.0}(1.1-1.0) = 1.01$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.8 and 1.0.

$$26.5-25.0 = 1.5$$
$$50.0-25.0 = 25.0$$

$$0.8 + \frac{1.5}{25.0}(1.0-0.8) = .81$$

3. The entry for "p" was then determined by vertical interpolation between .81 and 1.01.

$$189{,}600-150{,}000 = 39{,}600$$
$$200{,}000-150{,}000 = 50{,}000$$

$$1.01-\frac{39{,}600}{50{,}000}(1.01-.81) = 0.9$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 25.6 to 27.4 percent; the 90-percent confidence interval is from 25.1 to 27.9 percent; and the 95-percent confidence interval is from 24.7 to 28.3 percent.

Differences.—The standard errors shown are not directly applicable to differences

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the Louisville, Ky.-Ind., SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100	42.8	42.8	42.8	42.8	42.8	43.3
200	27.3	27.3	27.3	27.3	27.3	30.6
500	13.0	13.0	13.0	13.0	16.8	19.4
700	9.7	9.7	9.7	9.8	14.2	16.4
1,000	7.0	7.0	7.0	8.2	11.9	13.7
2,500	2.9	2.9	3.8	5.2	7.5	8.7
5,000	1.5	1.5	2.7	3.7	5.3	6.1
10,000	.7	.9	1.9	2.6	3.7	4.3
25,000	.3	.5	1.2	1.6	2.4	2.7
50,000	.15	.4	.8	1.2	1.7	1.9
100,000 ...	.07	.3	.6	.8	1.2	1.4
150,000 ...	.05	.2	.5	.7	1.0	1.1
200,000 ...	.04	.2	.4	.6	.8	1.0
250,000 ...	.03	.2	.4	.5	.7	.9
300,000 ...	.02	.2	.3	.5	.7	.8

[1] Standard errors are presented to nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to nearest one-hundredth of one percentage point.

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 error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; but, 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 part A of this report shows that in 1976 there were 98,300 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 48,100. Table I shows the standard error of 50,200 is approximately 1,810 and the standard error of 98,300 is approximately 2,360. Therefore, the standard error of the estimated difference of 48,100 is about:

$$2{,}970 = \sqrt{(1{,}810)^2 + (2{,}360)^2}$$

Consequently, the 68-percent confidence interval for the 48,100 difference is from 45,130 to 51,070 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 43,350 to 52,850 housing units, and the 95-percent confidence interval is from 42,160 to 54,040. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval 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 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 table II 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 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 of a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 2.9. The base of the distribution from which this median was determined is 189,600 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 189,600 is approximately 1.0 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 48.0 and 52.0.
3. From table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 80,600 owner-occupied housing units, or 42.5 percent, had one or two persons (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 38,000 owner-occupied housing units, or 20.0 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:

$$2.5 + (3.5-2.5)\left(\frac{48.0-42.5}{20.0}\right) = 2.8$$

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

$$2.5 + (3.5-2.5)\left(\frac{52.0-42.5}{20.0}\right) = 3.0$$

Thus, the 95-percent confidence interval ranges from 2.8 to 3.0 persons.

New York 1976

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Annual Housing Survey App-42
Selection of the sample App-42
Coverage improvement sample selection App-43
Coverage improvement for deficiency 1 ..... App-43
Coverage improvement for deficiency 2 ..... App-43
Coverage improvement for deficiencies 3-6 ..... App-44
Building loss sample selection ... App-44
1970 Census of Population and Housing App-44
ESTIMATION App-44
1976 housing inventory App-44
1970-1976 lost units App-45
Ratio estimation procedure of the 1970 Census of Population and Housing App-45
RELIABILITY OF THE ESTIMATES App-45
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-46
Coverage errors App-46
Rounding errors App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-48
Differences App-49
Illustration of the computation of the standard error of a difference App-49
Medians App-49
Illustration of the computation of the 95-percent confidence interval of a median .. App-50

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 20 SMSA's are based on data collected from the 1976 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. In each of the 20 SMSA's, the data were collected for the 12-month period from April 1976 through March 1977 with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and data for a second group of 21 SMSA's were collected from April 1975 through March 1976. The sample housing units for each group of the AHS SMSA's are interviewed on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the third group (1976-77) are: Houston, Tex., New York, N.Y., St. Louis, Mo.-Ill., and Seattle-Everett, Wash.

The remaining SMSA's in the third group are: Allentown-Bethlehem-Easton, Pa.-N.J., Baltimore, Md., Birmingham, Ala., Buffalo, N.Y., Cleveland, Ohio, Denver, Colo., Grand Rapids, Mich., Honolulu, Hawaii, Indianapolis, Ind., Las Vegas, Nev., Louisville, Ky.-Ind., Oklahoma City, Okla., Omaha, Nebr.-Iowa, Providence-Pawtucket-Warwick, R.I.-Mass., Raleigh, N.C., and Sacramento, Calif.

In this SMSA, 14,760 units were eligible for interview. Of these sample units, 1,309 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to the units eligible for interview, 1,052 units were visited but were not eligible for interview, because they were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit-issuing (Honolulu, Las Vegas, New York, and Sacramento) was selected from two sample frames—units enumerated in the 1970 Census of Population and Housing in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for the 16 SMSA's which are *not* 100-percent permit-issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe).

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of the sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA, since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and balance of the SMSA, since the sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total housing units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant housing unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent

associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to 1 of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—Family size	Renter—Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000-$5,999. . .		
$6,000-$9,999. . .		
$10,000-$14,999 .		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to 1 of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to 1 of the 4 vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. However, the housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of the interview, the units at each of the sample special places were listed and subsampled at a rate to produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since January 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit-issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then 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. At the time of the survey, those segments that did not have an expected size of four were further subdivided to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Coverage improvement sample selection.—The Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS-SMSA sample from the permit-issuing and new construction universes. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970, but completed after April 1, 1970.
2. Mobile homes placed in parks either missed during the 1970 census or established since the census.
3. Units missed in the 1970 census.
4. Units converted to residential use that were nonresidential at the time of the 1970 census.
5. Houses that have been moved onto their present site since the 1970 census.
6. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

Coverage improvement for deficiency 1.—A sample of new construction units whose permits were issued before January 1970, but completed after April 1970, was selected independently for each SMSA. The sampling was carried out in two stages for one-and-two-unit structures and in three stages for three-or-more-unit structures. These units were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units).

The first stage was a sample of permit offices and the second stage a sample of the 1969 permits within each of the selected permit offices. In the New York, N.Y., SMSA, Honolulu, Hawaii, SMSA, and Grand Rapids, Mich., SMSA, an additional sample of 1968 permits for three-or-more-unit structures was included in the second stage sampling. For the third stage, structures of size three or more were divided into clusters of an expected size of four units and a sample of clusters was selected. This procedure added an estimated 10,540 new construction units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiency 2—In permit-issuing areas, a sample of mobile homes placed in a park missed by the census or established after the census was selected in two stages. First, for each 1976-77 SMSA, a sample of tracts was selected and canvassed. All parks were listed and then matched back to the 1970 census to identify parks missed by the census and parks established after the census. Second, the parks were divided

into clusters of an expected size of four sites and a sample of clusters was selected and interviewed. Each of the sample units represented the same number of units that the regular AHS sample units represented. This procedure yielded no additional units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiencies 3-6.—The remaining missed units were sampled by one of two procedures. The first procedure was designed to represent units from the following types of missed structures (structures that had no chance of selection for the AHS):

1. Structures missed in the 1970 census.
2. Structures that were completely nonresidential in the 1970 census but contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and had a utility hookup, or were on the site but not occupied on April 1, 1970, or had no utility hookup but were occupied by persons with no usual residence elsewhere.
4. Houses that had been moved onto their present site since the 1970 census.

Initially, a subsample of regular AHS units was selected at a rate of 1 in 22.2152. Then succeeding structures, in a defined path of travel to the right of the structure containing the sample unit, were listed until eight structures (excluding the sample unit structure) were found that had been eligible to be selected for the AHS. Finally, the intervening structures that did not have a chance of selection in the AHS were identified and units within these structures were interviewed. In cases where the interviewer workload would have been too great, a representative subsample of units within these structures was selected. This procedure added an estimated 73,171 units to the coverage of the housing inventory of this SMSA.

The second procedure was designed to represent missed units from structures represented in the AHS. These missed units were:

1. Units missed in the 1970 census.
2. Units converted to residential use since the 1970 census in structures that contained some residential units in 1970.

First, a subsample of regular AHS units in multi-unit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multi-unit structures selected, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 14,651 units to the coverage of the housing inventory of this SMSA.

Building loss sample selection.—Some tables in this report show estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS interview). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample previously described. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

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 employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at the time of the interview (i.e., the 1976 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1976 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1976 housing inventory.—The AHS estimates of characteristics of the 1976 housing inventory employed a one-stage ratio estimation procedure. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 1,309 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units} + \text{Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 50 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of 1 or more of the different strata used in the stratification of the universe as previously illustrated). In addition, a

noninterview factor was computed for one noninterview cell for new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes, and one noninterview cell for sample units from both the nonpermit universe (if applicable) and the coverage improvement universe (if units were not included above).

The following ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding ratio estimation category.

The effect of this ratio estimation procedure was to somewhat reduce the variance due to the variation in sample size of strata used in the sample selection of the permit-issuing universe, thereby reducing the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

1970-1976 lost units.—The AHS estimate of characteristics of the 1970-1976 lost units employed a one-stage ratio estimation procedure which is similar to the ratio estimation procedure described previously. The file of 1970-1976 lost units was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error 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 housing inventory from 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 this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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-SMSA 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 can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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 both the 1970 Census of Population and Housing and the 1976 AHS-SMSA 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 included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census,** and PHC(E)-10, **Accu-**

racy of Data for Selected Housing Characteristics as Measured by Reinterviews.

AHS-SMSA.—For the 1976 AHS-SMSA 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 sample 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 accuracy of the AHS data collected from interviewed households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct 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.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview check. The results of this study are available and are presented in the following Census Bureau memorandum, "Response Error in the Annual Housing Survey as Measured by Reinterviews—Year III (1976-77) SMSA Sample."

Some of the results of this study are (note that these results are based on the reinterviews across all Year III SMSA's and not for any specific SMSA):

1. For attitudinal items which were not reconciled (i.e., after the question is answered in the reinterview, the enumerator does not present the previous response and then ask the respondent to decide upon the best answer), approximately 67 percent of the indices of inconsistency showed moderate levels of response disagreement while the remaining 33 percent showed high levels.
2. Some differences beyond those due to sampling error did occur. Most of the categories affected by bias were categories of attitudinal items.

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low, indices from 20—50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

The results of this study were based on sample data so 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 this study.

Coverage errors.—With respect to errors of coverage and estimation for missing data, the AHS new construction sample had deficiencies in the representation of conventional new construction. Due to time constraints, only those building permits issued more than 5 months before the survey ended were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. Also, permits were sampled on a monthly basis and were divided equally among the 12 panels in which interviews were conducted. Due to this procedure, some of the permits issued in November 1975 through October 1976 were not interviewed because they were assigned to panels in which the interviewing had already been completed. In this SMSA, 3.5 percent of the permits sampled were not interviewed because of this procedure. This percentage does not include permits issued during the last 5 months of the survey. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units since, due to the relatively short time span involved, it is possible that construction of these units was not completed at the time the survey was conducted and they were not eligible for interview. In addition to these deficiencies, new construction in special places that do not require building permits, such as military bases, are also not adequately represented.

The Coverage Improvement Program also had certain deficiencies. It appears that the listing procedure used to find units missed in the 1970 census, units converted from nonresidential to residential, houses moved onto their present site, and mobile homes placed outside parks, was not very efficient for finding nonresidential conversions. Such conversions were primarily in business districts, whereas the listing procedure started from a residential unit.

Deficiencies also exist in ED's where area sampling methods are used. It had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing.

Rounding errors.—For 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 or small medians, when these figures are derived from relatively large bases (e.g., median number of persons per household). 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 the survey.

Sampling errors for the AHS-SMSA sample.—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 schedules, instructions, and interviewers were used, estimates from each of the different samples would differ from each other.

The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variance in the estimates due to response and interviewer errors (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 error measured by the standard error, 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 in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from 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 following tables are approximations to the standard errors of various estimates shown in this report for this SMSA. 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 precise standard errors for any specific items.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1976 Housing Inventory and for Estimated Number of 1970-1976 Lost Units for the New York, N.Y., SMSA, for the Central City of the SMSA, and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard error[1]		
	SMSA	In central city	Not in central city
0	420	510	190
100	420	510	190
200	420	510	190
500	460	510	310
700	540	600	360
1,000	650	710	430
2,500	1,030	1,130	690
5,000	1,450	1,600	970
10,000	2,050	2,260	1,370
25,000	3,240	3,560	2,160
50,000	4,570	5,020	3,030
75,000	5,590	6,130	3,680
100,000	6,440	7,060	4,220
150,000	7,850	8,590	5,100
200,000	9,020	9,850	5,800
250,000	10,040	10,940	6,390
500,000	13,860	14,950	8,310
1,000,000	18,620	19,560	9,370
1,200,000	19,940	20,690	9,020
1,500,000	21,520	21,840	–
2,000,000	23,280	22,530	–
2,500,000	24,150	21,760	–
3,000,000	24,220	–	–
3,500,000	23,510	–	–
4,000,000	21,920	–	–

[1] For estimates pertaining to new construction, standard errors shown in table should be multiplied by a factor of 1.2.

Table I presents the standard errors applicable to estimates of characteristics of the 1976 housing inventory as well as estimates of characteristics of the 1970-1976 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in this table. The standard errors on the AHS estimates of the population in housing units shown in tables A-1, B-1, and C-1 of part A of this report are 87,770 for the total SMSA, 78,650 for the central city of the SMSA, and 38,920 for the balance of the SMSA.

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 II through IV present the standard errors of estimated percentages of the 1976 housing inventory as well as estimated percentages of the 1970-1976 lost units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables II through IV.

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 as 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.

For ratios, 100 (x/y), where x is not a subclass of y, tables II through 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:

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the New York, N.Y., SMSA

(68 chances out of 100)

Base of percetage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
500	45.8	45.8	45.8	45.8	45.8	45.9
700	37.6	37.6	37.6	37.6	37.6	38.8
1,000	29.7	29.7	29.7	29.7	29.7	32.5
2,500	14.4	14.4	14.4	14.4	17.8	20.5
5,000	7.8	7.8	7.8	8.7	12.6	14.5
10,000	4.0	4.0	4.5	6.2	8.9	10.3
25,000	1.7	1.7	2.8	3.9	5.6	6.5
50,000	.8	.9	2.0	2.8	4.0	4.6
75,000	.6	.7	1.6	2.3	3.2	3.8
100,000 ...	.4	.6	1.4	1.9	2.8	3.2
150,000 ...	.3	.5	1.2	1.6	2.3	2.7
200,000 ...	.2	.5	1.0	1.4	2.0	2.3
250,000 ...	.2	.4	.9	1.2	1.8	2.1
500,000 ...	.08	.3	.6	.9	1.3	1.5
1,000,000 ..	.04	.2	.4	.6	.9	1.0
1,200,000 ..	.04	.2	.4	.6	.8	.9
1,500,000 ..	.03	.2	.4	.5	.7	.8
2,000,000 ..	.02	.14	.3	.4	.6	.7
2,500,000 ..	.02	.13	.3	.4	.6	.6
3,000,000 ..	.01	.12	.3	.4	.5	.6
3,500,000 ..	.01	.11	.2	.3	.5	.5
4,000,000 ..	.01	.10	.2	.3	.4	.5

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point. For estimates pertaining to new construction, standard errors shown in table should be multiplied by a factor of 1.2.

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

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.—Table A-1 of part A of this report shows that in 1976 there were 1,548,100 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 21,690. The following interpolation procedure was used:

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

Size of estimate	Standard error
1,500,000	21,520
1,548,100	x
2,000,000	23,280

The entry for x_{II} is determined by vertically interpolating between 21,520 and 23,280.

1,548,100—1,500,000 = 48,100
2,000,000—1,500,000 = 500,000

$$21{,}520 + \frac{48{,}100}{500{,}000}(21{,}520-23{,}280) = 21{,}690$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 1,526,410 to 1,569,790 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 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 1,513,400 to 1,582,800 housing units with 90 percent confidence; and that the average estimate lies within the interval from 1,504,720 to 1,591,480 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 1,548,100 owner-occupied housing units, 351,200, or 22.7 percent, had two bedrooms. Interpolation in table II of the appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 22.7 percent is approximately 0.7 percentage points. The following interpolation procedure was used:

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

Base of percentage	Estimated percentage		
	10 or 90	22.7	25 or 75
1,500,000 ..	.5	a	.7
1,548,000 ..		p	
2,000,000 ..	.4	b	.6

1. The entry for cell "a" is determined by horizontal interpolation between 0.5 and 0.7.

22.7—10.0 = 12.7
25.0—10.0 = 15.0

$$0.5 + \frac{12.7}{15.0}(0.7-0.5) = .67$$

TABLE III. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Housing Units for the Central City of the New York, N.Y., SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
500	50.6	50.6	50.6	50.6	50.6	50.6
700	42.2	42.2	42.2	42.2	42.2	42.7
1,000	33.8	33.8	33.8	33.8	33.8	35.8
2,500	17.0	17.0	17.0	17.0	19.6	22.6
5,000	9.3	9.3	9.3	9.6	13.8	16.0
10,000	4.9	4.9	4.9	6.8	9.8	11.3
25,000	2.0	2.0	3.1	4.3	6.2	7.2
50,000	1.0	1.0	2.2	3.0	4.4	5.1
75,000	.7	.8	1.8	2.5	3.6	4.1
100,000 ...	.5	.7	1.6	2.1	3.1	3.6
150,000 ...	.3	.6	1.3	1.8	2.5	2.9
200,000 ...	.3	.5	1.1	1.5	2.2	2.5
250,000 ...	.2	.4	1.0	1.4	2.0	2.3
500,000 ...	.10	.3	.7	1.0	1.4	1.6
1,000,000 ..	.05	.2	.5	.7	1.0	1.1
1,200,000 ..	.04	.2	.4	.6	.9	1.0
1,500,000 ..	.03	.2	.4	.6	.8	.9
2,000,000 ..	.03	.2	.3	.5	.7	.8
2,500,000 ..	0.2	.14	.3	.4	.6	.7

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point. For estimates pertaining to new construction, standard errors shown in table should be multiplied by a factor of 1.2.

2. The entry for cell "b" is determined by horizontal interpolation between 0.4 and 0.6.

$$22.7-10.0 = 12.7$$
$$25.0-10.0 = 15.0$$

$$0.4 + \frac{12.7}{15.0}(0.6-0.4) = .57$$

3. The entry for "p" was then determined by vertical interpolation between .57 and .67.

$$1{,}548{,}100-1{,}500{,}000 = 48{,}100$$
$$2{,}000{,}000-1{,}500{,}000 = 500{,}000$$

$$.67-\frac{48{,}100}{500{,}000}(.67-.57) = 0.7$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 22.0 to 23.4 percent; the 90-percent confidence interval is from 21.6 to 23.8 percent; and the 95-percent confidence interval is from 21.3 to 24.1 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 error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; but, 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 part A of this report shows that in 1976 there were 712,900 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 361,700. Table I shows the standard error of 351,200 is approximately 11,590 and the standard error of 712,900 is approximately 15,890. Therefore, the standard error of the estimated difference of 361,700 is about:

$$19{,}670 = \sqrt{(15{,}890)^2 + (11{,}590)^2}$$

Consequently, the 68-percent confidence interval for the 361,700 difference is from 342,030 to 381,370 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 330,230 to 393,170 housing units, and the 95-percent confidence interval is from 322,360 to 401,040. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval 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 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

TABLE IV. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Housing Units for the Balance of the New York, N.Y., SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
200	48.5	48.5	48.5	48.5	48.5	48.5
500	27.4	27.4	27.4	27.4	27.4	30.7
700	21.2	21.2	21.2	21.2	22.5	25.9
1,000	15.9	15.9	15.9	15.9	18.8	21.7
2,500	7.0	7.0	7.0	8.2	11.9	13.7
5,000	3.6	3.6	4.2	5.8	8.4	9.7
10,000	1.8	1.8	3.0	4.1	5.9	6.9
25,000	.7	.9	1.9	2.6	3.8	4.3
50,000	.4	.6	1.3	1.8	2.7	3.1
75,000	.3	.5	1.1	1.5	2.2	2.5
100,000 ...	.2	.4	.9	1.3	1.9	2.2
150,000 ...	.13	.4	.8	1.1	1.5	1.8
200,000 ...	.09	.3	.7	.9	1.3	1.5
250,000 ...	.08	.3	.6	.8	1.2	1.4
500,000 ...	.04	.2	.4	.6	.8	1.0
1,000,000 ..	.02	.13	.3	.4	.6	.7
1,200,000 ..	.02	.12	.3	.4	.5	.6

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point. For estimates pertaining to new construction, standard errors shown in table should be multiplied by a factor of 1.2.

confidence limits of a median based on sample data:

1. From table II 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 of a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 3.2. The base of the distribution from which this median was determined is 1,548,100 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 1,548,100 is approximately 0.8 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 48.4 and 51.6.
3. From table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 581,100 owner-occupied housing units, or 37.5 percent, had one or two persons (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 280,100 owner-occupied housing units, or 18.1 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:

$$2.5 + (3.5-2.5)\left(\frac{48.4-37.5}{18.1}\right) = 3.1$$

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

$$2.5 + (3.5-2.5)\left(\frac{51.6-37.5}{18.1}\right) = 3.3$$

Thus, the 95-percent confidence interval ranges from 3.1 to 3.3 persons.

Oklahoma City 1976

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Annual Housing Survey App-42
Selection of the sample App-42
Coverage improvement sample selection App-43
Coverage improvement for deficiency 1 ..... App-43
Coverage improvement for deficiency 2 ..... App-43
Coverage improvement for deficiencies 3-6 ..... App-44
Building loss sample selection ... App-44
1970 Census of Population and Housing App-44
ESTIMATION App-44
1976 housing inventory App-44
1970-1976 lost units App-45
Ratio estimation procedure of the 1970 Census of Population and Housing App-45
RELIABILITY OF THE ESTIMATES App-45
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-46
Coverage errors App-46
Rounding errors App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-48
Differences App-49
Illustration of the computation of the standard error of a difference App-49
Medians App-49
Illustration of the computation of the 95-percent confidence interval of a median .. App-49

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 20 SMSA's are based on data collected from the 1976 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. In each of the 20 SMSA's, the data were collected for the 12-month period from April 1976 through March 1977 with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and data for a second group of 21 SMSA's were collected from April 1975 through March 1976. The sample housing units for each group of the AHS SMSA's are interviewed on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the third group (1976-77) are: Houston, Tex., New York, N.Y., St. Louis, Mo.-Ill., and Seattle-Everett, Wash.

The remaining SMSA's in the third group are: Allentown-Bethlehem-Easton, Pa.-N.J., Baltimore, Md., Birmingham, Ala., Buffalo, N.Y., Cleveland, Ohio, Denver, Colo., Grand Rapids, Mich., Honolulu, Hawaii, Indianapolis, Ind., Las Vegas, Nev., Louisville, Ky.-Ind., Oklahoma City, Okla., Omaha, Nebr.-Iowa, Providence-Pawtucket-Warwick, R.I.-Mass., Raleigh, N.C., and Sacramento, Calif.

In this SMSA, 4,784 units were eligible for interview. Of these sample units, 222 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to the units eligible for interview, 536 units were visited but were not eligible for interview, because they were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit-issuing (Honolulu, Las Vegas, New York, and Sacramento) was selected from two sample frames—units enumerated in the 1970 Census of Population and Housing in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for the 16 SMSA's which are *not* 100-percent permit-issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe).

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of the sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and balance of the SMSA, since the sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total housing units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant housing unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit

records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to 1 of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—Family size	Renter—Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000-$5,999. . .		
$6,000-$9,999. . .		
$10,000-$14,999 .		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to 1 of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to 1 of the 4 vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. However, the housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of the interview, the units at each of the sample special places were listed and subsampled at a rate to produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since January 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit-issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then 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. At the time of the survey, those segments that did not have an expected size of four were further subdivided to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Coverage improvement sample selection.—The Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS-SMSA sample from the permit-issuing and new construction universes. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970, but completed after April 1, 1970.
2. Mobile homes placed in parks either missed during the 1970 census or established since the census.
3. Units missed in the 1970 census.
4. Units converted to residential use that were nonresidential at the time of the 1970 census.
5. Houses that have been moved onto their present site since the 1970 census.
6. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

Coverage improvement for deficiency 1.—A sample of new construction units whose permits were issued before January 1970, but completed after April 1970, was selected independently for each SMSA. The sampling was carried out in two stages for one-and-two-unit structures and in three stages for three-or-more-unit structures. These units were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units).

The first stage was a sample of permit offices and the second stage a sample of the 1969 permits within each of the selected permit offices. In the New York, N.Y., SMSA, Honolulu, Hawaii, SMSA, and Grand Rapids, Mich., SMSA, an additional sample of 1968 permits for three-or-more-unit structures was included in the second stage sampling. For the third stage, structures of size three or more were divided into clusters of an expected size of four units and a sample of clusters was selected. This procedure added an estimated 2,382 new construction units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiency 2.—In permit-issuing areas, a sample of mobile homes placed in a park missed by the census or established after the census was selected in two stages. First, for each 1976-77 SMSA, a sample of tracts was selected and canvassed. All parks were listed and then matched back to the 1970 census to identify parks missed by the census and parks established after the census. Second, the parks were divided into clusters of an expected size of four

sites and a sample of clusters was selected and interviewed. Each of the sample units represented the same number of units that the regular AHS sample units represented. This procedure added an estimated 419 units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiencies 3-6.—The remaining missed units were sampled by one of two procedures. The first procedure was designed to represent units from the following types of missed structures (structures that had no chance of selection for the AHS):

1. Structures missed in the 1970 census.
2. Structures that were completely nonresidential in the 1970 census but contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and had a utility hookup, or were on the site but not occupied on April 1, 1970, or had no utility hookup but were occupied by persons with no usual residence elsewhere.
4. Houses that had been moved onto their present site since the 1970 census.

Initially, a subsample of regular AHS units was selected at a rate of 1 in 22.2152. Then succeeding structures, in a defined path of travel to the right of the structure containing the sample unit, were listed until eight structures (excluding the sample unit structure) were found that had been eligible to be selected for the AHS. Finally, the intervening structures that did not have a chance of selection in the AHS were identified and units within these structures were interviewed. In cases where the interviewer workload would have been too great, a representative subsample of units within these structures was selected. This procedure added an estimated 2,321 units to the coverage of the housing inventory of this SMSA.

The second procedure was designed to represent missed units from structures represented in the AHS. These missed units were:

1. Units missed in the 1970 census.
2. Units converted to residential use since the 1970 census in structures that contained some residential units in 1970.

First, a subsample of regular AHS units in multi-unit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multi-unit structures selected, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 532 units to the coverage of the housing inventory of this SMSA.

Building loss sample selection.—Some tables in this report show estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS interview). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample previously described. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

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 employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at the time of the interview (i.e., the 1976 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1976 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1976 housing inventory.—The AHS estimates of characteristics of the 1976 housing inventory employed a one-stage ratio estimation procedure. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 222 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units} + \text{Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 50 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of 1 or more of the different strata used in the stratification of the universe as previously illustrated). In addition, a

noninterview factor was computed for one noninterview cell for new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes, and one noninterview cell for sample units from both the nonpermit universe (if applicable) and the coverage improvement universe (if units were not included above).

The following ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding ratio estimation category.

The effect of this ratio estimation procedure was to somewhat reduce the variance due to the variation in sample size of strata used in the sample selection of the permit-issuing universe, thereby reducing the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

1970-1976 lost units.—The AHS estimate of characteristics of the 1970-1976 lost units employed a one-stage ratio estimation procedure which is similar to the ratio estimation procedure described previously. The file of 1970-1976 lost units was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error 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 housing inventory from 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 this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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-SMSA 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 can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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 both the 1970 Census of Population and Housing and the 1976 AHS-SMSA 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 included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series 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.

AHS-SMSA.—For the 1976 AHS-SMSA 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 sample 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 accuracy of the AHS data collected from interviewed households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct 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.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview check. The results of this study are available and are presented in the following Census Bureau memorandum, "Response Error in the Annual Housing Survey as Measured by Reinterviews—Year III (1976-1977) SMSA Sample."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year III SMSA's and not for any specific SMSA):

1. For attitudinal items which were not reconciled (i.e., after the question is answered in the reinterview, the enumerator does not present the previous response and then ask the respondent to decide upon the best answer), approximately 67 percent of the indices of inconsistency showed moderate levels of response disagreement while the remaining 33 percent showed high levels.
2. Some differences beyond those due to sampling error did occur. Most of the categories affected by bias were categories of attitudinal items.

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low; indices from 20—50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

The results of this study were based on sample data so 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 this study.

Coverage errors.—With respect to errors of coverage and estimation for missing data, the AHS new construction sample had deficiencies in the representation of conventional new construction. Due to time constraints, only those building permits issued more than 5 months before the survey ended were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. Also, permits were sampled on a monthly basis and were divided equally among the 12 panels in which interviews were conducted. Due to this procedure, some of the permits issued in November 1975 through October 1976 were not interviewed because they were assigned to panels in which the interviewing had already been completed. In this SMSA, 4.1 percent of the permits sampled were not interviewed because of this procedure. This percentage does not include permits issued during the last 5 months of the survey. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units since, due to the relatively short time-span involved, it is possible that construction of these units was not completed at the time the survey was conducted and they were not eligible for interview. In addition to these deficiencies, new construction in special places that do not require building permits, such as military bases, are also not adequately represented.

The Coverage Improvement Program also had certain deficiencies. It appears that the listing procedure used to find units missed in the 1970 census, units converted from nonresidential to residential, houses moved onto their present site, and mobile homes placed outside parks, was not very efficient for finding nonresidential conversions. Such conversions were primarily in business districts, whereas the listing procedure started from a residential unit.

Deficiencies also exist in ED's where area sampling methods are used. It had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing.

Rounding errors.—For 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 or small medians, when these figures are derived from relatively large bases (e.g., median number of persons per household). 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 the survey.

Sampling errors for the AHS-SMSA sample.—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 schedules, instructions, and interviewers were

used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variance in the estimates due to response and interviewer errors (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 error measured by the standard error, 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 in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from 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 following tables are approximations to the standard errors of various estimates shown in this report for this SMSA. 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 precise standard errors for any specific items.

Table I presents the standard errors applicable to estimates of characteristics of the 1976 housing inventory as well as estimates of characteristics of the 1970-1976 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in this table. The standard errors on the AHS estimates of the population in housing units shown in tables A-1, B-1, and C-1 of part A of this report are 9,580 for the total SMSA, 7,430 for the central city of the SMSA, and 9,190 for the balance of the SMSA.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1976 Housing Inventory and for Estimated Number of 1970-1976 Lost Units for the Oklahoma City, Okla., SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central city	Not in central city
0	70	70	80
100	90	80	90
200	120	120	130
500	190	190	200
700	230	220	240
1,000 ...	270	270	280
2,500 ...	430	420	450
5,000 ...	600	590	630
10,000 ..	850	830	880
25,000 ..	1,310	1,280	1,360
50,000 ..	1,790	1,700	1,830
75,000 ..	2,110	1,960	2,120
100,000 .	2,330	2,100	2,300
150,000 .	2,580	2,130	—
200,000 .	2,620	—	—
250,000 .	2,480	—	—

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.

Table II presents the standard errors of estimated percentages of the 1976 housing inventory as well as estimated percentages of the 1970-1976 lost units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in table II.

Included in tables I and II are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered as 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.

For ratios, 100 (x/y), where x is not a subclass of y, table II underestimates 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

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the Oklahoma City, Okla., SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100	44.6	44.6	44.6	44.6	44.6	44.8
200	28.7	28.7	28.7	28.7	28.7	31.7
500	13.9	13.9	13.9	13.9	17.4	20.1
700	10.3	10.3	10.3	10.3	14.7	17.0
1,000	7.4	7.4	7.4	8.5	12.3	14.2
2,500	3.1	3.1	3.9	5.4	7.8	9.0
5,000	1.6	1.6	2.8	3.8	5.5	6.3
10,000	.8	.9	2.0	2.7	3.9	4.5
25,000	.3	.6	1.2	1.7	2.5	2.8
50,000	.2	.4	.9	1.2	1.7	2.0
75,000	.11	.3	.7	1.0	1.4	1.6
100,000	.08	.3	.6	.9	1.2	1.4
150,000	.05	.2	.5	.7	1.0	1.2
200,000	.04	.2	.4	.6	.9	1.0
250,000	.03	.2	.4	.5	.8	.9

[1] Standard errors are presented to nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to nearest one-hundredth of one percentage point.

Illustration of the use of the standard error tables.—Table A-1 of part A of this report shows that in 1976 there were 160,500 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 2,590. The following interpolation procedure was used:

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

Size of estimate	Standard error
150,000	2,580
160,500	x
200,000	2,620

The entry for x is determined by vertically interpolating between 2,580 and 2,620.

$$160{,}500-150{,}000 = 10{,}500$$
$$200{,}000-150{,}000 = 50{,}000$$

$$2{,}580 + \frac{10{,}500}{50{,}000}(2{,}620-2{,}580) = 2{,}590$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 157,910 to 163,090 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 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 156,360 to 164,640 housing units with 90 percent confidence; and that the average estimate lies within the interval from 155,320 to 165,680 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 160,500 owner-occupied housing units, 48,500, or 30.2 percent, had two bedrooms. Interpolation in table II of the appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 30.2 percent is approximately 1.0 percentage points. The following interpolation procedure was used:

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

Base of percentage	Estimated percentage		
	25 or 75	30.2	50
150,000 .	1.0	a	1.2
160,500 .		p	
200,000 .	.9	b	1.0

1. The entry for cell "a" is determined by horizontal interpolation between 1.0 and 1.2.

$$30.2-25.0 = 5.2$$
$$50.0-25.0 = 25.0$$

$$1.0 + \frac{5.2}{25.0}(1.2-1.0) = 1.04$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.9 and 1.0.

$$30.2-25.0 = 5.2$$
$$50.0-25.0 = 25.0$$

$$0.9 + \frac{5.2}{25.0}(1.0-0.9) = .92$$

3. The entry for "p" was then determined by vertical interpolation between .92 and 1.04.

$$160{,}500-150{,}000 = 10{,}500$$
$$200{,}000-150{,}000 = 50{,}000$$

$$1.04-\frac{10{,}500}{50{,}000}(1.04-.92) = 1.0$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 29.2 to 31.2 percent; the 90-percent confidence interval is from 28.6 to 31.8 percent; and the 95-percent confidence interval is from 28.2 to 32.2 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 error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; but, 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 part A of this report shows that in 1976 there were 89,500 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 41,000. Table I shows the standard error of 48,500 is approximately 1,760 and the standard error of 89,500 is approximately 2,240. Therefore, the standard error of the estimated difference of 41,000 is about:

$$2{,}850 = \sqrt{(1{,}760)^2 + (2{,}240)^2}$$

Consequently, the 68-percent confidence interval for the difference is from 38,150 to 43,850 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 36,440 to 45,560 housing units, and the 95-percent confidence interval is from 35,300 to 46,700. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval 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 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 table II 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 of a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 2.5. The base of the distribution from which this median was determined is 160,500 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 160,500 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 table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 81,100 owner-occupied housing units, or 50.5 percent, had one or two persons (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 28,900 owner-occupied housing units, or 18.0 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:

$$2.5 + (3.5-2.5)\left(\frac{50.5-47.6}{18.0}\right) = 2.3$$

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

$$2.5 + (3.5-2.5)\left(\frac{52.4-50.5}{18.0}\right) = 2.6$$

Thus, the 95-percent confidence interval ranges from 2.3 to 2.6 persons.

Omaha 1976

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Annual Housing Survey App-42
Selection of the sample App-42
Coverage improvement sample selection App-43
Coverage improvement for deficiency 1 ..... App-43
Coverage improvement for deficiency 2 ..... App-43
Coverage improvement for deficiencies 3-6 ..... App-44
Building loss sample selection ... App-44
1970 Census of Population and Housing App-44
ESTIMATION App-44
1976 housing inventory App-44
1970-1976 lost units App-45
Ratio estimation procedure of the 1970 Census of Population and Housing App-45
RELIABILITY OF THE ESTIMATES App-45
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-46
Coverage errors App-46
Rounding errors App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-47
Differences App-48
Illustration of the computation of the standard error of a difference App-49
Medians App-49
Illustration of the computation of the 95-percent confidence interval of a median .. App-49

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 20 SMSA's are based on data collected from the 1976 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. In each of the 20 SMSA's, the data were collected for the 12-month period from April 1976 through March 1977 with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and data for a second group of 21 SMSA's were collected from April 1975 through March 1976. The sample housing units for each group of the AHS SMSA's are interviewed on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the third group (1976-77) are: Houston, Tex., New York, N.Y., St. Louis, Mo.-Ill., and Seattle-Everett, Wash.

The remaining SMSA's in the third group are: Allentown-Bethlehem-Easton, Pa.-N.J., Baltimore, Md., Birmingham, Ala., Buffalo, N.Y., Cleveland, Ohio, Denver, Colo., Grand Rapids, Mich., Honolulu, Hawaii, Indianapolis, Ind., Las Vegas, Nev., Louisville, Ky.-Ind., Oklahoma City, Okla., Omaha, Nebr.-Iowa, Providence-Pawtucket-Warwick, R.I.-Mass., Raleigh, N.C., and Sacramento, Calif.

In this SMSA, 4,765 units were eligible for interview. Of these sample units, 273 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to the units eligible for interview, 452 units were visited but were not eligible for interview, because they were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit-issuing (Honolulu, Las Vegas, New York, and Sacramento) was selected from two sample frames—units enumerated in the 1970 Census of Population and Housing in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for the 16 SMSA's which are *not* 100-percent permit-issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe).

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of the sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA, since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and balance of the SMSA, since the sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total housing units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant housing unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent

associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to 1 of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—Family size	Renter—Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000-$5,999. . .		
$6,000-$9,999. . .		
$10,000-$14,999 .		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to 1 of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to 1 of the 4 vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. However, the housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of the interview, the units at each of the sample special places were listed and subsampled at a rate to produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since January 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit-issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then 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. At the time of the survey, those segments that did not have an expected size of four were further subdivided to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Coverage improvement sample selection.—The Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS-SMSA sample from the permit-issuing and new construction universes. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970, but completed after April 1, 1970.
2. Mobile homes placed in parks either missed during the 1970 census or established since the census.
3. Units missed in the 1970 census.
4. Units converted to residential use that were nonresidential at the time of the 1970 census.
5. Houses that have been moved onto their present site since the 1970 census.
6. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

Coverage improvement for deficiency 1.—A sample of new construction units whose permits were issued before January 1970, but completed after April 1970, was selected independently for each SMSA. The sampling was carried out in two stages for one-and-two-unit structures and in three stages for three-or-more-unit structures. These units were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units).

The first stage was a sample of permit offices and the second stage a sample of the 1969 permits within each of the selected permit offices. In the New York, N.Y., SMSA, Honolulu, Hawaii, SMSA, and Grand Rapids, Mich., SMSA, an additional sample of 1968 permits for three-or-more-unit structures was included in the second stage sampling. For the third stage, structures of size three or more were divided into clusters of an expected size of four units and a sample of clusters was selected. This procedure added an estimated 1,706 new construction units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiency 2.—In permit-issuing areas, a sample of mobile homes placed in a park missed by the census or established after the census was selected in two stages. First, for each 1976-77 SMSA, a sample of tracts was selected and canvassed. All parks were listed and then matched back to the 1970 census to identify parks missed by the census and parks established after the census. Second, the parks were divided into clusters of an expected size of four sites and a sample of clusters was selected

and interviewed. Each of the sample units represented the same number of units that the regular AHS sample units represented. This procedure added an estimated 254 units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiencies 3-6.—The remaining missed units were sampled by one of two procedures. The first procedure was designed to represent units from the following types of missed structures (structures that had no chance of selection for the AHS):

1. Structures missed in the 1970 census.
2. Structures that were completely nonresidential in the 1970 census but contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and had a utility hookup, or were on the site but not occupied on April 1, 1970, or had no utility hookup but were occupied by persons with no usual residence elsewhere.
4. Houses that had been moved onto their present site since the 1970 census.

Initially, a subsample of regular AHS units was selected at a rate of 1 in 22.2152. Then succeeding structures, in a defined path of travel to the right of the structure containing the sample unit, were listed until eight structures (excluding the sample unit structure) were found that had been eligible to be selected for the AHS. Finally, the intervening structures that did not have a chance of selection in the AHS were identified and units within these structures were interviewed. In cases where the interviewer workload would have been too great, a representative subsample of units within these structures was selected. This procedure added an estimated 1,933 units to the coverage of the housing inventory of this SMSA.

The second procedure was designed to represent missed units from structures represented in the AHS. These missed units were:

1. Units missed in the 1970 census.
2. Units converted to residential use since the 1970 census in structures that contained some residential units in 1970.

First, a subsample of regular AHS units in multi-unit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multi-unit structures selected, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 330 units to the coverage of the housing inventory of this SMSA.

Building loss sample selection.—Some tables in this report show estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS interview). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample previously described. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

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 employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at the time of the interview (i.e., the 1976 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1976 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1976 housing inventory.—The AHS estimates of characteristics of the 1976 housing inventory employed a one-stage ratio estimation procedure. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 273 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units} + \text{Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 50 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of 1 or more of the different strata used in the stratification of the universe as previously illustrated). In addition, a noninterview factor was computed for one noninterview cell for new construc-

tion sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes, and one noninterview cell for sample units from both the nonpermit universe (if applicable) and the coverage improvement universe (if units were not included above).

The following ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding ratio estimation category.

The effect of this ratio estimation procedure was to somewhat reduce the variance due to the variation in sample size of strata used in the sample selection of the permit-issuing universe, thereby reducing the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

1970-1976 lost units.—The AHS estimate of characteristics of the 1970-1976 lost units employed a one-stage ratio estimation procedure which is similar to the ratio estimation procedure described previously. The file of 1970-1976 lost units was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error 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 housing inventory from 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 this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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-SMSA 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 can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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 both the 1970 Census of Population and Housing and the 1976 AHS-SMSA 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 included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series 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.**

AHS-SMSA.—For the 1976 AHS-SMSA 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 sample 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 accuracy of the AHS data collected from interviewed households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct 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.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview check. The results of this study are available and are presented in the following Census Bureau memorandum, "Response Error in the Annual Housing Survey as Measured by Reinterviews—Year III (1976-1977) SMSA Sample."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year III SMSA's and not for any specific SMSA):

1. For attitudinal items which were not reconciled (i.e., after the question is answered in the reinterview, the enumerator does not present the previous response and then ask the respondent to decide upon the best answer), approximately 67 percent of the indices of inconsistency showed moderate levels of response disagreement while the remaining 33 percent showed high levels.
2. Some differences beyond those due to sampling error did occur. Most of the categories affected by bias were categories of attitudinal items.

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low, indices from 20—50 are moderate, indicating that there is some problem with inconsistent reporting, those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

The results of this study were based on sample data so 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 this study.

Coverage errors.—With respect to errors of coverage and estimation for missing data, the AHS new construction sample had deficiencies in the representation of conventional new construction. Due to time constraints, only those building permits issued more than 5 months before the survey ended were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. Also, permits were sampled on a monthly basis and were divided equally among the 12 panels in which interviews were conducted. Due to this procedure, some of the permits issued in November 1975 through October 1976 were not interviewed because they were assigned to panels in which the interviewing had already been completed. In this SMSA, 5.3 percent of the permits sampled were not interviewed because of this procedure. This percentage does not include permits issued during the last 5 months of the survey. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units since, due to the relatively short time-span involved, it is possible that construction of these units was not completed at the time the survey was conducted and they were not eligible for interview. In addition to these deficiencies, new construction in special places that do not require building permits, such as military bases, are also not adequately represented.

The Coverage Improvement Program also had certain deficiencies. It appears that the listing procedure used to find units missed in the 1970 census, units converted from nonresidential to residential, houses moved onto their present site, and mobile homes placed outside parks, was not very efficient for finding nonresidential conversions. Such conversions were primarily in business districts, whereas the listing procedure started from a residential unit.

Deficiencies also exist in ED's where area sampling methods are used. It had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing.

Rounding errors.—For 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 or small medians, when these figures are derived from relatively large bases (e.g., median number of persons per household). 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 the survey.

Sampling errors for the AHS-SMSA sample.—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 schedules, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The stand-

ard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variance in the estimates due to response and interviewer errors (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 error measured by the standard error, 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 in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from 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 following tables are approximations to the standard errors of various estimates shown in this report for this SMSA. 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 precise standard errors for any specific items.

Table I presents the standard errors applicable to estimates of characteristics of the 1976 housing inventory as well as estimates of characteristics of the 1970-1976 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in this table. The standard errors on the AHS estimates of the population in housing units shown in tables A-1, B-1, and C-1 of part A of this report are 7,320 for the total SMSA, 4,430 for the central city of the SMSA, and 5,570 for the balance of the SMSA.

TABLE I. Standard errors for Estimated Number of Housing Units in the 1976 Housing Inventory and for Estimated Number of 1970-1976 Lost Units for the Omaha, Nebr.-Iowa, SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central city	Not in central city
0	50	50	50
100.	70	70	70
200.	100	100	100
500.	160	150	160
700.	180	180	180
1,000 . . .	220	220	220
2,500 . . .	350	340	350
5,000 . . .	490	480	490
10,000 . .	690	660	680
25,000 . .	1,060	980	1,050
50,000 . .	1,430	1,240	1,410
75,000 . .	1,660	1,300	1,630
100,000. .	1,800	1,190	–
150,000. .	1,910	–	–
200,000. .	1,800	–	–

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.

Table II presents the standard errors of estimated percentages of the 1976 housing inventory as well as estimated percentages of the 1970-1976 lost units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in table II.

Included in tables I and II are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered as 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.

For ratios, 100 (x/y), where x is not a subclass of y, table II underestimates 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.—Table A-1 of part A of this report shows that in 1976 there were 119,300 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is ap-

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the Omaha, Nebr.-Iowa, SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100.	32.8	32.8	32.8	32.8	32.8	35.0
200.	19.6	19.6	19.6	19.6	21.4	24.7
500.	8.9	8.9	8.9	9.4	13.5	15.6
700.	6.5	6.5	6.5	7.9	11.4	13.2
1,000	4.7	4.7	4.8	6.6	9.6	11.1
2,500	1.9	1.9	3.0	4.2	6.1	7.0
5,000	1.0	1.0	2.2	3.0	4.3	4.9
10,000	.5	.7	1.5	2.1	3.0	3.5
25,000	.2	.4	1.0	1.3	1.9	2.2
50,000	.10	.3	.7	.9	1.4	1.6
75,000	.07	.3	.6	.8	1.1	1.3
100,000. . . .	.05	.2	.5	.7	1.0	1.1
150,000. . . .	.03	.2	.4	.5	.8	.9
200,000. . . .	.02	.2	.3	.5	.7	.8

[1] Standard errors are presented to nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to nearest one-hundredth of one percentage point.

proximately 1,840. The following interpolation procedure was used:

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

Size of estimate	Standard error
100,000.	1,800
119,300.	x
150,000.	1,910

The entry for x is determined by vertically interpolating between 1,800 and 1,910.

$$119{,}300-100{,}000 = 19{,}300$$
$$150{,}000-100{,}000 = 50{,}000$$

$$1{,}800 + \frac{19{,}300}{50{,}000}(1{,}910-1{,}800)$$

$$= 1{,}840$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 117,460 to 121,140 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 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 116,360 to 122,240 housing units with 90 percent confidence; and that the average estimate lies within the interval from 115,620 to 122,980 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 119,300 owner-occupied housing units, 29,500, or 24.7 percent, had two bedrooms. Interpolation in table II of the appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 24.7 percent is approximately 0.9 percentage points. The following interpolation procedure was used:

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

Base of percentage	Estimated percentage		
	10 or 90	24.7	25 or 75
100,000. . . .	0.7	a	1.0
119,300. . . .		p	
150,000. . . .	0.5	b	0.8

1. The entry for cell "a" is determined by horizontal interpolation between 0.7 and 1.0

$$24.7-10.0 = 14.7$$
$$25.0-10.0 = 15.0$$

$$0.7 + \frac{14.7}{15.0}(1.0-0.7) = .99$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.5 and 0.8.

$$24.7-10.0 = 14.7$$
$$25.0-10.0 = 15.0$$

$$0.5 + \frac{14.7}{15.0}(0.8-0.5) = .79$$

3. The entry for "p" was then determined by vertical interpolation between .79 and .99.

$$119{,}300-100{,}000 = 19{,}300$$
$$150{,}000-100{,}000 = 50{,}000$$

$$.99 - \frac{19{,}300}{50{,}000}(.99-.79) = 0.9$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 23.8 to 25.6 percent; the 90-percent confidence interval is from 23.3 to 26.1 percent; and the 95-percent confidence interval is from 22.9 to 26.5 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 error of each estimate con-

sidered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; but, 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 part A of this report shows that in 1976 there were 61,600 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 32,100. Table I shows the standard error of 29,500 is approximately 1,130 and the standard error of 61,600 is approximately 1,540. Therefore, the standard error of the estimated difference of 32,100 is about:

$$1{,}910 = \sqrt{(1{,}130)^2 + (1{,}540)^2}$$

Consequently, the 68-percent confidence interval for the 32,100 difference is from 30,190 to 34,010 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 29,040 to 35,160 housing units, and the 95-percent confidence interval is from 28,280 to 35,920. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval 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 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 table II 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 of a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 3.0. The base of the distribution from which this median was determined is 119,300 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 119,300 is approximately 1.0 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 48.0 and 52.0.
3. From table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 49,700 owner-occupied housing units, or 41.7 percent, had one or two persons (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 20,900 owner-occupied housing units, or 17.5 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:

$$2.5 + (3.5-2.5)\left(\frac{48.0-41.7}{17.5}\right) = 2.9$$

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

$$2.5 + (3.5-2.5)\left(\frac{52.0-41.7}{17.5}\right) = 3.1$$

Thus, the 95-percent confidence interval ranges from 2.9 to 3.1 persons.

Providence-Pawtucket-Warwick 1976

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Annual Housing Survey App-42
Selection of the sample App-42
Coverage improvement sample selection App-43
Coverage improvement for deficiency 1 ..... App-43
Coverage improvement for deficiency 2 ..... App-43
Coverage improvement for deficiencies 3-6 ..... App-44
Building loss sample selection ... App-44
1970 Census of Population and Housing App-44
ESTIMATION App-44
1976 housing inventory App-44
1970-1976 lost units App-45
Ratio estimation procedure of the 1970 Census of Population and Housing App-45
RELIABILITY OF THE ESTIMATES App-45
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-46
Coverage errors App-46
Rounding errors App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-48
Differences App-49
Illustration of the computation of the standard error of a difference App-49
Medians App-49
Illustration of the computation of the 95-percent confidence interval of a median .. App-49

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 20 SMSA's are based on data collected from the 1976 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. In each of the 20 SMSA's, the data were collected for the 12-month period from April 1976 through March 1977 with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and data for a second group of 21 SMSA's were collected from April 1975 through March 1976. The sample housing units for each group of the AHS SMSA's are interviewed on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the third group (1976-77) are: Houston, Tex., New York, N.Y., St. Louis, Mo.-Ill., and Seattle-Everett, Wash.

The remaining SMSA's in the third group are: Allentown-Bethlehem-Easton, Pa.-N.J., Baltimore, Md., Birmingham, Ala., Buffalo, N.Y., Cleveland, Ohio, Denver, Colo., Grand Rapids, Mich., Honolulu, Hawaii, Indianapolis, Ind., Las Vegas, Nev., Louisville, Ky.-Ind., Oklahoma City, Okla., Omaha, Nebr.-Iowa, Providence-Pawtucket-Warwick, R.I.-Mass., Raleigh, N.C., and Sacramento, Calif.

In this SMSA, 4,898 units were eligible for interview. Of these sample units, 202 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to the units eligible for interview, 447 units were visited but were not eligible for interview, because they were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit-issuing (Honolulu, Las Vegas, New York, and Sacramento) was selected from two sample frames—units enumerated in the 1970 Census of Population and Housing in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for the 16 SMSA's which are *not* 100-percent permit-issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe).

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of the sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and balance of the SMSA, since the sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total housing units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant housing unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent

associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to 1 of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—Family size	Renter—Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000-$5,999. . .		
$6,000-$9,999. . .		
$10,000-$14,999 .		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to 1 of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to 1 of the 4 vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. However, the housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of the interview, the units at each of the sample special places were listed and subsampled at a rate to produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since January 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit-issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then 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. At the time of the survey, those segments that did not have an expected size of four were further subdivided to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Coverage improvement sample selection.—The Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS-SMSA sample from the permit-issuing and new construction universes. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970, but completed after April 1, 1970.
2. Mobile homes placed in parks either missed during the 1970 census or established since the census.
3. Units missed in the 1970 census.
4. Units converted to residential use that were nonresidential at the time of the 1970 census.
5. Houses that have been moved onto their present site since the 1970 census.
6. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

Coverage improvement for deficiency 1.—A sample of new construction units whose permits were issued before January 1970, but completed after April 1970, was selected independently for each SMSA. The sampling was carried out in two stages for one-and-two-unit structures and in three stages for three-or-more-unit structures. These units were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units).

The first stage was a sample of permit offices and the second stage a sample of the 1969 permits within each of the selected permit offices. In the New York, N.Y., SMSA, Honolulu, Hawaii, SMSA, and Grand Rapids, Mich., SMSA, an additional sample of 1968 permits for three-or-more-unit structures was included in the second stage sampling. For the third stage, structures of size three or more were divided into clusters of an expected size of four units and a sample of clusters was selected. This procedure added an estimated 522 new construction units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiency 2—In permit-issuing areas, a sample of mobile homes placed in a park missed by the census or established after the census was selected in two stages. First, for each 1976-77 SMSA, a sample of tracts was selected and canvassed. All parks were listed and then matched back to the 1970 census to identify parks missed by the census and parks established after the census. Second, the parks were divided into clusters of an expected size of four

sites and a sample of clusters was selected and interviewed. Each of the sample units represented the same number of units that the regular AHS sample units represented. This procedure added an estimated 746 units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiencies 3-6.—The remaining missed units were sampled by one of two procedures. The first procedure was designed to represent units from the following types of missed structures (structures that had no chance of selection for the AHS):

1. Structures missed in the 1970 census.
2. Structures that were completely nonresidential in the 1970 census but contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and had a utility hookup, or were on the site but not occupied on April 1, 1970, or had no utility hookup but were occupied by persons with no usual residence elsewhere.
4. Houses that had been moved onto their present site since the 1970 census.

Initially, a subsample of regular AHS units was selected at a rate of 1 in 22.2152. Then succeeding structures, in a defined path of travel to the right of the structure containing the sample unit, were listed until eight structures (excluding the sample unit structure) were found that had been eligible to be selected for the AHS. Finally, the intervening structures that did not have a chance of selection in the AHS were identified and units within these structures were interviewed. In cases where the interviewer workload would have been too great, a representative subsample of units within these structures was selected. This procedure added an estimated 10,089 units to the coverage of the housing inventory of this SMSA.

The second procedure was designed to represent missed units from structures represented in the AHS. These missed units were:

1. Units missed in the 1970 census.
2. Units converted to residential use since the 1970 census in structures that contained some residential units in 1970.

First, a subsample of regular AHS units in multi-unit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multi-unit structures selected, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 2,523 units to the coverage of the housing inventory of this SMSA.

Building loss sample selection.—Some tables in this report show estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS interview). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample previously described. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

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 employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at the time of the interview (i.e., the 1976 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1976 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1976 housing inventory.—The AHS estimates of characteristics of the 1976 housing inventory employed a one-stage ratio estimation procedure. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 202 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units} + \text{Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 50 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of 1 or more of the different strata used in the stratification of the universe as

previously illustrated). In addition, a noninterview factor was computed for one noninterview cell for new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes, and one noninterview cell for sample units from both the nonpermit universe (if applicable) and the coverage improvement universe (if units were not included above).

The following ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding ratio estimation category.

The effect of this ratio estimation procedure was to somewhat reduce the variance due to the variation in sample size of strata used in the sample selection of the permit-issuing universe, thereby reducing the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

1970-1976 lost units.—The AHS estimate of characteristics of the 1970-1976 lost units employed a one-stage ratio estimation procedure which is similar to the ratio estimation procedure described previously. The file of 1970-1976 lost units was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error 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 housing inventory from 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 this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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-SMSA 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 can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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 both the 1970 Census of Population and Housing and the 1976 AHS-SMSA 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 included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series 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.**

AHS-SMSA.—For the 1976 AHS-SMSA 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 sample 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 accuracy of the AHS data collected from interviewed households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct 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.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview check. The results of this study are available and are presented in the following Census Bureau memorandum, "Response Error in the Annual Housing Survey as Measured by Reinterviews—Year III (1976-1977) SMSA Sample."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year III SMSA's and not for any specific SMSA):

1. For attitudinal items which were not reconciled (i.e., after the question is answered in the reinterview, the enumerator does not present the previous response and then ask the respondent to decide upon the best answer), approximately 67 percent of the indices of inconsistency showed moderate levels of response disagreement while the remaining 33 percent showed high levels.
2. Some differences beyond those due to sampling error did occur. Most of the categories affected by bias were categories of attitudinal items.

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low; indices from 20—50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

The results of this study were based on sample data so 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 this study.

Coverage errors.—With respect to errors of coverage and estimation for missing data, the AHS new construction sample had deficiencies in the representation of conventional new construction. Due to time constraints, only those building permits issued more than 5 months before the survey ended were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. Also, permits were sampled on a monthly basis and were divided equally among the 12 panels in which interviews were conducted. Due to this procedure, some of the permits issued in November 1975 through October 1976 were not interviewed because they were assigned to panels in which the interviewing had already been completed. In this SMSA, 5.2 percent of the permits sampled were not interviewed because of this procedure. This percentage does not include permits issued during the last 5 months of the survey. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units since, due to the relatively short time-span involved, it is possible that construction of these units was not completed at the time the survey was conducted and they were not eligible for interview. In addition to these deficiencies, new construction in special places that do not require building permits, such as military bases, are also not adequately represented.

The Coverage Improvement Program also had certain deficiencies. It appears that the listing procedure used to find units missed in the 1970 census, units converted from nonresidential to residential, houses moved onto their present site, and mobile homes placed outside parks, was not very efficient for finding nonresidential conversions. Such conversions were primarily in business districts, whereas the listing procedure started from a residential unit.

Deficiencies also exist in ED's where area sampling methods are used. It had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing.

Rounding errors.—For 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 or small medians, when these figures are derived from relatively large bases (e.g., median number of persons per household). 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 the survey.

Sampling errors for the AHS-SMSA sample.—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 schedules, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variance in the estimates due to response and interviewer errors (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 error measured by the standard error, 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 in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from 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 following tables are approximations to the standard errors of various estimates shown in this report for this SMSA. 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 precise standard errors for any specific items.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1976 Housing Inventory and for Estimated Number of 1970-1976 Lost Units for the Providence-Pawtucket-Warwick, R.I.-Mass., SMSA, for the Central Cities, and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central city	Not in central city
0	80	80	80
100	90	90	90
200	130	130	120
500	200	200	200
700	240	240	230
1,000 . . .	280	280	280
2,500 . . .	440	440	440
5,000 . . .	630	620	620
10,000 . .	880	870	870
25,000 . .	1,370	1,300	1,330
50,000 . .	1,870	1,670	1,790
75,000 . .	2,210	1,810	2,070
100,000 .	2,450	1,780	2,240
150,000 .	2,740	—	2,330
200,000 .	2,840	—	2,120
250,000 .	2,770	—	—
300,000 .	2,500	—	—

Table I presents the standard errors applicable to estimates of characteristics of the 1976 housing inventory as well as estimates of characteristics of the 1970-1976 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in this table. The standard errors on the AHS estimates of the population in housing units shown in tables A-1, B-1, and C-1 of part A of this report are 11,390 for the total SMSA, 5,800 for the central city of the SMSA, and 8,990 for the balance of the SMSA.

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.

Table II presents the standard errors of estimated percentages of the 1976 housing inventory as well as estimated percentages of the 1970-1976 lost units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in table II.

Included in tables I and II are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered as 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.

For ratios, 100 (x/y), where x is not a subclass of y, table II underestimates 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:

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

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the Providence-Pawtucket-Warwick, R.I.-Mass., SMSA, for the Central Cities, and for the Balance of the SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100	44.3	44.3	44.3	44.3	44.3	44.6
200	28.5	28.5	28.5	28.5	28.5	31.5
500	13.7	13.7	13.7	13.7	17.3	20.0
700	10.2	10.2	10.2	10.2	14.6	16.9
1,000	7.4	7.4	7.4	8.5	12.2	14.1
2,500	3.1	3.1	3.9	5.4	7.7	8.9
5,000	1.6	1.6	2.7	3.8	5.5	6.3
10,000	0.8	0.9	1.9	2.7	3.9	4.5
25,000	0.3	0.6	1.2	1.7	2.4	2.8
50,000	0.2	0.4	0.9	1.2	1.7	2.0
75,000	0.11	0.3	0.7	1.0	1.4	1.6
100,000 . . .	0.08	0.3	0.6	0.8	1.2	1.4
150,000 . . .	0.05	0.2	0.5	0.7	1.0	1.2
200,000 . . .	0.04	0.2	0.4	0.6	0.9	1.0
250,000 . . .	0.03	0.2	0.4	0.5	0.8	0.9
300,000 . . .	0.03	0.2	0.4	0.5	0.7	0.8

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point.

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.—Table A-1 of part A of this report shows that in 1976 there were 183,100 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 2,810. The following interpolation procedure was used:

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

Size of estimates	Standard error
150,000	2,740
183,100	x
200,000	2,840

The entry for x is determined by vertically interpolating between 2,740 and 2,840.

$$183{,}100-150{,}000 = 33{,}100$$
$$200{,}000-150{,}000 = 50{,}000$$
$$2{,}740 + \frac{33{,}100}{50{,}000}(2{,}840-2{,}740)$$
$$= 2{,}810$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 180,290 to 185,910 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 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 178,600 to 187,600 housing units with 90 percent confidence; and that the average estimate lies within the interval from 177,480 to 188,720 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 183,100 owner-occupied housing units, 52,600, or 28.7 percent, had two bedrooms. Interpolation in table II of the appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 28.7 percent is approximately 1.0 percentage points. The following interpolation procedure was used:

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

Base of percentage	Estimated percentage		
	25 or 75	28.7	50
150,000 . . .	1.0	a	1.2
183,100 . . .		p	
200,000 . . .	0.9	b	1.0

1. The entry for cell "a" is determined by horizontal interpolation between 1.0 and 1.2.

$$28.7-25.0 = 3.7$$
$$50.0-25.0 = 25.0$$
$$1.0 + \frac{3.7}{25.0}(1.2-1.0) = 1.03$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.9 and 1.0.

$$28.7-25.0 = 3.7$$
$$50.0-25.0 = 25.0$$
$$0.9 + \frac{3.7}{25.0}(1.0-0.9 = .91$$

3. The entry for "p" was then determined by vertical interpolation between .91 and 1.03.

$$183{,}100-150{,}000 = 33{,}100$$
$$200{,}000-150{,}000 = 50{,}000$$
$$1.03-\frac{33{,}100}{50{,}000}(1.03-.91) = 1.0$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 27.7 to 29.7 percent; the 90-percent confidence interval is from 27.1 to 30.3 percent; and the 95-percent confidence interval is from 26.7 to 30.7 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 error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; but, 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 part A of this report shows that in 1976 there were 88,800 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 36,200. Table I shows the standard error of 52,600 is approximately 1,910 and the standard error of 88,800 is approximately 2,340. Therefore, the standard error of the estimated difference of 36,200 is about:

$$3{,}020 = \sqrt{(1{,}910)^2 + (2{,}340)^2}$$

Consequently, the 68-percent confidence interval for the 36,200 difference is from 33,180 to 39,220 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 31,370 to 41,030 housing units, and the 95-percent confidence interval is from 30,160 to 42,240. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval 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 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 table II 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 of a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 3.0. The base of the distribution from which this median was determined is 183,100 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 183,100 is approximately 1.1 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 47.8 and 52.2.
3. From table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 74,200 owner-occupied housing units, or 40.5 percent, had one or two persons (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 34,300 owner-occupied housing units, or 18.7 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:

$$2.5 + (3.5-2.5)\left(\frac{47.8-40.5}{18.7}\right) = 2.9$$

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

$$2.5 + (3.5-2.5)\left(\frac{52.2-40.5}{18.7}\right) = 3.1$$

Thus, the 95-percent confidence interval ranges from 2.9 to 3.1 persons.

Raleigh 1976

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Annual Housing Survey App-42
Selection of the sample App-42
Coverage improvement sample selection App-43
Coverage improvement for deficiency 1 ..... App-43
Coverage improvement for deficiency 2 ..... App-43
Coverage improvement for deficiencies 3-6 ..... App-44
Building loss sample selection ... App-44
1970 Census of Population and Housing App-44
ESTIMATION App-44
1976 housing inventory App-44
1970-1976 lost units App-45
Ratio estimation procedure of the 1970 Census of Population and Housing App-45
RELIABILITY OF THE ESTIMATES App-45
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-46
Coverage errors App-46
Rounding errors App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-48
Differences App-48
Illustration of the computation of the standard error of a difference App-49
Medians App-49
Illustration of the computation of the 95-percent confidence interval of a median .. App-49

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 20 SMSA's are based on data collected from the 1976 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. In each of the 20 SMSA's, the data were collected for the 12-month period from April 1976 through March 1977 with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and data for a second group of 21 SMSA's were collected from April 1975 through March 1976. The sample housing units for each group of the AHS SMSA's are interviewed on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the third group (1976-77) are: Houston, Tex., New York, N.Y., St. Louis, Mo.-Ill., and Seattle-Everett, Wash.

The remaining SMSA's in the third group are: Allentown-Bethlehem-Easton, Pa.-N.J., Baltimore, Md., Birmingham, Ala., Buffalo, N.Y., Cleveland, Ohio, Denver, Colo., Grand Rapids, Mich., Honolulu, Hawaii, Indianapolis, Ind., Las Vegas, Nev., Louisville, Ky.-Ind., Oklahoma City, Okla., Omaha, Nebr.-Iowa, Providence-Pawtucket-Warwick, R.I.-Mass., Raleigh, N.C., and Sacramento, Calif.

In this SMSA, 5,107 units were eligible for interview. Of these sample units, 120 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to the units eligible for interview, 647 units were visited but were not eligible for interview, because they were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit-issuing (Honolulu, Las Vegas, New York, and Sacramento) was selected from two sample frames—units enumerated in the 1970 Census of Population and Housing in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for the 16 SMSA's which are *not* 100-percent permit-issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe).

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of the sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA, since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and balance of the SMSA, since the sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total housing units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant housing unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit

records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to 1 of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—Family size	Renter—Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000-$5,999. . .		
$6,000-$9,999. . .		
$10,000-$14,999 .		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to 1 of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to 1 of the 4 vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. However, the housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of the interview, the units at each of the sample special places were listed and subsampled at a rate to produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since January 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit-issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then 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. At the time of the survey, those segments that did not have an expected size of four were further subdivided to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Coverage improvement sample selection.—The Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS-SMSA sample from the permit-issuing and new construction universes. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970, but completed after April 1, 1970.
2. Mobile homes placed in parks either missed during the 1970 census or established since the census.
3. Units missed in the 1970 census.
4. Units converted to residential use that were nonresidential at the time of the 1970 census.
5. Houses that have been moved onto their present site since the 1970 census.
6. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

Coverage improvement for deficiency 1.—A sample of new construction units whose permits were issued before January 1970, but completed after April 1970, was selected independently for each SMSA. The sampling was carried out in two stages for one-and-two-unit structures and in three stages for three-or-more-unit structures. These units were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units).

The first stage was a sample of permit offices and the second stage a sample of the 1969 permits within each of the selected permit offices. In the New York, N.Y., SMSA, Honolulu, Hawaii, SMSA, and Grand Rapids, Mich., SMSA, an additional sample of 1968 permits for three-or-more-unit structures was included in the second stage sampling. For the third stage, structures of size three or more were divided into clusters of an expected size of four units and a sample of clusters was selected. This procedure added an estimated 292 new construction units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiency 2.—In permit-issuing areas, a sample of mobile homes placed in a park missed by the census or established after the census was selected in two stages. First, for each 1976-77 SMSA, a sample of tracts was selected and canvassed. All parks were listed and then matched back to the 1970 census to identify parks missed by the

census and parks established after the census. Second, the parks were divided into clusters of an expected size of four sites and a sample of clusters was selected and interviewed. Each of the sample units represented the same number of units that the regular AHS sample units represented. This procedure yielded no additional units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiencies 3-6.—The remaining missed units were sampled by one of two procedures. The first procedure was designed to represent units from the following types of missed structures (structures that had no chance of selection for the AHS):

1. Structures missed in the 1970 census.
2. Structures that were completely nonresidential in the 1970 census but contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and had a utility hookup, or were on the site but not occupied on April 1, 1970, or had no utility hookup but were occupied by persons with no usual residence elsewhere.
4. Houses that had been moved onto their present site since the 1970 census.

Initially, a subsample of regular AHS units was selected at a rate of 1 in 22.2152. Then, succeeding structures, in a defined path of travel to the right of the structure containing the sample unit, were listed until eight structures (excluding the sample unit structure) were found that had been eligible to be selected for the AHS. Finally, the intervening structures that did not have a chance of selection in the AHS were identified and units within these structures were interviewed. In cases where the interviewer workload would have been too great, a representative subsample of units within these structures was selected. This procedure added an estimated 2,398 units to the coverage of the housing inventory of this SMSA.

The second procedure was designed to represent missed units from structures represented in the AHS. These missed units were:

1. Units missed in the 1970 census.
2. Units converted to residential use since the 1970 census in structures that contained some residential units in 1970.

First, a subsample of regular AHS units in multi-unit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multi-unit structures selected, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 465 units to the coverage of the housing inventory of this SMSA.

Building loss sample selection.—Some tables in this report show estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS interview). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample previously described. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

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 employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at the time of the interview (i.e., the 1976 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1976 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1976 housing inventory.—The AHS estimates of characteristics of the 1976 housing inventory employed a one-stage ratio estimation procedure. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 120 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units} + \text{Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 50 noninterview cells for sample housing units from the permit-

issuing universe (where the cells consisted of 1 or more of the different strata used in the stratification of the universe as previously illustrated). In addition, a noninterview factor was computed for one noninterview cell for new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes, and one noninterview cell for sample units from both the nonpermit universe (if applicable) and the coverage improvement universe (if units were not included above).

The following ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding ratio estimation category.

The effect of this ratio estimation procedure was to somewhat reduce the variance due to the variation in sample size of strata used in the sample selection of the permit-issuing universe, thereby reducing the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

1970-1976 lost units.—The AHS estimate of characteristics of the 1970-1976 lost units employed a one-stage ratio estimation procedure which is similar to the ratio estimation procedure described previously. The file of 1970-1976 lost units was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error 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 housing inventory from 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 this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

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-SMSA 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 can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

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 both the 1970 Census of Population and Housing and the 1976 AHS-SMSA 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 included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and

Research Program Reports, Series 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.**

AHS-SMSA.—For the 1976 AHS-SMSA 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 sample 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 accuracy of the AHS data collected from interviewed households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct 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.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview check. The results of this study are available and are presented in the following Census Bureau memorandum, "Response Error in the Annual Housing Survey as Measured by Reinterviews—Year III (1976-1977) SMSA Sample."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year III SMSA's and not for any specific SMSA):

1. For attitudinal items which were not reconciled (i.e., after the question is answered in the reinterview, the enumerator does not present the previous response and then ask the respondent to decide upon the best answer), approximately 67 percent of the indices of inconsistency showed moderate levels of response disagreement while the remaining 33 percent showed high levels.

2. Some differences beyond those due to sampling error did occur. Most of the categories affected by bias were categories of attitudinal items.

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low; indices from 20—50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

The results of this study were based on sample data so 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 this study.

Coverage errors.—With respect to errors of coverage and estimation for missing data, the AHS new construction sample had deficiencies in the representation of conventional new construction. Due to time constraints, only those building permits issued more than 5 months before the survey ended were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. Also, permits were sampled on a monthly basis and were divided equally among the 12 panels in which interviews were conducted. Due to this procedure, some of the permits issued in November 1975 through October 1976 were not interviewed because they were assigned to panels in which the interviewing had already been completed. In this SMSA, 5.2 percent of the permits sampled were not interviewed because of this procedure. This percentage does not include permits issued during the last 5 months of the survey. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units since, due to the relatively short time-span involved, it is possible that construction of these units was not completed at the time the survey was conducted and they were not eligible for interview. In addition to these deficiencies, new construction in special places that do not require building permits, such as military bases, are also not adequately represented.

The Coverage Improvement Program also had certain deficiencies. It appears that the listing procedure used to find units missed in the 1970 census, units converted from nonresidential to residential, houses moved onto their present site, and mobile homes placed outside parks, was not very efficient for finding nonresidential conversions. Such conversions were primarily in business districts, whereas the listing procedure started from a residential unit.

Deficiencies also exist in ED's where area sampling methods are used. It had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing.

Rounding errors.—For 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 or small medians, when these figures are derived from relatively large bases (e.g., median number of persons per household). 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 the survey.

Sampling errors for the AHS-SMSA sample.—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 schedules, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variance in the estimates due to response and interviewer errors (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 error measured by the standard error, 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 in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from 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 following tables are approximations to the standard errors of various estimates shown in this report for this SMSA. 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 precise standard errors for any specific items.

Table I presents the standard errors applicable to estimates of characteristics of the 1976 housing inventory as well as estimates of characteristics of the 1970-1976 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in this table. The standard errors on the AHS estimates of the population in housing units shown in tables A-1, B-1, and C-1 of part A of this report are 3,760 for the total SMSA, 2,290 for the central city of the SMSA, and 2,530 for the balance of the SMSA.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1976 Housing Inventory and for Estimated Number of 1970-1976 Lost Units for the Raleigh, N.C., SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central city	Not in central city
0	20	20	20
100.	50	50	50
200.	60	60	70
500.	100	100	110
700.	120	120	120
1,000 . . .	140	140	150
2,500 . . .	230	220	230
5,000 . . .	320	310	330
10,000 . .	450	430	460
25,000 . .	690	630	690
50,000 . .	920	760	890
75,000 . .	1,060	—	—

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.

Table II presents the standard errors of estimated percentages of the 1976 housing inventory as well as estimated percentages of the 1970-1976 lost units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in table II.

Included in tables I and II are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered as 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.

For ratios, 100 (x/y), where x is not a subclass of y, table II underestimates 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

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the Raleigh, N.C., SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100	18.2	18.2	18.2	18.2	20.4	23.6
200	10.0	10.0	10.0	10.0	14.4	16.7
500	4.2	4.2	4.6	6.3	9.1	10.5
700	3.1	3.1	3.9	5.3	7.7	8.9
1,000	2.2	2.2	3.2	4.5	6.4	7.4
2,500	0.9	0.9	2.1	2.8	4.1	4.7
5,000	0.4	0.7	1.5	2.0	2.9	3.3
10,000	0.2	0.5	1.0	1.4	2.0	2.4
25,000	0.09	0.3	0.6	0.9	1.3	1.5
50,000	0.04	0.2	0.5	0.6	0.9	1.1
75,000	0.03	0.15	0.4	0.5	0.7	0.9

[1] Standard errors are presented to nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to nearest one-hundredth of one percentage point.

Illustration of the use of the standard error tables.—Table A-1 of part A of this report shows that in 1976 there were 52,900 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 940. The following interpolation procedure was used:

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

Size of estimate	Standard error
50,000	920
52,900	x
75,000	1,060

The entry for x is determined by vertically interpolating between 920 and 1,060.

$$52{,}900-50{,}000 = 2{,}900$$
$$75{,}000-50{,}000 = 25{,}000$$

$$920 + \frac{2{,}900}{25{,}000}(1{,}060-920) = 940$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 51,960 to 53,840 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 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 51,400 to 54,400 housing units with 90 percent confidence; and that the average estimate lies within the interval from 51,020 to 54,780 housing units with 95 percent confidence.

Table 1 of part A also shows that of the 52,900 owner-occupied housing units, 11,200, or 21.2 percent, had two bedrooms. Interpolation in table II of the appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 21.2 percent is approximately 0.8 percentage points. The following interpolation procedure was used:

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

Base of percentage	Estimated percentage		
	10 or 90	21.2	25 or 75
50,000	0.6	a	0.9
52,900		p	
75,000	0.5	b	0.7

1. The entry for cell "a" is determined by horizontal interpolation between 0.6 and 0.9.

$$21.2-10.0 = 11.2$$
$$25.0-10.0 = 15.0$$

$$0.6 + \frac{11.2}{15.0}(0.9-0.6) = .82$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.5 and 0.7.

$$21.2-10.0 = 11.2$$
$$25.0-10.0 = 15.0$$

$$0.5 + \frac{11.2}{15.0}(0.7-0.5) = .65$$

3. The entry for "p" was then determined by vertical interpolation between .65 and .82.

$$52{,}900-50{,}000 = 2{,}900$$
$$75{,}000-50{,}000 = 25{,}000$$

$$.82 - \frac{2{,}900}{25{,}000}(.82-.65) = 0.8$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 20.4 to 22.0 percent; the 90-percent confidence interval is from 19.9 to 22.5 percent; and the 95-percent confidence interval is from 19.6 to 22.8 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 error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference

between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; but, 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 part A of this report shows that in 1976 there were 28,100 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 16,900. Table I shows the standard error of 11,200 is approximately 470 and the standard error of 28,100 is approximately 720. Therefore, the standard error of the estimated difference of 16,900 is about:

$$860 = \sqrt{(470)^2 + (720)^2}$$

Consequently, the 68-percent confidence interval for the 16,900 difference is from 16,040 to 17,760 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 15,520 to 18,280 housing units, and the 95-percent confidence interval is from 15,180 to 18,620. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval 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 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 table II 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 of a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 2.9. The base of the distribution from which this median was determined is 52,900 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 52,900 is approximately 1.1 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 47.8 and 52.2.
3. From table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 21,600 owner-occupied housing units, or 40.8 percent, had one or two persons (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 11,200 owner-occupied housing units, or 21.2 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:

$$2.5 + (3.5-2.5)\left(\frac{47.8-40.8}{21.2}\right) = 2.8$$

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

$$2.5 + (3.5-2.5)\left(\frac{52.2-40.8}{21.2}\right) = 3.0$$

Thus, the 95-percent confidence interval ranges from 2.8 to 3.0 persons.

Sacramento 1976

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Annual Housing Survey App-42
Selection of the sample App-42
Coverage improvement sample selection App-43
Coverage improvement for deficiency 1 ..... App-43
Coverage improvement for deficiency 2 ..... App-43
Coverage improvement for deficiencies 3-6 ..... App-44
Building loss sample selection ... App-44
1970 Census of Population and Housing App-44
ESTIMATION App-44
1976 housing inventory App-44
1970-1976 lost units App-45
Ratio estimation procedure of the 1970 Census of Population and Housing App-45
RELIABILITY OF THE ESTIMATES App-45
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-46
Coverage errors App-46
Rounding errors App-46
Sampling errors for the AHS-SMSA sample App-46
Illustration of the use of the standard error tables App-48
Differences App-49
Illustration of the computation of the standard error of a difference App-49
Medians App-49
Illustration of the computation of the 95-percent confidence interval of a median .. App-49

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 20 SMSA's are based on data collected from the 1976 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. In each of the 20 SMSA's, the data were collected for the 12-month period from April 1976 through March 1977 with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and data for a second group of 21 SMSA's were collected from April 1975 through March 1976. The sample housing units for each group of the AHS SMSA's are interviewed on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the third group (1976-77) are: Houston, Tex., New York, N.Y., St. Louis, Mo.-Ill., and Seattle-Everett, Wash.

The remaining SMSA's in the third group are: Allentown-Bethlehem-Easton, Pa.-N.J., Baltimore, Md., Birmingham, Ala., Buffalo, N.Y., Cleveland, Ohio, Denver, Colo., Grand Rapids, Mich., Honolulu, Hawaii, Indianapolis, Ind., Las Vegas, Nev., Louisville, Ky.-Ind., Oklahoma City, Okla., Omaha, Nebr.-Iowa, Providence-Pawtucket-Warwick, R.I.-Mass., Raleigh, N.C., and Sacramento, Calif.

In this SMSA, 4,932 units were eligible for interview. Of these sample units, 175 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to the units eligible for interview, 342 units were visited but were not eligible for interview, because they were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit-issuing (Honolulu, Las Vegas, New York, and Sacramento) was selected from two sample frames—units enumerated in the 1970 Census of Population and Housing in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for the 16 SMSA's which are *not* 100-percent permit-issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe).

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of the sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and balance of the SMSA, since the sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total housing units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant housing unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent

associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to 1 of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—Family size	Renter—Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000-$5,999. . .		
$6,000-$9,999. . .		
$10,000-$14,999 .		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to 1 of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to 1 of the 4 vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. However, the housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of the interview, the units at each of the sample special places were listed and subsampled at a rate to produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since January 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit-issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then 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. At the time of the survey, those segments that did not have an expected size of four were further subdivided to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Coverage improvement sample selection.—The Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS-SMSA sample from the permit-issuing and new construction universes. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970, but completed after April 1, 1970.
2. Mobile homes placed in parks either missed during the 1970 census or established since the census.
3. Units missed in the 1970 census.
4. Units converted to residential use that were nonresidential at the time of the 1970 census.
5. Houses that have been moved onto their present site since the 1970 census.
6. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

Coverage improvement for deficiency 1.—A sample of new construction units whose permits were issued before January 1970, but completed after April 1970, was selected independently for each SMSA. The sampling was carried out in two stages for one-and-two-unit structures and in three stages for three-or-more-unit structures. These units were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units).

The first stage was a sample of permit offices and the second stage a sample of the 1969 permits within each of the selected permit offices. In the New York, N.Y., SMSA, Honolulu, Hawaii, SMSA, and Grand Rapids, Mich., SMSA, an additional sample of 1968 permits for three-or-more-unit structures was included in the second stage sampling. For the third stage, structures of size three or more were divided into clusters of an expected size of four units and a sample of clusters was selected. This procedure added an estimated 10,795 new construction units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiency 2.—In permit-issuing areas, a sample of mobile homes placed in a park missed by the census or established after the census was selected in two stages. First, for each 1976-77 SMSA, a sample of tracts was selected and canvassed. All parks were listed and then matched back to the 1970 census to identify parks missed by the census and parks established after the census. Second, the parks were divided into clusters of an expected size of four sites and a sample of clusters was selected

and interviewed. Each of the sample units represented the same number of units that the regular AHS sample units represented. This procedure added an estimated 2,448 units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiencies 3-6.—The remaining missed units were sampled by one of two procedures. The first procedure was designed to represent units from the following types of missed structures (structures that had no chance of selection for the AHS):

1. Structures missed in the 1970 census.
2. Structures that were completely nonresidential in the 1970 census but contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and had a utility hookup, or were on the site but not occupied on April 1, 1970, or had no utility hookup but were occupied by persons with no usual residence elsewhere.
4. Houses that had been moved onto their present site since the 1970 census.

Initially, a subsample of regular AHS units was selected at a rate of 1 in 22.2152. Then succeeding structures, in a defined path of travel to the right of the structure containing the sample unit, were listed until eight structures (excluding the sample unit structure) were found that had been eligible to be selected for the AHS. Finally, the intervening structures that did not have a chance of selection in the AHS were identified and units within these structures were interviewed. In cases where the interviewer workload would have been too great, a representative subsample of units within these structures was selected. This procedure added an estimated 5,420 units to the coverage of the housing inventory of this SMSA.

The second procedure was designed to represent missed units from structures represented in the AHS. These missed units were:

1. Units missed in the 1970 census.
2. Units converted to residential use since the 1970 census in structures that contained some residential units in 1970.

First, a subsample of regular AHS units in multi-unit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multi-unit structures selected, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 196 units to the coverage of the housing inventory of this SMSA.

Building loss sample selection.—Some tables in this report show estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS interview). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample previously described. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

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 employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at the time of the interview (i.e., the 1976 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1976 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1976 housing inventory.—The AHS estimates of characteristics of the 1976 housing inventory employed a one-stage ratio estimation procedure. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 175 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units} + \text{Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 50 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of 1 or more of the different strata used in the stratification of the universe as previously illustrated). In addition, a noninterview factor was computed for one noninterview cell for new construction

sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes, and one noninterview cell for sample units from both the nonpermit universe (if applicable) and the coverage improvement universe (if units were not included above).

The following ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding ratio estimation category.

The effect of this ratio estimation procedure was to somewhat reduce the variance due to the variation in sample size of strata used in the sample selection of the permit-issuing universe, thereby reducing the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

1970-1976 lost units—The AHS estimate of characteristics of the 1970-1976 lost units employed a one-stage ratio estimation procedure which is similar to the ratio estimation procedure described previously. The file of 1970-1976 lost units was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error 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 housing inventory from 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 this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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-SMSA 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 can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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 both the 1970 Census of Population and Housing and the 1976 AHS-SMSA 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 included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series 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.

AHS-SMSA.—For the 1976 AHS-SMSA 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 sample 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 accuracy of the AHS data collected from interviewed households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct 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.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview check. The results of this study are available and are presented in the following Census Bureau memorandum, "Response Error in the Annual Housing Survey as Measured by Reinterviews—Year III (1976-1977) SMSA Sample."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year III SMSA's and not for any specific SMSA):

1. For attitudinal items which were not reconciled (i.e., after the question is answered in the reinterview, the enumerator does not present the previous response and then ask the respondent to decide upon the best answer), approximately 67 percent of the indices of inconsistency showed moderate levels of response disagreement while the remaining 33 percent showed high levels.
2. Some differences beyond those due to sampling error did occur. Most of the categories affected by bias were categories of attitudinal items.

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low; indices from 20—50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

The results of this study were based on sample data so 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 this study.

Coverage errors.—With respect to errors of coverage and estimation for missing data, the AHS new construction sample had deficiencies in the representation of conventional new construction. Due to time constraints, only those building permits issued more than 5 months before the survey ended were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. Also, permits were sampled on a monthly basis and were divided equally among the 12 panels in which interviews were conducted. Due to this procedure, some of the permits issued in November 1975 through October 1976 were not interviewed because they were assigned to panels in which the interviewing had already been completed. In this SMSA, 9.3 percent of the permits sampled were not interviewed because of this procedure. This percentage does not include permits issued during the last 5 months of the survey. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units since, due to the relatively short time-span involved, it is possible that construction of these units was not completed at the time the survey was conducted and they were not eligible for interview. In addition to these deficiencies, new construction in special places that do not require building permits, such as military bases, are also not adequately represented.

The Coverage Improvement Program also had certain deficiencies. It appears that the listing procedure used to find units missed in the 1970 census, units converted from nonresidential to residential, houses moved onto their present site, and mobile homes placed outside parks, was not very efficient for finding nonresidential conversions. Such conversions were primarily in business districts, whereas the listing procedure started from a residential unit.

Deficiencies also exist in ED's where area sampling methods are used. It had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing.

Rounding errors.—For 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 or small medians, when these figures are derived from relatively large bases (e.g., median number of persons per household). 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 the survey.

Sampling errors for the AHS-SMSA sample.—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 schedules, instructions, and interviewers were used, estimates from each of the different samples would differ from each other.

The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variance in the estimates due to response and interviewer errors (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 error measured by the standard error, 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 in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from 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 following tables are approximations to the standard errors of various estimates shown in this report for this SMSA. 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 precise standard errors for any specific items.

Table I presents the standard errors applicable to estimates of characteristics of the 1976 housing inventory as well as estimates of characteristics of the 1970-1976 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in this table. The standard errors on the AHS estimates of the population in housing units shown in tables A-1, B-1, and C-1 of part A of this report are 12,130 for the total SMSA, 9,200 for the central city of the SMSA, and 8,980 for the balance of the SMSA.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1976 Housing Inventory and for Estimated Number of 1970-1976 Lost Units for the Sacramento, Calif., SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central city	Not in central city
0	90	90	90
100	90	100	90
200	130	140	130
500	210	220	210
700	250	260	240
1,000 ...	290	310	290
2,500 ...	460	480	460
5,000 ...	660	680	650
10,000 ..	920	960	910
25,000 ..	1,440	1,510	1,400
50,000 ..	1,990	2,090	1,900
75,000 ..	2,380	2,510	2,230
100,000 .	2,680	2,830	2,440
150,000 .	3,100	–	2,650
200,000 .	3,360	–	2,610
250,000 .	3,490	–	–
300,000 .	3,510	–	–

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.

Table II presents the standard errors of estimated percentages of the 1976 housing inventory as well as estimated percentages of the 1970-1976 lost units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in table II.

Included in tables I and II are estimates of standard errors for estimates of zero and zero percent. These estimates of standard errors are considered as 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.

For ratios, 100 (x/y), where x is not a subclass of y, table II underestimates 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

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the Sacramento, Calif., SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100	46.5	46.5	46.5	46.5	46.5	46.6
200	30.3	30.3	30.3	30.3	30.3	32.9
500	14.8	14.8	14.8	14.8	18.0	20.8
700	11.0	11.0	11.0	11.0	15.2	17.6
1,000	8.0	8.0	8.0	8.8	12.8	14.7
2,500	3.4	3.4	4.1	5.6	8.1	9.3
5,000	1.7	1.7	2.9	4.0	5.7	6.6
10,000	.9	.9	2.0	2.8	4.0	4.7
25,000	.3	.6	1.3	1.8	2.6	2.9
50,000	.2	.4	.9	1.3	1.8	2.1
75,000	.12	.3	.7	1.0	1.5	1.7
100,000	.09	.3	.6	.9	1.3	1.5
150,000	.06	.2	.5	.7	1.0	1.2
200,000	.04	.2	.5	.6	.9	1.0
250,000	.03	.2	.4	.6	.8	.9
300,000	.03	.2	.4	.5	.7	.9

[1] Standard errors are presented to nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to nearest one-hundredth of one percentage point.

Illustration of the use of the standard error tables.—Table A-1 of part A of this report shows that in 1976 there were 198,700 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 3,350. The following interpolation procedure was used:

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

Size of estimate	Standard error
150,000	3,100
198,700	x
200,000	3,360

The entry for x is determined by vertically interpolating between 3,100 and 3,360.

$$198{,}700-150{,}000 = 48{,}700$$
$$200{,}000-150{,}000 = 50{,}000$$

$$3{,}100 + \frac{48{,}700}{50{,}000}(3{,}360-3{,}100) = 3{,}350$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 195,350 to 202,050 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 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 193,340 to 204,060 housing units with 90 percent confidence; and that the average estimate lies within the interval from 192,000 to 205,400 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 198,700 owner-occupied housing units, 48,000, or 24.2 percent, had two bedrooms. Interpolation in table II of the appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 24.2 percent is approximately 0.9 percentage points. The following interpolation procedure was used:

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

Base of percentage	Estimated percentage		
	10 or 90	24.2	25 or 75
150,000 .	.7	a	1.0
198,700 .		p	
200,000 .	.6	b.	.9

1. The entry for cell "a" is determined by horizontal interpolation between 0.7 and 1.0.

$$24.2-10.0 = 14.2$$
$$25.0-10.0 = 15.0$$

$$0.7 + \frac{14.2}{15.0}(1.0-0.7) = .98$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.6 and 0.9.

$$24.2-10.0 = 14.2$$
$$25.0-10.0 = 15.0$$

$$0.6 + \frac{14.2}{15.0}(0.9-0.6) = .88$$

3. The entry for "p" was then determined by vertical interpolation between .88 and .98.

$$198{,}700-150{,}000 = 48{,}700$$
$$200{,}000-150{,}000 = 50{,}000$$

$$.98-\frac{48{,}700}{50{,}000}(.98-.88) = 0.9$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 23.3 to 25.1 percent; the 90-percent confidence interval is from 22.8 to 25.6

percent; and the 95-percent confidence interval is from 22.4 to 26.0 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 error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; but, 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 part A of this report shows that in 1976 there were 105,500 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 57,500. Table I shows the standard error of 48,000 is approximately 1,950 and the standard error of 105,500 is approximately 2,730. Therefore, the standard error of the estimated difference of 57,500 is about:

$$3{,}350 = \sqrt{(1{,}950)^2 + (2{,}730)^2}$$

Consequently, the 68-percent confidence interval for the 57,500 difference is from 54,150 to 60,850 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 52,140 to 62,860 housing units, and the 95-percent confidence interval is from 50,800 to 64,200. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval 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 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 table II 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 of a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 2.7. The base of the distribution from which this median was determined is 198,700 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 198,700 is approximately 1.0 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 48.0 and 52.0.
3. From table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 92,600 owner-occupied housing units, or 46.6 percent, had one or two persons (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 34,300 owner-occupied housing units, or 17.3 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:

$$2.5 + (3.5-2.5)\left(\frac{48.0-46.6}{17.3}\right) = 2.6$$

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

$$2.5 + (3.5-2.5)\left(\frac{52.0-46.6}{17.3}\right) = 2.8$$

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

St. Louis 1976

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Annual Housing Survey . . . App-42
Selection of the sample. . . App-42
Coverage improvement sample selection . . App-43
Coverage improvement for deficiency 1 . . App-43
Coverage improvement for deficiency 2 . . App-43
Coverage improvement for deficiencies 3-6 . . App-44
Building loss sample selection . . App-44
1970 Census of Population and Housing. . App-44
ESTIMATION. . App-44
1976 housing inventory . . App-44
1970-1976 lost units . . App-45
Ratio estimation procedure of the 1970 Census of Population and Housing. . App-45
RELIABILITY OF THE ESTIMATES. . App-45
Nonsampling errors. . App-45
1970 census . . App-45
AHS-SMSA . . App-46
Coverage errors . . App-46
Rounding errors. . App-46
Sampling errors for the AHS-SMSA sample . . App-47
Illustration of the use of the standard error tables . . App-48
Differences. . App-49
Illustration of the computation of the standard error of a difference. . App-49
Medians . . App-50
Illustration of the computation of the 95-percent confidence interval of a median . . App-50

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 20 SMSA's are based on data collected from the 1976 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. In each of the 20 SMSA's, the data were collected for the 12-month period from April 1976 through March 1977 with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and data for a second group of 21 SMSA's were collected from April 1975 through March 1976. The sample housing units for each group of the AHS SMSA's are interviewed on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the third group (1976-77) are: Houston, Tex., New York, N.Y., St. Louis, Mo.-Ill., and Seattle-Everett, Wash.

The remaining SMSA's in the third group are: Allentown-Bethlehem-Easton, Pa.-N.J., Baltimore, Md., Birmingham, Ala., Buffalo, N.Y., Cleveland, Ohio, Denver, Colo., Grand Rapids, Mich., Honolulu, Hawaii, Indianapolis, Ind., Las Vegas, Nev., Louisville, Ky.-Ind., Oklahoma City, Okla., Omaha, Nebr.-Iowa, Providence-Pawtucket-Warwick, R.I.-Mass., Raleigh, N.C., and Sacramento, Calif.

In this SMSA, 15,798 units were eligible for interview. Of these sample units, 650 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to the units eligible for interview, 1,867 units were visited but were not eligible for interview, because they were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit-issuing (Honolulu, Las Vegas, New York, and Sacramento) was selected from two sample frames—units enumerated in the 1970 Census of Population and Housing in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for the 16 SMSA's which are *not* 100-percent permit-issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe).

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of the sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA, since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and balance of the SMSA, since the sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total housing units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant

housing unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to 1 of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—Family size	Renter—Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000-$5,999. . .		
$6,000-$9,999. . .		
$10,000-$14,999 .		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to 1 of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to 1 of the 4 vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected in such a way as to produce one-half of the desired sample size. However, the housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of the interview, the units at each of the sample special places were listed and subsampled at a rate to produce an expected four sample units thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since January 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit-issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then 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. At the time of the survey, those segments that did not have an expected size of four were further subdivided to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Coverage improvement sample selection.—The Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS-SMSA sample from the permit-issuing and new construction universes. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970, but completed after April 1, 1970.
2. Mobile homes placed in parks either missed during the 1970 census or established since the census.
3. Units missed in the 1970 census.
4. Units converted to residential use that were nonresidential at the time of the 1970 census.
5. Houses that have been moved onto their present site since the 1970 census.
6. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

Coverage improvement for deficiency 1.—A sample of new construction units whose permits were issued before January 1970, but completed after April 1970, was selected independently for each SMSA. The sampling was carried out in two stages for one-and-two-unit structures and in three stages for three-or-more-unit structures. These units were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units).

The first stage was a sample of permit offices and the second stage a sample of the 1969 permits within each of the selcted permit offices. In the New York, N.Y., SMSA, Honolulu, Hawaii, SMSA, and Grand Rapids, Mich., SMSA, an additional sample of 1968 permits for three-or-more-unit structures was included in the second stage sampling. For the third stage, structures of size three or more were divided into clusters of an expected size of four units and a sample of clusters was selected. This procedure added an estimated 2,077 units to the housing inventory of this SMSA.

Coverage improvement for deficiency 2.—In permit-issuing areas, a sample of mobile homes placed in a park missed by the census or established after the census was selected in two stages. First,

for each 1976-77 SMSA, a sample of tracts was selected and canvassed. All parks were listed and then matched back to the 1970 census to identify parks missed by the census and parks established after the census. Second, the parks were divided into clusters of an expected size of four sites and a sample of clusters was selected and interviewed. Each of the sample units represented the same number of units that the regular AHS sample units represented. This procedure added an estimated 2,158 units to the housing inventory of this SMSA.

Coverage improvement for deficiencies 3-6.—The remaining missed units were sampled by one of two procedures. The first procedure was designed to represent units from the following types of missed structures (structures that had no chance of selection for the AHS):

1. Structures missed in the 1970 census.
2. Structures that were completely nonresidential in the 1970 census but contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and had a utility hookup, or were on the site but not occupied on April 1, 1970, or had no utility hookup but were occupied by persons with no usual residence elsewhere.
4. Houses that had been moved onto their present site since the 1970 census.

Initially, a subsample of regular AHS units was selected at a rate of 1 in 22.2152. Then, succeeding structures in a defined path of travel to the right of the structure containing the sample unit were listed until eight structures (excluding the sample unit structure) were found that had been eligible to be selected for the AHS. Finally, the intervening structures that did not have a chance of selection in the AHS were identified and units within these structures were interviewed. In cases where the interviewer workload would have been too great, a representative subsample of units within these structures was selected. This procedure added an estimated 12,486 units to the housing inventory of this SMSA.

The second procedure was designed to represent missed units from structures represented in the AHS. These missed units were:

1. Units missed in the 1970 census.
2. Units converted to residential use since the 1970 census in structures that contained some residential units in 1970.

First, a subsample of regular AHS units in multi-unit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multi-unit structures selected, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 2,695 units to the housing inventory of this SMSA.

Building loss sample selection.—Some tables in this report show estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS interview). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample previously described. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

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 employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at the time of the interview (i.e., the 1976 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1976 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1976 housing inventory.—The AHS estimates of characteristics of the 1976 housing inventory employed a one-stage ratio estimation procedure. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 650 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units} + \text{Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 50 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated). In addition, a noninterview factor was computed for one noninterview cell for new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes, and one noninterview cell for sample units from both the nonpermit universe (if applicable) and the coverage improvement universe (if units were not included above).

The following ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding ratio estimation category.

The effect of this ratio estimation procedure was to somewhat reduce the variance due to the variation in sample size of strata used in the sample selection of the permit-issuing universe, thereby reducing the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

1970-1976 lost units.—The AHS estimate of characteristics of the 1970-1976 lost units employed a one-stage ratio estimation procedure which is similar to the ratio estimation procedure previously mentioned. The file of 1970-1976 lost units was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error 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 housing inventory from 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 this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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-SMSA 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 can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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 both the 1970 Census of Population and Housing and the 1976 AHS-SMSA 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 included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, series 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.**

AHS-SMSA.—For the 1976 AHS-SMSA 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 sample 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 accuracy of the AHS data collected from interviewed households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct 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.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview. The results of this study are available and are presented in the following Census Bureau memorandum, "Response Error in the Annual Housing Survey as Measured by Reinterviews—Year III (1976-1977) SMSA Sample."

Some of the results of this study are (note that these results are based on the reinterviews across all Year III SMSA's and not for any specific SMSA):

1. For attitudinal items which were not reconciled (i.e., after the question is answered in the reinterview, the enumerator does not present the previous response and then ask the respondent to decide upon the best answer), approximately 67 percent showed moderate levels of response disagreement while the remaining 33 percent showed high levels.

2. Some differences beyond those due to sampling error did occur. Most of the categories affected by bias were categories of attitudinal items.

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low, indices from 20—50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

The results of this study were based on sample data so 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 this study.

Coverage errors.—With respect to errors of coverage and estimation for missing data, the AHS new construction sample had deficiencies in the representation of conventional new construction. Due to time constraints, only those building permits issued more than five months before the survey ended were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. Also, permits were sampled on a monthly basis and were divided equally among the 12 panels in which interviews were conducted. Due to this procedure, some of the permits issued in November 1975 through October 1976 were not interviewed because they were assigned to panels in which the interviewing had already been completed. In this SMSA, 6.8 percent of the permits sampled were not interviewed because of this procedure. This percentage does not include permits issued during the last 5 months of the survey. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units, since, due to the relatively short time span involved, it is possible that construction of these units was not completed at the time the survey was conducted and they were not eligible for interview. In addition to these deficiencies, new construction in special places that do not require building permits, such as military bases, are also not adequately represented.

The Coverage Improvement Program also had certain deficiencies. It appears that the listing procedure used to find units missed in the 1970 census, units converted from nonresidential to residential, houses moved onto their present site, and mobile homes placed outside parks, was not very efficient for finding nonresidential conversions. Such conversions were primarily in business districts whereas the listing procedure started from a residential unit.

It is also felt that deficiencies exist in ED's where area sampling methods are used. It had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing.

Rounding errors.—For 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 or small medians, when these figures are derived from relatively large bases (e.g., median number of persons per household). 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 the survey.

Sampling errors for the AHS-SMSA sample.—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 schedules, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and, thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variance in the estimates due to response and interviewer errors (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 error measured by the standard error, 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 in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from 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 following tables are approximations to the standard errors of various estimates shown in this report for this SMSA. 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 precise standard errors for any specific items.

Table I presents the standard errors applicable to estimates of characteristics of the 1976 housing inventory as well as estimates of characteristics of the 1970-1976 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in this table.

The standard errors on the AHS estimates of population in housing units shown in tables A-1, B-1, and C-1 of part A of this report are 22,100 for the SMSA, 5,600 for the central city of this SMSA, and 21,030 for the balance of this SMSA.

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 II through IV present the standard errors of estimated percentages of the 1976 housing inventory as well as estimated percentages of the 1970-1976 lost units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables II through IV.

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 as 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.

For ratios, 100 (x/y), where x is not a subclass of y, tables II through 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:

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1976 Housing Inventory and for Estimated Number of 1970-1976 Lost Units for the St. Louis, Mo.-Ill., SMSA, for the Central City, and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central city	Not in central city
0	80	40	100
100	90	60	100
200	130	90	140
500	200	140	220
700	240	170	260
1,000 ...	280	200	310
2,500 ...	440	310	500
5,000 ...	630	440	700
10,000 ..	890	620	990
25,000 ..	1,400	950	1,560
50,000 ..	1,960	1,270	2,180
75,000 ..	2,380	1,460	2,630
100,000 .	2,730	1,570	3,000
150,000 .	3,280	1,610	3,580
250,000 .	4,090	–	4,370
500,000 .	5,210	–	5,160
600,000 .	5,440	–	5,140
800,000 .	5,610	–	–

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

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.—Table A-1 of part A of this report shows that in 1976 there were 252,700 renter-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 4,100. The following interpolation procedure was used:

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

Size of estimate	Standard error
250,000	4,090
252,700	x
500,000	5,210

The entry for x is determined as follows by vertically interpolating between 4,090 and 5,210.

$$252{,}700-250{,}000 = 2{,}700$$
$$500{,}000-250{,}000 = 250{,}000$$

$$4{,}090 + \frac{2{,}700}{250{,}000}(5{,}210-4{,}090) = 4{,}100$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 248,600 to 256,800 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 owner-occupied housing units lies within a range computed in this way would be correct for roughly 68

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the St. Louis, Mo.-Ill. SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100	44.2	44.2	44.2	44.2	44.2	44.5
200	28.4	28.4	28.4	28.4	28.4	31.5
500	13.7	13.7	13.7	13.7	17.2	19.9
700	10.2	10.2	10.2	10.2	14.6	16.8
1,000	7.3	7.3	7.3	8.4	12.2	14.1
2,500	3.1	3.1	3.9	5.3	7.7	8.9
5,000	1.6	1.6	2.7	3.8	5.5	6.3
10,000	.8	.9	1.9	2.7	3.9	4.5
25,000	.3	.6	1.2	1.7	2.4	2.8
50,000	.2	.4	.9	1.2	1.7	2.0
75,000	.11	.3	.7	1.0	1.4	1.6
100,000	.08	.3	.6	.8	1.2	1.4
150,000	.05	.2	.5	.7	1.0	1.1
250,000	.03	.2	.4	.5	.8	.9
500,000	.02	.13	.3	.4	.5	.6
600,000	.01	.11	.3	.3	.5	.6
800,000	.01	.10	.2	.3	.4	.5

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point.

Table III. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Housing Units for the Central City of the St. Louis, Mo.-Ill., SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100	28.4	28.4	28.4	28.4	28.4	31.5
200	16.6	16.6	16.6	16.6	19.3	22.3
500	7.4	7.4	7.4	8.5	12.2	14.1
700	5.4	5.4	5.4	7.1	10.3	11.9
1,000	3.8	3.8	4.3	6.0	8.6	10.0
2,500	1.6	1.6	2.7	3.8	5.5	6.3
5,000	.8	.9	1.9	2.7	3.9	4.5
10,000	.4	.6	1.4	1.9	2.7	3.2
25,000	.2	.4	.9	1.2	1.7	2.0
50,000	.08	.3	.6	.8	1.2	1.4
75,000	.05	.2	.5	.7	1.0	1.2
100,000	.04	.2	.4	.6	.9	1.0
150,000	.03	.2	.4	.5	.7	.8

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point.

percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples, lies within the interval from 246,140 to 259,260 housing units with 90 percent confidence; and that the average estimate lies within the interval from 244,500 to 260,900 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 252,700 renter-occupied housing units, 105,900, or 41.9 percent, had two bedrooms. Interpolation in table II of the appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 41.9 percent is approximately 0.9 percentage points. The following interpolation procedure was used:

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

Base of percentage	Estimated percentage		
	25 or 75	41.9	50
250,000 .	0.8	a	0.9
252,700 .		p	
500,000 .	0.5	b	0.6

1. The entry for cell "a" is determined by horizontal interpolation between 0.8 and 0.9.

$$41.9-25.0 = 16.9$$
$$50.0-25.0 = 25.0$$

$$0.8 + \frac{16.9}{25.0}(0.9-0.8) = 0.87$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.5 and 0.6.

$$41.9-25.0 = 16.9$$
$$50.0-25.0 = 25.0$$

$$0.5 + \frac{16.9}{25.0}(0.6-0.5) = 0.57$$

3. The entry for "p" was then determined by vertical interpolation between 0.57 and 0.87.

$$252{,}700-250{,}000 = 2{,}700$$
$$500{,}000-250{,}000 = 250{,}000$$

$$0.87 - \frac{2{,}700}{250{,}000}(0.87-0.57) = 0.9$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 41.0 to 42.8 percent; the 90-percent confidence interval is from 40.5 to 43.3 percent; and the 95-percent confidence interval is from 40.1 to 43.7 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 error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; but, 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 part A of this report shows that in 1976 there were 232,200 owner-occupied units with three bedrooms and 165,900 units with two bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 66,300. Table I shows the standard error of 165,000 is approximately 3,410 and the standard error of 232,200 is approximately 3,950. Therefore, the standard

TABLE IV. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Housing Units for the Balance of the St. Louis, Mo.-Ill., SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100	49.8	49.8	49.8	49.8	49.8	49.8
200	33.2	33.2	33.2	33.2	33.2	35.2
500	16.6	16.6	16.6	16.6	19.3	22.3
700	12.4	12.4	12.4	12.4	16.3	18.8
1,000	9.0	9.0	9.0	9.5	13.6	15.8
2,500	3.8	3.8	4.3	6.0	8.6	10.0
5,000	1.9	1.9	3.1	4.2	6.1	7.0
10,000	1.0	1.0	2.2	3.0	4.3	5.0
25,000	.4	.6	1.4	1.9	2.7	3.2
50,000	.2	.4	1.0	1.3	1.9	2.2
75,000	.13	.4	.8	1.1	1.6	1.8
100,000	.10	.3	.7	.9	1.4	1.6
150,000	.07	.3	.6	.8	1.1	1.3
250,000	.04	.2	.4	.6	.9	1.0
500,000	.02	.14	.3	.4	.6	.7
600,000	.02	.13	.3	.4	.6	.6

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point.

error of the estimated difference of 66,300 is about

$$5{,}220 = \sqrt{(3{,}410)^2 + (3{,}950)^2}$$

Consequently, the 68-percent confidence interval for the 66,300 difference is from 61,080 to 71,520 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 57,950 to 74,650 housing units, and the 95-percent confidence interval is from 55,860 to 76,740. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval 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 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 tables II through IV 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 of a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 2.8. The base of the distribution from which this median was determined is 518,800 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 518,800 is approximately 0.6 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 48.8 and 51.2.
3. From table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 225,200 owner-occupied housing units, or 43.4 percent, had one or two persons (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 95,500 owner-occupied housing units, or 18.4 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:

$$2.5 + (3.5-2.5)\left(\frac{48.8-43.4}{18.4}\right) = 2.8$$

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

$$2.5 + (3.5-2.5)\left(\frac{51.2-43.4}{18.4}\right) = 2.9$$

Thus, the 95-percent confidence interval ranges from 2.8 to 2.9 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 the paragraph on rounding errors in the nonsampling errors section).

Seattle-Everett 1976

Appendix B

Source and Reliability of the Estimates

SAMPLE DESIGN App-42
Annual Housing Survey App-42
Selection of the sample App-42
Coverage improvement sample selection App-43
Coverage improvement for deficiency 1 ..... App-43
Coverage improvement for deficiency 2 ..... App-43
Coverage improvement for deficiencies 3-6 ..... App-44
Building loss sample selection ... App-44
1970 Census of Population and Housing App-44
ESTIMATION App-44
1976 housing inventory App-44
1970-1976 lost units App-45
Ratio estimation procedure of the 1970 Census of Population and Housing App-45
RELIABILITY OF THE ESTIMATES App-45
Nonsampling errors App-45
1970 census App-45
AHS-SMSA App-46
Coverage errors App-46
Rounding errors App-46
Sampling errors for the AHS-SMSA sample App-47
Illustration of the use of the standard error tables App-48
Differences App-49
Illustration of the computation of the standard error of a difference App-49
Medians App-49
Illustration of the computation of the 95-percent confidence interval of a median .. App-50

SAMPLE DESIGN

Annual Housing Survey.—The estimates for each of the 20 SMSA's are based on data collected from the 1976 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. In each of the 20 SMSA's, the data were collected for the 12-month period from April 1976 through March 1977 with one-twelfth of the sample units being visited each month.

Data for the first group of 19 SMSA's were collected for the AHS from April 1974 through March 1975, and data for a second group of 21 SMSA's were collected from April 1975 through March 1976. The sample housing units for each group of the AHS SMSA's are interviewed on a rotating basis.

For each group of SMSA's, the largest SMSA from each of the four geographic regions is represented by a sample of about 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of about 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The largest SMSA's in the third group (1976-77) are: Houston, Tex., New York, N.Y., St. Louis, Mo.-Ill., and Seattle-Everett, Wash.

The remaining SMSA's in the third group are: Allentown-Bethlehem-Easton, Pa.-N.J., Baltimore, Md., Birmingham, Ala., Buffalo, N.Y., Cleveland, Ohio, Denver, Colo., Grand Rapids, Mich., Honolulu, Hawaii, Indianapolis, Ind., Las Vegas, Nev., Louisville, Ky.-Ind., Oklahoma City, Okla., Omaha, Nebr.-Iowa, Providence-Pawtucket-Warwick, R.I.-Mass., Raleigh, N.C., and Sacramento, Calif.

In this SMSA, 14,617 units were eligible for interview. Of these sample units, 811 interviews were not obtained because, for occupied sample units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to the units eligible for interview, 1,291 units were visited but were not eligible for interview, because they were found to be condemned, unfit, demolished, converted to group quarters use, etc.

Selection of the sample.—The sample for the SMSA's which are 100-percent permit-issuing (Honolulu, Las Vegas, New York, and Sacramento) was selected from two sample frames—units enumerated in the 1970 Census of Population and Housing in areas under the jurisdiction of permit-issuing offices (the permit-issuing universe) and units constructed in permit-issuing areas since the 1970 census (the new construction universe). In addition, the sample for the 16 SMSA's which are *not* 100-percent permit-issuing included a sample selected from a third frame—those units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe).

Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of the SMSA for each of the sample frames. The overall sampling rate used to select the sample for each SMSA was determined by the size of the sample. Thus, for the four largest SMSA's, the overall sampling rate differed by central city and balance of the SMSA, since the sample was divided equally between the central city and the balance of the SMSA. The remaining SMSA's had an overall sampling rate about the same for the sample selected from both the central city and balance of the SMSA, since the sample was distributed proportionately between the central city and the balance of the SMSA according to the distribution of total housing units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in permit-issuing areas of the SMSA during the 1970 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from the occupied and vacant housing unit records, the occupied housing unit records were stratified by race of head (non-Black/Black) and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units.

The occupied housing unit records were further stratified so that each unit was assigned to 1 of 50 strata according to its tenure, family size, and household income category as illustrated by the following table:

Household income	Tenure	
	Owner—Family size	Renter—Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000 . . .		
$3,000-$5,999. . .		
$6,000-$9,999. . .		
$10,000-$14,999 .		
$15,000 and over .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to 1 of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to 1 of the 4 vacant strata for either the central city or for the balance of the SMSA. A sample of housing unit records was then selected to produce one-half of the desired sample size. However, the housing unit record adjacent to each of the above sample housing unit records was also selected to be in sample, thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected to produce one-quarter of the desired sample size. However, at the time of the interview, the units at each of the sample special places were listed and subsampled at a rate to produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since January 1970 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

For those SMSA's which are not 100-percent permit-issuing, the remainder of the AHS sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for the universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then 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. At the time of the survey, those segments that did not have an expected size of four were further subdivided to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Coverage improvement sample selection.—The Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS-SMSA sample from the permit-issuing and new construction universes. The coverage deficiencies included the following units:

1. New construction from building permits issued prior to January 1970, but completed after April 1, 1970.
2. Mobile homes placed in parks either missed during the 1970 census or established since the census.
3. Units missed in the 1970 census.
4. Units converted to residential use that were nonresidential at the time of the 1970 census.
5. Houses that have been moved onto their present site since the 1970 census.
6. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

Coverage improvement for deficiency 1.—A sample of new construction units whose permits were issued before January 1970, but completed after April 1970, was selected independently for each SMSA. The sampling was carried out in two stages for one-and-two-unit structures and in three stages for three-or-more-unit structures. These units were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units).

The first stage was a sample of permit offices and the second stage a sample of the 1969 permits within each of the selected permit offices. In the New York, N.Y., SMSA, Honolulu, Hawaii, SMSA, and Grand Rapids, Mich., SMSA, an additional sample of 1968 permits for three-or-more-unit structures was included in the second stage sampling. For the third stage, structures of size three or more were divided into clusters of an expected size of four units and a sample of clusters was selected. This procedure added an estimated 4,200 new construction units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiency 2.—In permit-issuing areas, a sample of mobile homes placed in a park missed by the census or established after the census was selected in two stages. First, for each 1976-77 SMSA, a sample of tracts was selected and canvassed. All parks were listed and then matched back to the 1970 census to identify parks missed by the census and parks established after the

census. Second, the parks were divided into clusters of an expected size of four sites and a sample of clusters was selected and interviewed. Each of the sample units represented the same number of units that the regular AHS sample units represented. This procedure added an estimated 553 units to the coverage of the housing inventory of this SMSA.

Coverage improvement for deficiencies 3-6.—The remaining missed units were sampled by one of two procedures. The first procedure was designed to represent units from the following types of missed structures (structures that had no chance of selection for the AHS):

1. Structures missed in the 1970 census.

2. Structures that were completely nonresidential in the 1970 census but contain units converted to residential use.

3. Mobile homes that had been placed outside parks since the 1970 census and had a utility hookup, or were on the site but not occupied on April 1, 1970, or had no utility hookup but were occupied by persons with no usual residence elsewhere.

4. Houses that had been moved onto their present site since the 1970 census.

Initially, a subsample of regular AHS units was selected at a rate of 1 in 22.2152. Then succeeding structures, in a defined path of travel to the right of the structure containing the sample unit, were listed until eight structures (excluding the sample unit structure) were found that had been eligible to be selected for the AHS. Finally, the intervening structures that did not have a chance of selection in the AHS were identified and units within these structures were interviewed. In cases where the interviewer workload would have been too great, a representative subsample of units within these structures was selected. This procedure added an estimated 16,534 units to the coverage of the housing inventory of this SMSA.

The second procedure was designed to represent missed units from structures represented in the AHS. These missed units were:

1. Units missed in the 1970 census.

2. Units converted to residential use since the 1970 census in structures that contained some residential units in 1970.

First, a subsample of regular AHS units in multi-unit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multi-unit structures selected, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 953 units to the coverage of the housing inventory of this SMSA.

Building loss sample selection.—Some tables in this report show estimates of characteristics of housing units removed from the housing inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS interview). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample previously described. Since the addresses of these sample units were known to exist in 1970, the AHS interviewers could determine which of the sample units were no longer considered to be part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For each of the SMSA's with a nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

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 employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

ESTIMATION

The AHS sample produced two types of estimates for each SMSA: Estimates pertaining to characteristics of the housing inventory at the time of the interview (i.e., the 1976 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1970 (i.e., 1970-1976 lost units). Each type of estimate employed separate, although similar, estimation procedures.

1976 housing inventory.—The AHS estimates of characteristics of the 1976 housing inventory employed a one-stage ratio estimation procedure. Prior to the implementation of the ratio estimation procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 811 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units} + \text{Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within each sector of each SMSA, a noninterview factor was computed separately for 50 noninterview cells for sample housing units from the permit-

issuing universe (where the cells consisted of 1 or more of the different strata used in the stratification of the universe as previously illustrated). In addition, a noninterview factor was computed for one noninterview cell for new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes, and one noninterview cell for sample units from both the nonpermit universe (if applicable) and the coverage improvement universe (if units were not included above).

The following ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor for each cell was equal to the following:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

For each SMSA, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all the AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding ratio estimation category.

The effect of this ratio estimation procedure was to somewhat reduce the variance due to the variation in sample size of strata used in the sample selection of the permit-issuing universe, thereby reducing the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

1970-1976 lost units.—The AHS estimate of characteristics of the 1970-1976 lost units employed a one-stage ratio estimation procedure which is similar to the ratio estimation procedure described previously. The file of 1970-1976 lost units was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. Some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole structure losses and part structure losses. The effect of this estimation procedure was to reduce the sampling error 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 housing inventory from 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 this ratio estimation procedure can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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-SMSA 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 can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

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 both the 1970 Census of Population and Housing and the 1976 AHS-SMSA 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 included space errors, definitional errors, and occupancy errors. The "content" errors measured the accuracy of the data collected for enumerated housing units. These errors were measured by reinterviews, record checks, and other surveys.

The detailed results of these studies, as well as the methodology employed, are presented in the 1970 Census of Population and Housing Evaluation and Research Program Reports, Series 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.**

AHS-SMSA.—For the 1976 AHS-SMSA 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 sample 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 accuracy of the AHS data collected from interviewed households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct 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.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview check. The results of this study are available and are presented in the following Census Bureau memorandum, "Response Error in the Annual Housing Survey as Measured by Reinterviews—Year III (1976-1977) SMSA Sample."

Some of the results of this study are presented below (note that these results are based on the reinterviews across all Year III SMSA's and not for any specific SMSA):

1. For attitudinal items which were not reconciled (i.e., after the question is answered in the reinterview, the enumerator does not present the previous response and then ask the respondent to decide upon the best answer), approximately 67 percent of the indices of inconsistency showed moderate levels of response disagreement while the remaining 33 percent showed high levels.

2. Some differences beyond those due to sampling error did occur. Most of the categories affected by bias were categories of attitudinal items.

The range for evaluating inconsistency is from 0—100. The rule of thumb is that indices below 20 are low; indices from 20—50 are moderate, indicating that there is some problem with inconsistent reporting; those over 50 are high, indicating that improvements are needed in the method used to collect these data or that the category concepts themselves are ambiguous.

The results of this study were based on sample data so 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 this study.

Coverage errors.—With respect to errors of coverage and estimation for missing data, the AHS new construction sample had deficiencies in the representation of conventional new construction. Due to time constraints, only those building permits issued more than 5 months before the survey ended were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. Also, permits were sampled on a monthly basis and were divided equally among the 12 panels in which interviews were conducted. Due to this procedure, some of the permits issued in November 1975 through October 1976 were not interviewed because they were assigned to panels in which the interviewing had already been completed. In this SMSA, 13.9 percent of the permits sampled were not interviewed because of this procedure. This percentage does not include permits issued during the last 5 months of the survey. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units, since, due to the relatively short time span involved, it is possible that construction of these units was not completed at the time the survey was conducted and they were not eligible for interview. In addition to these deficiencies, new construction in special places that do not require building permits, such as military bases, are also not adequately represented.

The Coverage Improvement Program also had certain deficiencies. It appears that the listing procedure used to find units missed in the 1970 census, units converted from nonresidential to residential, houses moved onto their present site, and mobile homes placed outside parks, was not very efficient for finding nonresidential conversions. Such conversions were primarily in business districts, whereas the listing procedure started from a residential unit.

Deficiencies also exist in ED's where area sampling methods are used. It had been assumed that all units located inside these ED's would be represented in the sample. However, it has been estimated that the 1976 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing.

Rounding errors.—For 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 or small medians, when these figures are derived from relatively large bases (e.g., median number of persons per household). 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 the survey.

Sampling errors for the AHS-SMSA sample.—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 schedules, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to provide a measure of this variation among the estimates from the possible samples and thus, is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variance in the estimates due to response and interviewer errors (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 error measured by the standard error, 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 in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from 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 following tables are approximations to the standard errors of various estimates shown in this report for this SMSA. 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 precise standard errors for any specific items.

Table I presents the standard errors applicable to estimates of characteristics of the 1976 housing inventory as well as estimates of characteristics of the 1970-1976 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in this table. The standard errors on the AHS estimates of the population in housing units shown in tables A-1, B-1, and C-1 of part A of this report are 8,610 for the total SMSA, 4,540 for the central cities of the SMSA, and 7,290 for the balance of the SMSA.

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 II through IV present the standard errors of estimated percentages of the 1976 housing inventory as well as estimated percentages of the 1970-1976 lost units. Two-way interpolation should be used to determine standard errors for estimated percentages not specifically shown in tables II through IV.

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 as 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.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1976 Housing Inventory and for Estimated Number of 1970-1976 Lost Units for the Seattle-Everett, Wash., SMSA, for the Central Cities, and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard error[1]		
	SMSA	In central cities	Not in central cities
0	40	40	50
100.	70	60	70
200.	90	90	100
500.	150	140	160
700.	180	170	190
1,000 . . .	210	200	220
2,500 . . .	330	310	350
5,000 . . .	470	440	490
10,000 . .	660	620	700
25,000 . . .	1,030	950	1,080
50,000 . .	1,440	1,290	1,490
75,000 . .	1,730	1,510	1,770
100,000. .	1,970	1,650	1,980
150,000. .	2,320	1,790	2,250
200,000. .	2,570	1,760	2,390
250,000. .	2,760	—	2,420
300,000. .	2,880	—	2,340
400,000. .	2,970	—	—
500,000. .	2,880	—	—

[1] For estimates pertaining to new construction, the standard error shown in this table should be multiplied by a factor of 1.3 for estimates for the total SMSA, 1.2 for estimates for the central cities of the SMSA, and 1.0 for the balance of the SMSA.

For ratios, 100 (x/y), where x is not a subclass of y, tables II through 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.—Table A-1 of part A of this report shows that in 1976 there were 337,700 owner-occupied housing units in this SMSA. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 2,910. The following interpolation procedure was used:

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

Size of estimate	Standard error
300,000.	2,880
337,700.	x
400,000.	2,970

The entry for x is determined by vertically interpolating between 2,880 and 2,970.

$$337{,}700-300{,}000 = 37{,}700$$
$$400{,}000-300{,}000 = 100{,}000$$

$$2{,}880+\frac{37{,}700}{100{,}000}(2{,}970-2{,}880)$$
$$= 2{,}910$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 334,790 to 340,610 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1976 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 333,040 to 342,360 housing units with 90 percent confidence; and that the average estimate lies within the interval from 331,880 to 343,520 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 337,700 owner-occupied housing units, 78,800, or 23.3 percent, had two bedrooms. Interpolation in table II of the appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 23.3 percent is approximately 0.5 percentage points. The following interpolation procedure was used:

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

Base of percentage	Estimated percentage		
	10 or 90	23.3	25 or 75
300,000. . . .	.4	a	.5
337,700. . . .		p	
400,000. . . .	.3	b	.5

1. The entry for cell "a" is determined by horizontal interpolation between 0.4 and 0.5.

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Units for the Seattle-Everett, Wash., SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100.	30.6	30.6	30.6	30.6	30.6	33.2
200.	18.1	18.1	18.1	18.1	20.3	23.5
500.	8.1	8.1	8.1	8.9	12.9	14.9
700.	5.9	5.9	5.9	7.5	10.9	12.6
1,000	4.2	4.2	4.6	6.3	9.1	10.5
2,500	1.7	1.7	2.9	4.0	5.8	6.6
5,000	.9	.9	2.0	2.8	4.1	4.7
10,000	.4	.7	1.4	2.0	2.9	3.3
25,000	.2	.4	.9	1.3	1.8	2.1
50,000	.09	.3	.6	.9	1.3	1.5
75,000	.06	.2	.5	.7	1.1	1.2
100,000. . . .	.04	.2	.5	.6	.9	1.1
150,000. . . .	.03	.2	.4	.5	.7	.9
200,000. . . .	.02	.15	.3	.4	.6	.7
250,000. . . .	.02	.13	.3	.4	.6	.7
300,000. . . .	.01	.12	.3	.4	.5	.6
400,000. . . .	.01	.10	.2	.3	.5	.5
500,000. . . .	.01	.09	.2	.3	.4	.5

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point. For estimates pertaining to new construction, the standard error shown in this table should be multiplied by a factor of 1.3 for estimates from the total SMSA.

$$23.3-10.0 = 13.3$$
$$25.0-10.0 = 15.0$$

$$0.4 + \frac{13.3}{15.0}(0.5-0.4) = .49$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.3 and 0.5.

$$23.3-10.0 = 13.3$$
$$25.0-10.0 = 15.0$$

$$0.3 + \frac{13.3}{15.0}(0.5-0.3) = .48$$

3. The entry for "p" was then determined by vertical interpolation between .48 and .49

$$337{,}700-300{,}000 = 37{,}700$$
$$400{,}000-300{,}000 = 100{,}000$$

$$.49 - \frac{37{,}700}{100{,}000}(.49-.48) = 0.5$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 22.8 to 23.8 percent; the 90-percent confidence interval is from 22.5 to 24.1 percent; and the 95-percent confidence interval is from 22.3 to 24.3 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 error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. However, if there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error; but, 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 part A of this report shows that in 1976 there were 154,800 owner-occupied units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied units with two bedrooms and owner-occupied units with three bedrooms is 76,000. Table I shows the standard error of 78,800 is approximately 1,770 and the standard error of 154,800 is approximately 2,340. Therefore, the standard error of the estimated difference of 76,000 is about:

$$2{,}930 = \sqrt{(1{,}770)^2 + (2{,}340)^2}$$

Consequently, the 68-percent confidence interval for the 76,000 difference is from 73,070 to 78,930 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference 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 71,310 to 80,690 housing units, and the 95-percent confidence interval is from 70,140 to 81,860. Thus, we can conclude with 95 percent confidence that the number of 1976 owner-occupied units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 95-percent confidence interval 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 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:

TABLE III. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Housing Units for the Central Cities of the Seattle-Everett, Wash., SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100.	28.2	28.2	28.2	28.2	28.2	31.3
200.	16.4	16.4	16.4	16.4	19.2	22.1
500.	7.3	7.3	7.3	8.4	12.1	14.0
700.	5.3	5.3	5.3	7.1	10.2	11.8
1,000	3.8	3.8	4.3	5.9	8.6	9.9
2,500	1.5	1.5	2.7	3.8	5.4	6.3
5,000	.8	.9	1.9	2.7	3.8	4.4
10,000	.4	.6	1.4	1.9	2.7	3.1
25,000	.2	.4	.9	1.2	1.7	2.0
50,000	0.8	.3	.6	.8	1.2	1.4
75,000	.05	.2	.5	.7	1.0	1.1
100,000. . . .	.04	.2	.4	.6	.9	1.0
150,000. . . .	.03	.2	.4	.5	.7	.8
200,000. . . .	.02	.14	.3	.4	.6	.7

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point. For estimates pertaining to new construction, the standard error shown in this table should be multiplied by a factor of 1.2 for estimates from the central cities of the SMSA.

TABLE IV. Standard Errors for Estimated Percentages of Housing Units in the 1976 Housing Inventory and for Estimated Percentages of 1970-1976 Lost Housing Units for the Balance of the Seattle-Everett, Wash., SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100.	33.1	33.1	33.1	33.1	33.1	35.2
200.	19.8	19.8	19.8	19.8	21.5	24.9
500.	9.0	9.0	9.0	9.4	13.6	15.7
700.	6.6	6.6	6.6	8.0	11.5	13.3
1,000	4.7	4.7	4.8	6.7	9.6	11.1
2,500	1.9	1.9	3.1	4.2	6.1	7.0
5,000	1.0	1.0	2.2	3.0	4.3	5.0
10,000	.5	.7	1.5	2.1	3.0	3.5
25,000	.2	.4	1.0	1.3	1.9	2.2
50,000	.10	.3	.7	.9	1.4	1.6
75,000	.07	.3	.6	.8	1.1	1.3
100,000. . . .	.05	.2	.5	.7	1.0	1.1
150,000. . . .	.03	.2	.4	.5	.8	.9
200,000. . . .	.02	.2	.3	.5	.7	.8
250,000. . . .	.02	.14	.3	.4	.6	.7
300,000. . . .	.02	.13	.3	.4	.6	.6

[1] Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than fifteen-hundredths of one percentage point; in those cases, the standard error is shown to the nearest one-hundredth of one percentage point.

1. From table II 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 of a median.—Table A-1 of part A of this report shows the median number of persons for owner-occupied housing units is 2.7. The base of the distribution from which this median was determined is 337,700 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 337,700 is approximately 0.6 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 48.8 and 51.2.
3. From table A-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 157,200 owner-occupied housing units, or 46.6 percent, had one or two persons (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 57,900 owner-occupied housing units, or 17.1 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:

$$2.5 + (3.5-2.5)\left(\frac{48.8-46.6}{17.1}\right) = 2.6$$

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

$$2.5 + (3.5-2.5)\left(\frac{51.2-46.6}{17.1}\right) = 2.8$$

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