

# 1978 Metropolitan Appendix B: Source and Accuracy

## Table of Contents

Atlanta .....	3
Cincinnati .....	13
Columbus.....	22
Kansas City .....	31
New Orleans .....	40
Newport News-Hampton .....	49
Paterson-Clifton-Passaic .....	58
Philadelphia .....	67
Riverside-San Bernardino-Ontario .....	77
Rochester .....	86
San Antonio .....	95
San Diego .....	104
San Francisco-Oakland .....	113
Springfield-Chicopee-Holyoke .....	123

Atlanta 1978

# Appendix B

## Source and Reliability of the Estimates

<b>SAMPLE DESIGN</b> . . . . .	App-40	<b>Coverage improvement for deficiencies 3-6</b> . . . . .	App-42	<b>AHS-SMSA</b> . . . . .	App-44
<b>Annual Housing Survey</b> . . . . .	App-40	<b>1970 Census of Population and Housing</b> . . . . .	App-43	<b>Coverage errors</b> . . . . .	App-44
<b>Designation of sample housing units for the 1978 survey</b> . . . . .	App-40	<b>ESTIMATION</b> . . . . .	App-43	<b>Rounding errors</b> . . . . .	App-44
<b>Selection of the 1975 AHS-SMSA sample</b> . . . . .	App-40	<b>1978 housing inventory</b> . . . . .	App-43	<b>Sampling errors for the AHS-SMSA sample</b> . . . . .	App-44
<b>1975-1978 additions to the housing inventory</b> . . . . .	App-41	<b>1975-1978 lost units</b> . . . . .	App-43	<b>Illustration of the use of the standard error tables</b> . . . . .	App-45
<b>Sample selection for the 1978 Coverage Improvement Program</b> . . . . .	App-42	<b>1975 estimation procedure</b> . . . . .	App-43	<b>Differences</b> . . . . .	App-47
<b>Coverage improvement for deficiency 1</b> . . . . .	App-42	<b>Ratio estimation procedure of the 1970 Census of Population and Housing</b> . . . . .	App-43	<b>Illustration of the computation of the standard error of a difference</b> . . . . .	App-47
<b>Coverage improvement for deficiency 2</b> . . . . .	App-42	<b>RELIABILITY OF THE ESTIMATES</b> . . . . .	App-43	<b>Medians</b> . . . . .	App-48
		<b>Nonsampling errors</b> . . . . .	App-44	<b>Illustration of the computation of the 95-percent confidence interval of a median</b> . . . . .	App-48
		<b>1970 census</b> . . . . .	App-44		

### SAMPLE DESIGN

**Annual Housing Survey**—The estimates for each of the 15 SMSA's are based on data collected from the 1978 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 15 SMSA's, the data were collected for the 12-month period from April 1978 through March 1979 with one-twelfth of the sample units being visited each month.

Each group of SMSA's is being interviewed for the AHS on a rotating basis with this group (AA-1 SMSA's) being the second to be revisited. All of these SMSA's were enumerated for the first time in 1975.

For each group of SMSA's, the largest SMSA from 3 of the 4 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 the remaining 12 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 AA-1 group (1978-1979) are: Atlanta, Ga.; Philadelphia, Pa.-N.J.; and San Francisco-Oakland, Calif.

The remaining SMSA's in the AA-1 group are: Cincinnati, Ohio-Ky.-Ind.; Colorado Springs, Colo.; Columbus, Ohio; Kansas City, Mo.-Kans.; New Orleans, La.; Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; Rochester, N.Y.; San Antonio, Tex.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In this SMSA, 13,597 housing units were eligible for interview. Of these sample units, 676 interviews were not obtained because, for occupied sample units, the occupants refused to be interviewed, 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 units eligible for interview, 1,264 units were visited but were not eligible for interview because they were condemned, unfit, demolished, converted to group quarters use, etc.

**Designation of sample housing units for the 1978 survey**—The sample housing units designated to be interviewed in the 1978 survey consisted of the following categories which are described in detail in the succeeding sections.

1. All sample housing units that were interviewed in the 1975 survey.
2. All sample housing units that were type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of the survey but which could become eligible in the future) in the 1975 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1978 AHS questionnaire, page App-20.)
3. All sample housing units that were selected from the list of new construction building permits issued since the 1975 survey. (This sample represented the housing units built in permit-issuing areas, since the 1975 survey.)
4. All sample housing units that were added to sample segments in the nonpermit universe since the 1975 survey. (This sample represented additions to the housing inventory in nonpermit-issuing areas since the 1975 survey.)
5. All sample housing units that were selected as part of the 1978 Coverage Improvement Program. (This sample represented most of the housing units which, until 1978, did not have a chance of selection.)

**Selection of the 1975 AHS-SMSA sample**—The sample for the SMSA's which are 100-percent permit-issuing 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 those 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 non-permit universe). The following SMSA's are 100-percent permit-issuing: Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and San Francisco-Oakland, Calif. The remaining 10 SMSA's contain a sample from 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 three 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 the total housing units in each sector.

The major portion of the sample in each SMSA 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 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 1 2 3 4 5+	Renter— Family size 1 2 3 4 5+
Under \$3,000 . . . . .		
\$3,000 to \$5,999 . . . . .		
\$6,000 to \$9,999 . . . . .		
\$10,000 to \$14,999 . . . . .		
\$15,000 and over . . . . .		

Thus, for this SMSA, the occupied housing unit records from the permit-issuing 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 selection procedure was then instituted that would produce one-half of the desired sample size. However, whenever a record was selected to be in sample, the housing unit record adjacent to it on the file 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 by a procedure that produced one-quarter of the desired sample size. However, at the time of the survey, the units at each of the special places were listed and subsampled at a rate which produced 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 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:

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

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. All units in existence at the time of interview in these selected segments are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

**1975-1978 additions to the housing inventory**—In the permit-issuing universe, a sample of new construction building permits,

issued since the 1975 survey, was selected to represent housing units built in permit-issuing areas since the 1975 survey. Sampling procedures were identical to those used in selecting the 1970-1975 new construction sample, which were described previously. In the nonpermit universe, sample segments were dependently recanvassed, using listing sheets from 1975, to identify any units missed in the 1975 survey or any units added since the 1975 survey.

**Sample selection for the 1978 Coverage Improvement Program**—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 1970 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 for each of the AA-1 SMSA's except Newport News-Hampton, Va. Two different procedures were used. For the first procedure, the sampling was carried out in two stages for one- and two-unit structures and in three stages for three-or-more-unit structures. Sample units selected from one- and two-unit structures were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units), while sample units selected from three-or-more-unit structures were sampled at one-half the rate of 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. Since these permits were not available for Newport News-Hampton, Va., SMSA, these coverage improvement procedures were not instituted for this SMSA. For the third stage, structures of size three or more were divided into clusters of an expected size of two units and a sample of clusters was selected. This procedure was employed in the following SMSA's: Colorado Springs, Colo.; Rochester, N.Y.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In the remaining 11 SMSA's, units whose permits were issued before January 1970, but which were completed after April 1970, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. These units were then sampled at one-third the rate of regular AHS units. These procedures added an estimated 11,002 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 that was missed by the census or established after the census was selected in two stages. First, for each 1978-1979 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 1,120 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 now contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and have a utility hookup, or were on the site during the present survey 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 AHS sample units was selected from the permit-issuing universe at a rate of 1 in 24. 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 15,720 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 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 above, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 1,113 units to the coverage of the housing inventory of this SMSA.

**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 1978 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1975 (i.e., 1975-1978 lost units). Each type of estimate employed separate, although similar, estimation procedures.

**1978 housing inventory**—The AHS estimates of characteristics of the 1978 housing inventory employed a 1-stage ratio estimation procedure in all SMSA's. 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 non-interviews previously mentioned. This noninterview adjustment was done separately for occupied and vacant housing units. 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 described). In addition, within each sector separate noninterview factors were computed for one noninterview cell for conventional new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one non-interview cell for mobile homes from both the nonpermit universe and the coverage improvement universe, and one noninterview cell for other sample units from both the nonpermit universe and the coverage improvement universe, (if units were not included in any of the previous cells).

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.

**1975-1978 lost units**—The 1975-1978 lost unit estimates employed the three-stage ratio estimation procedure used to produce the AHS-SMSA estimates of the 1975 housing inventory, as was described in the 1975 Current Housing Report, Series H-170, **Housing Characteristics for Selected Metropolitan Areas**. Since the 1975-1978 lost units existed, by definition, in the 1975 housing inventory, there was a 1975 housing inventory weight associated with each 1975-1978 lost unit. This weight was used to tabulate the estimates of the characteristics of the 1975-1978 lost units.

**1975 estimation procedure**—This report presents data on the housing characteristics of the 1975 housing inventory from the 1975 Annual Housing Survey SMSA sample. The AHS-SMSA estimation procedure employed a three-stage ratio estimation process. A detailed description of this ratio estimation procedure can be found in the AHS Series H-170 reports for 1975.

**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 the 1970 Census of Population and Housing and the 1975 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 the extent to which occupancy status was erroneously reported. 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 1978 AHS-SMSA sample, a reinterview program was not conducted. However, a study was conducted for the 1975 AHS-SMSA sample. The results of which are presented in the Census Bureau memorandum, "Reinterview Results for Annual Housing Survey—SMSA Sample: 1975."

**Coverage errors**—In 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. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units. 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, in which case, they would

not have been 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 correct deficiencies 3-6 (see the coverage improvement section of this appendix) was not very effective in 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 1975 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing. It should be noted that since these ED's were recanvassed for the 1978 survey, the number of missed housing units may be considerably less for 1978.

**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 statistics 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 sampling error of a survey estimate provides a measure of the variation among the estimates from all 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.

One common measure of the sampling error is the standard error. As calculated for this report, the standard error reflects the variation in the estimates due to sampling and nonsampling errors, but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors measured by the standard error, biases, and any 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 item. Standard errors applicable to estimates of characteristics of the 1975 housing inventory can be found in the AHS Series H-170 reports for 1975.

Table I presents the standard errors applicable to estimates of characteristics of the 1978 housing inventory as well as estimates of characteristics of the 1975-1978 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 14,350 for the total SMSA, 5,910 for the central city of the SMSA, and 12,820 for the balance of the SMSA.

The reliability of an estimated percentage, computed by using the 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 1978 housing inventory as well as estimated percentages of the 1975-1978 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

**TABLE I. Standard Errors for Estimated Number of Housing Units in the 1978 Housing Inventory and for Estimated Number of 1975-1978 Lost Units for the Atlanta, Ga., SMSA, for the Central City of the SMSA and for the Balance (Not in Central City) of the SMSA.**

(68 chances out of 100)

Size of estimate	Standard error <sup>1</sup>		
	SMSA	In central city	Not in central city
0	60	30	80
100	80	60	90
200	110	80	120
500	180	130	190
700	210	150	230
1,000	250	180	280
2,500	390	290	440
5,000	550	410	620
10,000	780	570	870
25,000	1,230	880	1,360
50,000	1,720	1,210	1,900
75,000	2,090	1,430	2,300
100,000	2,400	1,590	2,620
150,000	2,880	1,780	3,120
200,000	3,270	1,850	3,490
250,000	3,590	—	3,780
300,000	3,860	—	4,000
400,000	4,270	—	4,280
500,000	4,570	—	4,370
600,000	4,760	—	—
700,000	4,870	—	—

<sup>1</sup> For estimates pertaining to new construction, the standard errors shown in the table should be multiplied by a factor of 1.2 for the total SMSA, 1.3 for the central city, and 1.0 for the balance (not in central city) estimates.

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  
 $\sigma_x$  = the standard error of the numerator  
 $\sigma_y$  = 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 1978 there were 329,800 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 3,980. 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 .....	3,860
329,800 .....	x
400,000 .....	4,270

The entry for x is determined as follows by vertically interpolating between 3,860 and 4,270:

$$329,800 - 300,000 = 29,800$$

$$400,000 - 300,000 = 100,000$$

$$3,860 + \frac{29,800}{100,000} (4,270 - 3,860) = 3,980$$

**TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1978 Housing Inventory and for Estimated Percentages of 1975-1978 Lost Units for the Atlanta, Ga., SMSA**

(68 chances out of 100)

Base of percentage	Estimated percentage <sup>1</sup>					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100.....	38.0	38.0	38.0	38.0	38.0	39.2
200.....	23.5	23.5	23.5	23.5	24.0	27.7
500.....	10.9	10.9	10.9	10.9	15.2	17.5
700.....	8.1	8.1	8.1	8.9	12.8	14.8
1,000.....	5.8	5.8	5.8	7.4	10.7	12.4
2,500.....	2.4	2.4	3.4	4.7	6.8	7.8
5,000.....	1.2	1.2	2.4	3.3	4.8	5.5
10,000.....	0.6	0.8	1.7	2.4	3.4	3.9
25,000.....	0.2	0.5	1.1	1.5	2.1	2.5
50,000.....	0.12	0.3	0.8	1.1	1.5	1.8
75,000.....	0.08	0.3	0.6	0.9	1.2	1.4
100,000.....	0.06	0.2	0.5	0.7	1.1	1.2
150,000.....	0.04	0.2	0.4	0.6	0.9	1.0
200,000.....	0.03	0.2	0.4	0.5	0.8	0.9
250,000.....	0.02	0.2	0.3	0.5	0.7	0.8
300,000.....	0.02	0.14	0.3	0.4	0.6	0.7
400,000.....	0.02	0.12	0.3	0.4	0.5	0.6
500,000.....	0.01	0.11	0.2	0.3	0.5	0.6
600,000.....	0.01	0.10	0.2	0.3	0.4	0.5
700,000.....	0.01	0.09	0.2	0.3	0.4	0.5

<sup>1</sup> 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 errors shown in the table should be multiplied by a factor of 1.2.

Consequently, the 68-percent confidence interval, as shown by these data, is from 325,820 to 333,780 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1978 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 323,432 to 336,168 housing units with 90 percent confidence; and that the average estimate lies within the interval from 321,840 to 337,760 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 329,800 owner-occupied housing units, 69,000 or 20.9 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 20.9 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	20.9	25 or 75
300,000 .....	0.4	a	0.6
329,800 .....		p	
400,000 .....	0.4	b	0.5

1. The entry for cell "a" is determined by horizontal interpolation between 0.4 and 0.6.

$$20.9 - 10.0 = 10.9$$

$$25.0 - 10.0 = 15.0$$

$$0.4 + \frac{10.9}{15.0} (0.6 - 0.4) = 0.5$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.4 and 0.5.

$$20.9 - 10.0 = 10.9$$

$$25.0 - 10.0 = 15.0$$

$$0.4 + \frac{10.9}{15.0} (0.5 - 0.4) = 0.5$$

3. The entry for "p" was then determined by vertical interpolation between 0.5 and 0.5.

$$329,800 - 300,000 = 29,800$$

$$400,000 - 300,000 = 100,000$$

$$0.5 + \frac{29,800}{100,000} (0.5 - 0.5) = 0.5$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 20.4 to 21.4 percent; the 90-percent confidence interval is from 20.1 to 21.7 percent; and the 95-percent confidence interval is from 19.9 to 21.9 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 characteristics 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. Due to the overlap of the 1975 and 1978 AHS-SMSA samples a positive correlation should be expected when making comparisons between 1975 and 1978 characteristics.

**Illustration of the computation of the standard error of a difference**—Table A-1 of part A of this report shows that in 1978 there were 176,700 owner-occupied housing units with

three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied housing units with two bedrooms and owner-occupied housing units with three bedrooms is 107,700. Table I shows the standard error of 176,700 is approximately 3,090, and the standard error of 69,000 is approximately 2,000. Therefore, the standard error of the estimated difference of 107,700 is about:

$$3,680 = \sqrt{(3,090)^2 + (2,000)^2}$$

Consequently, the 68-percent confidence interval for the 107,700 difference is from 104,020 to 111,380 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 101,812 to 113,588 housing units, and the 95-percent confidence interval is from 100,340 to 115,060. Thus, we can conclude with 95 percent confidence that the

**TABLE III. Standard Errors for Estimated Percentages of Housing Units in the 1978 Housing Inventory and for Estimated Percentages of the 1975-1978 Lost Housing Units for the Central City of the Atlanta, Ga. SMSA**

(68 chances out of 100)

Base of percentage	Estimated percentage <sup>1</sup>					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100. ....	25.0	25.0	25.0	25.0	25.0	28.9
200. ....	14.3	14.3	14.3	14.3	17.7	20.4
500. ....	6.2	6.2	6.2	7.7	11.2	12.9
700. ....	4.5	4.5	4.8	6.5	9.4	10.9
1,000. ....	3.2	3.2	4.0	5.5	7.9	9.1
2,500. ....	1.3	1.3	2.5	3.5	5.0	5.8
5,000. ....	0.7	0.8	1.8	2.4	3.5	4.1
10,000. ....	0.3	0.6	1.3	1.7	2.5	2.9
25,000. ....	0.13	0.4	0.8	1.1	1.6	1.8
50,000. ....	0.07	0.3	0.6	0.8	1.1	1.3
75,000. ....	0.04	0.2	0.5	0.6	0.9	1.1
100,000. ....	0.03	0.2	0.4	0.5	0.8	0.9
150,000. ....	0.02	0.15	0.3	0.4	0.6	0.7
200,000. ....	0.02	0.13	0.3	0.4	0.6	0.6

<sup>1</sup> 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 errors shown in the table should be multiplied by a factor of 1.3.

**TABLE IV. Standard Errors for Estimated Percentages of Housing Units in the 1978 Housing Inventory and for Estimated Percentages of 1975-1978 Lost Housing Units for the Balance (Not in Central City) of the SMSA**

(68 chances out of 100)

Base of percentage	Estimated percentage <sup>1</sup>					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100. ....	43.2	43.2	43.2	43.2	43.2	43.6
200. ....	27.6	27.6	27.6	27.6	27.6	30.8
500. ....	13.2	13.2	13.2	13.2	16.9	19.5
700. ....	9.8	9.8	9.8	9.9	14.3	16.5
1,000. ....	7.1	7.1	7.1	8.3	11.9	13.8
2,500. ....	3.0	3.0	3.8	5.2	7.6	8.7
5,000. ....	1.5	1.5	2.7	3.7	5.3	6.2
10,000. ....	0.8	0.9	1.9	2.6	3.8	4.4
25,000. ....	0.3	0.5	1.2	1.7	2.4	2.8
50,000. ....	0.15	0.4	0.9	1.2	1.7	2.0
75,000. ....	0.10	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	1.1
200,000. ....	0.04	0.2	0.4	0.6	0.8	1.0
250,000. ....	0.03	0.2	0.4	0.5	0.8	0.9
300,000. ....	0.03	0.2	0.3	0.5	0.7	0.8
400,000. ....	0.02	0.14	0.3	0.4	0.6	0.7
500,000. ....	0.02	0.12	0.3	0.4	0.5	0.6

<sup>1</sup> 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 errors shown in the table should be multiplied by a factor of 1.0.

number of 1978 owner-occupied housing units with three bedrooms is greater than the number of owner-occupied housing 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 characteristics, 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

\*B-D

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 329,800 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 329,800 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 133,300 owner-occupied housing units or 40.4 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 69,300 owner-occupied housing units, or 21.0 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:

$$2.5 + (3.5-2.5) \left( \frac{48.6-40.4}{21.0} \right) = 2.9$$

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

$$2.5 + (3.5-2.5) \left( \frac{51.4-40.4}{21.0} \right) = 3.0$$

Thus, the 95-percent confidence interval ranges from 2.9 to 3.0 persons.

Cincinnati 1978

# Appendix B

## Source and Reliability of the Estimates

<b>SAMPLE DESIGN</b> . . . . .	App-40	<b>Coverage improvement for deficiencies 3-6</b> . . . . .	App-42	<b>AHS-SMSA</b> . . . . .	App-44
<b>Annual Housing Survey</b> . . . . .	App-40	<b>1970 Census of Population and Housing</b> . . . . .	App-43	<b>Coverage errors</b> . . . . .	App-44
<b>Designation of sample housing units for the 1978 survey</b> . . . . .	App-40	<b>ESTIMATION</b> . . . . .	App-43	<b>Rounding errors</b> . . . . .	App-44
<b>Selection of the 1975 AHS-SMSA sample</b> . . . . .	App-40	<b>1978 housing inventory</b> . . . . .	App-43	<b>Sampling errors for the AHS-SMSA sample</b> . . . . .	App-44
<b>1975-1978 additions to the housing inventory</b> . . . . .	App-42	<b>1975-1978 lost units</b> . . . . .	App-43	<b>Illustration of the use of the standard error tables</b> . . . . .	App-45
<b>Sample selection for the 1978 Coverage Improvement Program</b> . . . . .	App-42	<b>1975 estimation procedure</b> . . . . .	App-43	<b>Differences</b> . . . . .	App-46
<b>Coverage improvement for deficiency 1</b> . . . . .	App-42	<b>Ratio estimation procedure of the 1970 Census of Population and Housing</b> . . . . .	App-43	<b>Illustration of the computation of the standard error of a difference</b> . . . . .	App-46
<b>Coverage improvement for deficiency 2</b> . . . . .	App-42	<b>RELIABILITY OF THE ESTIMATES</b> . . . . .	App-43	<b>Medians</b> . . . . .	App-47
		<b>Nonsampling errors</b> . . . . .	App-44	<b>Illustration of the computation of the 95-percent confidence interval of a median</b> . . . . .	App-47
		<b>1970 census</b> . . . . .	App-44		

### SAMPLE DESIGN

**Annual Housing Survey**—The estimates for each of the 15 SMSA's are based on data collected from the 1978 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 15 SMSA's, the data were collected for the 12-month period from April 1978 through March 1979 with one-twelfth of the sample units being visited each month.

Each group of SMSA's is being interviewed for the AHS on a rotating basis with this group (AA-1 SMSA's) being the second to be revisited. All of these SMSA's were enumerated for the first time in 1975.

For each group of SMSA's, the largest SMSA from 3 of the 4 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 the remaining 12 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 AA-1 group (1978-1979) are: Atlanta, Ga.; Philadelphia, Pa.-N.J.; and San Francisco-Oakland, Calif.

The remaining SMSA's in the AA-1 group are: Cincinnati, Ohio-Ky.-Ind.; Colorado Springs, Colo.; Columbus, Ohio; Kansas City, Mo.-Kans.; New Orleans, La.; Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; Rochester, N.Y.; San Antonio, Tex.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In this SMSA, 4,794 housing units were eligible for interview. Of these sample units, 243 interviews were not obtained because, for occupied sample units, the occupants refused to be interviewed, 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 units eligible for interview, 386 units were visited but were not eligible for interview because they were condemned, unfit, demolished, converted to group quarters use, etc.

**Designation of sample housing units for the 1978 survey**—The sample housing units designated to be interviewed in the 1978 survey consisted of the following categories which are described in detail in the succeeding sections.

1. All sample housing units that were interviewed in the 1975 survey.
2. All sample housing units that were type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of the survey but which could become eligible in the future) in the 1975 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1978 AHS questionnaire, page App-20.)
3. All sample housing units that were selected from the list of new construction building permits issued since the 1975 survey. (This sample represented the housing units built in permit-issuing areas, since the 1975 survey.)
4. All sample housing units that were added to sample segments in the nonpermit universe since the 1975 survey. (This sample represented additions to the housing inventory in nonpermit-issuing areas since the 1975 survey.)
5. All sample housing units that were selected as part of the 1978 Coverage Improvement Program. (This sample represented most of the housing units which, until 1978, did not have a chance of selection.)

**Selection of the 1975 AHS-SMSA sample**—The sample for the SMSA's which are 100-percent permit-issuing 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 those 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 non-permit universe). The following SMSA's are 100-percent permit-issuing: Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and San Francisco-Oakland, Calif. The remaining 10 SMSA's contain a sample from 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 three largest SMSA's, the overall sampling rate differed by central city and the 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 the 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 the total housing units in each sector.

The major portion of the sample in each SMSA 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 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 to \$5,999 . . . . .		
\$6,000 to \$9,999 . . . . .		
\$10,000 to \$14,999 . . . . .		
\$15,000 and over . . . . .		

Thus, for this SMSA, the occupied housing unit records from the permit-issuing 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 selection procedure was then instituted that would produce one-half of the desired sample size. However, whenever a record was selected to be in sample, the housing unit record adjacent to it on the file 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 by a procedure that produced one-quarter of the desired sample size. However, at the time of the survey, the units at each of the special places were listed and subsampled at a rate which produced 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 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:

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

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. All units in existence at the time of interview in these selected segments are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

**1975-1978 additions to the housing inventory**—In the permit-issuing universe, a sample of new construction building permits, issued since the 1975 survey, was selected to represent housing units built in permit-issuing areas since the 1975 survey. Sampling procedures were identical to those used in selecting the 1970-1975 new construction sample, which were described previously. In the nonpermit universe, sample segments were dependently recanvassed, using listing sheets from 1975, to identify any units missed in the 1975 survey or any units added since the 1975 survey.

**Sample selection for the 1978 Coverage Improvement Program**—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 1970 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 for each of the AA-1 SMSA's except Newport News-Hampton, Va. Two different procedures were used. For the first procedure, the sampling was carried out in two stages for one- and two-unit structures and in three stages for three-or-more-unit structures. Sample units selected from one- and two-unit structures were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units), while sample units selected from three-or-more-unit structures were sampled at one-half the rate of 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. Since these permits were not available for Newport News-Hampton, Va., SMSA, these coverage improvement procedures were not instituted for this SMSA. For the third stage, structures of size three or more were divided into clusters of an expected size of two units and a sample of clusters was selected. This procedure was employed in the following SMSA's: Colorado Springs, Colo.; Rochester, N.Y.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In the remaining 11 SMSA's, units whose permits were issued before January 1970, but which were completed after April 1970, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. These units were then sampled at one-third the rate of regular AHS units. These procedures added an estimated 2,153 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 parks that were missed by the census or established after the census was selected in two stages. First, for each 1978-1979 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 2,072 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 now contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and have a utility hookup, or were on the site during the present survey 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 AHS sample units was selected from the permit-issuing universe at a rate of 1 in 24. 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 16,447 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 AHS units in multiunit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multiunit structures selected above, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 5,437 units to the coverage of the housing inventory of this SMSA.

**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 1978 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1975 (i.e., 1975-1978 lost units). Each type of estimate employed separate, although similar, estimation procedures.

**1978 housing inventory**—The AHS estimates of characteristics of the 1978 housing inventory employed a 1-stage ratio estimation procedure in all SMSA's. 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 noninterviews previously mentioned. This noninterview adjustment was done separately for occupied and vacant housing units. 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 described). In addition, within each sector separate noninterview factors were computed for one noninterview cell for conventional new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes from both the nonpermit universe and the coverage improvement universe, and one noninterview cell for other sample units from both the nonpermit universe and the coverage improvement universe, (if units were not included in any of the previous cells).

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.

**1975-1978 lost units**—The 1975-1978 lost unit estimates employed the three-stage ratio estimation procedure used to produce the AHS-SMSA estimates of the 1975 housing inventory, as was described in the 1975 Current Housing Report, Series H-170, *Housing Characteristics for Selected Metropolitan Areas*. Since the 1975-1978 lost units existed, by definition, in the 1975 housing inventory, there was a 1975 housing inventory weight associated with each 1975-1978 lost unit. This weight was used to tabulate the estimates of the characteristics of the 1975-1978 lost units.

**1975 estimation procedure**—This report presents data on the housing characteristics of the 1975 housing inventory from the 1975 Annual Housing Survey SMSA sample. The AHS-SMSA estimation procedure employed a three-stage ratio estimation process. A detailed description of this ratio estimation procedure can be found in the AHS Series H-170 reports for 1975.

**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 1, *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 the 1970 Census of Population and Housing and the 1975 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 the extent to which occupancy status was erroneously reported. 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 1978 AHS-SMSA sample, a reinterview program was not conducted. However, a study was conducted for the 1975 AHS-SMSA sample. The results of which are presented in the Census Bureau memorandum, "Reinterview Results for Annual Housing Survey—SMSA Sample: 1975."

**Coverage errors**—In 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. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units. 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, in which case, they would not have been 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 correct deficiencies 3-6 (see the coverage improvement section of this appendix) was not very effective in 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 1975 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing. It should be noted that since these ED's were recanvassed for the 1978 survey, the number of missed housing units may be considerably less for 1978.

**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 statistics 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 questionnaires, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The sampling error of a survey estimate provides a measure of the variation among the estimates from all 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.

One common measure of the sampling error is the standard error. As calculated for this report, the standard error reflects the variation in the estimates due to sampling and nonsampling errors, but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors measured by the standard error, biases, and any 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 item. Standard errors applicable to estimates of characteristics of the 1975 housing inventory can be found in the AHS Series H-170 reports for 1975.

Table I presents the standard errors applicable to estimates of characteristics of the 1978 housing inventory as well as estimates of characteristics of the 1975-1978 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 16,940 for the total SMSA, 7,440 for the central city of the SMSA, and 15,150 for the balance of the SMSA.

The reliability of an estimated percentage, computed by using the 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 1978 housing inventory as well as estimated percentages of the 1975-1978 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  
 $\sigma_x$  = the standard error of the numerator  
 $\sigma_y$  = 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 1978 there were 305,500 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 4,600. 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 .....	4,600
305,500 .....	x
400,000 .....	4,640

The entry for "x" is determined as follows by vertically interpolating between 4,600 and 4,640:

$$305,500 - 300,000 = 5,500$$

$$400,000 - 300,000 = 100,000$$

$$4,600 + \frac{5,500}{100,000} (4,640 - 4,600) = 4,600$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 300,900 to 310,100 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1978 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 298,140 to 312,860 housing units with 90 percent confidence; and that the average estimate lies within the interval from 290,800 to 320,220 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 305,500 owner-occupied housing units, 148,100, or 48.5 percent, had three bedrooms. Interpolation in table II of this appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 48.5 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	48.5	50
300,000 . . . . .	0.9	a	1.0
305,500 . . . . .		p	
400,000 . . . . .	0.8	b	0.9

1. The entry for cell "a" is determined by horizontal interpolation between 0.9 and 1.0.

$$48.5 - 25.0 = 23.5$$

$$50.0 - 25.0 = 25.0$$

$$0.9 + \frac{23.5}{25.0} (1.0 - 0.9) = 1.0$$

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

$$48.5 - 25.0 = 23.5$$

$$50.0 - 25.0 = 25.0$$

$$0.8 + \frac{23.5}{25.0} (0.9 - 0.8) = 0.9$$

3. The entry for "p" was then determined by vertical interpolation between 1.0 and 0.9.

$$305,500 - 300,000 = 5,500$$

$$400,000 - 300,000 = 100,000$$

$$1.0 + \frac{5,500}{100,000} (0.9 - 1.0) = 1.0$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 47.5 to 49.5 percent; the 90-percent confidence interval is from 46.9 to 50.1 percent; and the 95-percent confidence interval is from 46.5 to 50.5 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 characteristics 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. Due to the overlap of the 1975 and 1978 AHS-SMSA samples a positive correlation should be expected when making comparisons between the 1975 and 1978 characteristics.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1978 Housing Inventory and for Estimated Number of 1975-1978 Lost Units for the Cincinnati, Ohio-Ky.-Ind., SMSA, for the Central City, and for the Balance (Not in Central City) of the SMSA

Size of estimate	(68 chances out of 100)		
	Standard error <sup>1</sup>		
	SMSA	In central city	Not in central city
0 . . . . .	120	110	130
100 . . . . .	120	110	130
200 . . . . .	150	150	160
500 . . . . .	240	240	250
700 . . . . .	290	280	300
1,000 . . . . .	350	330	350
2,500 . . . . .	550	520	560
5,000 . . . . .	770	740	790
10,000 . . . . .	1,090	1,030	1,110
25,000 . . . . .	1,700	1,590	1,720
50,000 . . . . .	2,360	2,130	2,370
75,000 . . . . .	2,840	2,450	2,810
100,000 . . . . .	3,220	2,640	3,140
150,000 . . . . .	3,780	2,720	3,580
200,000 . . . . .	4,170	2,400	3,800
250,000 . . . . .	4,440	—	3,840
300,000 . . . . .	4,600	—	3,700
400,000 . . . . .	4,640	—	2,750
500,000 . . . . .	4,320	—	—
600,000 . . . . .	3,540	—	—

<sup>1</sup>For estimates pertaining to new construction, the standard errors shown in the table should be multiplied by a factor of 1.1 for the total SMSA, 1.1 for the central city, and 1.1 for the balance (not in central city) estimates.

*Illustration of the computation of the standard error of a difference*—Table A-1 of part A of this report shows that in 1978 there were 76,300 owner-occupied housing units with two bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied housing units with two bedrooms and owner-occupied housing units with three bedrooms is 71,800. Table I shows the standard error of 76,300 is approximately 2,860, and the standard error of 148,100 is approximately 3,760. Therefore, the standard error of the estimated difference of 71,800 is about:

$$4,720 = \sqrt{(2,860)^2 + (3,760)^2}$$

Consequently, the 68-percent confidence interval for the 71,800 difference is from 67,080 to 76,520 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 64,248 to 79,352 housing units, and the 95-percent confidence interval is from 62,360 to 81,240. Thus, we can conclude with 95 percent confidence that the number of 1978 owner-occupied housing units with three bedrooms is greater than the number of owner-occupied housing 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

**TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1978 Housing Inventory and for Estimated Percentages of 1975-1978 Lost Units for the Cincinnati, Ohio-Ky.-Ind., SMSA, for the Central City, and for the Balance (Not in Central City) of the SMSA**

(68 chances out of 100)

Base of percentage	Estimated percentage <sup>1</sup>					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
200.....	35.5	36.5	38.5	38.5	38.5	39.6
500.....	20.0	20.0	20.0	20.0	21.7	25.0
700.....	15.2	15.2	15.2	15.2	18.3	21.2
1,000.....	11.1	11.1	11.1	11.1	15.3	17.7
2,500.....	4.8	4.8	4.9	6.7	9.7	11.2
5,000.....	2.4	2.4	3.5	4.7	6.9	7.9
10,000.....	1.2	1.2	2.4	3.4	4.8	5.6
25,000.....	0.5	0.7	1.5	2.1	3.1	3.5
50,000.....	0.2	0.5	1.1	1.5	2.2	2.5
75,000.....	0.2	0.4	0.9	1.2	1.8	2.0
100,000.....	0.13	0.4	0.8	1.1	1.5	1.6
150,000.....	0.08	0.3	0.6	0.9	1.3	1.4
200,000.....	0.06	0.2	0.5	0.8	1.1	1.3
250,000.....	0.05	0.2	0.5	0.7	1.0	1.1
300,000.....	0.04	0.2	0.4	0.6	0.9	1.0
400,000.....	0.03	0.2	0.4	0.5	0.8	0.9
500,000.....	0.03	0.2	0.3	0.5	0.7	0.8
600,000.....	0.02	0.14	0.3	0.4	0.6	0.7

<sup>1</sup> 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 errors shown in the table should be multiplied by a factor of 1.1.

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 characteristics, 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 305,500 housing units.

1. Interpolation in table II shows that the standard error of 50 percent of a base of 305,500 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 125,100 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 57,000 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:

$$2.5 + (3.5 - 2.5) \left( \frac{48.0 - 41.0}{18.7} \right) = 2.9$$

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

$$2.5 + (3.5 - 2.5) \left( \frac{52.0 - 41.0}{18.7} \right) = 3.1$$

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

Columbus 1978

# Appendix B

## Source and Reliability of the Estimates

<b>SAMPLE DESIGN</b> .....	App-40	<b>Coverage improvement for deficiencies 3-6</b> .....	App-42	<b>AHS-SMSA</b> .....	App-44
Annual Housing Survey .....	App-40	1970 Census of Population and Housing .....	App-43	Coverage errors .....	App-44
Designation of sample housing units for the 1978 survey .....	App-40	<b>ESTIMATION</b> .....	App-43	Rounding errors .....	App-44
Selection of the 1975 AHS-SMSA sample .....	App-40	1978 housing inventory .....	App-43	Sampling errors for the AHS-SMSA sample .....	App-44
1975-1978 additions to the housing inventory .....	App-42	1975-1978 lost units .....	App-43	Illustration of the use of the standard error tables .....	App-45
Sample selection for the 1978 Coverage Improvement Program. Coverage improvement for deficiency 1 .....	App-42	1975 estimation procedure .....	App-43	Differences .....	App-46
Coverage improvement for deficiency 2 .....	App-42	Ratio estimation procedure of the 1970 Census of Population and Housing .....	App-43	Illustration of the computation of the standard error of a difference .....	App-46
		<b>RELIABILITY OF THE ESTIMATES</b> .....	App-44	Medians .....	App-47
		Nonsampling errors .....	App-44	Illustration of the computation of the 95-percent confidence interval of a median .....	App-47
		1970 census .....	App-44		

### SAMPLE DESIGN

**Annual Housing Survey**—The estimates for each of the 15 SMSA's are based on data collected from the 1978 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 15 SMSA's, the data were collected for the 12-month period from April 1978 through March 1979 with one-twelfth of the sample units being visited each month.

Each group of SMSA's is being interviewed for the AHS on a rotating basis with this group (AA-1 SMSA's) being the second to be revisited. All of these SMSA's were enumerated for the first time in 1975.

For each group of SMSA's, the largest SMSA from 3 of the 4 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 the remaining 12 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 AA-1 group (1978-1979) are: Atlanta, Ga.; Philadelphia, Pa.-N.J.; and San Francisco-Oakland, Calif.

The remaining SMSA's in the AA-1 group are: Cincinnati, Ohio-Ky.-Ind.; Colorado Springs, Colo.; Columbus, Ohio; Kansas City, Mo.-Kans.; New Orleans, La.; Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; Rochester, N.Y.; San Antonio, Tex.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In this SMSA, 4,827 housing units were eligible for interview. Of these sample units, 268 interviews were not obtained because, for occupied sample units, the occupants refused to be interviewed, 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 units eligible for interview, 381 units were visited but were not eligible for interview because they were condemned, unfit, demolished, converted to group quarters use, etc.

**Designation of sample housing units for the 1978 survey**—The sample housing units designated to be interviewed in the 1978 survey consisted of the following categories which are described in detail in the succeeding sections.

1. All sample housing units that were interviewed in the 1975 survey.
2. All sample housing units that were type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of the survey but which could become eligible in the future) in the 1975 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1978 AHS questionnaire, page App-20.)
3. All sample housing units that were selected from the list of new construction building permits issued since the 1975 survey. (This sample represented the housing units built in permit-issuing areas, since the 1975 survey.)
4. All sample housing units that were added to sample segments in the nonpermit universe since the 1975 survey. (This sample represented additions to the housing inventory in nonpermit-issuing areas since the 1975 survey.)
5. All sample housing units that were selected as part of the 1978 Coverage Improvement Program. (This sample represented most of the housing units which, until 1978, did not have a chance of selection.)

**Selection of the 1975 AHS-SMSA sample**—The sample for the SMSA's which are 100-percent permit-issuing 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 those 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 non-permit universe). The following SMSA's are 100-percent permit-issuing: Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and San Francisco-Oakland, Calif. The remaining 10 SMSA's contain a sample from 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 three largest SMSA's, the overall sampling rate differed by central city and the 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 the 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 the total housing units in each sector.

The major portion of the sample in each SMSA 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 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 to \$5,999 . .		
\$6,000 to \$9,999 . .		
\$10,000 to \$14,999 .		
\$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from the permit-issuing 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 selection procedure was then instituted that would produce one-half of the desired sample size. However, whenever a record was selected to be in sample, the housing unit record adjacent to it on the file 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 by a procedure that produced one-quarter of the desired sample size. However, at the time of the survey, the units at each of the special places were listed and subsampled at a rate which produced 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 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:

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

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. All units in existence at the time of interview in these selected segments are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

**1975-1978 additions to the housing inventory**—In the permit-issuing universe, a sample of new construction building permits, issued since the 1975 survey, was selected to represent housing units built in permit-issuing areas since the 1975 survey. Sampling procedures were identical to those used in selecting the 1970-1975 new construction sample, which were described previously. In the nonpermit universe, sample segments were dependently recanvassed, using listing sheets from 1975, to identify any units missed in the 1975 survey or any units added since the 1975 survey.

**Sample selection for the 1978 Coverage Improvement Program**—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 1970 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 for each of the AA-1 SMSA's except Newport News-Hampton, Va. Two different procedures were used. For the first procedure, the sampling was carried out in two stages for one- and two-unit structures and in three stages for three-or-more-unit structures. Sample units selected from one- and two-unit structures were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units), while sample units selected from three-or-more-unit structures were sampled at one-half the rate of 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. Since these permits were not available for Newport News-Hampton, Va., SMSA, these coverage improvement procedures were not instituted for this SMSA. For the third stage, structures of size three or more were divided into clusters of an expected size of two units and a sample of clusters was selected. This procedure was employed in the following SMSA's: Colorado Springs, Colo.; Rochester, N.Y.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In the remaining 11 SMSA's, units whose permits were issued before January 1970, but which were completed after April 1970, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. These units were then sampled at one-third the rate of regular AHS units. These procedures added an estimated

3,070 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 parks that were missed by the census or established after the census was selected in two stages. First, for each 1978-1979 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 878 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 now contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and have a utility hookup, or were on the site during the present survey 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 AHS sample units was selected from the permit-issuing universe at a rate of 1 in 24. 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 9,321 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 AHS units in multiunit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multiunit structures selected above, all units were listed and matched to the 1970 census. Any missed units

were then assigned for interview. This procedure added an estimated 210 units to the coverage of the housing inventory of this SMSA.

**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 1978 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1975 (i.e., 1975-1978 lost units). Each type of estimate employed separate, although similar, estimation procedures.

**1978 housing inventory**—The AHS estimates of characteristics of the 1978 housing inventory employed a 1-stage ratio estimation procedure in all SMSA's. 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 noninterviews previously mentioned. This noninterview adjustment was done separately for occupied and vacant housing units. 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 described). In addition, within each sector separate noninterview factors were computed for one noninterview cell for conventional new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes from both the nonpermit universe and the coverage improvement universe, and one noninterview cell for other sample units from both the nonpermit universe and the coverage improvement universe, (if units were not included in any of the previous cells).

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.

**1975-1978 lost units**—The 1975-1978 lost unit estimates employed the three-stage ratio estimation procedure used to produce the AHS-SMSA estimates of the 1975 housing inventory, as was described in the 1975 Current Housing Report, Series H-170, **Housing Characteristics for Selected Metropolitan Areas**. Since the 1975-1978 lost units existed, by definition, in the 1975 housing inventory, there was a 1975 housing inventory weight associated with each 1975-1978 lost unit. This weight was used to tabulate the estimates of the characteristics of the 1975-1978 lost units.

**1975 estimation procedure**—This report presents data on the housing characteristics of the 1975 housing inventory from the 1975 Annual Housing Survey SMSA sample. The AHS-SMSA estimation procedure employed a three-stage ratio estimation process. A detailed description of this ratio estimation procedure can be found in the AHS Series H-170 reports for 1975.

**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 1, *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 the 1970 Census of Population and Housing and the 1975 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 the extent to which occupancy status was erroneously reported. 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 1978 AHS-SMSA sample, a reinterview program was not conducted. However, a study was conducted for the 1975 AHS-SMSA sample. The results of which are presented in the Census Bureau memorandum, "Reinterview Results for Annual Housing Survey—SMSA Sample: 1975."

**Coverage errors**—In 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. However, these permits issued during the last 5

months of the survey do not necessarily represent missed housing units. 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, in which case, they would not have been 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 correct deficiencies 3-6 (see the coverage improvement section of this appendix) was not very effective in 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 1975 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing. It should be noted that since these ED's were recanvassed for the 1978 survey, the number of missed housing units may be considerably less for 1978.

**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 statistics 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 questionnaires, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The sampling error of a survey estimate provides a measure of the variation among the estimates from all 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.

One common measure of the sampling error is the standard error. As calculated for this report, the standard error reflects the variation in the estimates due to sampling and nonsampling errors, but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors measured by the standard error, biases, and any 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 item. Standard errors applicable to estimates of characteristics of the 1975 housing inventory can be found in the AHS Series H-170 reports for 1975.

Table I presents the standard errors applicable to estimates of characteristics of the 1978 housing inventory as well as estimates of characteristics of the 1975-1978 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 13,610 for the total SMSA, 7,720 for the central city of the SMSA, and 11,130 for the balance of the SMSA.

The reliability of an estimated percentage, computed by using the 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 1978 housing inventory as well as estimated percentages of the 1975-1978 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  
 $\sigma_x$  = the standard error of the numerator  
 $\sigma_y$  = 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 1978 there were 198,800 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 3,480. 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 .....	3,130
198,800 .....	
200,000 .....	3,490

The entry for "x" is determined as follows by vertically interpolating between 3,130 and 3,490:

$$\begin{aligned} 198,800 - 150,000 &= 48,800 \\ 200,000 - 150,000 &= 50,000 \\ 3,130 + \frac{48,800}{50,000} (3,490 - 3,130) &= 3,480 \end{aligned}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 195,320 to 202,280 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1978 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 193,232 to 204,368 housing units with 90 percent confidence; and that the average estimate lies within the interval from 191,840 to 205,760 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 198,800 owner-occupied housing units, 39,800, or 20.0 percent, had two

bedrooms. Interpolation in table II of this appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 20.0 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	20.0	25 or 75
150,000 . . . . .	0.7	a	1.0
198,800 . . . . .		p	
200,000 . . . . .	0.6	b	0.9

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

$$20.0 - 10.0 = 10.0$$

$$25.0 - 10.0 = 15.0$$

$$0.7 + \frac{10.0}{15.0} (1.0 - 0.7) = 0.9$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.6 and 0.9

$$20.0 - 10.0 = 10.0$$

$$25.0 - 10.0 = 15.0$$

$$0.6 + \frac{10.0}{15.0} (0.9 - 0.6) = 0.8$$

3. The entry for "p" was then determined by vertical interpolation between 0.9 and 0.8.

$$198,800 - 150,000 = 48,800$$

$$200,000 - 150,000 = 50,000$$

$$0.9 + \frac{48,800}{50,000} (0.8 - 0.9) = 0.8$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 19.2 to 20.8 percent; the 90-percent confidence interval is from 18.7 to 21.3 percent; and the 95-percent confidence interval is from 18.4 to 21.6 percent.

*Differences*—The standard errors shown are not directly applicable to differences between two sample estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard error of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristics 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. Due to the overlap of the 1975 and 1978 AHS-SMSA samples a positive correlation should be expected when making comparisons between the 1975 and 1978 characteristics.

*Illustration of the computation of the standard error of a difference*—Table A-1 of part A of this report shows that in 1978 there were 110,700 owner-occupied housing units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied housing units with two bedrooms and owner-occupied housing units with three bedrooms is 70,900. Table I shows the standard error of 110,700 is approximately 2,750, and the standard error of 39,800 is approximately 1,710. Therefore, the standard error of the estimated difference of 70,900 is about:

$$3,240 = \sqrt{(2,750)^2 + (1,710)^2}$$

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1978 Housing Inventory and for Estimated Number of 1975-1978 Lost Units for the Columbus, Ohio, SMSA, for the Central City, and for the Balance (Not in Central City) of the SMSA

(68 chances out of 100)

Size of estimate	Standard error <sup>1</sup>		
	SMSA	In central city	Not in central city
0 . . . . .	80	80	80
100 . . . . .	90	90	90
200 . . . . .	130	130	130
500 . . . . .	200	200	200
700 . . . . .	240	230	240
1,000 . . . . .	280	280	290
2,500 . . . . .	440	440	460
5,000 . . . . .	630	620	650
10,000 . . . . .	890	870	910
25,000 . . . . .	1,390	1,350	1,430
50,000 . . . . .	1,930	1,840	1,990
75,000 . . . . .	2,330	2,160	2,400
100,000 . . . . .	2,650	2,380	2,730
150,000 . . . . .	3,130	2,610	3,240
200,000 . . . . .	3,490	2,630	—
250,000 . . . . .	3,750	2,420	—
300,000 . . . . .	3,930	—	—
400,000 . . . . .	4,110	—	—

<sup>1</sup> For estimates pertaining to new construction, the standard errors shown in the table should be multiplied by a factor of 1.2 for the total SMSA, 1.3 for the central city, and 1.1 for the balance (not in central city) estimates.

**TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1978 Housing Inventory and for Estimated Percentages of 1975-1978 Lost Units for the Columbus, Ohio, SMSA, for the Central City, and for the Balance (Not in Central City) of the SMSA**

(68 chances out of 100)

Base of percentage	Estimated percentage <sup>1</sup>					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100. . . . .	45.6	45.6	45.6	45.6	45.6	45.8
200. . . . .	29.6	29.6	29.6	29.6	29.6	32.4
500. . . . .	14.4	14.4	14.4	14.4	17.7	20.5
700. . . . .	10.7	10.7	10.7	10.7	15.0	17.3
1,000 . . . . .	7.7	7.7	7.7	8.7	12.5	14.5
2,500 . . . . .	3.2	3.2	4.0	5.5	7.9	9.2
5,000 . . . . .	1.7	1.7	2.8	3.9	5.6	6.5
10,000 . . . . .	0.8	0.9	2.0	2.7	4.0	4.6
25,000 . . . . .	0.3	0.6	1.3	1.7	2.5	2.9
50,000 . . . . .	0.2	0.4	0.9	1.2	1.8	2.0
75,000 . . . . .	0.11	0.3	0.7	1.0	1.4	1.7
100,000. . . . .	0.08	0.3	0.6	0.9	1.3	1.4
150,000. . . . .	0.06	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
400,000. . . . .	0.02	0.14	0.3	0.4	0.6	0.7

<sup>1</sup> 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 errors shown in the table should be multiplied by a factor of 1.2 for the total SMSA, 1.3 for the central city, and 1.1 for the balance (not in central city).

Consequently, the 68-percent confidence interval for the 70,900 difference is from 67,660 to 74,140 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 65,716 to 76,084 housing units, and the 95-percent confidence interval is from 64,420 to 77,380. Thus, we can conclude with 95 percent confidence that the number of 1978 owner-occupied housing units with three bedrooms is greater than the number of owner-occupied housing 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 deter-

mine 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 characteristics, 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 198,800 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 198,800 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-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 91,300 owner-occupied housing units or 45.9 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 38,300 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:

$$2.5 + (3.5 - 2.5) \left( \frac{48.0 - 45.9}{19.3} \right) = 2.6$$

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

$$2.5 + (3.5 - 2.5) \left( \frac{52.0 - 45.9}{19.3} \right) = 2.8$$

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

# Kansas City 1978

# Appendix B

## Source and Reliability of the Estimates

<b>SAMPLE DESIGN</b> . . . . .	App-40	<b>Coverage improvement for deficiencies 3-6</b> . . . . .	App-42	<b>AHS-SMSA</b> . . . . .	App-44
<b>Annual Housing Survey</b> . . . . .	App-40	<b>1970 Census of Population and Housing</b> . . . . .	App-43	<b>Coverage errors</b> . . . . .	App-44
<b>Designation of sample housing units for the 1978 survey</b> . . . . .	App-40	<b>ESTIMATION</b> . . . . .	App-43	<b>Rounding errors</b> . . . . .	App-44
<b>Selection of the 1975 AHS-SMSA sample</b> . . . . .	App-40	<b>1978 housing inventory</b> . . . . .	App-43	<b>Sampling errors for the AHS-SMSA sample</b> . . . . .	App-44
<b>1975-1978 additions to the housing inventory</b> . . . . .	App-42	<b>1975-1978 lost units</b> . . . . .	App-43	<b>Illustration of the use of the standard error tables</b> . . . . .	App-45
<b>Sample selection for the 1978 Coverage Improvement Program.</b> . . . . .	App-42	<b>1975 estimation procedure</b> . . . . .	App-43	<b>Differences</b> . . . . .	App-46
<b>Coverage improvement for deficiency 1</b> . . . . .	App-42	<b>Ratio estimation procedure of the 1970 Census of Population and Housing</b> . . . . .	App-43	<b>Illustration of the computation of the standard error of a difference</b> . . . . .	App-46
<b>Coverage improvement for deficiency 2</b> . . . . .	App-42	<b>RELIABILITY OF THE ESTIMATES</b> . . . . .	App-43	<b>Medians</b> . . . . .	App-47
		<b>Nonsampling errors</b> . . . . .	App-44	<b>Illustration of the computation of the 95-percent confidence interval of a median</b> . . . . .	App-47
		<b>1970 census</b> . . . . .	App-44		

### SAMPLE DESIGN

**Annual Housing Survey**—The estimates for each of the 15 SMSA's are based on data collected from the 1978 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 15 SMSA's, the data were collected for the 12-month period from April 1978 through March 1979 with one-twelfth of the sample units being visited each month.

Each group of SMSA's is being interviewed for the AHS on a rotating basis with this group (AA-1 SMSA's) being the second to be revisited. All of these SMSA's were enumerated for the first time in 1975.

For each group of SMSA's, the largest SMSA from 3 of the 4 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 the remaining 12 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 AA-1 group (1978-1979) are Atlanta, Ga.; Philadelphia, Pa.-N.J.; and San Francisco-Oakland, Calif.

The remaining SMSA's in the AA-1 group are Cincinnati, Ohio-Ky.-Ind.; Colorado Springs, Colo.; Columbus, Ohio; Kansas City, Mo.-Kans.; New Orleans, La.; Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; Rochester, N.Y.; San Antonio, Tex.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In this SMSA, 4,971 housing units were eligible for interview. Of these sample units, 244 interviews were not obtained because, for occupied sample units, the occupants refused to be interviewed, 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 units eligible for interview, 403 units were visited but were not eligible for interview because they were condemned, unfit, demolished, converted to group quarters use, etc.

**Designation of sample housing units for the 1978 survey**—The sample housing units designated to be interviewed in the 1978 survey consisted of the following categories which are described in detail in the succeeding sections.

1. All sample housing units that were interviewed in the 1975 survey.
2. All sample housing units that were type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of the survey but which could become eligible in the future) in the 1975 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1978 AHS questionnaire, page App-20.)
3. All sample housing units that were selected from the list of new construction building permits issued since the 1975 survey. (This sample represented the housing units built in permit-issuing areas, since the 1975 survey.)
4. All sample housing units that were added to sample segments in the nonpermit universe since the 1975 survey. (This sample represented additions to the housing inventory in nonpermit-issuing areas since the 1975 survey.)
5. All sample housing units that were selected as part of the 1978 Coverage Improvement Program. (This sample represented most of the housing units which, until 1978, did not have a chance of selection.)

**Selection of the 1975 AHS-SMSA sample**—The sample for the SMSA's which are 100-percent permit-issuing 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 those 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 non-permit universe). The following SMSA's are 100-percent permit-issuing: Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and San Francisco-Oakland, Calif. The remaining 10 SMSA's contain a sample from 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 three largest SMSA's, the overall sampling rate differed by central city and the 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 the 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 the total housing units in each sector.

The major portion of the sample in each SMSA 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 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 to \$5,999 . .		
\$6,000 to \$9,999 . .		
\$10,000 to \$14,999 .		
\$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from the permit-issuing 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 selection procedure was then instituted that would produce one-half of the desired sample size. However, whenever a record was selected to be in sample, the housing unit record adjacent to it on the file 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 by a procedure that produced one-quarter of the desired sample size. However, at the time of the survey, the units at each of the special places were listed and subsampled at a rate which produced 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 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:

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

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. All units in existence at the time of interview in these selected segments are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

**1975-1978 additions to the housing inventory**—In the permit-issuing universe, a sample of new construction building permits, issued since the 1975 survey, was selected to represent housing units built in permit-issuing areas since the 1975 survey. Sampling procedures were identical to those used in selecting the 1970-1975 new construction sample, which were described previously. In the nonpermit universe, sample segments were dependently recanvassed, using listing sheets from 1975, to identify any units missed in the 1975 survey or any units added since the 1975 survey.

**Sample selection for the 1978 Coverage Improvement Program**—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 1970 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 for each of the AA-1 SMSA's except Newport News-Hampton, Va. Two different procedures were used. For the first procedure, the sampling was carried out in two stages for one- and two-unit structures and in three stages for three-or-more-unit structures. Sample units selected from one- and two-unit structures were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units), while sample units selected from three-or-more-unit structures were sampled at one-half the rate of 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. Since these permits were not available for Newport News-Hampton, Va., SMSA, these coverage improvement procedures were not instituted for this SMSA. For the third stage, structures of size three or more were divided into clusters of an expected size of two units and a sample of clusters was selected. This procedure was employed in the following SMSA's: Colorado Springs, Colo.; Rochester, N.Y.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In the remaining 11 SMSA's, units whose permits were issued before January 1970, but which were completed after April 1970, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. These units were then sampled at one-third the rate of regular AHS units. These procedures added an estimated 4,868 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 parks that were missed by the census or established after the census was selected in two stages. First, for each 1978-1979 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 1,945 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 now contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and have a utility hookup, or were on the site during the present survey 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 AHS sample units was selected from the permit-issuing universe at a rate of 1 in 24. 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 8,065 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 AHS units in multiunit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multiunit structures selected above, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 1,165 units to the coverage of the housing inventory of this SMSA.

**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 1978 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1975 (i.e., 1975-1978 lost units). Each type of estimate employed separate, although similar, estimation procedures.

**1978 housing inventory**—The AHS estimates of characteristics of the 1978 housing inventory employed a 1-stage ratio estimation procedure in all SMSA's. 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 noninterviews previously mentioned. This noninterview adjustment was done separately for occupied and vacant housing units. 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 described). In addition, within each sector separate noninterview factors were computed for one noninterview cell for conventional new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes from both the nonpermit universe and the coverage improvement universe, and one noninterview cell for other sample units from both the nonpermit universe and the coverage improvement universe (if units were not included in any of the previous cells).

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.

**1975-1978 lost units**—The 1975-1978 lost unit estimates employed the three-stage ratio estimation procedure used to produce the AHS-SMSA estimates of the 1975 housing inventory, as was described in the 1975 Current Housing Report, Series H-170, *Housing Characteristics for Selected Metropolitan Areas*. Since the 1975-1978 lost units existed, by definition, in the 1975 housing inventory, there was a 1975 housing inventory weight associated with each 1975-1978 lost unit. This weight was used to tabulate the estimates of the characteristics of the 1975-1978 lost units.

**1975 estimation procedure**—This report presents data on the housing characteristics of the 1975 housing inventory from the 1975 Annual Housing Survey SMSA sample. The AHS-SMSA estimation procedure employed a three-stage ratio estimation process. A detailed description of this ratio estimation procedure can be found in the AHS Series H-170 reports for 1975.

**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 the 1970 Census of Population and Housing and the 1975 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 the extent to which occupancy status was erroneously reported. 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 1978 AHS-SMSA sample, a reinterview program was not conducted. However, a study was conducted for the 1975 AHS-SMSA sample. The results of which are presented in the Census Bureau memorandum, "Reinterview Results for Annual Housing Survey—SMSA Sample: 1975."

**Coverage errors**—In 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. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units. 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, in which case, they would not have been 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 correct deficiencies 3-6 (see the coverage improvement section of this appendix) was not very effective in 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 1975 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing. It should be noted that since these ED's were recanvassed for the 1978 survey, the number of missed housing units may be considerably less for 1978.

**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 statistics 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 questionnaires, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The sampling error of a survey estimate provides a measure of the variation among the estimates from all 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.

One common measure of the sampling error is the standard error. As calculated for this report, the standard error reflects the variation in the estimates due to sampling and nonsampling errors, but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors measured by the standard error, biases, and any 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 item. Standard errors applicable to estimates of characteristics of the 1975 housing inventory can be found in the AHS Series H-170 reports for 1975.

Table I presents the standard errors applicable to estimates of characteristics of the 1978 housing inventory as well as estimates of characteristics of the 1975-1978 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 17,110 for the total SMSA, 10,800 for the central city of the SMSA, and 13,200 for the balance of the SMSA.

The reliability of an estimated percentage, computed by using the 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 1978 housing inventory as well as estimated percentages of the 1975-1978 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  
 $\sigma_x$  = the standard error of the numerator  
 $\sigma_y$  = 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 1978 there were 303,800 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 5,040. 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 .....	5,030
303,800 .....	x
400,000 .....	5,390

The entry for "x" is determined as follows by vertically interpolating between 5,030 and 5,390:

$$303,800 - 300,000 = 3,800$$

$$400,000 - 300,000 = 100,000$$

$$5,030 + \frac{3,800}{100,000} (5,390 - 5,030) = 5,040$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 298,760 to 308,840 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1978 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 295,736 to 311,864 housing units with 90 percent confidence; and that the average estimate lies within the interval from 293,720 to 313,880 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 303,800 owner-occupied housing units, 157,800, or 51.9 percent, had three bedrooms. Interpolation in table II of this appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 51.9 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		
	50	51.9	25 or 75
300,000 .....	1.0	a	0.9
303,800 .....		p	
400,000 .....	0.9	b	0.7

1. The entry for cell "a" is determined by horizontal interpolation between 1.0 and 0.9.

$$51.9 - 50.0 = 1.9$$

$$75.0 - 50.0 = 25.0$$

$$1.0 + \frac{1.9}{25.0} (0.9 - 1.0) = 1.0$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.9 and 0.7.

$$51.9 - 50.0 = 1.9$$

$$75.0 - 50.0 = 25.0$$

$$0.9 + \frac{1.9}{25.0} (0.7 - 0.9) = 0.9$$

3. The entry for "p" was then determined by vertical interpolation between 1.0 and 0.9.

$$303,800 - 300,000 = 3,800$$

$$400,000 - 300,000 = 100,000$$

$$1.0 + \frac{3,800}{100,000} (0.9 - 1.0) = 1.0$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 50.9 to 52.9 percent; the 90-percent confidence interval is from 50.3 to 53.5 percent; and the 95-percent confidence interval is from 49.9 to 53.9 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 characteristics 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. Due to the overlap of the 1975 and 1978 AHS-SMSA samples a positive correlation should be expected when making comparisons between the 1975 and 1978 characteristics.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1978 Housing Inventory and for Estimated Number of 1975-1978 Lost Units for the Kansas City, Mo.-Kans., SMSA, for the Central City, and for the Balance (Not in Central City) of the SMSA

(68 chances out of 100)

Size of estimate	Standard error <sup>1</sup>		
	SMSA	In central city	Not in central city
0 .....	120	120	110
100 .....	120	120	110
200 .....	150	150	150
500 .....	240	240	240
700 .....	290	280	280
1,000 .....	350	340	340
2,500 .....	550	540	530
5,000 .....	770	760	750
10,000 .....	1,090	1,070	1,060
25,000 .....	1,710	1,670	1,650
50,000 .....	2,390	2,310	2,280
75,000 .....	2,890	2,780	2,720
100,000 .....	3,290	3,150	3,050
150,000 .....	3,910	3,700	3,520
200,000 .....	4,390	4,080	3,810
250,000 .....	4,750	—	3,940
300,000 .....	5,030	—	3,940
400,000 .....	5,390	—	3,520
500,000 .....	5,510	—	—

<sup>1</sup> For estimates pertaining to new construction, the standard errors shown in the table should be multiplied by a factor of 1.3 for the total SMSA, 1.1 for the central city, and 1.2 for the balance (not in central city) estimates.

*Illustration of the computation of the standard error of a difference*—Table A-1 of part A of this report shows that in 1978 there were 78,300 owner-occupied housing units with two bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied housing units with two bedrooms and owner-occupied housing units with three bedrooms is 79,500. Table I shows the standard error of 78,300 is approximately 2,940, and the standard error of 157,800 is approximately 3,980. Therefore, the standard error of the estimated difference of 79,500 is about:

$$4,950 = \sqrt{(2,940)^2 + (3,980)^2}$$

Consequently, the 68-percent confidence interval for the 79,500 difference is from 74,550 to 84,450 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 71,580 to 87,420 housing units, and the 95-percent confidence interval is from 69,600 to 89,400. Thus, we can conclude with 95 percent confidence that the number of 1978 owner-occupied housing units with three bedrooms is greater than the number of owner-occupied housing 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 characteristics, 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.

TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1978 Housing Inventory and for Estimated Percentages of 1975-1978 Lost Units for the Kansas City, Mo.-Kans., SMSA, for the Central City, and for the Balance (Not in Central City) of the SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage <sup>1</sup>					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100 . . . . .	54.5	54.5	54.5	54.5	54.5	54.7
200 . . . . .	37.5	37.5	37.5	37.5	37.5	38.7
500 . . . . .	19.3	19.3	19.3	19.3	21.2	24.5
700 . . . . .	14.6	14.6	14.6	14.6	17.9	20.7
1,000 . . . . .	10.7	10.7	10.7	10.7	15.0	17.3
2,500 . . . . .	4.6	4.6	4.8	6.6	9.5	10.9
5,000 . . . . .	2.3	2.3	3.4	4.6	6.7	7.7
10,000 . . . . .	1.2	1.2	2.4	3.3	4.7	5.5
25,000 . . . . .	0.5	0.7	1.5	2.1	3.0	3.5
50,000 . . . . .	0.2	0.5	1.1	1.5	2.1	2.4
75,000 . . . . .	0.2	0.4	0.9	1.2	1.7	2.0
100,000 . . . . .	0.12	0.3	0.8	1.0	1.5	1.7
150,000 . . . . .	0.08	0.3	0.6	0.8	1.2	1.4
200,000 . . . . .	0.06	0.2	0.5	0.7	1.1	1.2
250,000 . . . . .	0.05	0.2	0.5	0.7	0.9	1.1
300,000 . . . . .	0.04	0.2	0.4	0.6	0.9	1.0
400,000 . . . . .	0.03	0.2	0.4	0.5	0.7	0.9
500,000 . . . . .	0.02	0.15	0.3	0.5	0.7	0.8

<sup>1</sup> 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 errors shown in the table should be multiplied by a factor of 1.3 for the total SMSA, 1.1 for the central city, and 1.2 for the balance (not in central city).

*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 2.6. The base of the distribution from which this median was determined is 303,800 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 303,800 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-1 of part A, it can be seen that 44,600 owner-occupied housing units, or 14.7 percent, had one person (for purposes of calculating the median, the category of one person is considered to be from 0.5 to 1.5 persons); that an additional 101,700 owner-occupied housing units, or 33.5 percent, had two persons (i.e., 1.5 to 2.5 persons); and that 56,100, 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:

$$2.5 + (3.5 - 2.5) \left( \frac{48.0 - 48.2}{18.5} \right) = 2.5$$

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

$$2.5 + (3.5 - 2.5) \left( \frac{52.0 - 48.2}{18.5} \right) = 2.7$$

Thus, the 95-percent confidence interval ranges from 2.5 to 2.7 persons.

New Orleans 1978

# Appendix B

## Source and Reliability of the Estimates

<b>SAMPLE DESIGN</b> . . . . .	App-40	<b>Coverage improvement for deficiencies 3-6</b> . . . . .	App-42	<b>AHS-SMSA</b> . . . . .	App-44
Annual Housing Survey . . . . .	App-40	1970 Census of Population and Housing . . . . .	App-43	Coverage errors . . . . .	App-44
Designation of sample housing units for the 1978 survey . . . . .	App-40	<b>ESTIMATION</b> . . . . .	App-43	Rounding errors . . . . .	App-44
Selection of the 1975 AHS-SMSA sample . . . . .	App-40	1978 housing inventory . . . . .	App-43	<b>Sampling errors for the AHS-SMSA sample</b> . . . . .	App-44
1975-1978 additions to the housing inventory . . . . .	App-42	1975-1978 lost units . . . . .	App-43	Illustration of the use of the standard error tables . . . . .	App-45
Sample selection for the 1978 Coverage Improvement Program . . . . .	App-42	1975 estimation procedure . . . . .	App-43	Differences . . . . .	App-47
Coverage improvement for deficiency 1 . . . . .	App-42	Ratio estimation procedure of the 1970 Census of Population and Housing . . . . .	App-43	Illustration of the computation of the standard error of a difference . . . . .	App-47
Coverage improvement for deficiency 2 . . . . .	App-42	<b>RELIABILITY OF THE ESTIMATES</b> . . . . .	App-44	Medians . . . . .	App-47
		Nonsampling errors . . . . .	App-44	Illustration of the computation of the 95-percent confidence interval of a median . . . . .	App-47
		1970 census . . . . .	App-44		

### SAMPLE DESIGN

**Annual Housing Survey**—The estimates for each of the 15 SMSA's are based on data collected from the 1978 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 15 SMSA's, the data were collected for the 12-month period from April 1978 through March 1979 with one-twelfth of the sample units being visited each month.

Each group of SMSA's is being interviewed for the AHS on a rotating basis with this group (AA-1 SMSA's) being the second to be revisited. All of these SMSA's were enumerated for the first time in 1975.

For each group of SMSA's, the largest SMSA from 3 of the 4 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 the remaining 12 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 AA-1 group (1978-1979) are: Atlanta, Ga.; Philadelphia, Pa.-N.J.; and San Francisco-Oakland, Calif.

The remaining SMSA's in the AA-1 group are: Cincinnati, Ohio-Ky.-Ind.; Colorado Springs, Colo.; Columbus, Ohio; Kansas City, Mo.-Kans.; New Orleans, La.; Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; Rochester, N.Y.; San Antonio, Tex.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In this SMSA, 4,895 housing units were eligible for interview. Of these sample units, 208 interviews were not obtained because, for occupied sample units, the occupants refused to be interviewed, 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 units eligible for interview, 447 units were visited but were not eligible for interview because they were condemned, unfit, demolished, converted to group quarters use, etc.

**Designation of sample housing units for the 1978 survey**—The sample housing units designated to be interviewed in the 1978 survey consisted of the following categories which are described in detail in the succeeding sections.

1. All sample housing units that were interviewed in the 1975 survey.
2. All sample housing units that were type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of the survey but which could become eligible in the future) in the 1975 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1978 AHS questionnaire, page App-20.)
3. All sample housing units that were selected from the list of new construction building permits issued since the 1975 survey. (This sample represented the housing units built in permit-issuing areas, since the 1975 survey.)
4. All sample housing units that were added to sample segments in the nonpermit universe since the 1975 survey. (This sample represented additions to the housing inventory in nonpermit-issuing areas since the 1975 survey.)
5. All sample housing units that were selected as part of the 1978 Coverage Improvement Program. (This sample represented most of the housing units which, until 1978, did not have a chance of selection.)

**Selection of the 1975 AHS-SMSA sample**—The sample for the SMSA's which are 100-percent permit-issuing 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 those 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 non-permit universe). The following SMSA's are 100-percent permit-issuing: Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and San Francisco-Oakland, Calif. The remaining 10 SMSA's contain a sample from 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 three largest SMSA's, the overall sampling rate differed by central city and the 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 the 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 the total housing units in each sector.

The major portion of the sample in each SMSA 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 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 to \$5,999 . . . . .		
\$6,000 to \$9,999 . . . . .		
\$10,000 to \$14,999 . . . . .		
\$15,000 and over . . . . .		

Thus, for this SMSA, the occupied housing unit records from the permit-issuing 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 selection procedure was then instituted that would produce one-half of the desired sample size. However, whenever a record was selected to be in sample, the housing unit record adjacent to it on the file 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 by a procedure that produced one-quarter of the desired sample size. However, at the time of the survey, the units at each of the special places were listed and subsampled at a rate which produced 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 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:

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

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. All units in existence at the time of interview in these selected segments are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

**1975-1978 additions to the housing inventory**—In the permit-issuing universe, a sample of new construction building permits, issued since the 1975 survey, was selected to represent housing units built in permit-issuing areas since the 1975 survey. Sampling procedures were identical to those used in selecting the 1970-1975 new construction sample, which were described previously. In the nonpermit universe, sample segments were dependently recanvassed, using listing sheets from 1975, to identify any units missed in the 1975 survey or any units added since the 1975 survey.

**Sample selection for the 1978 Coverage Improvement Program**—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 1970 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 for each of the AA-1 SMSA's except Newport News-Hampton, Va. Two different procedures were used. For the first procedure, the sampling was carried out in two stages for one- and two-unit structures and in three stages for three-or-more-unit structures. Sample units selected from one- and two-unit structures were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units), while sample units selected from three-or-more-unit structures were sampled at one-half the rate of 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. Since these permits were not available for Newport News-Hampton, Va., SMSA, these coverage improvement procedures were not instituted for this SMSA. For the third stage, structures of size three or more were divided into clusters of an expected size of two units and a sample of clusters was selected. This procedure was employed in the following SMSA's: Colorado Springs, Colo.; Rochester, N.Y.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In the remaining 11 SMSA's, units whose permits were issued before January 1970, but which were completed after April 1970, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. These units were then sampled at one-third the rate of regular AHS units. These procedures added an estimated

993 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 parks that were missed by the census or established after the census was selected in two stages. First, for each 1978-1979 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 325 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 now contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and have a utility hookup, or were on the site during the present survey 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 AHS sample units was selected from the permit-issuing universe at a rate of 1 in 24. 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 8,898 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 AHS units in multiunit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multiunit structures selected above, all units were listed and matched to the 1970 census. Any missed units

were then assigned for interview. This procedure added an estimated 1,406 units to the coverage of the housing inventory of this SMSA.

**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 1, *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 1978 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1975 (i.e., 1975-1978 lost units). Each type of estimate employed separate, although similar, estimation procedures.

**1978 housing inventory**—The AHS estimates of characteristics of the 1978 housing inventory employed a 1-stage ratio estimation procedure in all SMSA's. 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 noninterviews previously mentioned. This noninterview adjustment was done separately for occupied and vacant housing units. 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 described). In addition, within each sector separate noninterview factors were computed for one noninterview cell for conventional new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes from both the nonpermit universe and the coverage improvement universe, and one noninterview cell for other sample units from both the nonpermit universe and the coverage improvement universe, (if units were not included in any of the previous cells).

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.

**1975-1978 lost units**—The 1975-1978 lost unit estimates employed the three-stage ratio estimation procedure used to produce the AHS-SMSA estimates of the 1975 housing inventory, as was described in the 1975 Current Housing Report, Series H-170, *Housing Characteristics for Selected Metropolitan Areas*. Since the 1975-1978 lost units existed, by definition, in the 1975 housing inventory, there was a 1975 housing inventory weight associated with each 1975-1978 lost unit. This weight was used to tabulate the estimates of the characteristics of the 1975-1978 lost units.

**1975 estimation procedure**—This report presents data on the housing characteristics of the 1975 housing inventory from the 1975 Annual Housing Survey SMSA sample. The AHS-SMSA estimation procedure employed a three-stage ratio estimation process. A detailed description of this ratio estimation procedure can be found in the AHS Series H-170 reports for 1975.

**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 1, *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 1, *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 the 1970 Census of Population and Housing and the 1975 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 the extent to which occupancy status was erroneously reported. 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 1978 AHS-SMSA sample, a reinterview program was not conducted. However, a study was conducted for the 1975 AHS-SMSA sample. The results of which are presented in the Census Bureau memorandum, "Reinterview Results for Annual Housing Survey—SMSA Sample: 1975."

**Coverage errors**—In 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. However, these permits issued during the last 5 months of the survey do not necessarily represent missed

housing units. 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, in which case, they would not have been 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 correct deficiencies 3-6 (see the coverage improvement section of this appendix) was not very effective in 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 1975 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing. It should be noted that since these ED's were recanvassed for the 1978 survey, the number of missed housing units may be considerably less for 1978.

**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 statistics 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 questionnaires, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The sampling error of a survey estimate provides a measure of the variation among the estimates from all 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.

One common measure of the sampling error is the standard error. As calculated for this report, the standard error reflects the variation in the estimates due to sampling and nonsampling errors, but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors measured by the standard error, biases, and any 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 item. Standard errors applicable to estimates of characteristics of the 1975 housing inventory can be found in the AHS Series H-170 reports for 1975.

Table I presents the standard errors applicable to estimates of characteristics of the 1978 housing inventory as well as estimates of characteristics of the 1975-1978 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 17,520 for the total SMSA, 9,730 for the central city of the SMSA, and 14,480 for the balance of the SMSA.

The reliability of an estimated percentage, computed by using the 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 1978 housing inventory as well as estimated percentages of the 1975-1978 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

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1978 Housing Inventory and for Estimated Number of 1975-1978 Lost Units for the New Orleans, La., SMSA, for the Central City, and for the Balance (Not in Central City) of the SMSA

(68 chances out of 100)

Size of estimate	Standard error <sup>1</sup>		
	SMSA	In central city	Not in central city
0 .....	100	90	110
100 .....	100	100	110
200 .....	140	140	150
500 .....	230	220	240
700 .....	270	260	280
1,000 .....	320	310	330
2,500 .....	510	480	530
5,000 .....	720	680	750
10,000 .....	1,020	960	1,050
25,000 .....	1,600	1,480	1,670
50,000 .....	2,230	2,000	2,370
75,000 .....	2,710	2,340	2,910
100,000 .....	3,090	2,560	3,370
150,000 .....	3,700	2,770	4,150
200,000 .....	4,180	2,720	4,820
250,000 .....	4,560	2,370	5,410
300,000 .....	4,860	—	—
400,000 .....	5,310	—	—
500,000 .....	5,570	—	—

<sup>1</sup> For estimates pertaining to new construction, the standard errors shown in the table should be multiplied by a factor of 1.1.

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  
 $\sigma_x$  = the standard error of the numerator  
 $\sigma_y$  = 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 1978 there were 211,300 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 approximately 4,270. 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 .....	4,180
211,300 .....	x
250,000 .....	4,560

The entry for "x" is determined as follows by vertically interpolating between 4,180 and 4,560:

$$211,300 - 200,000 = 11,300$$

$$250,000 - 200,000 = 50,000$$

$$4,180 + \frac{11,300}{50,000} (4,560 - 4,180) = 4,270$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 207,030 to 215,570 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1978 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 204,468 to 218,132 housing units with 90 percent confidence; and that the average estimate lies within the interval from 202,760 to 219,840 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 211,300 owner-occupied housing units, 50,700, or 24.0 percent, had two bedrooms. Interpolation in table II of this appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 24.0 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		
	10 or 90	24.0	25 or 75
200,000 .....	0.7	a	1.0
211,300 .....		p	
250,000 .....	0.6	b	0.9

Table II. Standard Errors for Estimated Percentages of Housing Units in the 1978 Housing Inventory and for Estimated Percentages of 1975-1978 Lost Units for New Orleans, La., SMSA, for the Central City, and for the Balance (Not in Central City) of the SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage <sup>1</sup>					
	0 or 100	1 or 99	.5 or 95	10 or 90	25 or 75	50
200 .....	35.7	35.7	35.7	35.7	35.7	37.2
500 .....	18.2	18.2	18.2	18.2	20.4	23.6
700 .....	13.7	13.7	13.7	13.7	17.2	19.9
1,000 .....	10.0	10.0	10.0	10.0	14.4	16.7
2,500 .....	4.2	4.2	4.6	6.3	9.1	10.5
5,000 .....	2.2	2.2	3.2	4.5	6.5	7.4
10,000 .....	1.1	1.1	2.3	3.2	4.6	5.3
25,000 .....	0.4	0.7	1.5	2.0	2.9	3.3
50,000 .....	0.2	0.5	1.0	1.4	2.0	2.4
75,000 .....	0.15	0.4	0.8	1.2	1.7	1.9
100,000 .....	0.11	0.3	0.7	1.0	1.4	1.7
150,000 .....	0.07	0.3	0.6	0.8	1.2	1.4
200,000 .....	0.06	0.2	0.5	0.7	1.0	1.2
250,000 .....	0.04	0.2	0.5	0.6	0.9	1.1
300,000 .....	0.04	0.2	0.4	0.6	0.8	1.0
400,000 .....	0.03	0.2	0.4	0.5	0.7	0.8
500,000 .....	0.02	0.15	0.3	0.4	0.6	0.7

<sup>1</sup> 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 errors shown in the table should be multiplied by a factor of 1.1.

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

$$24.0 - 10.0 = 14.0$$

$$25.0 - 10.0 = 15.0$$

$$0.7 + \frac{14.0}{15.0} (1.0 - 0.7) = 1.0$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.6 and 0.9.

$$24.0 - 10.0 = 14.0$$

$$25.0 - 10.0 = 15.0$$

$$0.6 + \frac{14.0}{15.0} (0.9 - 0.6) = 0.9$$

3. The entry for "p" was then determined by vertical interpolation between 1.0 and 0.9.

$$211,300 - 200,000 = 11,300$$

$$250,000 - 200,000 = 50,000$$

$$1.0 + \frac{11,300}{50,000} (0.9 - 1.0) = 1.0$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 23.0 to 25.0 percent; the 90-percent

confidence interval is from 22.4 to 25.6 percent; and the 95-percent confidence interval is from 22.0 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 characteristics 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. Due to the overlap of the 1975 and 1978 AHS-SMSA samples a positive correlation should be expected when making comparisons between the 1975 and 1978 characteristics.

**Illustration of the computation of the standard error of a difference**—Table A-1 of part A of this report shows that in 1978 there were 109,200 owner-occupied housing units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied housing units with two bedrooms and owner-occupied housing units with three bedrooms is 58,500. Table I shows the standard error of 109,200 is approximately 3,200, and the standard error of 50,700 is approximately 2,240. Therefore, the standard error of the estimated difference of 58,500 is about:

$$3,910 = \sqrt{(3,200)^2 + (2,240)^2}$$

Consequently, the 68-percent confidence interval for the 58,500 difference is from 54,590 to 62,410 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 52,244 to 64,756 housing units, and the 95-percent confidence interval is from 50,680 to 66,320. Thus, we can conclude with 95 percent confidence that the number of 1978 owner-occupied housing units with three bedrooms is greater than the number of owner-occupied housing 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 characteristics, 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 211,300 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 211,300 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 87,800 owner-occupied housing units or 41.6 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 42,400 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:

$$2.5 + (3.5 - 2.5) \left( \frac{47.6 - 41.6}{20.1} \right) = 2.8$$

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

$$2.5 + (3.5 - 2.5) \left( \frac{52.4 - 41.6}{20.1} \right) = 3.0$$

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

# Newport News-Hampton 1978

# Appendix B

## Source and Reliability of the Estimates

SAMPLE DESIGN .....	App-40	Coverage improvement for deficiencies 3-6 .....	App-42	AHS-SMSA .....	App-44
Annual Housing Survey .....	App-40	1970 Census of Population and Housing .....	App-43	Coverage errors .....	App-44
Designation of sample housing units for the 1978 survey .....	App-40	ESTIMATION .....	App-43	Rounding errors .....	App-44
Selection of the 1975 AHS-SMSA sample .....	App-40	1978 housing inventory .....	App-43	Sampling errors for the AHS-SMSA sample .....	App-44
1975-1978 additions to the housing inventory .....	App-42	1975-1978 lost units .....	App-43	Illustration of the use of the standard error tables .....	App-45
Sample selection for the 1978 Coverage Improvement Program .....	App-42	1975 estimation procedure .....	App-43	Differences .....	App-46
Coverage improvement for deficiency 1 .....	App-42	Ratio estimation procedure of the 1970 Census of Population and Housing .....	App-43	Illustration of the computation of the standard error of a difference .....	App-47
Coverage improvement for deficiency 2 .....	App-42	RELIABILITY OF THE ESTIMATES .....	App-43	Medians .....	App-47
		Nonsampling errors .....	App-44	Illustration of the computation of the 95-percent confidence interval of a median .....	App-47
		1970 census .....	App-44		

### SAMPLE DESIGN

**Annual Housing Survey**—The estimates for each of the 15 SMSA's are based on data collected from the 1978 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 15 SMSA's, the data were collected for the 12-month period from April 1978 through March 1979 with one-twelfth of the sample units being visited each month.

Each group of SMSA's is being interviewed for the AHS on a rotating basis with this group (AA-1 SMSA's) being the second to be revisited. All of these SMSA's were enumerated for the first time in 1975.

For each group of SMSA's, the largest SMSA from 3 of the 4 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 the remaining 12 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 AA-1 group (1978-1979) are: Atlanta, Ga.; Philadelphia, Pa.-N.J.; and San Francisco-Oakland, Calif.

The remaining SMSA's in the AA-1 group are: Cincinnati, Ohio-Ky.-Ind.; Colorado Springs, Colo.; Columbus, Ohio; Kansas City, Mo.-Kans.; New Orleans, La.; Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; Rochester, N.Y.; San Antonio, Tex.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In this SMSA, 4,187 housing units were eligible for interview. Of these sample units, 138 interviews were not obtained because, for occupied sample units, the occupants refused to be interviewed, 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 units eligible for interview, 290 units were visited but were not eligible for interview because they were condemned, unfit, demolished, converted to group quarters use, etc.

**Designation of sample housing units for the 1978 survey**—The sample housing units designated to be interviewed in the 1978 survey consisted of the following categories which are described in detail in the succeeding sections.

1. All sample housing units that were interviewed in the 1975 survey.
2. All sample housing units that were type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of the survey but which could become eligible in the future) in the 1975 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1978 AHS questionnaire, page App-20.)
3. All sample housing units that were selected from the list of new construction building permits issued since the 1975 survey. (This sample represented the housing units built in permit-issuing areas, since the 1975 survey.)
4. All sample housing units that were added to sample segments in the nonpermit universe since the 1975 survey. (This sample represented additions to the housing inventory in nonpermit-issuing areas since the 1975 survey.)
5. All sample housing units that were selected as part of the 1978 Coverage Improvement Program. (This sample represented most of the housing units which, until 1978, did not have a chance of selection.)

**Selection of the 1975 AHS-SMSA sample**—The sample for the SMSA's which are 100-percent permit-issuing 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 those 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 non-permit universe). The following SMSA's are 100-percent permit-issuing: Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and San Francisco-Oakland, Calif. The remaining 10 SMSA's contain a sample from 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 three largest SMSA's, the overall sampling rate differed by central city and the 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 the 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 the total housing units in each sector.

The major portion of the sample in each SMSA 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 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 to \$5,999 . .		
\$6,000 to \$9,999 . .		
\$10,000 to \$14,999 .		
\$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from the permit-issuing 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 selection procedure was then instituted that would produce one-half of the desired sample size. However, whenever a record was selected to be in sample, the housing unit record adjacent to it on the file 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 by a procedure that produced one-quarter of the desired sample size. However, at the time of the survey, the units at each of the special places were listed and subsampled at a rate which produced 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 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:

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

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. All units in existence at the time of interview in these selected segments are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

**1975-1978 additions to the housing inventory**—In the permit-issuing universe, a sample of new construction building permits, issued since the 1975 survey, was selected to represent housing units built in permit-issuing areas since the 1975 survey. Sampling procedures were identical to those used in selecting the 1970-1975 new construction sample, which were described previously. In the nonpermit universe, sample segments were dependently recanvassed, using listing sheets from 1975, to identify any units missed in the 1975 survey or any units added since the 1975 survey.

**Sample selection for the 1978 Coverage Improvement Program**—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 1970 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 for each of the AA-1 SMSA's except Newport News-Hampton, Va. Two different procedures were used. For the first procedure, the sampling was carried out in two stages for one- and two-unit structures and in three stages for three-or-more-unit structures. Sample units selected from one- and two-unit structures were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units), while sample units selected from three-or-more-unit structures were sampled at one-half the rate of 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. Since these permits were not available for Newport News-Hampton, Va., SMSA, these coverage improvement procedures were not instituted for this SMSA. For the third stage, structures of size three or more were divided into clusters of an expected size of two units and a sample of clusters was selected. This procedure was employed in the following SMSA's: Colorado Springs, Colo.; Rochester, N.Y.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In the remaining 11 SMSA's, units whose permits were issued before January 1970, but which were completed after April 1970, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census.

**Coverage improvement for deficiency 2**—In permit-issuing areas, a sample of mobile homes placed in parks that were missed by the census or established after the census was selected in two stages. First, for each 1978-1979 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 133 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 now contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and have a utility hookup, or were on the site during the present survey 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 AHS sample units was selected from the permit-issuing universe at a rate of 1 in 24. 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 1,516 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 AHS units in multiunit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multiunit structures selected above, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 81 units to the coverage of the housing inventory of this SMSA.

**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 1978 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1975 (i.e., 1975-1978 lost units). Each type of estimate employed separate, although similar, estimation procedures.

**1978 housing inventory**—The AHS estimates of characteristics of the 1978 housing inventory employed a 1-stage ratio estimation procedure in all SMSA's. 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 noninterviews previously mentioned. This noninterview adjustment was done separately for occupied and vacant housing units. 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 described). In addition, within each sector separate noninterview factors were computed for one noninterview cell for conventional new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes from both the nonpermit universe and the coverage improvement universe, and one noninterview cell for other sample units from both the nonpermit universe and the coverage improvement universe, (if units were not included in any of the previous cells).

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.

**1975-1978 lost units**—The 1975-1978 lost unit estimates employed the three-stage ratio estimation procedure used to produce the AHS-SMSA estimates of the 1975 housing inventory, as was described in the 1975 Current Housing Report, Series H-170, **Housing Characteristics for Selected Metropolitan Areas**. Since the 1975-1978 lost units existed, by definition, in the 1975 housing inventory, there was a 1975 housing inventory weight associated with each 1975-1978 lost unit. This weight was used to tabulate the estimates of the characteristics of the 1975-1978 lost units.

**1975 estimation procedure**—This report presents data on the housing characteristics of the 1975 housing inventory from the 1975 Annual Housing Survey SMSA sample. The AHS-SMSA estimation procedure employed a three-stage ratio estimation process. A detailed description of this ratio estimation procedure can be found in the AHS Series H-170 reports for 1975.

**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 1, *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 the 1970 Census of Population and Housing and the 1975 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 the extent to which occupancy status was erroneously reported. 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 1978 AHS-SMSA sample, a reinterview program was not conducted. However, a study was conducted for the 1975 AHS-SMSA sample. The results of which are presented in the Census Bureau memorandum, "Reinterview Results for Annual Housing Survey—SMSA Sample: 1975."

**Coverage errors**—In 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. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units. 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, in which case, they would not have been 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 correct deficiencies 3-6 (see the coverage improvement section of this appendix) was not very effective in 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 1975 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing. It should be noted that since these ED's were recanvassed for the 1978 survey, the number of missed housing units may be considerably less for 1978.

**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 statistics 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 questionnaires, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The sampling error of a survey estimate provides a measure of the variation among the estimates from all 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.

One common measure of the sampling error is the standard error. As calculated for this report, the standard error reflects the variation in the estimates due to sampling and nonsampling errors, but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors measured by the standard error, biases, and any 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 item. Standard errors applicable to estimates of characteristics of the 1975 housing inventory can be found in the AHS Series H-170 reports for 1975.

Table I presents the standard errors applicable to estimates of characteristics of the 1978 housing inventory as well as estimates of characteristics of the 1975-1978 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 4,930 for the total SMSA, 4,580 for the central city of the SMSA, and 11,780 for the balance of the SMSA.

The reliability of an estimated percentage, computed by using the 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 1978 housing inventory as well as estimated percentages of the 1975-1978 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 I. Standard Errors for Estimated Number of Housing Units in the 1978 Housing Inventory and for Estimated Number of 1975-1978 Lost Units for the Newport News-Hampton, Va., SMSA, for the Central City, and for the Balance (Not in Central City) of the SMSA**

(68 chances out of 100)

Size of estimate	Standard error <sup>1</sup>		
	SMSA	In central city	Not in central city
0 .....	30	30	20
100 .....	50	50	50
200 .....	80	80	70
500 .....	120	120	110
700 .....	140	140	130
1,000 .....	170	170	160
2,500 .....	270	270	250
5,000 .....	380	370	350
10,000 .....	530	520	510
25,000 .....	820	790	840
50,000 .....	1,110	1,040	—
75,000 .....	1,310	1,170	—
100,000 .....	1,440	1,210	—
150,000 .....	1,590	—	—

<sup>1</sup>For estimates pertaining to new construction, the standard errors shown in the table should be multiplied by a factor of 1.2 for the total SMSA and for the central city, and 1.0 for the balance (not in central city) estimates.

where: x = the numerator of the ratio  
 y = the denominator of the ratio  
 $\sigma_x$  = the standard error of the numerator  
 $\sigma_y$  = 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 1978 there were 58,000 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 1,170. 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
50,000 .....	1,110
58,000 .....	x
75,000 .....	1,310

The entry for "x" is determined as follows by vertically interpolating between 1,110 and 1,310:

$$\begin{aligned} 58,000 - 50,000 &= 8,000 \\ 75,000 - 50,000 &= 25,000 \\ 1,110 + \frac{8,000}{25,000} (1,310 - 1,110) &= 1,170 \end{aligned}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 56,830 to 59,170 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1978 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 56,128 to 59,872 housing units with 90 percent confidence; and that the average estimate lies within the interval from 55,660 to 60,340 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 58,000 owner-occupied housing units, 10,100, or 17.4 percent, had two bedrooms. Interpolation in table II of this appendix (i.e., interpolation on both the base and percent) shows that the

**TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1978 Housing Inventory and for Estimated Percentages of 1975-1978 Lost Units for the Newport News-Hampton, Va., SMSA, for the Central City, and for the Balance (Not in Central City) of the SMSA**

(68 chances out of 100)

Base of percentage	Estimated percentage <sup>1</sup>					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100 . . . . .	22.3	22.3	22.3	22.3	23.2	26.8
200 . . . . .	12.6	12.6	12.6	12.6	16.4	19.0
500 . . . . .	5.4	5.4	5.4	7.2	10.4	12.0
700 . . . . .	3.9	3.9	4.4	6.1	8.8	10.1
1,000 . . . . .	2.8	2.8	3.7	5.1	7.3	8.5
2,500 . . . . .	1.1	1.1	2.3	3.2	4.6	5.4
5,000 . . . . .	0.6	0.8	1.7	2.3	3.3	3.8
10,000 . . . . .	0.3	0.5	1.2	1.6	2.3	2.7
25,000 . . . . .	0.11	0.3	0.7	1.0	1.5	1.7
50,000 . . . . .	0.06	0.2	0.5	0.7	1.0	1.2
75,000 . . . . .	0.04	0.2	0.4	0.6	0.8	1.0
100,000 . . . . .	0.03	0.2	0.4	0.5	0.7	0.8
150,000 . . . . .	0.02	0.14	0.3	0.4	0.6	0.7

<sup>1</sup> 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 errors shown in the table should be multiplied by a factor of 1.2 for the total SMSA and for the central city, and 1.0 for the balance (not in central city).

standard error of the 17.4 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	17.4	25 or 75
50,000 . . . . .	0.7	a	1.0
58,000 . . . . .		p	
75,000 . . . . .	0.6	b	0.8

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

$$\begin{aligned} 17.4 - 10.0 &= 7.4 \\ 25.0 - 10.0 &= 15.0 \\ 0.7 + \frac{7.4}{15.0} (1.0 - 0.7) &= 0.8 \end{aligned}$$

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

$$\begin{aligned} 17.4 - 10.0 &= 7.4 \\ 25.0 - 10.0 &= 15.0 \\ 0.6 + \frac{7.4}{15.0} (0.8 - 0.6) &= 0.7 \end{aligned}$$

3. The entry for "p" was then determined by vertical interpolation between 0.8 and 0.7.

$$\begin{aligned} 58,000 - 50,000 &= 8,000 \\ 75,000 - 50,000 &= 25,000 \\ 0.8 + \frac{8,000}{25,000} (0.7 - 0.8) &= 0.8 \end{aligned}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 16.6 to 18.2 percent; the 90-percent confidence interval is from 16.1 to 18.7 percent; and the 95-percent confidence interval is from 15.8 to 19.0 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 characteristics 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. Due to the overlap of the 1975 and 1978 AHS-SMSA samples a positive correlation should be expected when making comparisons between the 1975 and 1978 characteristics.

*Illustration of the computation of the standard error of a difference*—Table A-1 of part A of this report shows that in 1978 there were 33,400 owner-occupied housing units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied housing units with two bedrooms and owner-occupied housing units with three bedrooms is 23,300. Table I shows the standard error of 33,400 is approximately 920, and the standard error of 10,100 is approximately 530. Therefore, the standard error of the estimated difference of 23,300 is about:

$$1,060 = \sqrt{(920)^2 + (530)^2}$$

Consequently, the 68-percent confidence interval for the 23,300 difference is from 22,240 to 24,360 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 21,604 to 24,996 housing units, and the 95-percent confidence interval is from 21,180 to 25,420. Thus, we can conclude with 95 percent confidence that the number of 1978 owner-occupied housing units with three bedrooms is greater than the number of owner-occupied housing 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 characteristics, 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 58,000 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 58,000 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 24,500 owner-occupied housing units or 42.2 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 13,200 owner-occupied housing units, or 22.8 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:

$$2.5 + (3.5-2.5) \left( \frac{47.8-42.2}{22.8} \right) = 2.7$$

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

$$2.5 + (3.5-2.5) \left( \frac{52.2-42.2}{22.8} \right) = 2.9$$

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

Paterson-Clifton-Passaic 1978

# Appendix B

## Source and Reliability of the Estimates

<b>SAMPLE DESIGN</b> . . . . .	App-40	<b>Coverage improvement for deficiencies 3-6</b> . . . . .	App-42	<b>AHS-SMSA</b> . . . . .	App-44
<b>Annual Housing Survey</b> . . . . .	App-40	<b>1970 Census of Population and Housing</b> . . . . .	App-43	<b>Coverage errors</b> . . . . .	App-44
<b>Designation of sample housing units for the 1978 survey</b> . . . . .	App-40	<b>ESTIMATION</b> . . . . .	App-43	<b>Rounding errors</b> . . . . .	App-44
<b>Selection of the 1975 AHS-SMSA sample</b> . . . . .	App-40	<b>1978 housing inventory</b> . . . . .	App-43	<b>Sampling errors for the AHS-SMSA sample</b> . . . . .	App-44
<b>1975-1978 additions to the housing inventory</b> . . . . .	App-42	<b>1975-1978 lost units</b> . . . . .	App-43	<b>Illustration of the use of the standard error tables</b> . . . . .	App-45
<b>Sample selection for the 1978 Coverage Improvement Program.</b> . . . . .	App-42	<b>1975 estimation procedure</b> . . . . .	App-43	<b>Differences</b> . . . . .	App-46
<b>Coverage improvement for deficiency 1</b> . . . . .	App-42	<b>Ratio estimation procedure of the 1970 Census of Population and Housing</b> . . . . .	App-43	<b>Illustration of the computation of the standard error of a difference</b> . . . . .	App-46
<b>Coverage improvement for deficiency 2</b> . . . . .	App-42	<b>RELIABILITY OF THE ESTIMATES</b> . . . . .	App-43	<b>Medians</b> . . . . .	App-47
		<b>Nonsampling errors</b> . . . . .	App-44	<b>Illustration of the computation of the 95-percent confidence interval of a median</b> . . . . .	App-47
		<b>1970 census</b> . . . . .	App-44		

### SAMPLE DESIGN

**Annual Housing Survey**—The estimates for each of the 15 SMSA's are based on data collected from the 1978 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 15 SMSA's, the data were collected for the 12-month period from April 1978 through March 1979 with one-twelfth of the sample units being visited each month.

Each group of SMSA's is being interviewed for the AHS on a rotating basis with this group (AA-1 SMSA's) being the second to be revisited. All of these SMSA's were enumerated for the first time in 1975.

For each group of SMSA's, the largest SMSA from 3 of the 4 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 the remaining 12 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 AA-1 group (1978-1979) are Atlanta, Ga.; Philadelphia, Pa.-N.J.; and San Francisco-Oakland, Calif.

The remaining SMSA's in the AA-1 group are Cincinnati, Ohio-Ky.-Ind.; Colorado Springs, Colo.; Columbus, Ohio; Kansas City, Mo.-Kans.; New Orleans, La.; Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; Rochester, N.Y.; San Antonio, Tex.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In this SMSA, 4,816 housing units were eligible for interview. Of these sample units, 298 interviews were not obtained because, for occupied sample units, the occupants refused to be interviewed, 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 units eligible for interview, 166 units were visited but were not eligible for interview because they were condemned, unfit, demolished, converted to group quarters use, etc.

**Designation of sample housing units for the 1978 survey**—The sample housing units designated to be interviewed in the 1978 survey consisted of the following categories which are described in detail in the succeeding sections.

1. All sample housing units that were interviewed in the 1975 survey.
2. All sample housing units that were type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of the survey but which could become eligible in the future) in the 1975 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1978 AHS questionnaire, page App-20.)
3. All sample housing units that were selected from the list of new construction building permits issued since the 1975 survey. (This sample represented the housing units built in permit-issuing areas, since the 1975 survey.)
4. All sample housing units that were added to sample segments in the nonpermit universe since the 1975 survey. (This sample represented additions to the housing inventory in nonpermit-issuing areas since the 1975 survey.)
5. All sample housing units that were selected as part of the 1978 Coverage Improvement Program. (This sample represented most of the housing units which, until 1978, did not have a chance of selection.)

**Selection of the 1975 AHS-SMSA sample**—The sample for the SMSA's which are 100-percent permit-issuing 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 those 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 non-permit universe). The following SMSA's are 100-percent permit-issuing: Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and San Francisco-Oakland, Calif. The remaining 10 SMSA's contain a sample from 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 three largest SMSA's, the overall sampling rate differed by central city and the 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 the 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 the total housing units in each sector.

The major portion of the sample in each SMSA 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 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 to \$5,999 . .		
\$6,000 to \$9,999 . .		
\$10,000 to \$14,999 .		
\$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from the permit-issuing 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 selection procedure was then instituted that would produce one-half of the desired sample size. However, whenever a record was selected to be in sample, the housing unit record adjacent to it on the file 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 by a procedure that produced one-quarter of the desired sample size. However, at the time of the survey, the units at each of the special places were listed and subsampled at a rate which produced 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 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:

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

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. All units in existence at the time of interview in these selected segments are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

**1975-1978 additions to the housing inventory**—In the permit-issuing universe, a sample of new construction building permits, issued since the 1975 survey, was selected to represent housing units built in permit-issuing areas since the 1975 survey. Sampling procedures were identical to those used in selecting the 1970-1975 new construction sample, which were described previously. In the nonpermit universe, sample segments were dependently recanvassed, using listing sheets from 1975, to identify any units missed in the 1975 survey or any units added since the 1975 survey.

**Sample selection for the 1978 Coverage Improvement Program**—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 1970 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 for each of the AA-1 SMSA's except Newport News-Hampton, Va. Two different procedures were used. For the first procedure, the sampling was carried out in two stages for one- and two-unit structures and in three stages for three-or-more-unit structures. Sample units selected from one- and two-unit structures were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units), while sample units selected from three-or-more-unit structures were sampled at one-half the rate of 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. Since these permits were not available for Newport News-Hampton, Va., SMSA, these coverage improvement procedures were not instituted for this SMSA. For the third stage, structures of size three or more were divided into clusters of an expected size of two units, and a sample of clusters was selected. This procedure was employed in the following SMSA's: Colorado Springs, Colo.; Rochester, N.Y.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In the remaining 11 SMSA's, units whose permits were issued before January 1970, but which were completed after April 1970, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. These units were then sampled at one-third the rate of regular AHS units. These procedures added an estimated 1,749 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 parks that were missed by the census or established after the census was selected in two stages. First, for each 1978-1979 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 now contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and have a utility hookup, or were on the site during the present survey 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 AHS sample units was selected from the permit-issuing universe at a rate of 1 in 24. 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 16,595 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 AHS units in multiunit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multiunit structures selected above, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 3,420 units to the coverage of the housing inventory of this SMSA.

**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 1978 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1975 (i.e., 1975-1978 lost units). Each type of estimate employed separate, although similar, estimation procedures.

**1978 housing inventory**—The AHS estimates of characteristics of the 1978 housing inventory employed a 1-stage ratio estimation procedure in all SMSA's. 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 noninterviews previously mentioned. This noninterview adjustment was done separately for occupied and vacant housing units. 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 described). In addition, within each sector separate noninterview factors were computed for one noninterview cell for conventional new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes from both the nonpermit universe and the coverage improvement universe, and one noninterview cell for other sample units from both the nonpermit universe and the coverage improvement universe (if units were not included in any of the previous cells).

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.

**1975-1978 lost units**—The 1975-1978 lost unit estimates employed the three-stage ratio estimation procedure used to produce the AHS-SMSA estimates of the 1975 housing inventory, as was described in the 1975 Current Housing Report, Series H-170, *Housing Characteristics for Selected Metropolitan Areas*. Since the 1975-1978 lost units existed, by definition, in the 1975 housing inventory, there was a 1975 housing inventory weight associated with each 1975-1978 lost unit. This weight was used to tabulate the estimates of the characteristics of the 1975-1978 lost units.

**1975 estimation procedure**—This report presents data on the housing characteristics of the 1975 housing inventory from the 1975 Annual Housing Survey SMSA sample. The AHS-SMSA estimation procedure employed a three-stage ratio estimation process. A detailed description of this ratio estimation procedure can be found in the AHS Series H-170 reports for 1975.

**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 the 1970 Census of Population and Housing and the 1975 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 the extent to which occupancy status was erroneously reported. 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 1978 AHS-SMSA sample, a reinterview program was not conducted. However, a study was conducted for the 1975 AHS-SMSA sample. The results of which are presented in the Census Bureau memorandum, "Reinterview Results for Annual Housing Survey—SMSA Sample: 1975."

**Coverage errors**—In 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. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units. 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, in which case, they would not have been 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 correct deficiencies 3-6 (see the coverage improvement section of this appendix) was not very effective in 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 1975 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing. It should be noted that since these ED's were recanvassed for the 1978 survey, the number of missed housing units may be considerably less for 1978.

**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 statistics 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 questionnaires, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The sampling error of a survey estimate provides a measure of the variation among the estimates from all 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.

One common measure of the sampling error is the standard error. As calculated for this report, the standard error reflects the variation in the estimates due to sampling and nonsampling errors, but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors measured by the standard error, biases, and any 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 item. Standard errors applicable to estimates of characteristics of the 1975 housing inventory can be found in the AHS Series H-170 reports for 1975.

Table I presents the standard errors applicable to estimates of characteristics of the 1978 housing inventory as well as estimates of characteristics of the 1975-1978 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 15,360 for the total SMSA, 8,460 for the central city of the SMSA, and 12,760 for the balance of the SMSA.

The reliability of an estimated percentage, computed by using the 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 1978 housing inventory as well as estimated percentages of the 1975-1978 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  
 $\sigma_x$  = the standard error of the numerator  
 $\sigma_y$  = 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 1978 there were 291,200 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 4,160. 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 .....	4,100
291,200 .....	x
300,000 .....	4,170

The entry for "x" is determined as follows by vertically interpolating between 4,100 and 4,170:

$$291,200 - 250,000 = 41,200$$

$$300,000 - 250,000 = 50,000$$

$$4,100 + \frac{41,200}{50,000} (4,170 - 4,100) = 4,160$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 287,040 to 295,360 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1978 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 284,544 to 297,856 housing units with 90 percent confidence and that the average estimate lies within the interval from 282,880 to 299,520 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 291,200 owner-occupied housing units, 67,200, or 23.1 percent, had two bedrooms. Interpolation in table II of this 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
250,000 .....	0.6	a	0.9
291,200 .....		p	
300,000 .....	0.6	b	0.8

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

$$23.1 - 10.0 = 13.1$$

$$25.0 - 10.0 = 15.0$$

$$0.6 + \frac{13.1}{15.0} (0.9 - 0.6) = 0.9$$

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

$$23.1 - 10.0 = 13.1$$

$$25.0 - 10.0 = 15.0$$

$$0.6 + \frac{13.1}{15.0} (0.8 - 0.6) = 0.8$$

3. The entry for "p" was then determined by vertical interpolation between 0.9 and 0.8.

$$291,200 - 250,000 = 41,200$$

$$300,000 - 250,000 = 50,000$$

$$0.9 + \frac{41,200}{50,000} (0.8 - 0.9) = 0.8$$

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 characteristics 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. Due to the overlap of the 1975 and 1978 AHS-SMSA samples, a positive correlation should be expected when making comparisons between the 1975 and 1978 characteristics.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1978 Housing Inventory and for Estimated Number of 1975-1978 Lost Units for the Paterson-Clifton-Passaic, N.J., SMSA, for the Central City, and for the Balance (Not in Central City) of the SMSA

(68 chances out of 100)

Size of estimate	Standard error <sup>1</sup>		
	SMSA	In central city	Not in central city
0 .....	110	100	110
100 .....	110	100	110
200 .....	150	140	150
500 .....	240	230	240
700 .....	280	270	280
1,000 .....	340	320	340
2,500 .....	530	500	530
5,000 .....	750	710	750
10,000 .....	1,060	980	1,060
25,000 .....	1,650	1,480	1,650
50,000 .....	2,280	1,900	2,260
75,000 .....	2,730	2,060	2,680
100,000 .....	3,080	2,030	3,000
150,000 .....	3,580	1,260	3,410
200,000 .....	3,910	—	3,610
250,000 .....	4,100	—	3,630
300,000 .....	4,170	—	3,480
400,000 .....	3,970	—	2,500
500,000 .....	3,240	—	—

<sup>1</sup> For estimates pertaining to new construction, the standard errors shown in the table should be multiplied by a factor of 1.1.

*Illustration of the computation of the standard error of a difference*—Table A-1 of part A of this report shows that in 1978 there were 139,300 owner-occupied housing units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied housing units with two bedrooms and owner-occupied housing units with three bedrooms is 72,100. Table I shows the standard error of 139,300 is approximately 3,470, and the standard error of 67,200 is approximately 2,590. Therefore, the standard error of the estimated difference of 72,100 is about:

$$4,330 = \sqrt{(3,470)^2 + (2,590)^2}$$

Consequently, the 68-percent confidence interval for the 72,100 difference is from 67,770 to 76,430 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 65,172 to 79,028 housing units, and the 95-percent confidence interval is from 63,440 to 80,760. Thus, we can conclude with

95 percent confidence that the number of 1978 owner-occupied housing units with three bedrooms is greater than the number of owner-occupied housing 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 II. Standard Errors for Estimated Percentages of Housing Units in the 1978 Housing Inventory and for Estimated Percentages of 1975-1978 Lost Units for the Paterson-Clifton-Passaic, N.J., SMSA, for the Central City, and for the Balance (Not in Central City) of the SMSA**

(68 chances out of 100)

Base of percentage	Estimated percentage <sup>1</sup>					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
200.....	36.4	36.4	36.4	36.4	36.4	37.8
500.....	18.6	18.6	18.6	18.6	20.7	23.9
700.....	14.1	14.1	14.1	14.1	17.5	20.2
1,000.....	10.3	10.3	10.3	10.3	14.7	16.9
2,500.....	4.4	4.4	4.7	6.4	9.3	10.7
5,000.....	2.2	2.2	3.3	4.5	6.6	7.6
10,000.....	1.1	1.1	2.3	3.2	4.6	5.4
25,000.....	0.5	0.7	1.5	2.0	2.9	3.4
50,000.....	0.2	0.5	1.0	1.4	2.1	2.4
75,000.....	0.15	0.4	0.9	1.2	1.7	2.0
100,000.....	0.11	0.3	0.7	1.0	1.5	1.7
150,000.....	0.08	0.3	0.6	0.8	1.2	1.4
200,000.....	0.06	0.2	0.5	0.7	1.0	1.2
250,000.....	0.05	0.2	0.5	0.6	0.9	1.1
300,000.....	0.04	0.2	0.4	0.6	0.8	1.0
400,000.....	0.03	0.2	0.4	0.5	0.7	0.8
500,000.....	0.02	0.2	0.3	0.5	0.7	0.8

<sup>1</sup> 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 errors shown in the table should be multiplied by a factor of 1.1.

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 characteristics, 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 3.1. The base of the distribution from which this median was determined is 291,200 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 291,200 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 108,100 owner-occupied housing units, or 37.1 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 58,900 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:

$$2.5 + (3.5 - 2.5) \left( \frac{48.0 - 37.1}{20.2} \right) = 3.0$$

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

$$2.5 + (3.5 - 2.5) \left( \frac{52.0 - 37.1}{20.2} \right) = 3.2$$

Thus, the 95-percent confidence interval ranges from 3.0 to 3.2 persons.

Philadelphia 1978

# Appendix B

## Source and Reliability of the Estimates

<b>SAMPLE DESIGN</b> . . . . .	<b>App-40</b>	<b>Coverage improvement for deficiencies 3-6</b> . . . . .	<b>App-42</b>	<b>AHS-SMSA</b> . . . . .	<b>App-44</b>
Annual Housing Survey . . . . .	App-40	1970 Census of Population and Housing . . . . .	App-42	Coverage errors . . . . .	App-44
Designation of sample housing units for the 1978 survey . . . . .	App-40	<b>ESTIMATION</b> . . . . .	<b>App-43</b>	Rounding errors . . . . .	App-44
Selection of the 1975 AHS-SMSA sample . . . . .	App-40	1978 housing inventory . . . . .	App-43	Sampling errors for the AHS-SMSA sample . . . . .	App-44
1975-1978 additions to the housing inventory . . . . .	App-41	1975-1978 lost units . . . . .	App-43	Illustration of the use of the standard error tables . . . . .	App-46
Sample selection for the 1978 Coverage Improvement Program . . . . .	App-42	1975 estimation procedure . . . . .	App-43	Differences . . . . .	App-46
Coverage improvement for deficiency 1 . . . . .	App-42	Ratio estimation procedure of the 1970 Census of Population and Housing . . . . .	App-43	Illustration of the computation of the standard error of a difference . . . . .	App-46
Coverage improvement for deficiency 2 . . . . .	App-42	<b>RELIABILITY OF THE ESTIMATES</b> . . . . .	<b>App-43</b>	Medians . . . . .	App-47
		Nonsampling errors . . . . .	App-44	Illustration of the computation of the 95-percent confidence interval of a median . . . . .	App-48
		1970 census . . . . .	App-44		

### SAMPLE DESIGN

**Annual Housing Survey**—The estimates for each of the 15 SMSA's are based on data collected from the 1978 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 15 SMSA's, the data were collected for the 12-month period from April 1978 through March 1979 with one-twelfth of the sample units being visited each month.

Each group of SMSA's is being interviewed for the AHS on a rotating basis with this group (AA-1 SMSA's) being the second to be revisited. All of these SMSA's were enumerated for the first time in 1975.

For each group of SMSA's, the largest SMSA from 3 of the 4 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 the remaining 12 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 AA-1 group (1978-1979) are: Atlanta, Ga.; Philadelphia, Pa.-N.J.; and San Francisco-Oakland, Calif.

The remaining SMSA's in the AA-1 group are: Cincinnati, Ohio-Ky.-Ind.; Colorado Springs, Colo.; Columbus, Ohio; Kansas City, Mo.-Kans.; New Orleans, La.; Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; Rochester, N.Y.; San Antonio, Tex.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In this SMSA, 14,294 housing units were eligible for interview. Of these sample units, 807 interviews were not obtained because, for occupied sample units, the occupants refused to be interviewed, 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 units eligible for interview, 1,226 units were visited but were not eligible for interview because they were condemned, unfit, demolished, converted to group quarters use, etc.

**Designation of sample housing units for the 1978 survey**—The sample housing units designated to be interviewed in the 1978 survey consisted of the following categories which are described in detail in the succeeding sections.

1. All sample housing units that were interviewed in the 1975 survey.
2. All sample housing units that were type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of the survey but which could become eligible in the future) in the 1975 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1978 AHS questionnaire, page App-20.)
3. All sample housing units that were selected from the list of new construction building permits issued since the 1975 survey. (This sample represented the housing units built in permit-issuing areas, since the 1975 survey.)
4. All sample housing units that were added to sample segments in the nonpermit universe since the 1975 survey. (This sample represented additions to the housing inventory in nonpermit-issuing areas since the 1975 survey.)
5. All sample housing units that were selected as part of the 1978 Coverage Improvement Program. (This sample represented most of the housing units which, until 1978, did not have a chance of selection.)

**Selection of the 1975 AHS-SMSA sample**—The sample for the SMSA's which are 100-percent permit-issuing 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 those 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 non-permit universe). The following SMSA's are 100-percent permit-issuing: Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and San Francisco-Oakland, Calif. The remaining 10 SMSA's contain a sample from 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 three largest SMSA's, the overall sampling rate differed by central city and the 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 the 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 the total housing units in each sector.

The major portion of the sample in each SMSA 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 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 to \$5,999 . .		
\$6,000 to \$9,999 . .		
\$10,000 to \$14,999 .		
\$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from the permit-issuing 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 selection procedure was then instituted that would produce one-half of the desired sample size. However, whenever a record was selected to be in sample, the housing unit record adjacent to it on the file 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 by a procedure that produced one-quarter of the desired sample size. However, at the time of the survey, the units at each of the special places were listed and subsampled at a rate which produced 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 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:

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

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. All units in existence at the time of interview in these selected segments are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

**1975-1978 additions to the housing inventory**—In the permit-issuing universe, a sample of new construction building permits, issued since the 1975 survey, was selected to represent housing

units built in permit-issuing areas since the 1975 survey. Sampling procedures were identical to those used in selecting the 1970-1975 new construction sample, which were described previously. In the nonpermit universe, sample segments were dependently recanvassed, using listing sheets from 1975, to identify any units missed in the 1975 survey or any units added since the 1975 survey.

**Sample selection for the 1978 Coverage Improvement Program**—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 1970 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 for each of the AA-1 SMSA's except Newport News-Hampton, Va. Two different procedures were used. For the first procedure, the sampling was carried out in two stages for one- and two-unit structures and in three stages for three-or-more-unit structures. Sample units selected from one- and two-unit structures were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units), while sample units selected from three-or-more-unit structures were sampled at one-half the rate of 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. Since these permits were not available for Newport News-Hampton, Va., SMSA, these coverage improvement procedures were not instituted for this SMSA. For the third stage, structures of size three or more were divided into clusters of an expected size of two units and a sample of clusters was selected. This procedure was employed in the following SMSA's: Colorado Springs, Colo.; Rochester, N.Y.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In the remaining 11 SMSA's, units whose permits were issued before January 1970, but which were completed after April 1970, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. These units were then sampled at one-third the rate of regular AHS units. These procedures added an estimated 13,905 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 parks that were missed by

the census or established after the census was selected in two stages. First, for each 1978-1979 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 4,025 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 now contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and have a utility hookup, or were on the site during the present survey 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 AHS sample units was selected from the permit-issuing universe at a rate of 1 in 24. 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 39,443 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 AHS units in multiunit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multiunit structures selected above, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 6,553 units to the coverage of the housing inventory of this SMSA.

**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 1, **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 1978 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1975 (i.e., 1975-1978 lost units). Each type of estimate employed separate, although similar, estimation procedures.

**1978 housing inventory**—The AHS estimates of characteristics of the 1978 housing inventory employed a 1-stage ratio estimation procedure in all SMSA's. 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 noninterviews previously mentioned. This noninterview adjustment was done separately for occupied and vacant housing units. 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 described). In addition, within each sector separate noninterview factors were computed for one noninterview cell for conventional new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes from both the nonpermit universe and the coverage improvement universe, and one noninterview cell for other sample units from both the nonpermit universe and the coverage improvement universe, (if units were not included in any of the previous cells).

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.

**1975-1978 lost units**—The 1975-1978 lost unit estimates employed the three-stage ratio estimation procedure used to produce the AHS-SMSA estimates of the 1975 housing inventory, as was described in the 1975 Current Housing Report, Series H-170, **Housing Characteristics for Selected Metropolitan Areas**. Since the 1975-1978 lost units existed, by definition, in the 1975 housing inventory, there was a 1975 housing inventory weight associated with each 1975-1978 lost unit. This weight was used to tabulate the estimates of the characteristics of the 1975-1978 lost units.

**1975 estimation procedure**—This report presents data on the housing characteristics of the 1975 housing inventory from the 1975 Annual Housing Survey SMSA sample. The AHS-SMSA estimation procedure employed a three-stage ratio estimation process. A detailed description of this ratio estimation procedure can be found in the AHS Series H-170 reports for 1975.

**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 1, **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 1, *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 the 1970 Census of Population and Housing and the 1975 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 the extent to which occupancy status was erroneously reported. 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 Reinterview*.

**AHS-SMSA**—For the 1978 AHS-SMSA sample, a reinterview program was not conducted. However, a study was conducted for the 1975 AHS-SMSA sample. The results of which are presented in the Census Bureau memorandum, "Reinterview Results for Annual Housing Survey--SMSA Sample: 1975."

**Coverage errors**—In 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. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units. 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, in which case, they would not have been 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 correct deficiencies 3-6 (see the coverage improvement section of this appendix) was not very effective in 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 1975 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing. It should be noted that since these ED's were recanvassed for the 1978 survey, the number of missed housing units may be considerably less for 1978.

**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 statistics 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 questionnaires, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The sampling error of a survey estimate provides a measure of the variation among the estimates from all 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.

One common measure of the sampling error is the standard error. As calculated for this report, the standard error reflects the variation in the estimates due to sampling and nonsampling errors, but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors measured by the standard error, biases, and any 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 item. Standard errors applicable to estimates of characteristics of the 1975 housing inventory can be found in the AHS Series H-170 reports for 1975.

Table I presents the standard errors applicable to estimates of characteristics of the 1978 housing inventory as well as estimates of characteristics of the 1975-1978 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 36,090 for the total SMSA, 18,010 for the central city of the SMSA, and 31,100 for the balance of the SMSA.

The reliability of an estimated percentage, computed by using the 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 1978 housing inventory as well as estimated percentages of the 1975-1978 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 1978 Housing Inventory and for Estimated Number of 1975-1978 Lost Units for the Philadelphia, Pa.-N.J., SMSA, for the Central City of the SMSA and for the Balance (Not in Central City) of the SMSA**

(68 chances out of 100)

Size of estimate	Standard error <sup>1</sup>		
	SMSA	In central city	Not in central city
0	150	110	180
100	150	110	180
200	170	150	190
500	270	240	300
700	320	280	360
1,000	380	340	430
2,500	600	530	670
5,000	860	750	950
10,000	1,210	1,060	1,340
25,000	1,910	1,650	2,120
50,000	2,690	2,300	2,980
75,000	3,280	2,780	3,630
100,000	3,780	3,150	4,170
150,000	4,600	3,730	5,050
200,000	5,270	4,150	5,760
250,000	5,860	4,460	6,370
300,000	6,370	4,680	6,890
400,000	7,260	4,880	7,760
500,000	8,010	4,810	8,450
600,000	8,640	4,450	9,000
700,000	9,200	3,710	9,430
800,000	9,680	—	9,770
900,000	10,110	—	10,020
1,000,000	10,490	—	10,190
1,100,000	10,810	—	10,280
1,200,000	11,100	—	10,290
1,300,000	11,350	—	—
1,400,000	11,560	—	—
1,500,000	11,730	—	—
1,750,000	12,020	—	—

<sup>1</sup> For estimates pertaining to new construction, the standard errors shown in the table should be multiplied by a factor of 1.2 for the total SMSA, 1.2 for the central city, and 1.1 for the balance (not in central city) estimates. For estimates pertaining to lost units the standard errors shown should be multiplied by a factor of 1.2 for the total SMSA, 1.0 for the central city, and 1.2 for the balance (not in central city) estimates.

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

where:  $x$  = the numerator of the ratio  
 $y$  = the denominator of the ratio  
 $\sigma_x$  = the standard error of the numerator  
 $\sigma_y$  = 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 1978 there were 1,097,900 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 10,800. 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
1,000,000 .....	10,490
1,097,900 .....	x
1,100,000 .....	10,810

The entry for "x" is determined as follows by vertically interpolating between 10,490 and 10,810:

$$\begin{aligned}
 1,097,900 - 1,000,000 &= 97,900 \\
 1,100,000 - 1,000,000 &= 100,000 \\
 10,490 + \frac{97,900}{100,000} (10,810 - 10,490) &= 10,800
 \end{aligned}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 1,087,100 to 1,108,700 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1978 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 1,080,620 to 1,115,180 housing units with 90 percent confidence; and that the average estimate lies within the interval from 1,076,300 to 1,119,500 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 1,097,900 owner-occupied housing units, 168,200 or 15.3 percent, had two bedrooms. Interpolation in table II of this appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 15.3 percent is approximately 0.4 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	15.3	25 or 75
1,000,000 .....	0.4	a	0.5
1,097,900 .....		p	
1,100,000 .....	0.3	b	0.5

1. The entry for cell "a" is determined by horizontal interpolation between 0.4 and 0.5.

$$\begin{aligned}
 15.3 - 10.0 &= 5.3 \\
 25.0 - 10.0 &= 15.0 \\
 0.4 + \frac{5.3}{15.0} (0.5 - 0.4) &= 0.4
 \end{aligned}$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.3 and 0.5.

$$\begin{aligned}
 15.3 - 10.0 &= 5.3 \\
 25.0 - 10.0 &= 15.0 \\
 0.3 + \frac{5.3}{15.0} (0.5 - 0.3) &= 0.4
 \end{aligned}$$

3. The entry for "p" was then determined by vertical interpolation between 0.4 and 0.4.

$$\begin{aligned}
 1,097,900 - 1,000,000 &= 97,900 \\
 1,100,000 - 1,000,000 &= 100,000 \\
 0.4 + \frac{97,900}{100,000} (0.4 - 0.4) &= 0.4
 \end{aligned}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 14.9 to 15.7 percent; the 90-percent confidence interval is from 14.7 to 15.9 percent; and the 95-percent confidence interval is from 14.5 to 16.1 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 characteristics 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. Due to the overlap of the 1975 and 1978 AHS-SMSA samples a positive correlation should be expected when making comparisons between the 1975 and 1978 characteristics.

*Illustration of the computation of the standard error of a difference*—Table A-1 of part A of this report shows that in 1978 there were 614,100 owner-occupied housing units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied housing units with two bedrooms and owner-occupied housing units with three bedrooms is 445,900. Table I shows the standard error of 614,100 is approximately 8,720, and the standard error of 168,200 is approximately 4,840. Therefore, the standard error of the estimated difference of 445,900 is about.

$$9,970 = \sqrt{(8,720)^2 + (4,840)^2}$$

Consequently, the 68-percent confidence interval for the 445,900 difference is from 435,930 to 455,870 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 429,948 to 461,852 housing units, and the 95-percent confidence interval is from 425,960 to 465,840. Thus, we can conclude with 95 percent confidence that the number of 1978 owner-occupied housing units with three

**TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1978 Housing Inventory and for Estimated Percentages of 1975-1978 Lost Units for the Philadelphia, Pa.-N.J., SMSA**

(68 chances out of 100)

Base of percentage	Estimated percentage <sup>1</sup>					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
200	42.3	42.3	42.3	42.3	42.3	42.8
500	22.6	22.6	22.6	22.6	23.4	27.1
700	17.3	17.3	17.3	17.3	19.8	22.9
1,000	12.8	12.8	12.8	12.8	16.6	19.1
2,500	5.5	5.5	5.5	7.3	10.5	12.1
5,000	2.8	2.8	3.7	5.1	7.4	8.6
10,000	1.4	1.4	2.6	3.6	5.2	6.0
25,000	0.6	0.8	1.7	2.3	3.3	3.8
50,000	0.3	0.5	1.2	1.6	2.3	2.7
75,000	0.2	0.4	1.0	1.3	1.9	2.2
100,000	0.15	0.4	0.8	1.1	1.7	1.9
150,000	0.10	0.3	0.7	0.9	1.4	1.6
200,000	0.07	0.3	0.6	0.8	1.2	1.4
250,000	0.06	0.2	0.5	0.7	1.0	1.2
300,000	0.05	0.2	0.5	0.7	1.0	1.1
400,000	0.04	0.2	0.4	0.6	0.8	1.0
500,000	0.03	0.2	0.4	0.5	0.7	0.9
600,000	0.02	0.2	0.3	0.5	0.7	0.8
700,000	0.02	0.14	0.3	0.4	0.6	0.7
800,000	0.02	0.13	0.3	0.4	0.6	0.7
900,000	0.02	0.13	0.3	0.4	0.6	0.6
1,000,000	0.01	0.12	0.3	0.4	0.5	0.6
1,100,000	0.01	0.11	0.3	0.3	0.5	0.6
1,200,000	0.01	0.11	0.2	0.3	0.5	0.6
1,300,000	0.01	0.11	0.2	0.3	0.5	0.5
1,400,000	0.01	0.10	0.2	0.3	0.4	0.5
1,500,000	0.01	0.10	0.2	0.3	0.4	0.5
1,750,000	0.01	0.09	0.2	0.3	0.4	0.5

<sup>1</sup> 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 errors shown in the table should be multiplied by a factor of 1.2. For estimates pertaining to lost units, the standard errors shown in the table should be multiplied by a factor of 1.2.

bedrooms is greater than the number of owner-occupied housing 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 characteristics, read off the confidence interval corresponding to the two points established in step 2.

**TABLE III. Standard Errors for Estimated Percentages of Housing Units in the 1978 Housing Inventory and for Estimated Percentages of 1975-1978 Lost Housing Units for the Central City of the SMSA**

(68 chances out of 100)

Base of percentage	Estimated percentage <sup>1</sup>					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
200	36.1	36.1	36.1	36.1	36.1	37.5
500	18.4	18.4	18.4	18.4	20.6	23.7
700	13.9	13.9	13.9	13.9	17.4	20.1
1,000	10.1	10.1	10.1	10.1	14.5	16.8
2,500	4.3	4.3	4.6	6.4	9.2	10.6
5,000	2.2	2.2	3.3	4.5	6.5	7.5
10,000	1.1	1.1	2.3	3.2	4.6	5.3
25,000	0.4	0.7	1.5	2.0	2.9	3.4
50,000	0.2	0.5	1.0	1.4	2.1	2.4
75,000	0.15	0.4	0.8	1.2	1.7	1.9
100,000	0.11	0.3	0.7	1.0	1.5	1.7
150,000	0.08	0.3	0.6	0.8	1.2	1.4
200,000	0.06	0.2	0.5	0.7	1.0	1.2
250,000	0.05	0.2	0.5	0.6	0.9	1.1
300,000	0.04	0.2	0.4	0.6	0.8	1.0
400,000	0.03	0.2	0.4	0.5	0.7	0.8
500,000	0.02	0.15	0.3	0.5	0.7	0.8
600,000	0.02	0.14	0.3	0.4	0.6	0.7
700,000	0.02	0.13	0.3	0.4	0.5	0.6

<sup>1</sup> 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 errors shown in the table should be multiplied by a factor of 1.2.

**TABLE IV. Standard Errors for Estimated Percentages of Housing Units in the 1978 Housing Inventory and for Estimated Percentages of 1975-1978 Lost Housing Units for the Balance (Not in Central City) of the SMSA**

(68 chances out of 100)

Base of percentage	Estimated percentage <sup>1</sup>					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
200.....	47.6	47.6	47.6	47.6	47.6	47.6
500.....	26.6	26.6	26.6	26.6	26.6	30.1
700.....	20.6	20.6	20.6	20.6	22.1	25.5
1,000.....	15.4	15.4	15.4	15.4	18.5	21.3
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	9.5
10,000.....	1.8	1.8	2.9	4.0	5.8	6.7
25,000.....	0.7	0.8	1.9	2.6	3.7	4.3
50,000.....	0.4	0.6	1.3	1.8	2.6	3.0
75,000.....	0.2	0.5	1.1	1.5	2.1	2.5
100,000.....	0.2	0.4	0.9	1.3	1.8	2.1
150,000.....	0.12	0.3	0.8	1.0	1.5	1.7
200,000.....	0.09	0.3	0.7	0.9	1.3	1.5
250,000.....	0.07	0.3	0.6	0.8	1.2	1.3
300,000.....	0.06	0.2	0.5	0.7	1.1	1.2
400,000.....	0.05	0.2	0.5	0.6	0.9	1.1
500,000.....	0.04	0.2	0.4	0.6	0.8	1.0
600,000.....	0.03	0.2	0.4	0.5	0.8	0.9
700,000.....	0.03	0.2	0.4	0.5	0.7	0.8
800,000.....	0.02	0.15	0.3	0.5	0.7	0.8
900,000.....	0.02	0.14	0.3	0.4	0.6	0.7
1,000,000.....	0.02	0.13	0.3	0.4	0.6	0.7
1,100,000.....	0.02	0.13	0.3	0.4	0.6	0.6
1,200,000.....	0.02	0.12	0.3	0.4	0.5	0.6

<sup>1</sup> 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 errors shown in the table should be multiplied by a factor of 1.1. For estimates pertaining to lost units, the standard errors shown in the table should be multiplied by a factor of 1.2.

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 1,097,900 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 1,097,900 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 453,800 owner-occupied housing units, or 41.3 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 210,000 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:

$$2.5 + (3.5 - 2.5) \left( \frac{48.8 - 41.3}{19.1} \right) = 2.9$$

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

$$2.5 + (3.5 - 2.5) \left( \frac{51.2 - 41.3}{19.1} \right) = 3.0$$

Thus, the 95-percent confidence interval ranges from 2.9 to 3.0 persons.

Riverside-San Bernardino-Ontario 1978

# Appendix B

## Source and Reliability of the Estimates

<b>SAMPLE DESIGN</b> . . . . .	App-40	Coverage improvement for deficiencies 3-6 . . . . .	App-42	<b>AHS-SMSA</b> . . . . .	App-44
Annual Housing Survey . . . . .	App-40	1970 Census of Population and Housing . . . . .	App-43	Coverage errors . . . . .	App-44
Designation of sample housing units for the 1978 survey . . . . .	App-40	<b>ESTIMATION</b> . . . . .	App-43	Rounding errors . . . . .	App-44
Selection of the 1975 AHS-SMSA sample . . . . .	App-40	1978 housing inventory . . . . .	App-43	Sampling errors for the AHS-SMSA sample . . . . .	App-44
1975-1978 additions to the housing inventory . . . . .	App-42	1975-1978 lost units . . . . .	App-43	Illustration of the use of the standard error tables . . . . .	App-46
Sample selection for the 1978 Coverage Improvement Program . . . . .	App-42	1975 estimation procedure . . . . .	App-43	Differences . . . . .	App-47
Coverage improvement for deficiency 1 . . . . .	App-42	Ratio estimation procedure of the 1970 Census of Population and Housing . . . . .	App-43	Illustration of the computation of the standard error of a difference . . . . .	App-47
Coverage improvement for deficiency 2 . . . . .	App-42	<b>RELIABILITY OF THE ESTIMATES</b> . . . . .	App-44	Medians . . . . .	App-47
		Nonsampling errors . . . . .	App-44	Illustration of the computation of the 95-percent confidence interval of a median . . . . .	App-47
		1970 census . . . . .	App-44		

### SAMPLE DESIGN

**Annual Housing Survey**—The estimates for each of the 15 SMSA's are based on data collected from the 1978 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 15 SMSA's, the data were collected for the 12-month period from April 1978 through March 1979 with one-twelfth of the sample units being visited each month.

Each group of SMSA's is being interviewed for the AHS on a rotating basis with this group (AA-1 SMSA's) being the second to be revisited. All of these SMSA's were enumerated for the first time in 1975.

For each group of SMSA's, the largest SMSA from 3 of the 4 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 the remaining 12 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 AA-1 group (1978-1979) are: Atlanta, Ga.; Philadelphia, Pa.-N.J.; and San Francisco-Oakland, Calif.

The remaining SMSA's in the AA-1 group are: Cincinnati, Ohio-Ky.-Ind.; Colorado Springs, Colo.; Columbus, Ohio; Kansas City, Mo.-Kans.; New Orleans, La.; Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; Rochester, N.Y.; San Antonio, Tex.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In this SMSA, 5,828 housing units were eligible for interview. Of these sample units, 308 interviews were not obtained because, for occupied sample units, the occupants refused to be interviewed, 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 units eligible for interview, 570 units were visited but were not eligible for interview because they were condemned, unfit, demolished, converted to group quarters use, etc.

**Designation of sample housing units for the 1978 survey**—The sample housing units designated to be interviewed in the 1978 survey consisted of the following categories which are described in detail in the succeeding sections.

1. All sample housing units that were interviewed in the 1975 survey.
2. All sample housing units that were type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of the survey but which could become eligible in the future) in the 1975 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1978 AHS questionnaire, page App-20.)
3. All sample housing units that were selected from the list of new construction building permits issued since the 1975 survey. (This sample represented the housing units built in permit-issuing areas, since the 1975 survey.)
4. All sample housing units that were added to sample segments in the nonpermit universe since the 1975 survey. (This sample represented additions to the housing inventory in nonpermit-issuing areas since the 1975 survey.)
5. All sample housing units that were selected as part of the 1978 Coverage Improvement Program. (This sample represented most of the housing units which, until 1978, did not have a chance of selection.)

**Selection of the 1975 AHS-SMSA sample**—The sample for the SMSA's which are 100-percent permit-issuing 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 those 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 non-permit universe). The following SMSA's are 100-percent permit-issuing: Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and San Francisco-Oakland, Calif. The remaining 10 SMSA's contain a sample from 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 three largest SMSA's, the overall sampling rate differed by central city and the 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 the 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 the total housing units in each sector.

The major portion of the sample in each SMSA 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 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 to \$5,999 . .		
\$6,000 to \$9,999 . .		
\$10,000 to \$14,999 .		
\$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from the permit-issuing 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 selection procedure was then instituted that would produce one-half of the desired sample size. However, whenever a record was selected to be in sample, the housing unit record adjacent to it on the file 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 by a procedure that produced one-quarter of the desired sample size. However, at the time of the survey, the units at each of the special places were listed and subsampled at a rate which produced 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 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:

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

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. All units in existence at the time of interview in these selected segments are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

**1975-1978 additions to the housing inventory**—In the permit-issuing universe, a sample of new construction building permits, issued since the 1975 survey, was selected to represent housing units built in permit-issuing areas since the 1975 survey. Sampling procedures were identical to those used in selecting the 1970-1975 new construction sample, which were described previously. In the nonpermit universe, sample segments were dependently recanvassed, using listing sheets from 1975, to identify any units missed in the 1975 survey or any units added since the 1975 survey.

**Sample selection for the 1978 Coverage Improvement Program**—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 1970 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 for each of the AA-1 SMSA's except Newport News-Hampton, Va. Two different procedures were used. For the first procedure, the sampling was carried out in two stages for one- and two-unit structures and in three stages for three-or-more-unit structures. Sample units selected from one- and two-unit structures were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units), while sample units selected from three-or-more-unit structures were sampled at one-half the rate of 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. Since these permits were not available for Newport News-Hampton, Va., SMSA, these coverage improvement procedures were not instituted for this SMSA. For the third stage, structures of size three or more were divided into clusters of an expected size of two units and a sample of clusters was selected. This procedure was employed in the following SMSA's: Colorado Springs, Colo.; Rochester, N.Y.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In the remaining 11 SMSA's, units whose permits were issued before January 1970, but which were completed after April 1970, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. These units were then sampled at one-third the rate of regular AHS units. These procedures added an estimated

910 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 parks that were missed by the census or established after the census was selected in two stages. First, for each 1978-1979 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 14,755 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 now contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and have a utility hookup, or were on the site during the present survey 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 AHS sample units was selected from the permit-issuing universe at a rate of 1 in 24. 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 13,683 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 AHS units in multiunit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multiunit structures selected above, all units were listed and matched to the 1970 census. Any missed units

were then assigned for interview. This procedure added an estimated 946 units to the coverage of the housing inventory of this SMSA.

**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 1978 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1975 (i.e., 1975-1978 lost units). Each type of estimate employed separate, although similar, estimation procedures.

**1978 housing inventory**—The AHS estimates of characteristics of the 1978 housing inventory employed a 1-stage ratio estimation procedure in all SMSA's. 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 noninterviews previously mentioned. This noninterview adjustment was done separately for occupied and vacant housing units. 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 described). In addition, within each sector separate noninterview factors were computed for one noninterview cell for conventional new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes from both the nonpermit universe and the coverage improvement universe, and one noninterview cell for other sample units from both the nonpermit universe and the coverage improvement universe, (if units were not included in any of the previous cells).

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.

**1975-1978 lost units**—The 1975-1978 lost unit estimates employed the three-stage ratio estimation procedure used to produce the AHS-SMSA estimates of the 1975 housing inventory, as was described in the 1975 Current Housing Report, Series H-170, *Housing Characteristics for Selected Metropolitan Areas*. Since the 1975-1978 lost units existed, by definition, in the 1975 housing inventory, there was a 1975 housing inventory weight associated with each 1975-1978 lost unit. This weight was used to tabulate the estimates of the characteristics of the 1975-1978 lost units.

**1975 estimation procedure**—This report presents data on the housing characteristics of the 1975 housing inventory from the 1975 Annual Housing Survey SMSA sample. The AHS-SMSA estimation procedure employed a three-stage ratio estimation process. A detailed description of this ratio estimation procedure can be found in the AHS Series H-170 reports for 1975.

**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 the 1970 Census of Population and Housing and the 1975 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 the extent to which occupancy status was erroneously reported. 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 1978 AHS-SMSA sample, a reinterview program was not conducted. However, a study was conducted for the 1975 AHS-SMSA sample. The results of which are presented in the Census Bureau memorandum, "Reinterview Results for Annual Housing Survey—SMSA Sample: 1975."

**Coverage errors**—In 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. However, these permits issued during the last 5

months of the survey do not necessarily represent missed housing units. 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, in which case, they would not have been 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 correct deficiencies 3-6 (see the coverage improvement section of this appendix) was not very effective in 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 1975 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing. It should be noted that since these ED's were recanvassed for the 1978 survey, the number of missed housing units may be considerably less for 1978.

**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 statistics 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 questionnaires, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The sampling error of a survey estimate provides a measure of the variation among the estimates from all 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.

One common measure of the sampling error is the standard error. As calculated for this report, the standard error reflects the variation in the estimates due to sampling and nonsampling errors, but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors measured by the standard error, biases, and any 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 item. Standard errors applicable to estimates of characteristics of the 1975 housing inventory can be found in the AHS Series H-170 reports for 1975.

Table I presents the standard errors applicable to estimates of characteristics of the 1978 housing inventory as well as estimates of characteristics of the 1975-1978 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 20,110 for the total SMSA, 9,710 for the central cities of the SMSA, and 17,520 for the balance of the SMSA.

The reliability of an estimated percentage, computed by using the 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 1978 housing inventory as well as estimated percentages of the 1975-1978 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

**TABLE I. Standard Errors for Estimated Number of Housing Units in the 1978 Housing Inventory and for Estimated Number of 1975-1978 Lost Units for the San Bernardino-Riverside-Ontario, Calif., SMSA, for the Central Cities, and for the Balance (Not in Central Cities) of the SMSA**

(68 chances out of 100)

Size of estimate	Standard error <sup>1</sup>		
	SMSA	In central cities	Not in central cities
0	110	120	110
100	110	120	110
200	150	150	150
500	240	240	240
700	280	280	280
1,000	340	340	340
2,500	540	540	530
5,000	760	750	750
10,000	1,070	1,060	1,060
25,000	1,670	1,620	1,660
50,000	2,330	2,150	2,310
75,000	2,820	2,460	2,770
100,000	3,210	2,620	3,130
150,000	3,810	2,600	3,670
200,000	4,260	—	4,040
250,000	4,600	—	4,290
300,000	4,850	—	4,420
400,000	5,160	—	4,420
500,000	5,210	—	4,020
600,000	5,030	—	—

<sup>1</sup>For estimates pertaining to new construction, the standard errors shown in the table should be multiplied by a factor of 1.2 for the total SMSA, 1.1 for the central cities, and 1.2 for the balance (not in central cities) estimates.

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  
 $\sigma_x$  = the standard error of the numerator  
 $\sigma_y$  = 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 1978 there were 318,800 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 4,910. 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 .....	4,850
318,800 .....	x
400,000 .....	5,160

The entry for "x" is determined as follows by vertically interpolating between 4,850 and 5,160:

$$\begin{aligned}
 318,800 - 300,000 &= 18,800 \\
 400,000 - 300,000 &= 100,000 \\
 4,850 + \frac{18,800}{100,000} (5,160 - 4,850) &= 4,910
 \end{aligned}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 313,890 to 323,710 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1978 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 310,944 to 326,656 housing units with 90 percent confidence; and that the average estimate lies within the interval from 308,980 to 328,620 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 318,800 owner-occupied housing units, 99,600, or 31.2 percent, had two bedrooms. Interpolation in table II of this appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 31.2 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		
	25 or 75	31.2	50
300,000 .....	0.9	a	1.0
318,800 .....		p	
400,000 .....	0.7	b	0.9

**TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1978 Housing Inventory and for Estimated Percentages of 1975-1978 Lost Units for the San Bernardino-Riverside-Ontario, Calif., SMSA, for the Central Cities, and for the Balance (Not in Central Cities) of the SMSA**

(68 chances out of 100)

Base of percentage	Estimated percentage <sup>1</sup>					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
200 .....	36.8	36.8	36.8	36.8	36.8	38.1
500 .....	18.9	18.9	18.9	18.9	20.9	24.1
700 .....	14.2	14.2	14.2	14.2	17.6	20.4
1,000 .....	10.4	10.4	10.4	10.4	14.8	17.0
2,500 .....	4.4	4.4	4.7	6.5	9.3	10.8
5,000 .....	2.3	2.3	3.3	4.6	6.6	7.6
10,000 .....	1.1	1.1	2.3	3.2	4.7	5.4
25,000 .....	0.5	0.7	1.5	2.0	3.0	3.4
50,000 .....	0.2	0.5	1.1	1.4	2.1	2.4
75,000 .....	0.15	0.4	0.9	1.2	1.7	2.0
100,000 .....	0.12	0.3	0.7	1.0	1.5	1.7
150,000 .....	0.08	0.3	0.6	0.8	1.2	1.4
200,000 .....	0.06	0.2	0.5	0.7	1.0	1.2
250,000 .....	0.05	0.2	0.5	0.6	0.9	1.1
300,000 .....	0.04	0.2	0.4	0.6	0.9	1.0
400,000 .....	0.03	0.2	0.4	0.5	0.7	0.9
500,000 .....	0.02	0.15	0.3	0.5	0.7	0.8
600,000 .....	0.02	0.14	0.3	0.4	0.6	0.7

<sup>1</sup>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 errors shown in the table should be multiplied by a factor of 1.2 for the total SMSA, 1.1 for the central cities, and 1.2 for the balance (not in central cities).

1. The entry for cell "a" is determined by horizontal interpolation between 0.9 and 1.0.

$$\begin{aligned}
 31.2 - 25.0 &= 6.2 \\
 50.0 - 25.0 &= 25.0
 \end{aligned}$$

$$0.9 + \frac{6.2}{25.0} (1.0 - 0.9) = 0.9$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.7 and 0.9.

$$\begin{aligned}
 31.2 - 25.0 &= 6.2 \\
 50.0 - 25.0 &= 25.0
 \end{aligned}$$

$$0.7 + \frac{6.2}{25.0} (0.9 - 0.7) = 0.7$$

3. The entry for "p" was then determined by vertical interpolation between 0.9 and 0.7.

$$\begin{aligned} 318,800 - 300,000 &= 18,800 \\ 400,000 - 300,000 &= 100,000 \\ 0.9 + \frac{18,800}{100,000} (0.7 - 0.9) &= 0.9 \end{aligned}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 30.3 to 32.1 percent; the 90-percent confidence interval is from 29.8 to 32.6 percent; and the 95-percent confidence interval is from 29.4 to 33.0 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 characteristics 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. Due to the overlap of the 1975 and 1978 AHS-SMSA samples a positive correlation should be expected when making comparisons between the 1975 and 1978 characteristics.

*Illustration of the computation of the standard error of a difference*—Table A-1 of part A of this report shows that in 1978 there were 139,800 owner-occupied housing units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied housing units with two bedrooms and owner-occupied housing units with three bedrooms is 40,200. Table I shows the standard error of 139,800 is approximately 3,690, and the standard error of 99,600 is approximately 3,200. Therefore, the standard error of the estimated difference of 40,200 is about:

$$4,880 = \sqrt{(3,690)^2 + (3,200)^2}$$

Consequently, the 68-percent confidence interval for the 40,200 difference is from 35,320 to 45,080 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 32,392 to 48,008 housing units, and the 95-percent confidence interval is from 30,440 to 49,960. Thus, we can conclude with 95 percent confidence that the number of 1978 owner-occupied housing units with three bedrooms is greater than the number of owner-occupied housing 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 characteristics, 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 318,800 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 318,800 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-1 of part A, it can be seen that 48,700 owner-occupied housing units, or 15.3 percent, had one person (for purposes of calculating the median, the category of one person is considered to be from 0.5 to 1.5 persons); that an additional 114,200 owner-occupied housing units, or 35.8 percent, had two persons (i.e., 1.5 to 2.5 persons); and that 54,200, 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:

$$1.5 + (2.5 - 1.5) \left( \frac{48.0 - 15.3}{35.8} \right) = 2.4$$

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

$$2.5 + (3.5 - 2.5) \left( \frac{52.0 - 51.1}{17.0} \right) = 2.6$$

Thus, the 95-percent confidence interval ranges from 2.3 to 2.6 persons.

Rochester 1978

# Appendix B

## Source and Reliability of the Estimates

<b>SAMPLE DESIGN</b> .....	App-40	<b>Coverage improvement for</b>	<b>AHS-SMSA</b> .....	App-44
Annual Housing Survey .....	App-40	deficiencies 3-6 .....	Coverage errors .....	App-44
Designation of sample housing		1970 Census of Population and	Rounding errors .....	App-44
units for the 1978 survey .....	App-40	Housing .....	Sampling errors for the	
Selection of the 1975		<b>ESTIMATION</b> .....	AHS-SMSA sample .....	App-44
AHS-SMSA sample .....	App-40	1978 housing inventory .....	Illustration of the use of the	
1975-1978 additions to the		1975-1978 lost units .....	standard error tables .....	App-45
housing inventory .....	App-41	1975 estimation procedure .....	Differences .....	App-46
Sample selection for the 1978		Ratio estimation procedure of the	Illustration of the computa-	
Coverage Improvement Program	App-42	1970 Census of Population and	tion of the standard error of	
Coverage improvement for		Housing .....	a difference .....	App-46
deficiency 1 .....	App-42	<b>RELIABILITY OF THE ESTIMATES</b>	Medians .....	App-47
Coverage improvement for		Nonsampling errors .....	Illustration of the computa-	
deficiency 2 .....	App-42	1970 census .....	tion of the 95-percent con-	
			fidence interval of a median	App-47

### SAMPLE DESIGN

**Annual Housing Survey**—The estimates for each of the 15 SMSA's are based on data collected from the 1978 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 15 SMSA's, the data were collected for the 12-month period from April 1978 through March 1979 with one-twelfth of the sample units being visited each month.

Each group of SMSA's is being interviewed for the AHS on a rotating basis with this group (AA-1 SMSA's) being the second to be revisited. All of these SMSA's were enumerated for the first time in 1975.

For each group of SMSA's, the largest SMSA from 3 of the 4 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 the remaining 12 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 AA-1 group (1978-1979) are: Atlanta, Ga.; Philadelphia, Pa.-N.J.; and San Francisco-Oakland, Calif.

The remaining SMSA's in the AA-1 group are: Cincinnati, Ohio-Ky.-Ind.; Colorado Springs, Colo.; Columbus, Ohio; Kansas City, Mo.-Kans.; New Orleans, La.; Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; Rochester, N.Y.; San Antonio, Tex.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In this SMSA, 4,909 housing units were eligible for interview. Of these sample units, 216 interviews were not obtained because, for occupied sample units, the occupants refused to be interviewed, 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 units eligible for interview, 425 units were visited but were not eligible for interview because they were condemned, unfit, demolished, converted to group quarters use, etc.

**Designation of sample housing units for the 1978 survey**—The sample housing units designated to be interviewed in the 1978 survey consisted of the following categories which are described in detail in the succeeding sections.

1. All sample housing units that were interviewed in the 1975 survey.
2. All sample housing units that were type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of the survey but which could become eligible in the future) in the 1975 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1978 AHS questionnaire, page App-20.)
3. All sample housing units that were selected from the list of new construction building permits issued since the 1975 survey. (This sample represented the housing units built in permit-issuing areas, since the 1975 survey.)
4. All sample housing units that were added to sample segments in the nonpermit universe since the 1975 survey. (This sample represented additions to the housing inventory in nonpermit-issuing areas since the 1975 survey.)
5. All sample housing units that were selected as part of the 1978 Coverage Improvement Program. (This sample represented most of the housing units which, until 1978, did not have a chance of selection.)

**Selection of the 1975 AHS-SMSA sample**—The sample for the SMSA's which are 100-percent permit-issuing 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 those 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 non-permit universe). The following SMSA's are 100-percent permit-issuing: Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and San Francisco-Oakland, Calif. The remaining 10 SMSA's contain a sample from 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 three largest SMSA's, the overall sampling rate differed by central city and the 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 the 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 the total housing units in each sector.

The major portion of the sample in each SMSA 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 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 to \$5,999 . . . . .		
\$6,000 to \$9,999 . . . . .		
\$10,000 to \$14,999 . . . . .		
\$15,000 and over . . . . .		

Thus, for this SMSA, the occupied housing unit records from the permit-issuing 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 selection procedure was then instituted that would produce one-half of the desired sample size. However, whenever a record was selected to be in sample, the housing unit record adjacent to it on the file 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 by a procedure that produced one-quarter of the desired sample size. However, at the time of the survey, the units at each of the special places were listed and subsampled at a rate which produced 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 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:

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

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. All units in existence at the time of interview in these selected segments are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

**1975-1978 additions to the housing inventory**—In the permit-issuing universe, a sample of new construction building permits,

issued since the 1975 survey, was selected to represent housing units built in permit-issuing areas since the 1975 survey. Sampling procedures were identical to those used in selecting the 1970-1975 new construction sample, which were described previously. In the nonpermit universe, sample segments were dependently recanvassed, using listing sheets from 1975, to identify any units missed in the 1975 survey or any units added since the 1975 survey.

**Sample selection for the 1978 Coverage Improvement Program**—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 1970 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 for each of the AA-1 SMSA's except Newport News-Hampton, Va. Two different procedures were used. For the first procedure, the sampling was carried out in two stages for one- and two-unit structures and in three stages for three-or-more-unit structures. Sample units selected from one- and two-unit structures were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units), while sample units selected from three-or-more-unit structures were sampled at one-half the rate of 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. Since these permits were not available for Newport News-Hampton, Va., SMSA, these coverage improvement procedures were not instituted for this SMSA. For the third stage, structures of size three or more were divided into clusters of an expected size of two units and a sample of clusters was selected. This procedure was employed in the following SMSA's: Colorado Springs, Colo.; Rochester, N.Y.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In the remaining 11 SMSA's, units whose permits were issued before January 1970, but which were completed after April 1970, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. These units were then sampled at one-third the rate of regular AHS units. These procedures added an estimated 8,406 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 parks that were missed by the census or established after the census was selected in two stages. First, for each 1978-1979 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 792 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 now contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and have a utility hookup, or were on the site during the present survey 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 AHS sample units was selected from the permit-issuing universe at a rate of 1 in 24. 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 5,455 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 AHS units in multiunit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multiunit structures selected above, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 2,080 units to the coverage of the housing inventory of this SMSA.

**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 1978 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1975 (i.e., 1975-1978 lost units). Each type of estimate employed separate, although similar, estimation procedures.

**1978 housing inventory**—The AHS estimates of characteristics of the 1978 housing inventory employed a 1-stage ratio estimation procedure in all SMSA's. 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 noninterviews previously mentioned. This noninterview adjustment was done separately for occupied and vacant housing units. 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 described). In addition, within each sector separate noninterview factors were computed for one noninterview cell for conventional new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes from both the nonpermit universe and the coverage improvement universe, and one noninterview cell for other sample units from both the nonpermit universe and the coverage improvement universe, (if units were not included in any of the previous cells).

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.

**1975-1978 lost units**—The 1975-1978 lost unit estimates employed the three-stage ratio estimation procedure used to produce the AHS-SMSA estimates of the 1975 housing inventory, as was described in the 1975 Current Housing Report, Series H-170, **Housing Characteristics for Selected Metropolitan Areas**. Since the 1975-1978 lost units existed, by definition, in the 1975 housing inventory, there was a 1975 housing inventory weight associated with each 1975-1978 lost unit. This weight was used to tabulate the estimates of the characteristics of the 1975-1978 lost units.

**1975 estimation procedure**—This report presents data on the housing characteristics of the 1975 housing inventory from the 1975 Annual Housing Survey SMSA sample. The AHS-SMSA estimation procedure employed a three-stage ratio estimation process. A detailed description of this ratio estimation procedure can be found in the AHS Series H-170 reports for 1975.

**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 the 1970 Census of Population and Housing and the 1975 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 the extent to which occupancy status was erroneously reported. 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 1978 AHS-SMSA sample, a reinterview program was not conducted. However, a study was conducted for the 1975 AHS-SMSA sample. The results of which are presented in the Census Bureau memorandum, "Reinterview Results for Annual Housing Survey—SMSA Sample: 1975."

**Coverage errors**—In 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. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units. 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, in which case, they would not have been 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 correct deficiencies 3-6 (see the coverage improvement section of this appendix) was not very effective in 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 1975 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing. It should be noted that since these ED's were recanvassed for the 1978 survey, the number of missed housing units may be considerably less for 1978.

**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 statistics 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 questionnaires, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The sampling error of a survey estimate provides a measure of the variation among the estimates from all 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.

One common measure of the sampling error is the standard error. As calculated for this report, the standard error reflects the variation in the estimates due to sampling and nonsampling errors, but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors measured by the standard error, biases, and any 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 item. Standard errors applicable to estimates of characteristics of the 1975 housing inventory can be found in the AHS Series H-170 reports for 1975.

Table I presents the standard errors applicable to estimates of characteristics of the 1978 housing inventory as well as estimates of characteristics of the 1975-1978 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 11,560 for the total SMSA, 5,900 for the central city of the SMSA, and 9,860 for the balance of the SMSA.

The reliability of an estimated percentage, computed by using the 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 1978 housing inventory as well as estimated percentages of the 1975-1978 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  
 $\sigma_x$  = the standard error of the numerator  
 $\sigma_y$  = 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 1978 there were 200,800 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 3,130. 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 .....	3,130
200,800 .....	x
250,000 .....	3,230

The entry for "x" is determined as follows by vertically interpolating between 3,130 and 3,230:

$$\begin{aligned} 200,800 - 200,000 &= 800 \\ 250,000 - 200,000 &= 50,000 \\ 3,130 + \frac{800}{50,000} (3,230 - 3,130) &= 3,130 \end{aligned}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 197,670 to 203,930 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1978 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 195,792 to 205,808 housing units with 90 percent confidence; and that the average estimate lies within the interval from 194,540 to 207,060 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 200,800 owner-occupied housing units, 35,200, or 17.5 percent, had two bedrooms. Interpolation in table II of this appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 17.5 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	17.5	25 or 75
200,000 . . . . .	0.6	a	0.9
200,800 . . . . .		p	
250,000 . . . . .	0.5	b	0.8

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

$$17.5 - 10.0 = 7.5$$

$$25.0 - 10.0 = 15.0$$

$$0.6 + \frac{7.5}{15.0} (0.9 - 0.6) = 0.8$$

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

$$17.5 - 10.0 = 7.5$$

$$25.0 - 10.0 = 15.0$$

$$0.5 + \frac{7.5}{15.0} (0.8 - 0.5) = 0.7$$

3. The entry for "p" was then determined by vertical interpolation between 0.8 and 0.7.

$$200,800 - 200,000 = 800$$

$$250,000 - 200,000 = 50,000$$

$$0.8 + \frac{800}{50,000} (0.7 - 0.8) = 0.8$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 16.7 to 18.3 percent; the 90-percent confidence interval is from 16.2 to 18.8 percent; and the 95-percent confidence interval is from 15.9 to 19.1 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 characteristics 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. Due to the overlap of the 1975 and 1978 AHS-SMSA samples a positive correlation should be expected when making comparisons between the 1975 and 1978 characteristics.

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1978 Housing Inventory and for Estimated Number of 1975-1978 Lost Units for the Rochester, N.Y., SMSA, for the Central City, and for the Balance (Not in Central City) of the SMSA

(68 chances out of 100)

Size of estimate	Standard error <sup>1</sup>		
	SMSA	In central city	Not in central city
0 . . . . .	80	70	70
100 . . . . .	90	90	90
200 . . . . .	130	120	120
500 . . . . .	200	190	190
700 . . . . .	230	230	230
1,000 . . . . .	280	270	270
2,500 . . . . .	440	420	430
5,000 . . . . .	630	600	610
10,000 . . . . .	880	830	850
25,000 . . . . .	1,370	1,250	1,320
50,000 . . . . .	1,890	1,620	1,810
75,000 . . . . .	2,250	1,770	2,140
100,000 . . . . .	2,530	1,780	2,380
150,000 . . . . .	2,910	1,270	2,680
200,000 . . . . .	3,130	—	2,800
250,000 . . . . .	3,230	—	2,750
300,000 . . . . .	3,200	—	—
400,000 . . . . .	2,780	—	—

<sup>1</sup> For estimates pertaining to new construction, the standard errors shown in the table should be multiplied by a factor of 1.2 for the total SMSA, 1.4 for the central city, and 1.1 for the balance (not in central city) estimates.

*Illustration of the computation of standard error of a difference*—Table A-1 of part A of this report shows that in 1978 there were 100,800 owner-occupied housing units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied housing units with two bedrooms and owner-occupied housing units with three bedrooms is 65,600. Table I shows the standard error of 100,800 is approximately 2,540, and the standard error of 35,200 is approximately 1,580. Therefore, the standard error of the estimated difference of 65,600 is about:

$$2,990 = \sqrt{(2,540)^2 + (1,580)^2}$$

Consequently, the 68-percent confidence interval for the 65,600 difference is from 62,610 to 68,590 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 60,816 to 70,384 housing units, and the 95-percent confidence interval is from 59,620 to 71,580. Thus, we can conclude with 95 percent confidence that the number of 1978 owner-occupied housing units with three bedrooms is greater than the number of owner-occupied housing 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 characteristics, 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.

**TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1978 Housing Inventory and for Estimated Percentages of 1975-1978 Lost Units for the Rochester, N.Y., SMSA, for the Central City, and for the Balance (Not in Central City) of the SMSA**

(68 chances out of 100)

Base of percentage	Estimated percentage <sup>1</sup>					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100. ....	44.1	44.1	44.1	44.1	44.1	44.4
200. ....	28.3	28.3	28.3	28.3	28.3	31.4
500. ....	13.6	13.6	13.6	13.6	17.2	19.9
700. ....	10.1	10.1	10.1	10.1	14.5	16.8
1,000. ....	7.3	7.3	7.3	8.4	12.2	14.0
2,500. ....	3.1	3.1	3.9	5.3	7.7	8.9
5,000. ....	1.6	1.6	2.7	3.8	5.4	6.3
10,000. ....	0.8	0.9	1.9	2.7	3.8	4.4
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	1.1
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
400,000. ....	0.02	0.14	0.3	0.4	0.6	0.7

<sup>1</sup> 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 the new construction, the standard errors shown in the table should be multiplied by a factor of 1.2 for the total SMSA, 1.4 for the central city, and 1.1 for the balance (not in central city).

*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 200,800 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 200,800 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-1 of part A, it can be seen by cumulating the frequencies for the first two categories that 81,400 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 37,900 owner-occupied housing units, or 18.9 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:

$$2.5 + (3.5 - 2.5) \left( \frac{48.0 - 40.5}{18.9} \right) = 2.9$$

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

$$2.5 + (3.5 - 2.5) \left( \frac{52.0 - 40.5}{18.9} \right) = 3.1$$

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

San Antonio 1978

# Appendix B

## Source and Reliability of the Estimates

<b>SAMPLE DESIGN</b> .....	App-40	<b>Coverage improvement for deficiencies 3-6</b> .....	App-42	<b>AHS-SMSA</b> .....	App-44
Annual Housing Survey .....	App-40	1970 Census of Population and Housing .....	App-43	Coverage errors .....	App-44
Designation of sample housing units for the 1978 survey .....	App-40	<b>ESTIMATION</b> .....	App-43	Rounding errors .....	App-44
Selection of the 1975 AHS-SMSA sample .....	App-40	1978 housing inventory .....	App-43	Sampling errors for the AHS-SMSA sample .....	App-44
1975-1978 additions to the housing inventory .....	App-42	1975-1978 lost units .....	App-43	Illustration of the use of the standard error tables .....	App-45
Sample selection for the 1978 Coverage Improvement Program .....	App-42	1975 estimation procedure .....	App-43	Differences .....	App-47
Coverage improvement for deficiency 1 .....	App-42	Ratio estimation procedure of the 1970 Census of Population and Housing .....	App-43	Illustration of the computation of the standard error of a difference .....	App-47
Coverage improvement for deficiency 2 .....	App-42	<b>RELIABILITY OF THE ESTIMATES</b> .....	App-44	Medians .....	App-47
		Nonsampling errors .....	App-44	Illustration of the computation of the 95-percent confidence interval of a median .....	App-47
		1970 census .....	App-44		

### SAMPLE DESIGN

**Annual Housing Survey**—The estimates for each of the 15 SMSA's are based on data collected from the 1978 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 15 SMSA's, the data were collected for the 12-month period from April 1978 through March 1979 with one-twelfth of the sample units being visited each month.

Each group of SMSA's is being interviewed for the AHS on a rotating basis with this group (AA-1 SMSA's) being the second to be revisited. All of these SMSA's were enumerated for the first time in 1975.

For each group of SMSA's, the largest SMSA from 3 of the 4 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 the remaining 12 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 AA-1 group (1978-1979) are Atlanta, Ga.; Philadelphia, Pa.-N.J.; and San Francisco-Oakland, Calif.

The remaining SMSA's in the AA-1 group are Cincinnati, Ohio-Ky.-Ind.; Colorado Springs, Colo.; Columbus, Ohio; Kansas City, Mo.-Kans.; New Orleans, La.; Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; Rochester, N.Y.; San Antonio, Tex.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In this SMSA, 5,130 housing units were eligible for interview. Of these sample units, 193 interviews were not obtained because, for occupied sample units, the occupants refused to be interviewed, 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 units eligible for interview, 349 units were visited but were not eligible for interview because they were condemned, unfit, demolished, converted to group quarters use, etc.

**Designation of sample housing units for the 1978 survey**—The sample housing units designated to be interviewed in the 1978 survey consisted of the following categories which are described in detail in the succeeding sections.

1. All sample housing units that were interviewed in the 1975 survey.
2. All sample housing units that were type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of the survey but which could become eligible in the future) in the 1975 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1978 AHS questionnaire, page App-20.)
3. All sample housing units that were selected from the list of new construction building permits issued since the 1975 survey. (This sample represented the housing units built in permit-issuing areas, since the 1975 survey.)
4. All sample housing units that were added to sample segments in the nonpermit universe since the 1975 survey. (This sample represented additions to the housing inventory in nonpermit-issuing areas since the 1975 survey.)
5. All sample housing units that were selected as part of the 1978 Coverage Improvement Program. (This sample represented most of the housing units which, until 1978, did not have a chance of selection.)

**Selection of the 1975 AHS-SMSA sample**—The sample for the SMSA's which are 100-percent permit-issuing 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 those 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 non-permit universe). The following SMSA's are 100-percent permit-issuing: Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and San Francisco-Oakland, Calif. The remaining 10 SMSA's contain a sample from 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 three largest SMSA's, the overall sampling rate differed by central city and the 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 the 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 the total housing units in each sector.

The major portion of the sample in each SMSA 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 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 to \$5,999 . . . . .		
\$6,000 to \$9,999 . . . . .		
\$10,000 to \$14,999 . . . . .		
\$15,000 and over . . . . .		

Thus, for this SMSA, the occupied housing unit records from the permit-issuing 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 selection procedure was then instituted that would produce one-half of the desired sample size. However, whenever a record was selected to be in sample, the housing unit record adjacent to it on the file 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 by a procedure that produced one-quarter of the desired sample size. However, at the time of the survey, the units at each of the special places were listed and subsampled at a rate which produced 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 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:

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

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. All units in existence at the time of interview in these selected segments are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

**1975-1978 additions to the housing inventory**—In the permit-issuing universe, a sample of new construction building permits, issued since the 1975 survey, was selected to represent housing units built in permit-issuing areas since the 1975 survey. Sampling procedures were identical to those used in selecting the 1970-1975 new construction sample, which were described previously. In the nonpermit universe, sample segments were dependently recanvassed, using listing sheets from 1975, to identify any units missed in the 1975 survey or any units added since the 1975 survey.

**Sample selection for the 1978 Coverage Improvement Program**—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 1970 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 for each of the AA-1 SMSA's except Newport News-Hampton, Va. Two different procedures were used. For the first procedure, the sampling was carried out in two stages for one- and two-unit structures and in three stages for three-or-more-unit structures. Sample units selected from one- and two-unit structures were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units), while sample units selected from three-or-more-unit structures were sampled at one-half the rate of 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. Since these permits were not available for Newport News-Hampton, Va., SMSA, these coverage improvement procedures were not instituted for this SMSA. For the third stage, structures of size three or more were divided into clusters of an expected size of two units and a sample of clusters was selected. This procedure was employed in the following SMSA's: Colorado Springs, Colo.; Rochester, N.Y.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In the remaining 11 SMSA's, units whose permits were issued before January 1970, but which were completed after April 1970, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. These units were then sampled at one-third the rate

of regular AHS units. These procedures added an estimated 1,633 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 parks that were missed by the census or established after the census was selected in two stages. First, for each 1978-1979 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 242 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 now contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and have a utility hookup, or were on the site during the present survey 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 AHS sample units was selected from the permit-issuing universe at a rate of 1 in 24. 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 2,690 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 AHS units in multiunit structures of less than 10 units was selected from the permit-issuing universe.

Second, for the multiunit structures selected above, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated zero units to the coverage of the housing inventory of this SMSA.

**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 1978 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1975 (i.e., 1975-1978 lost units). Each type of estimate employed separate, although similar, estimation procedures.

**1978 housing inventory**—The AHS estimates of characteristics of the 1978 housing inventory employed a 1-stage ratio estimation procedure in all SMSA's. 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 noninterviews previously mentioned. This noninterview adjustment was done separately for occupied and vacant housing units. 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 described). In addition, within each sector separate noninterview factors were computed for one noninterview cell for conventional new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes from both the nonpermit universe and the coverage improvement universe, and one noninterview cell for other sample units from both the nonpermit universe and the coverage improvement universe, (if units were not included in any of the previous cells).

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.

**1975-1978 lost units**—The 1975-1978 lost unit estimates employed the three-stage ratio estimation procedure used to produce the AHS-SMSA estimates of the 1975 housing inventory, as was described in the 1975 Current Housing Report, Series H-170, *Housing Characteristics for Selected Metropolitan Areas*. Since the 1975-1978 lost units existed, by definition, in the 1975 housing inventory, there was a 1975 housing inventory weight associated with each 1975-1978 lost unit. This weight was used to tabulate the estimates of the characteristics of the 1975-1978 lost units.

**1975 estimation procedure**—This report presents data on the housing characteristics of the 1975 housing inventory from the 1975 Annual Housing Survey SMSA sample. The AHS-SMSA estimation procedure employed a three-stage ratio estimation process. A detailed description of this ratio estimation procedure can be found in the AHS Series H-170 reports for 1975.

**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 the 1970 Census of Population and Housing and the 1975 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 the extent to which occupancy status was erroneously reported. 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 1978 AHS-SMSA sample, a reinterview program was not conducted. However, a study was conducted for the 1975 AHS-SMSA sample. The results of which are presented in the Census Bureau memorandum, "Reinterview Results for Annual Housing Survey—SMSA Sample: 1975."

**Coverage errors**—In 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. However, these permits issued during the last 5

months of the survey do not necessarily represent missed housing units. 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, in which case, they would not have been 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 correct deficiencies 3-6 (see the coverage improvement section of this appendix) was not very effective in 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 1975 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing. It should be noted that since these ED's were recanvassed for the 1978 survey, the number of missed housing units may be considerably less for 1978.

**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 statistics 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 questionnaires, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The sampling error of a survey estimate provides a measure of the variation among the estimates from all 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.

One common measure of the sampling error is the standard error. As calculated for this report, the standard error reflects the variation in the estimates due to sampling and nonsampling errors, but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors measured by the standard error, biases, and any 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 item. Standard errors applicable to estimates of characteristics of the 1975 housing inventory can be found in the AHS Series H-170 reports for 1975.

Table I presents the standard errors applicable to estimates of characteristics of the 1978 housing inventory as well as estimates of characteristics of the 1975-1978 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 22,030 for the total SMSA, 8,790 for the central city of the SMSA, and 20,030 for the balance of the SMSA.

The reliability of an estimated percentage, computed by using the 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 1978 housing inventory as well as estimated percentages of the 1975-1978 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

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1978 Housing Inventory and for Estimated Number of 1975-1978 Lost Units for the San Antonio, Tex., SMSA, for the Central City, and for the Balance (Not in Central City) of the SMSA

(68 chances out of 100)

Size of estimate	Standard error <sup>1</sup>		
	SMSA	In central city	Not in central city
0	90	80	70
100	90	90	80
200	130	120	120
500	210	200	190
700	240	230	220
1,000	290	280	270
2,500	460	440	450
5,000	660	620	690
10,000	930	870	1,120
25,000	1,510	1,340	2,290
50,000	2,190	1,810	4,200
75,000	2,760	2,120	6,090
100,000	3,270	2,320	7,980
150,000	4,210	2,520	—
200,000	5,080	2,470	—
250,000	5,920	2,160	—
300,000	6,740	—	—
400,000	8,330	—	—

<sup>1</sup> For estimates pertaining to new construction, the standard errors shown in the table should be multiplied by a factor of 1.3 for the total SMSA, 1.2 for the central city, and 1.3 for the balance (not in central city) estimates.

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  
 $\sigma_x$  = the standard error of the numerator  
 $\sigma_y$  = 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 1978 there were 189,200 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 4,890. 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 .....	4,210
189,200 .....	x
200,000 .....	5,080

The entry for "x" is determined as follows by vertically interpolating between 4,210 and 5,080:

$$189,200 - 150,000 = 39,200$$

$$200,000 - 150,000 = 50,000$$

$$4,210 + \frac{39,200}{50,000} (5,080 - 4,210) = 4,890$$

**TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1978 Housing Inventory and for Estimated Percentages of 1975-1978 Lost Units for the San Antonio, Tex., SMSA, for the Central City, and for the Balance (Not in Central City) of the SMSA**

(68 chances out of 100)

Base of percentage	Estimated percentage <sup>1</sup>					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100.....	46.0	46.0	46.0	46.0	46.0	46.1
200.....	29.9	29.9	29.9	29.9	29.9	32.6
500.....	14.5	14.5	14.5	14.5	17.9	20.6
700.....	10.8	10.8	10.8	10.8	15.1	17.4
1,000.....	7.8	7.8	7.8	8.8	12.6	14.6
2,500.....	3.3	3.3	4.0	5.5	8.0	9.2
5,000.....	1.7	1.7	2.8	3.9	5.6	6.5
10,000.....	0.8	0.9	2.0	2.8	4.0	4.6
25,000.....	0.3	0.6	1.3	1.8	2.5	2.9
50,000.....	0.2	0.4	0.9	1.2	1.8	2.1
75,000.....	0.11	0.3	0.7	1.0	1.5	1.7
100,000.....	0.09	0.3	0.6	0.9	1.3	1.5
150,000.....	0.06	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	0.6	0.8	0.9
300,000.....	0.03	0.2	0.4	0.5	0.7	0.8
400,000.....	0.02	0.15	0.3	0.4	0.6	0.7

<sup>1</sup> 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 errors shown in the table should be multiplied by a factor of 1.3 for the total SMSA, 1.2 for the central city, and 1.3 for the balance (not in central city) estimates.

Consequently, the 68-percent confidence interval, as shown by these data, is from 184,310 to 194,090 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1978 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 181,376 to 197,024 housing units with 90 percent confidence; and that the average estimate lies within the interval from 179,420 to 198,980 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 189,200 owner-occupied housing units, 52,100, or 27.5 percent, had two bedrooms. Interpolation in table II of this appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 27.5 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	27.5	50
150,000 .....	1.0	a	1.2
189,200 .....		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.

$$27.5 - 25.0 = 2.5$$

$$50.0 - 25.0 = 25.0$$

$$1.0 + \frac{2.5}{25.0} (1.2 - 1.0) = 1.0$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.9 and 1.0.

$$27.5 - 25.0 = 2.5$$

$$50.0 - 25.0 = 25.0$$

$$0.9 + \frac{2.5}{25.0} (1.0 - 0.9) = 0.9$$

3. The entry for "p" was then determined by vertical interpolation between 1.0 and 0.9.

$$189,200 - 150,000 = 39,200$$

$$200,000 - 150,000 = 50,000$$

$$1.0 + \frac{39,200}{50,000} (0.9 - 1.0) = 0.9$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 26.6 to 28.4 percent; the 90-percent confidence interval is from 26.1 to 28.9 percent; and the 95-percent confidence interval is from 25.7 to 29.3 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 characteristics 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. Due to the overlap of the 1975 and 1978 AHS-SMSA samples a positive correlation should be expected when making comparisons between the 1975 and 1978 characteristics.

*Illustration of the computation of the standard error of a difference*—Table A-1 of part A of this report shows that in 1978 there were 103,900 owner-occupied housing units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied housing units with two bedrooms and owner-occupied housing units with three bedrooms is 51,800. Table I shows the standard error of 103,900 is approximately 3,340, and the standard error of 52,100 is approximately 2,240. Therefore, the standard error of the estimated difference of 51,800 is about:

$$4,020 = \sqrt{(3,340)^2 + (2,240)^2}$$

Consequently, the 68-percent confidence interval for the 51,800 difference is from 47,780 to 55,820 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 45,368 to 58,232 housing units, and the 95-percent confidence interval is from 43,760 to 59,840. Thus, we can conclude with 95 percent confidence that the number of 1978 owner-occupied housing units with three bedrooms is greater than the number of owner-occupied housing 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 characteristics, 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 189,200 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 189,200 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 82,100 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 34,500 owner-occupied housing units, or 18.2 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:

$$2.5 + (3.5 - 2.5) \left( \frac{48.0 - 43.4}{18.2} \right) = 2.8$$

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

$$2.5 + (3.5 - 2.5) \left( \frac{52.0 - 43.4}{18.2} \right) = 3.0$$

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

San Diego 1978

# Appendix B

## Source and Reliability of the Estimates

<b>SAMPLE DESIGN</b> .....	App-40	Coverage improvement for deficiencies 3-6 .....	App-42	AHS-SMSA .....	App-44
Annual Housing Survey .....	App-40	1970 Census of Population and Housing .....	App-43	Coverage errors .....	App-44
Designation of sample housing units for the 1978 survey .....	App-40	<b>ESTIMATION</b> .....	App-43	Rounding errors .....	App-44
Selection of the 1975 AHS-SMSA sample .....	App-40	1978 housing inventory .....	App-43	Sampling errors for the AHS-SMSA sample .....	App-44
1975-1978 additions to the housing inventory .....	App-42	1975-1978 lost units .....	App-43	Illustration of the use of the standard error tables .....	App-46
Sample selection for the 1978 Coverage Improvement Program .....	App-42	1975 estimation procedure .....	App-43	Differences .....	App-47
Coverage improvement for deficiency 1 .....	App-42	Ratio estimation procedure of the 1970 Census of Population and Housing .....	App-43	Illustration of the computation of the standard error of a difference .....	App-47
Coverage improvement for deficiency 2 .....	App-42	<b>RELIABILITY OF THE ESTIMATES.</b> .....	App-44	Medians .....	App-47
		Nonsampling errors .....	App-44	Illustration of the computation of the 95-percent confidence interval of a median .....	App-47
		1970 census .....	App-44		

### SAMPLE DESIGN

**Annual Housing Survey**—The estimates for each of the 15 SMSA's are based on data collected from the 1978 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 15 SMSA's, the data were collected for the 12-month period from April 1978 through March 1979 with one-twelfth of the sample units being visited each month.

Each group of SMSA's is being interviewed for the AHS on a rotating basis with this group (AA-1 SMSA's) being the second to be revisited. All of these SMSA's were enumerated for the first time in 1975.

For each group of SMSA's, the largest SMSA from 3 of the 4 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 the remaining 12 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 AA-1 group (1978-1979) are: Atlanta, Ga.; Philadelphia, Pa.-N.J.; and San Francisco-Oakland, Calif.

The remaining SMSA's in the AA-1 group are: Cincinnati, Ohio-Ky.-Ind.; Colorado Springs, Colo.; Columbus, Ohio; Kansas City, Mo.-Kans.; New Orleans, La.; Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; Rochester, N.Y.; San Antonio, Tex.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In this SMSA, 5,424 housing units were eligible for interview. Of these sample units, 243 interviews were not obtained because, for occupied sample units, the occupants refused to be interviewed, 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 units eligible for interview, 417 units were visited but were not eligible for interview because they were condemned, unfit, demolished, converted to group quarters use, etc.

**Designation of sample housing units for the 1978 survey**—The sample housing units designated to be interviewed in the 1978 survey consisted of the following categories which are described in detail in the succeeding sections.

1. All sample housing units that were interviewed in the 1975 survey.
2. All sample housing units that were type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of the survey but which could become eligible in the future) in the 1975 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1978 AHS questionnaire, page App-20.)
3. All sample housing units that were selected from the list of new construction building permits issued since the 1975 survey. (This sample represented the housing units built in permit-issuing areas, since the 1975 survey.)
4. All sample housing units that were added to sample segments in the nonpermit universe since the 1975 survey. (This sample represented additions to the housing inventory in nonpermit-issuing areas since the 1975 survey.)
5. All sample housing units that were selected as part of the 1978 Coverage Improvement Program. (This sample represented most of the housing units which, until 1978, did not have a chance of selection.)

**Selection of the 1975 AHS-SMSA sample**—The sample for the SMSA's which are 100-percent permit-issuing 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 those 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 non-permit universe). The following SMSA's are 100-percent permit-issuing: Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and San Francisco-Oakland, Calif. The remaining 10 SMSA's contain a sample from 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 three largest SMSA's, the overall sampling rate differed by central city and the 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 the 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 the total housing units in each sector.

The major portion of the sample in each SMSA 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 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 to \$5,999 . .		
\$6,000 to \$9,999 . .		
\$10,000 to \$14,999 .		
\$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from the permit-issuing 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 selection procedure was then instituted that would produce one-half of the desired sample size. However, whenever a record was selected to be in sample, the housing unit record adjacent to it on the file 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 by a procedure that produced one-quarter of the desired sample size. However, at the time of the survey, the units at each of the special places were listed and subsampled at a rate which produced 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 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:

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

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. All units in existence at the time of interview in these selected segments are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

**1975-1978 additions to the housing inventory**—In the permit-issuing universe, a sample of new construction building permits, issued since the 1975 survey, was selected to represent housing units built in permit-issuing areas since the 1975 survey. Sampling procedures were identical to those used in selecting the 1970-1975 new construction sample, which were described previously. In the nonpermit universe, sample segments were dependently recanvassed, using listing sheets from 1975, to identify any units missed in the 1975 survey or any units added since the 1975 survey.

**Sample selection for the 1978 Coverage Improvement Program**—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 1970 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 for each of the AA-1 SMSA's except Newport News-Hampton, Va. Two different procedures were used. For the first procedure, the sampling was carried out in two stages for one- and two-unit structures and in three stages for three-or-more-unit structures. Sample units selected from one- and two-unit structures were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units), while sample units selected from three-or-more-unit structures were sampled at one-half the rate of 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. Since these permits were not available for Newport News-Hampton, Va., SMSA, these coverage improvement procedures were not instituted for this SMSA. For the third stage, structures of size three or more were divided into clusters of an expected size of two units and a sample of clusters was selected. This procedure was employed in the following SMSA's: Colorado Springs, Colo.; Rochester, N.Y.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In the remaining 11 SMSA's, units whose permits were issued before January 1970, but which were completed after April 1970, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. These units were then sampled at one-third the rate of regular AHS units. These procedures added an estimated

9,178 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 parks that were missed by the census or established after the census was selected in two stages. First, for each 1978-1979 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 10,201 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 now contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and have a utility hookup, or were on the site during the present survey 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 AHS sample units was selected from the permit-issuing universe at a rate of 1 in 24. 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 13,738 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 AHS units in multiunit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multiunit structures selected above, all units were listed and matched to the 1970 census. Any missed units

were then assigned for interview. This procedure added an estimated 402 units to the coverage of the housing inventory of this SMSA.

**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 1, **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 1978 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1975 (i.e., 1975-1978 lost units). Each type of estimate employed separate, although similar, estimation procedures.

**1978 housing inventory**—The AHS estimates of characteristics of the 1978 housing inventory employed a 1-stage ratio estimation procedure in all SMSA's. 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 noninterviews previously mentioned. This noninterview adjustment was done separately for occupied and vacant housing units. 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 described). In addition, within each sector separate noninterview factors were computed for one noninterview cell for conventional new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes from both the nonpermit universe and the coverage improvement universe, and one noninterview cell for other sample units from both the nonpermit universe and the coverage improvement universe, (if units were not included in any of the previous cells).

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.

**1975-1978 lost units**—The 1975-1978 lost unit estimates employed the three-stage ratio estimation procedure used to produce the AHS-SMSA estimates of the 1975 housing inventory, as was described in the 1975 Current Housing Report, Series H-170, **Housing Characteristics for Selected Metropolitan Areas**. Since the 1975-1978 lost units existed, by definition, in the 1975 housing inventory, there was a 1975 housing inventory weight associated with each 1975-1978 lost unit. This weight was used to tabulate the estimates of the characteristics of the 1975-1978 lost units.

**1975 estimation procedure**—This report presents data on the housing characteristics of the 1975 housing inventory from the 1975 Annual Housing Survey SMSA sample. The AHS-SMSA estimation procedure employed a three-stage ratio estimation process. A detailed description of this ratio estimation procedure can be found in the AHS Series H-170 reports for 1975.

**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 1, **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 1, *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 the 1970 Census of Population and Housing and the 1975 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 the extent to which occupancy status was erroneously reported. 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 1978 AHS-SMSA sample, a reinterview program was not conducted. However, a study was conducted for the 1975 AHS-SMSA sample. The results of which are presented in the Census Bureau memorandum, "Reinterview Results for Annual Housing Survey—SMSA Sample: 1975."

**Coverage errors**—In 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. However, these permits issued during the last 5 months of the survey do not necessarily represent missed

housing units. 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, in which case, they would not have been 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 correct deficiencies 3-6 (see the coverage improvement section of this appendix) was not very effective in 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 1975 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing. It should be noted that since these ED's were recanvassed for the 1978 survey, the number of missed housing units may be considerably less for 1978.

**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 statistics 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 questionnaires, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The sampling error of a survey estimate provides a measure of the variation among the estimates from all 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.

One common measure of the sampling error is the standard error. As calculated for this report, the standard error reflects the variation in the estimates due to sampling and nonsampling errors, but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors measured by the standard error, biases, and any 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 item. Standard errors applicable to estimates of characteristics of the 1975 housing inventory can be found in the AHS Series H-170 reports for 1975.

Table I presents the standard errors applicable to estimates of characteristics of the 1978 housing inventory as well as estimates of characteristics of the 1975-1978 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 25,800 for the total SMSA, 19,080 for the central city of the SMSA, and 17,360 for the balance of the SMSA.

The reliability of an estimated percentage, computed by using the 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 1978 housing inventory as well as estimated percentages of the 1975-1978 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

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1978 Housing Inventory and for Estimated Number of 1975-1978 Lost Units for the San Diego, Calif., SMSA, for the Central City, and for the Balance (Not in Central City) of the SMSA

(68 chances out of 100)

Size of estimate	Standard error <sup>1</sup>		
	SMSA	In central city	Not in central city
0	150	140	140
100	150	140	140
200	170	170	170
500	270	260	270
700	320	310	320
1,000	380	370	380
2,500	610	580	600
5,000	860	830	840
10,000	1,210	1,160	1,190
25,000	1,910	1,830	1,880
50,000	2,680	2,560	2,640
75,000	3,270	3,100	3,220
100,000	3,750	3,530	3,690
150,000	4,550	4,220	4,470
200,000	5,190	4,750	5,100
250,000	5,740	5,160	5,630
300,000	6,210	5,490	6,080
400,000	6,990	—	6,830
500,000	7,620	—	—
600,000	8,120	—	—
700,000	8,510	—	—

<sup>1</sup>For estimates pertaining to new construction, the standard errors shown in the table should be multiplied by a factor of 1.2 for the total SMSA, 1.1 for the central city, and 1.2 for the balance (not in central city) estimates.

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  
 $\sigma_x$  = the standard error of the numerator  
 $\sigma_y$  = the standard error of the denominator

**TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1978 Housing Inventory and for Estimated Percentages of 1975-1978 Lost Units for the San Diego, Calif., SMSA, for the Central City, and for the Balance (Not in Central City) of the SMSA**

(68 chances out of 100)

Base of percentage	Estimated percentage <sup>1</sup>					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
200 . . . . .	42.4	42.4	42.4	42.4	42.4	42.9
500 . . . . .	22.7	22.7	22.7	22.7	23.5	27.1
700 . . . . .	17.4	17.4	17.4	17.4	19.9	22.9
1,000 . . . . .	12.8	12.8	12.8	12.8	16.6	19.2
2,500 . . . . .	5.6	5.6	5.6	7.3	10.5	12.1
5,000 . . . . .	2.9	2.9	3.7	5.1	7.4	8.6
10,000 . . . . .	1.5	1.5	2.6	3.6	5.3	6.1
25,000 . . . . .	0.6	0.8	1.7	2.3	3.3	3.8
50,000 . . . . .	0.3	0.5	1.2	1.6	2.3	2.7
75,000 . . . . .	0.2	0.4	1.0	1.3	1.9	2.2
100,000 . . . . .	0.15	0.4	0.8	1.2	1.7	1.9
150,000 . . . . .	0.10	0.3	0.7	0.9	1.4	1.6
200,000 . . . . .	0.07	0.3	0.6	0.8	1.2	1.4
250,000 . . . . .	0.06	0.2	0.5	0.7	1.1	1.2
300,000 . . . . .	0.05	0.2	0.5	0.7	1.0	1.1
400,000 . . . . .	0.04	0.2	0.4	0.6	0.8	1.0
500,000 . . . . .	0.03	0.2	0.4	0.5	0.7	0.9
600,000 . . . . .	0.02	0.2	0.3	0.5	0.7	0.8
700,000 . . . . .	0.02	0.14	0.3	0.4	0.6	0.7

<sup>1</sup>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 errors shown in the table should be multiplied by a factor of 1.2 for the total SMSA, 1.1 for the central city, and 1.2 for the balance (not in central city).

*Illustration of the use of the standard error tables*—Table A-1 of part A of this report shows that in 1978 there were 346,900 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 6,580. 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 . . . . .	6,210
346,900 . . . . .	x
400,000 . . . . .	6,990

The entry for "x" is determined as follows by vertically interpolating between 6,210 and 6,990:

$$\begin{aligned} 346,900 - 300,000 &= 46,900 \\ 400,000 - 300,000 &= 100,000 \end{aligned}$$

$$6,210 + \frac{46,900}{100,000} (6,990 - 6,210) = 6,580$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 340,320 to 353,480 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1978 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 336,372 to 357,428 housing units with 90 percent confidence; and that the average estimate lies within the interval from 333,740 to 360,060 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 346,900 owner-occupied housing units, 99,400, or 28.7 percent, had two bedrooms. Interpolation in table II of this appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 28.7 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	28.7	50
300,000 . . . . .	1.0	a	1.1
346,900 . . . . .		p	
400,000 . . . . .	0.8	b	1.0

1. The entry for cell "a" is determined by horizontal interpolation between 1.0 and 1.1.

$$\begin{aligned} 28.7 - 25.0 &= 3.7 \\ 50.0 - 25.0 &= 25.0 \end{aligned}$$

$$1.0 + \frac{3.7}{25.0} (1.1 - 1.0) = 1.0$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.8 and 1.0.

$$\begin{aligned} 28.7 - 25.0 &= 3.7 \\ 50.0 - 25.0 &= 25.0 \end{aligned}$$

$$0.8 + \frac{3.7}{25.0} (1.0 - 0.8) = 0.8$$

3. The entry for "p" was then determined by vertical interpolation between 1.0 and 0.8.

$$\begin{aligned} 346,900 - 300,000 &= 46,900 \\ 400,000 - 300,000 &= 100,000 \\ 1.0 + \frac{46,900}{100,000} (0.8 - 1.0) &= 0.9 \end{aligned}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 27.8 to 29.6 percent; the 90-percent confidence interval is from 27.3 to 30.1 percent; and the 95-percent confidence interval is from 26.9 to 30.5 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 characteristics 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. Due to the overlap of the 1975 and 1978 AHS-SMSA samples a positive correlation should be expected when making comparisons between the 1975 and 1978 characteristics.

*Illustration of the computation of the standard error of a difference*—Table A-1 of part A of this report shows that in 1978 there were 158,600 owner-occupied housing units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied housing units with two bedrooms and owner-occupied housing units with three bedrooms is 59,200. Table I shows the standard error of 158,600 is approximately 4,660, and the standard error of 99,400 is approximately 3,740. Therefore, the standard error of the estimated difference of 59,200 is about:

$$5,980 = \sqrt{(4,660)^2 + (3,740)^2}$$

Consequently, the 68-percent confidence interval for the 59,200 difference is from 53,220 to 65,180 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 49,632 to 68,768 housing units, and the 95-percent confidence interval is from 47,240 to 71,160. Thus, we can conclude with 95 percent confidence that the number of 1978 owner-occupied housing units with three bedrooms is greater than the number of owner-occupied housing 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 characteristics, 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 346,900 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 346,900 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 that 47,800 owner-occupied housing units, or 13.8 percent, had one person (for purposes of calculating the median, the category of one person is considered to be from 0.5 to 1.5 persons); that an additional 128,600 owner-occupied housing units, or 37.1 percent, had two persons (i.e., 1.5 to 2.5 persons); and that 58,900, 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:

$$1.5 + (2.5 - 1.5) \left( \frac{47.8 - 13.8}{37.1} \right) = 2.4$$

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

$$2.5 + (3.5 - 2.5) \left( \frac{52.2 - 50.9}{17.0} \right) = 2.6$$

Thus, the 95-percent confidence interval ranges from 2.4 to 2.6 persons.

San Francisco-Oakland 1978

# Appendix B

## Source and Reliability of the Estimates

<b>SAMPLE DESIGN</b> .....	App-40	<b>Coverage improvement for deficiencies 3-6</b> .....	App-42	<b>AHS-SMSA</b> .....	App-44
Annual Housing Survey .....	App-40	<b>1970 Census of Population and Housing</b> .....	App-42	<b>Coverage errors</b> .....	App-44
Designation of sample housing units for the 1978 survey .....	App-40	<b>ESTIMATION</b> .....	App-43	<b>Rounding errors</b> .....	App-44
Selection of the 1975 AHS-SMSA sample .....	App-40	<b>1978 housing inventory</b> .....	App-43	<b>Sampling errors for the AHS-SMSA sample</b> .....	App-44
1975-1978 additions to the housing inventory .....	App-41	<b>1975-1978 lost units</b> .....	App-43	<b>Illustration of the use of the standard error tables</b> .....	App-46
Sample selection for the 1978 Coverage Improvement Program .....	App-42	<b>1975 estimation procedure</b> .....	App-43	<b>Differences</b> .....	App-46
Coverage improvement for deficiency 1 .....	App-42	<b>Ratio estimation procedure of the 1970 Census of Population and Housing</b> .....	App-43	<b>Illustration of the computation of the standard error of a difference</b> .....	App-46
Coverage improvement for deficiency 2 .....	App-42	<b>RELIABILITY OF THE ESTIMATES</b> .....	App-43	<b>Medians</b> .....	App-47
		<b>Nonsampling errors</b> .....	App-44	<b>Illustration of the computation of the 95-percent confidence interval of a median</b> .....	App-48
		<b>1970 census</b> .....	App-44		

### SAMPLE DESIGN

**Annual Housing Survey**—The estimates for each of the 15 SMSA's are based on data collected from the 1978 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 15 SMSA's, the data were collected for the 12-month period from April 1978 through March 1979 with one-twelfth of the sample units being visited each month.

Each group of SMSA's is being interviewed for the AHS on a rotating basis with this group (AA-1 SMSA's) being the second to be revisited. All of these SMSA's were enumerated for the first time in 1975.

For each group of SMSA's, the largest SMSA from 3 of the 4 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 the remaining 12 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 AA-1 group (1978-1979) are: Atlanta, Ga.; Philadelphia, Pa.-N.J.; and San Francisco-Oakland, Calif.

The remaining SMSA's in the AA-1 group are: Cincinnati, Ohio-Ky.-Ind.; Colorado Springs, Colo.; Columbus, Ohio; Kansas City, Mo.-Kans.; New Orleans, La.; Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; Rochester, N.Y.; San Antonio, Tex.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In this SMSA, 14,261 housing units were eligible for interview. Of these sample units, 958 interviews were not obtained because, for occupied sample units, the occupants refused to be interviewed, 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 units eligible for interview, 950 units were visited but were not eligible for interview because they were condemned, unfit, demolished, converted to group quarters use, etc.

**Designation of sample housing units for the 1978 survey**—The sample housing units designated to be interviewed in the 1978 survey consisted of the following categories which are described in detail in the succeeding sections.

1. All sample housing units that were interviewed in the 1975 survey.
2. All sample housing units that were type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of the survey but which could become eligible in the future) in the 1975 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1978 AHS questionnaire, page App-20.)
3. All sample housing units that were selected from the list of new construction building permits issued since the 1975 survey. (This sample represented the housing units built in permit-issuing areas, since the 1975 survey.)
4. All sample housing units that were added to sample segments in the nonpermit universe since the 1975 survey. (This sample represented additions to the housing inventory in nonpermit-issuing areas since the 1975 survey.)
5. All sample housing units that were selected as part of the 1978 Coverage Improvement Program. (This sample represented most of the housing units which, until 1978, did not have a chance of selection.)

**Selection of the 1975 AHS-SMSA sample**—The sample for the SMSA's which are 100-percent permit-issuing 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 those 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 non-permit universe). The following SMSA's are 100-percent permit-issuing: Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and San Francisco-Oakland, Calif. The remaining 10 SMSA's contain a sample from 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 three 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 the total housing units in each sector.

The major portion of the sample in each SMSA 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 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 to \$5,999 . . . . .		
\$6,000 to \$9,999 . . . . .		
\$10,000 to \$14,999 . . . . .		
\$15,000 and over . . . . .		

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 selection procedure was then instituted that would produce one-half of the desired sample size. However, whenever a record was selected to be in sample, the housing unit record adjacent to it on the file 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 by a procedure that produced one-quarter of the desired sample size. However, at the time of the survey, the units at each of the special places were listed and subsampled at a rate which produced 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 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:

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

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. All units in existence at the time of interview in these selected segments are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

**1975-1978 additions to the housing inventory**—In the permit-issuing universe, a sample of new construction building permits, issued since the 1975 survey, was selected to represent housing

Thus, for this SMSA, the occupied housing unit records from the permit-issuing universe were assigned to 1 of 100 strata for

units built in permit-issuing areas since the 1975 survey. Sampling procedures were identical to those used in selecting the 1970-1975 new construction sample, which were described previously. In the nonpermit universe, sample segments were dependently recanvassed, using listing sheets from 1975, to identify any units missed in the 1975 survey or any units added since the 1975 survey.

**Sample selection for the 1978 Coverage Improvement Program**—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 1970 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 for each of the AA-1 SMSA's except Newport News-Hampton, Va. Two different procedures were used. For the first procedure, the sampling was carried out in two stages for one- and two-unit structures and in three stages for three-or-more-unit structures. Sample units selected from one- and two-unit structures were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units), while sample units selected from three-or-more-unit structures were sampled at one-half the rate of 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. Since these permits were not available for Newport News-Hampton, Va., SMSA, these coverage improvement procedures were not instituted for this SMSA. For the third stage, structures of size three or more were divided into clusters of an expected size of two units and a sample of clusters was selected. This procedure was employed in the following SMSA's: Colorado Springs, Colo.; Rochester, N.Y.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In the remaining 11 SMSA's, units whose permits were issued before January 1970, but which were completed after April 1970, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. These units were then sampled at one-third the rate of regular AHS units. These procedures added an estimated 8,534 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 that was missed by

the census or established after the census was selected in two stages. First, for each 1978-1979 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 9,060 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 now contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and have a utility hookup, or were on the site during the present survey 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 AHS sample units was selected from the permit-issuing universe at a rate of 1 in 24. 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 18,255 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 AHS units in multiunit structures of less than 10 units was selected from the permit-issuing universe. Second, for the multiunit structures selected above, all units were listed and matched to the 1970 census. Any missed units were then assigned for interview. This procedure added an estimated 1,489 units to the coverage of the housing inventory of this SMSA.

**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 1978 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1975 (i.e., 1975-1978 lost units). Each type of estimate employed separate, although similar, estimation procedures.

**1978 housing inventory**—The AHS estimates of characteristics of the 1978 housing inventory employed a 1-stage ratio estimation procedure in all SMSA's. 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 noninterviews previously mentioned. This noninterview adjustment was done separately for occupied and vacant housing units. 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 described). In addition, within each sector separate noninterview factors were computed for one noninterview cell for conventional new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes from both the nonpermit universe and the coverage improvement universe, and one noninterview cell for other sample units from both the nonpermit universe and the coverage improvement universe, (if units were not included in any of the previous cells).

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.

**1975-1978 lost units**—The 1975-1978 lost unit estimates employed the three-stage ratio estimation procedure used to produce the AHS-SMSA estimates of the 1975 housing inventory, as was described in the 1975 Current Housing Report, Series H-170, *Housing Characteristics for Selected Metropolitan Areas*. Since the 1975-1978 lost units existed, by definition, in the 1975 housing inventory, there was a 1975 housing inventory weight associated with each 1975-1978 lost unit. This weight was used to tabulate the estimates of the characteristics of the 1975-1978 lost units.

**1975 estimation procedure**—This report presents data on the housing characteristics of the 1975 housing inventory from the 1975 Annual Housing Survey SMSA sample. The AHS-SMSA estimation procedure employed a three-stage ratio estimation process. A detailed description of this ratio estimation procedure can be found in the AHS Series H-170 reports for 1975.

**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 1, **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 the 1970 Census of Population and Housing and the 1975 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 the extent to which occupancy status was erroneously reported. 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 Reinterview.**

**AHS-SMSA**—For the 1978 AHS-SMSA sample, a reinterview program was not conducted. However, a study was conducted for the 1975 AHS-SMSA sample. The results of which are presented in the Census Bureau memorandum, "Reinterview Results for Annual Housing Survey—SMSA Sample: 1975."

**Coverage errors**—In 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. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units. 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, in which case, they would not have been 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 correct deficiencies 3-6 (see the coverage improvement section of this appendix) was not very effective in 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 1975 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing. It should be noted that since these ED's were recanvassed for the 1978 survey, the number of missed housing units may be considerably less for 1978.

**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 statistics 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 sampling error of a survey estimate provides a measure of the variation among the estimates from all 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.

One common measure of the sampling error is the standard error. As calculated for this report, the standard error reflects the variation in the estimates due to sampling and nonsampling errors, but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors measured by the standard error, biases, and any 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 item. Standard errors applicable to estimates of characteristics of the 1975 housing inventory can be found in the AHS Series H-170 reports for 1975.

Table I presents the standard errors applicable to estimates of characteristics of the 1978 housing inventory as well as estimates of characteristics of the 1975-1978 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 25,720 for the total SMSA, 10,960 for the central city of the SMSA, and 23,200 for the balance of the SMSA.

The reliability of an estimated percentage, computed by using the 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 1978 housing inventory as well as estimated percentages of the 1975-1978 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 1978 Housing Inventory and for Estimated Number of 1975-1978 Lost Units for the San Francisco-Oakland, Calif., SMSA, for the Central City of the SMSA and for the Balance (Not in Central City) of the SMSA

(68 chances out of 100)

Size of estimate	Standard error <sup>1</sup>		
	SMSA	In central city	Not in central city
0	100	70	120
100	100	80	120
200	140	120	150
500	220	190	240
700	260	220	290
1,000	320	270	340
2,500	500	420	540
5,000	700	590	770
10,000	990	840	1,090
25,000	1,570	1,310	1,710
50,000	2,210	1,810	2,400
75,000	2,690	2,170	2,920
100,000	3,090	2,450	3,350
150,000	3,750	2,860	4,050
200,000	4,280	3,130	4,610
250,000	4,740	3,290	5,080
300,000	5,130	3,360	5,480
400,000	5,800	3,250	6,140
500,000	6,330	2,750	6,640
600,000	6,770	—	7,010
700,000	7,130	—	7,290
800,000	7,420	—	7,480
900,000	7,560	—	7,580
1,000,000	7,820	—	—
1,110,000	7,940	—	—
1,200,000	8,020	—	—
1,300,000	8,040	—	—
1,400,000	8,020	—	—

<sup>1</sup>For estimates pertaining to new construction the standard errors shown in the table should be multiplied by a factor of 1.2 for the total SMSA, 1.1 for the central city, and 1.1 for the balance (not in central city) estimates.

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

where:  $x$  = the numerator of the ratio  
 $y$  = the denominator of the ratio  
 $\sigma_x$  = the standard error of the numerator  
 $\sigma_y$  = 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 1978 there were 661,000 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 6,990. 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
600,000 .....	6,770
661,000 .....	x
700,000 .....	7,130

The entry for "x" is determined as follows by vertically interpolating between 6,770 and 7,130:

$$\begin{aligned}
 661,000 - 600,000 &= 61,000 \\
 700,000 - 600,000 &= 100,000 \\
 6,770 + \frac{61,000}{100,000} (7,130 - 6,770) &= 6,990
 \end{aligned}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 654,010 to 667,990 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1978 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 649,816 to 672,184 housing units with 90 percent confidence; and that the average estimate lies within the interval from 647,020 to 674,980 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 661,000 owner-occupied housing units, 188,400, or 28.5 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 28.5 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		
	25 or 75	28.5	50
600,000 .....	0.6	a	0.6
661,000 .....		p	
700,000 .....	0.5	b	0.6

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

$$\begin{aligned}
 28.5 - 25.0 &= 3.5 \\
 50.0 - 25.0 &= 25.0 \\
 0.6 + \frac{3.5}{25.0} (0.6 - 0.6) &= 0.6
 \end{aligned}$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.5 and 0.6.

$$\begin{aligned}
 28.5 - 25.0 &= 3.5 \\
 50.0 - 25.0 &= 25.0 \\
 0.5 + \frac{3.5}{25.0} (0.6 - 0.5) &= 0.5
 \end{aligned}$$

3. The entry for "p" was then determined by vertical interpolation between 0.6 and 0.5.

$$\begin{aligned}
 661,000 - 600,000 &= 61,000 \\
 700,000 - 600,000 &= 100,000 \\
 0.6 + \frac{61,000}{100,000} (0.5 - 0.6) &= 0.5
 \end{aligned}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 28.0 to 29.0 percent; the 90-percent confidence interval is from 27.7 to 29.3 percent; and the 95-percent confidence interval is from 27.5 to 29.5 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 characteristics 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. Due to the overlap of the 1975 and 1978 AHS-SMSA samples a positive correlation should be expected when making comparisons between 1975 and 1978 characteristics.

*Illustration of the computation of the standard error of a difference*—Table A-1 of part A of this report shows that in 1978 there were 304,500 owner-occupied housing units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied housing units with two bedrooms and owner-occupied housing units with three bedrooms is 116,100. Table I shows the standard error of 304,500 is approximately 5,160, and the standard error of 188,400 is approximately 4,160. Therefore, the standard error of the estimated difference of 116,100 is about:

$$6,630 = \sqrt{(5,160)^2 + (4,160)^2}$$

**TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1978 Housing Inventory and for Estimated Percentages of 1975-1978 Lost Units for the San Francisco-Oakland, Calif., SMSA**

(68 chances out of 100)

Base of percentage	Estimated percentage <sup>1</sup>					
	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.....	0.4	0.6	1.4	1.9	2.7	3.2
50,000.....	0.2	0.4	1.0	1.3	1.9	2.2
75,000.....	0.13	0.4	0.8	1.1	1.6	1.8
100,000.....	0.10	0.3	0.7	0.9	1.4	1.6
150,000.....	0.07	0.3	0.6	0.8	1.1	1.3
200,000.....	0.05	0.2	0.5	0.7	1.0	1.1
250,000.....	0.04	0.2	0.4	0.6	0.9	1.0
300,000.....	0.03	0.2	0.4	0.5	0.8	0.9
400,000.....	0.02	0.2	0.3	0.5	0.7	0.8
500,000.....	0.02	0.14	0.3	0.4	0.6	0.7
600,000.....	0.02	0.13	0.3	0.4	0.6	0.6
700,000.....	0.01	0.12	0.3	0.4	0.5	0.6
800,000.....	0.01	0.11	0.2	0.3	0.5	0.6
900,000.....	0.01	0.10	0.2	0.3	0.5	0.5
1,000,000.....	0.01	0.10	0.2	0.3	0.4	0.5
1,100,000.....	0.01	0.09	0.2	0.3	0.4	0.5
1,200,000.....	0.01	0.09	0.2	0.3	0.4	0.5
1,300,000.....	0.01	0.09	0.2	0.3	0.4	0.4
1,400,000.....	0.01	0.08	0.2	0.3	0.4	0.4

<sup>1</sup>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 errors shown in the table should be multiplied by a factor of 1.2.

Consequently, the 68-percent confidence interval for the 116,100 difference is from 109,470 to 122,730 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 105,492 to 126,708 housing units, and the 95-percent confidence interval is from 102,840 to 129,360. Thus, we can conclude with 95 percent confidence that the number of 1978 owner-occupied housing units with three bedrooms is greater than the number of owner-occupied housing units with two bedrooms since the 95-percent confidence interval does not include zero or negative values.

**TABLE III. Standard Errors for Estimated Percentages of Housing Units in the 1978 Housing Inventory and for Estimated Percentages of 1975-1978 Lost Housing Units for the Central City of the San Francisco-Oakland, Calif., SMSA**

(68 chances out of 100)

Base of percentage	Estimated percentage <sup>1</sup>					
	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.....	0.7	0.8	1.8	2.5	3.7	4.2
25,000.....	0.3	0.5	1.2	1.6	2.3	2.7
50,000.....	0.14	0.4	0.8	1.1	1.6	1.9
75,000.....	0.10	0.3	0.7	0.9	1.3	1.5
100,000.....	0.07	0.3	0.6	0.8	1.2	1.3
150,000.....	0.05	0.2	0.5	0.7	0.9	1.1
200,000.....	0.04	0.2	0.4	0.6	0.8	0.9
250,000.....	0.03	0.2	0.4	0.5	0.7	0.8
300,000.....	0.02	0.15	0.3	0.5	0.7	0.8
400,000.....	0.02	0.13	0.3	0.4	0.6	0.7
500,000.....	0.01	0.12	0.3	0.4	0.5	0.6

<sup>1</sup>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 errors shown in the table should be multiplied by a factor of 1.1.

**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 characteristics, 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 661,000 housing units.

1. Interpolation in table II shows that the standard error of 50 percent of a base of 661,000 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 that 102,000 owner-occupied housing units, or 15.4 percent, had one person (for purposes of calculating the median the category of one person is considered to be from 0.5 to 1.5 persons) and that an additional 222,200 owner-occupied housing units, or 35.8 percent, had two persons (i.e., 1.5 to 2.5 persons).

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

$$1.5 + (2.5-1.5) \left( \frac{48.8-15.4}{33.6} \right) = 2.5$$

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

$$1.5 + (2.5-1.5) \left( \frac{51.2-15.4}{33.6} \right) = 2.6$$

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

TABLE IV. Standard Errors for Estimated Percentages of Housing Units in the 1978 Housing Inventory and for Estimated Percentages of 1975-1978 Lost Housing Units for the Balance (Not in Central City) of the SMSA

(68 chances out of 100)

Base of percentage	Estimated percentage <sup>1</sup>					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
200. . . . .	37.2	37.2	37.2	37.2	37.2	38.5
500. . . . .	19.1	19.1	19.1	19.1	21.1	24.3
700. . . . .	14.5	14.5	14.5	14.5	17.8	20.6
1,000 . . . . .	10.6	10.6	10.6	10.6	14.9	17.2
2,500 . . . . .	4.5	4.5	4.7	6.5	9.4	10.9
5,000 . . . . .	2.3	2.3	3.4	4.6	6.7	7.7
10,000 . . . . .	1.2	1.2	2.4	3.3	4.7	5.4
25,000 . . . . .	0.5	0.7	1.5	2.1	3.0	3.4
50,000 . . . . .	0.2	0.5	1.1	1.5	2.1	2.4
75,000 . . . . .	0.2	0.4	0.9	1.2	1.7	2.0
100,000 . . . . .	0.12	0.3	0.7	1.0	1.5	1.7
150,000 . . . . .	0.08	0.3	0.6	0.8	1.2	1.4
200,000 . . . . .	0.06	0.2	0.5	0.7	1.1	1.2
250,000 . . . . .	0.05	0.2	0.5	0.7	0.9	1.1
300,000 . . . . .	0.04	0.2	0.4	0.6	0.9	1.0
400,000 . . . . .	0.03	0.2	0.4	0.5	0.7	0.9
500,000 . . . . .	0.02	0.15	0.3	0.5	0.7	0.8
600,000 . . . . .	0.02	0.14	0.3	0.4	0.6	0.7
700,000 . . . . .	0.02	0.13	0.3	0.4	0.6	0.7
800,000 . . . . .	0.01	0.12	0.3	0.4	0.5	0.6
900,000 . . . . .	0.01	0.11	0.2	0.3	0.5	0.6

<sup>1</sup> 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 errors shown in the table should be multiplied by a factor of 1.2.

Springfield-Chicopee-Holyoke 1978

# Appendix B

## Source and Reliability of the Estimates

<b>SAMPLE DESIGN</b> .....	App-40	<b>Coverage improvement for deficiencies 3-6</b> .....	App-42	<b>AHS-SMSA</b> .....	App-44
Annual Housing Survey .....	App-40	1970 Census of Population and Housing .....	App-43	Coverage errors .....	App-44
Designation of sample housing units for the 1978 survey .....	App-40	<b>ESTIMATION</b> .....	App-43	Rounding errors .....	App-44
Selection of the 1975 AHS-SMSA sample .....	App-40	1978 housing inventory .....	App-43	Sampling errors for the AHS-SMSA sample .....	App-44
1975-1978 additions to the housing inventory .....	App-42	1975-1978 lost units .....	App-43	Illustration of the use of the standard error tables .....	App-45
Sample selection for the 1978 Coverage Improvement Program .....	App-42	1975 estimation procedure .....	App-43	Differences .....	App-47
Coverage improvement for deficiency 1 .....	App-42	Ratio estimation procedure of the 1970 Census of Population and Housing .....	App-43	Illustration of the computation of the standard error of a difference .....	App-47
Coverage improvement for deficiency 2 .....	App-42	<b>RELIABILITY OF THE ESTIMATES</b> .....	App-44	Medians .....	App-47
		Nonsampling errors .....	App-44	Illustration of the computation of the 95-percent confidence interval of a median .....	App-47
		1970 census .....	App-44		

### SAMPLE DESIGN

**Annual Housing Survey**—The estimates for each of the 15 SMSA's are based on data collected from the 1978 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 15 SMSA's, the data were collected for the 12-month period from April 1978 through March 1979 with one-twelfth of the sample units being visited each month.

Each group of SMSA's is being interviewed for the AHS on a rotating basis with this group (AA-1 SMSA's) being the second to be revisited. All of these SMSA's were enumerated for the first time in 1975.

For each group of SMSA's, the largest SMSA from 3 of the 4 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 the remaining 12 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 AA-1 group (1978-1979) are: Atlanta, Ga.; Philadelphia, Pa.-N.J.; and San Francisco-Oakland, Calif.

The remaining SMSA's in the AA-1 group are: Cincinnati, Ohio-Ky.-Ind.; Colorado Springs, Colo.; Columbus, Ohio; Kansas City, Mo.-Kans.; New Orleans, La.; Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; Rochester, N.Y.; San Antonio, Tex.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In this SMSA, 5,362 housing units were eligible for interview. Of these sample units, 153 interviews were not obtained because, for occupied sample units, the occupants refused to be interviewed, 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 units eligible for interview, 449 units were visited but were not eligible for interview because they were condemned, unfit, demolished, converted to group quarters use, etc.

**Designation of sample housing units for the 1978 survey**—The sample housing units designated to be interviewed in the 1978 survey consisted of the following categories which are described in detail in the succeeding sections.

1. All sample housing units that were interviewed in the 1975 survey.
2. All sample housing units that were type A noninterviews (i.e., units eligible to be interviewed) or type B noninterviews (i.e., units not eligible for interview at the time of the survey but which could become eligible in the future) in the 1975 survey. (For a list of reasons for type A and type B noninterviews, see the facsimile of the 1978 AHS questionnaire, page App-20.)
3. All sample housing units that were selected from the list of new construction building permits issued since the 1975 survey. (This sample represented the housing units built in permit-issuing areas, since the 1975 survey.)
4. All sample housing units that were added to sample segments in the nonpermit universe since the 1975 survey. (This sample represented additions to the housing inventory in nonpermit-issuing areas since the 1975 survey.)
5. All sample housing units that were selected as part of the 1978 Coverage Improvement Program. (This sample represented most of the housing units which, until 1978, did not have a chance of selection.)

**Selection of the 1975 AHS-SMSA sample**—The sample for the SMSA's which are 100-percent permit-issuing 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 those 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 non-permit universe). The following SMSA's are 100-percent permit-issuing: Newport News-Hampton, Va.; Paterson-Clifton-Passaic, N.J.; San Bernardino-Riverside-Ontario, Calif.; San Diego, Calif.; and San Francisco-Oakland, Calif. The remaining 10 SMSA's contain a sample from 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 three largest SMSA's, the overall sampling rate differed by central city and the 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 the 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 the total housing units in each sector.

The major portion of the sample in each SMSA 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 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:

Thus, for this SMSA, the occupied housing unit records from the permit-issuing 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 selection procedure was then instituted that would produce one-half of the desired sample size. However, whenever a record was selected to be in sample, the housing unit record adjacent to it on the file 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 by a procedure that produced one-quarter of the desired sample size. However, at the time of the survey, the units at each of the special places were listed and subsampled at a rate which produced 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 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:

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

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. All units in existence at the time of interview in these selected segments are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

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

**1975-1978 additions to the housing inventory**—In the permit-issuing universe, a sample of new construction building permits, issued since the 1975 survey, was selected to represent housing units built in permit-issuing areas since the 1975 survey. Sampling procedures were identical to those used in selecting the 1970-1975 new construction sample, which were described previously. In the nonpermit universe, sample segments were dependently recanvassed, using listing sheets from 1975, to identify any units missed in the 1975 survey or any units added since the 1975 survey.

**Sample selection for the 1978 Coverage Improvement Program**—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 1970 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 for each of the AA-1 SMSA's except Newport News-Hampton, Va. Two different procedures were used. For the first procedure, the sampling was carried out in two stages for one- and two-unit structures and in three stages for three-or-more-unit structures. Sample units selected from one- and two-unit structures were sampled at one-fourth the rate of units originally selected for the AHS-SMSA sample (regular AHS units), while sample units selected from three-or-more-unit structures were sampled at one-half the rate of 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. Since these permits were not available for Newport News-Hampton, Va., SMSA, these coverage improvement procedures were not instituted for this SMSA. For the third stage, structures of size three or more were divided into clusters of an expected size of two units and a sample of clusters was selected. This procedure was employed in the following SMSA's: Colorado Springs, Colo.; Rochester, N.Y.; and Springfield-Chicopee-Holyoke, Mass.-Conn.

In the remaining 11 SMSA's, units whose permits were issued before January 1970, but which were completed after April 1970, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. These units were then sampled at one-third the rate of regular AHS units. These procedures added an estimated

1,804 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 parks that were missed by the census or established after the census was selected in two stages. First, for each 1978-1979 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 179 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 now contain units converted to residential use.
3. Mobile homes that had been placed outside parks since the 1970 census and have a utility hookup, or were on the site during the present survey 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 AHS sample units was selected from the permit-issuing universe at a rate of 1 in 24. 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 1,482 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 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 above, all units were listed and matched to the 1970 census. Any missed units

were then assigned for interview. This procedure added an estimated 1,178 units to the coverage of the housing inventory of this SMSA.

**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 1978 housing inventory) and estimates pertaining to characteristics of units removed from the housing inventory since 1975 (i.e., 1975-1978 lost units). Each type of estimate employed separate, although similar, estimation procedures.

**1978 housing inventory**—The AHS estimates of characteristics of the 1978 housing inventory employed a 1-stage ratio estimation procedure in all SMSA's. 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 noninterviews previously mentioned. This noninterview adjustment was done separately for occupied and vacant housing units. 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 described). In addition, within each sector separate noninterview factors were computed for one noninterview cell for conventional new construction sample housing units from both the permit-issuing universe and the coverage improvement universe, one noninterview cell for mobile homes from both the nonpermit universe and the coverage improvement universe, and one noninterview cell for other sample units from both the nonpermit universe and the coverage improvement universe, (if units were not included in any of the previous cells).

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.

**1975-1978 lost units**—The 1975-1978 lost unit estimates employed the three-stage ratio estimation procedure used to produce the AHS-SMSA estimates of the 1975 housing inventory, as was described in the 1975 Current Housing Report, Series H-170, **Housing Characteristics for Selected Metropolitan Areas**. Since the 1975-1978 lost units existed, by definition, in the 1975 housing inventory, there was a 1975 housing inventory weight associated with each 1975-1978 lost unit. This weight was used to tabulate the estimates of the characteristics of the 1975-1978 lost units.

**1975 estimation procedure**—This report presents data on the housing characteristics of the 1975 housing inventory from the 1975 Annual Housing Survey SMSA sample. The AHS-SMSA estimation procedure employed a three-stage ratio estimation process. A detailed description of this ratio estimation procedure can be found in the AHS Series H-170 reports for 1975.

**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 the 1970 Census of Population and Housing and the 1975 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 the extent to which occupancy status was erroneously reported. 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 1978 AHS-SMSA sample, a reinterview program was not conducted. However, a study was conducted for the 1975 AHS-SMSA sample. The results of which are presented in the Census Bureau memorandum, "Reinterview Results for Annual Housing Survey—SMSA Sample: 1975."

**Coverage errors**—In 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. However, these permits issued during the last 5 months of the survey do not necessarily represent missed housing units. 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, in which case, they would not have been 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 correct deficiencies 3-6 (see the coverage improvement section of this appendix) was not very effective in 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 1975 AHS sample missed as much as 2 percent of all housing units in these ED's because they were not listed during the canvassing. It should be noted that since these ED's were recanvassed for the 1978 survey, the number of missed housing units may be considerably less for 1978.

**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 statistics 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 questionnaires, instructions, and interviewers were used, estimates from each of the different samples would differ from each other. The sampling error of a survey estimate provides a measure of the variation among the estimates from all 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.

One common measure of the sampling error is the standard error. As calculated for this report, the standard error reflects the variation in the estimates due to sampling and nonsampling errors, but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors measured by the standard error, biases, and any 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 item. Standard errors applicable to estimates of characteristics of the 1975 housing inventory can be found in the AHS Series H-170 reports for 1975.

Table I presents the standard errors applicable to estimates of characteristics of the 1978 housing inventory as well as estimates of characteristics of the 1975-1978 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 6,280 for the total SMSA, 4,500 for the central cities of the SMSA, and 4,320 for the balance of the SMSA.

The reliability of an estimated percentage, computed by using the 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 1978 housing inventory as well as estimated percentages of the 1975-1978 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

**TABLE I. Standard Errors for Estimated Number of Housing Units in the 1978 Housing Inventory and for Estimated Number of 1975-1978 Lost Units for the Springfield-Chicopee-Holyoke, Mass.-Conn., SMSA, for the Central Cities, and for the Balance (Not in Central Cities) of the SMSA**

(68 chances out of 100)

Size of estimate	Standard error <sup>1</sup>		
	SMSA	In central cities	Not in central cities
0 .....	40	40	40
100 .....	70	60	60
200 .....	90	90	90
500 .....	150	140	140
700 .....	170	170	170
1,000 .....	210	200	200
2,500 .....	330	310	320
5,000 .....	460	440	450
10,000 .....	650	610	630
25,000 .....	1,000	930	950
50,000 .....	1,360	1,210	1,230
75,000 .....	1,590	1,350	1,360
100,000 .....	1,750	1,380	1,380
150,000 .....	1,900	—	—
200,000 .....	1,880	—	—

<sup>1</sup> For estimates pertaining to new construction, the standard errors shown in the table should be multiplied by a factor of 1.2 for the total SMSA, 1.3 for the central cities, and 1.1 for the balance (not in central cities) estimates.

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  
 $\sigma_x$  = the standard error of the numerator  
 $\sigma_y$  = 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 1978 there were 107,100 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 1,770. The following interpolation procedure was used.

**TABLE II. Standard Errors for Estimated Percentages of Housing Units in the 1978 Housing Inventory and for Estimated Percentages of 1975-1978 Lost Units for the Springfield-Chicopee-Holyoke, Mass.-Conn., SMSA, for the Central Cities, and for the Balance (Not in Central Cities) of the SMSA**

(68 chances out of 100)

Base of percentage	Estimated percentage <sup>1</sup>					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100. ....	30.3	30.3	30.3	30.3	30.3	33.0
200. ....	17.9	17.9	17.9	17.9	20.2	23.3
500. ....	8.0	8.0	8.0	8.8	12.8	14.7
700. ....	5.9	5.9	5.9	7.5	10.8	12.5
1,000. ....	4.2	4.2	4.5	6.3	9.0	10.4
2,500. ....	1.7	1.7	2.9	4.0	5.7	6.6
5,000. ....	0.9	0.9	2.0	2.8	4.0	4.7
10,000. ....	0.4	0.7	1.4	2.0	2.9	3.3
25,000. ....	0.2	0.4	0.9	1.3	1.8	2.1
50,000. ....	0.09	0.3	0.6	0.9	1.3	1.5
75,000. ....	0.06	0.2	0.5	0.7	1.0	1.2
100,000. ....	0.04	0.2	0.5	0.6	0.9	1.0
150,000. ....	0.03	0.2	0.4	0.5	0.7	0.9
200,000. ....	0.02	0.15	0.3	0.4	0.6	0.7

<sup>1</sup>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 errors shown in the table should be multiplied by a factor of 1.2 for the total SMSA, 1.3 for the central cities, and 1.1 for the balance (not in central cities).

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. ....	1,750
107,100. ....	x
150,000. ....	1,900

The entry for "x" is determined as follows by vertically interpolating between 1,750 and 1,900:

$$107,100 - 100,000 = 7,100$$

$$150,000 - 100,000 = 50,000$$

$$1,750 + \frac{7,100}{50,000} (1,900 - 1,750) = 1,770$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 105,330 to 108,870 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of 1978 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 104,268 to 109,932 housing units with 90 percent confidence; and that the average estimate lies within the interval from 103,560 to 110,640 housing units with 95 percent confidence.

Table A-1 of part A also shows that of the 107,100 owner-occupied housing units, 29,200, or 27.3 percent, had two bedrooms. Interpolation in table II of this appendix (i.e., interpolation on both the base and percent) shows that the standard error of the 27.3 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	27.3	50
100,000. ....	0.9	a	1.0
107,100. ....		p	
150,000. ....	0.7	b	0.9

1. The entry for cell "a" is determined by horizontal interpolation between 0.9 and 1.0.

$$27.3 - 25.0 = 2.3$$

$$50.0 - 25.0 = 25.0$$

$$0.9 + \frac{2.3}{25.0} (1.0 - 0.9) = 0.9$$

2. The entry for cell "b" is determined by horizontal interpolation between 0.7 and 0.9.

$$27.3 - 25.0 = 2.3$$

$$50.0 - 25.0 = 25.0$$

$$0.7 + \frac{2.3}{25.0} (0.9 - 0.7) = 0.7$$

3. The entry for "p" was then determined by vertical interpolation between 0.9 and 0.7.

$$107,100 - 100,000 = 7,100$$

$$150,000 - 100,000 = 50,000$$

$$0.9 + \frac{7,100}{50,000} (0.7 - 0.9) = 0.9$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 26.4 to 28.2 percent; the 90-percent confidence interval is from 25.9 to 28.7 percent; and the 95-percent confidence interval is from 25.5 to 29.1 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 characteristics 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. Due to the overlap of the 1975 and 1978 AHS-SMSA samples a positive correlation should be expected when making comparisons between the 1975 and 1978 characteristics.

*Illustration of the computation of the standard error of a difference*—Table A-1 of part A of this report shows that in 1978 there were 52,300 owner-occupied housing units with three bedrooms in this SMSA. Thus, the apparent difference, as shown by these data, between owner-occupied housing units with two bedrooms and owner-occupied housing units with three bedrooms is 23,100. Table I shows the standard error of 52,300 is approximately 1,380, and the standard error of 29,200 is approximately 1,060. Therefore, the standard error of the estimated difference of 23,100 is about:

$$1,740 = \sqrt{(1,380)^2 + (1,060)^2}$$

Consequently, the 68-percent confidence interval for the 23,100 difference is from 21,360 to 24,840 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 20,316 to 25,884 housing units, and the 95-percent confidence interval is from 19,620 to 26,580. Thus, we can conclude with 95 percent confidence that the number of 1978 owner-occupied housing units with three bedrooms is greater than the number of owner-occupied housing 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 proce-

dure 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 characteristics, 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 107,100 housing units.

1. Interpolation in table II shows that the standard error of 50 percent on a base of 107,100 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 46,000 owner-occupied housing units, or 43.0 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 19,300 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:

$$2.5 + (3.5 - 2.5) \left( \frac{48.0 - 43.0}{18.0} \right) = 2.8$$

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

$$2.5 + (3.5 - 2.5) \left( \frac{52.0 - 43.0}{18.0} \right) = 3.0$$

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