Appendix B. Source and Accuracy of the Data

SAMPLE DESIGN App-53
American Housing Survey Metropolitan Sample App-53
Designation of sample housing units for the 1988 survey App-54
1988 AHS-MS original sample selection for the 1970-based area sample of the metropolitan areas App-54
Sample selection for the Coverage Improvement Program App-55
1984 AHS-MS sample reduction App-56
AHS-MS sample selection for the 1980-based area sample of the metropolitan areas App-56

ESTIMATION App-57
1988 housing inventory App-57
Type M noninterview adjustment App-57
Type A noninterview adjustment App-57
Ratio estimation procedures App-58

ACCURACY OF THE DATA App-59
Nonsampling errors App-59
AHS-MS content errors App-59
Coverage errors App-60
Rounding errors App-60
Sampling errors for the AHS-MS App-61
Illustration of the use of the standard error tables App-61
Differences App-62
Illustration of the computation of the standard error of a difference App-62
Medians App-62
Illustration of the computation of the 90-percent confidence interval of a median App-63
Standard error tables App-64

SAMPLE DESIGN

American Housing Survey Metropolitan Sample. The estimates for each of the 11 metropolitan areas in this report series (H-170-88) are based on data collected from the 1988 American Housing Survey Metropolitan Sample (AHS-MS) which was conducted by the Bureau of the Census acting as collection agent for the Department of Housing and Urban Development.

The sample areas covered for metropolitan areas that remained in the AHS-MS sample after survey year 1983 are consistent with the 1983 Office of Management and Budget (OMB) definitions of a metropolitan statistical area (MSA), consolidated metropolitan statistical area (CMSA), or primary metropolitan statistical area (PMSA). In some instances, a given metropolitan area is a combination of primary metropolitan statistical areas and will be referred to as PMSA's. In addition to adding new areas to some metropolitan samples in order to comply with the 1983 definitional changes, some new metropolitan areas have been added. Thus, each of the 1988 metropolitan areas will fall into one of three categories—

a. Areas of the same geographic area as defined for surveys prior to 1984 (i.e., areas in which the 1970 OMB definition of a standard metropolitan statistical area is the same as the 1983 MSA, PMSA, or CMSA definition, 1970-based area)—Buffalo, NY CMSA; Cleveland, OH PMSA; Indianapolis, IN MSA; and Milwaukee, WI PMSA;

b. Areas consisting of new area in addition to the 1970-based area—Birmingham, AL MSA; Memphis, TN-AR-MS MSA; Norfolk-Virginia Beach-Newport News, VA MSA; Oklahoma City, OK MSA; Providence-Pawtucket-Warwick, RI-MA PMSA's; and the Salt Lake City, UT MSA;

c. Areas that were in sample for the first time in 1984—San Jose, CA PMSA.

The metropolitan areas selected for the 1988 AHS-MS are interviewed on a rotating basis once every 4 years. Each metropolitan area had an expected sample size of 4,250 housing units uniformly distributed throughout nine panels (panels 4-12). Due to budget constraints, panel 4 was dropped from sample in all metropolitan areas, and interviewing took place from May 1988 to December 1988. As a result, the expected sample sizes were lower than the original goal of 4,250 sample units.

In this metropolitan area, 3,414 AHS-MS housing units were eligible for interview. Of these sample housing units, 142 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 the AHS-MS housing units eligible for interview, 360 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 1988 survey. The sample housing units designated to be interviewed in the 1988 survey consisted of the following categories which are described in detail in the following sections.

Housing units which were in the 1970-based area include the following:

a. All sample housing units that were interviewed in the previous survey and remained in sample after the 1988 reduction. This sample includes housing units that were selected as part of the 1976-1981 Coverage Improvement Program. These coverage improvement cases represented most of the housing units which, until these procedures were implemented, did not have a chance of selection.

b. 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 previous survey and remained in sample after the 1988 reduction. For a list of reasons for type A noninterviews, see the facsimile of the 1988 AHS questionnaire, page App-24.

c. All sample housing units that were selected from a listing of new residential construction building permits issued since the previous survey and remained in sample after the 1988 reduction. This sample represented the housing units built in permit-issuing areas since the previous survey.

d. All sample housing units that were added to sample segments in the nonpermit universe since the previous survey and remained in sample after the 1988 reduction. This sample represented additions to the housing inventory in nonpermit-issuing areas since the previous survey.

e. In the Birmingham, AL MSA; Memphis, TN-AR-MS MSA; Norfolk-Virginia Beach-Newport News, VA MSA; and the Providence-Pawtucket-Warwick, RI-MA PMSA's, all housing units selected from the 1980 Census of Population and Housing.

Housing units within new areas added to the metropolitan area in 1980 and for metropolitan areas that were in sample for the first time in 1984 (1980-based area):

a. All housing units selected from the 1980 Census of Population and Housing.

b. All housing units that were selected from a list of new residential construction building permits issued. This sample represented the housing units built in permit-issuing areas since the 1980 census.

c. All sample housing units that were added to sample segments in the nonpermit universe. This sample represented additions to the housing inventory in nonpermit-issuing areas since the 1980 census.

The following table shows the percent of sample that is 1970-based and 1980-based for each metropolitan area:

Metropolitan area	Percent 1970-based	Percent 1980-based
Birmingham, AL MSA	91.8	8.2
Buffalo, NY CMSA	100.0	0.0
Cleveland, OH PMSA	100.0	0.0
Indianapolis, IN MSA	100.0	0.0
Memphis, TN-AR-MS MSA	92.1	7.9
Milwaukee, WI PMSA	100.0	0.0
Norfolk-Virginia Beach-Newport News, VA MSA	26.9	73.1
Oklahoma City, OK MSA	88.3	11.7
Providence-Pawtucket-Warwick, RI-MA PMSA's	93.2	6.8
Salt Lake City, UT MSA	83.4	16.6
San Jose, CA PMSA	0.0	100.0

1988 AHS-MS original sample selection for the 1970-based area sample of the metropolitan areas. The original sample for the 1970-based area of the metropolitan areas which, in 1970, were 100-percent permit-issuing was selected from two frames: housing units enumerated in the 1970 Census of Population and Housing in areas under the jurisdiction of permit-issuing areas (the 1970-based permit-issuing universe); and housing units constructed in permit-issuing areas since the 1970 census (the 1970-based new construction universe). In addition, the sample for those metropolitan areas which were not 100-percent permit-issuing in 1970 included a sample selected from a third frame—those housing units located in areas not under the jurisdiction of permit-issuing offices (the 1970-based nonpermit universe). In 1970, the Norfolk-Virginia Beach-Newport News, VA MSA was the only metropolitan area that was 100-percent permit-issuing.

Sampling operations, described in the following paragraphs, were performed separately within the central city and balance (outside the central city) using the 1970 OMB definitions of the central city of each metropolitan area for each of the sample frames. The overall sampling rate used to select the sample for each metropolitan area was determined by the size of the sample. Each metropolitan area had an overall sampling rate about the same for both the central city and the balance, since the sample was distributed proportionately between the central city and balance of the metropolitan area according to the corresponding distribution of total housing units.

The major portion of the sample in each of the metropolitan areas was selected from a file which represented the 20-percent sample of housing units enumerated in permit-issuing areas of the metropolitan areas during the 1970 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and housing 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 the head of household (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 (renter/owner), family size, and family income category as illustrated by the following table:

Family 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, 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 metropolitan areas. A sample selection procedure was then instituted that would produce one-half of the desired sample. 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 metropolitan areas. 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 housing 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 the metropolitan area 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 the metropolitan area. Under clerically selected procedures, 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. In February 1984, the new construction sampling operation for the 1970-based and 1980-based areas were combined into one computerized system. The universe sampled in the computerized system will be referred to in the estimation section as the 1980-based permit universe. Under these procedures, prior to sample selection, the list of permits was chronologically stratified by the date of issue, State, 1980 central city and balance, county or minor civil division, and permit office. Clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at twice the overall sampling rate. The housing units within each of the clusters were then subsampled so that two of the four housing units originally selected were kept in sample.

For those metropolitan areas which were 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 nonpermit universe was the selection of a sample of census enumeration districts. Prior to this sample selec tion, the ED's were stratified by census tract within the central city and within the balance of the metropolitan area. The probability of selection of an ED was proportionate to the following:

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

The sample ED's were then divided into segments (i.e., small land areas with well-defined boundaries having an expected size of four, or a multiple of four, housing units). At the time of the survey, those segments that did not have an expected size of four were further subdivided to produce an expected four sample housing units. The next step was the selection of one of these segments within each sample ED. All housing units in existence at the time of interview in these selected segments were eligible for sample. Thus, housing units enumerated in the 1970 census as well as housing units built since the 1970 census were included.

Sample selection for the Coverage Improvement Program. The Coverage Improvement Program was undertaken to correct certain deficiencies in the AHS-MS from the 1970 permit-issuing area universe and the 1970 new construction universe within the 1970-based area. The coverage deficiencies included the following units:

a. New construction from building permits issued prior to January 1970, but completed after April 1, 1970.

b. Mobile homes placed in parks either missed during the 1970 census or established since the 1970 census.

c. Housing units missed in the 1970 census.

d. Housing units converted to residential use that were nonresidential at the time of the 1970 census.

e. Houses that have been moved onto their present site since the 1970 census.

f. Mobile homes placed outside parks since the 1970 census or vacant at the time of the 1970 census.

For a detailed description of the coverage improvement sample selection process, see earlier reports in the H-170 series for the years 1976 through 1981.

1984 AHS-MS sample reduction. The 1984 AHS-MS sample reduction dropped units from sample. The universes involved were the 1970-based permit-issuing universe, the 1970-based new construction universe, and the 1970-based nonpermit universe. These procedures involved dropping individual housing units from the permit-issuing universe, whole clusters from the new construction universe, and whole segments from the nonpermit universe.

The reduction was implemented to achieve two criteria:

(1) A sample size of 4,250 in all metropolitan areas.

(2) A sample having an equal number of owners and renters.

In order to achieve these results, each unit was classified according to the original panel number (the original sample was divided into 12 panels, with one-twelfth of the sample being in each panel) and 1984 tenure (each housing unit was given a 1984 tenure based on the previous year's tenure status). In order to simplify field procedures, panels 1-3 (i.e., a random one-fourth of the original sample) were dropped from sample whenever possible. More sample reductions were implemented separately for each 1984 tenure group (using different selection rates) across the remaining panels.

AHS-MS sample selection for the 1980-based area sample of the metropolitan areas. The sample for new areas added to the 1970-metropolitan areas, and metropolitan areas in sample for the first time which, in 1980, were 100-percent permit-issuing, was selected from two frames—housing units enumerated in the 1980 Census of Population and Housing in areas under the jurisdiction of permit-issuing areas (the 1980-base permit-issuing universe) and housing units constructed in permit-issuing areas since the 1980 census (the 1980-based new construction universe). In addition, the sample for those metropolitan areas which were not 100-percent permit-issuing in 1980 included a sample from a third frame—those housing units not under the jurisdiction of permit-issuing offices (the 1980-based nonpermit universe). In 1980, the Birmingham, AL MSA; Memphis, TN-AR-MS MSA; and Oklahoma City, OK MSA were the only metropolitan areas that added new areas which were not 100-percent permit-issuing.

In order to satisfy confidentiality requirements in the Birmingham, AL MSA; Memphis, TN-AR-MS MSA; Norfolk-Virginia Beach-Newport News, VA MSA; and the Providence-Pawtucket-Warwick, RI-MA PMSA's, it was necessary to supplement the existing sample within the 1970-based area for each metropolitan area. The additional sample housing units were selected separately for each metropolitan area from the 1980-based permit-issuing universe.

The major portion of the sample in each metropolitan area was selected from a file which represented all the housing units enumerated in permit-issuing areas of the metropolitan area during the 1980 Census of Population and Housing. This file contained records for occupied housing units, vacant housing units, and housing units in group quarters. Sampling operations were done separately for noninstitutionalized group quarters and for all other housing units in permit-issuing areas. In addition, in order that an equal number of owner and renter housing units were selected in each metropolitan area, a selection rate that differed by tenure group was used. Before the sample was selected, the housing units that were not classified as group quarters were stratified into 60 categories by tenure, contract rent, value, and number of rooms as illustrated by the following table:

Contract rent and value	Number of rooms		
	1-3	4-5	6+
RENTER			
Contract rent:			
Less than $100			
$100 to $149			
$150 to $199			
$200 to $249			
$250 to $299			
$300 to $349			
$350 to $399			
$400 or more			
Not available			
OWNER			
Value:			
Less than $20,000			
$20,000 to $29,999			
$30,000 to $34,999			
$35,000 to $39,999			
$40,000 to $49,999			
$50,000 to $64,999			
$65,000 to $79,999			
$80,000 to $99,999			
$100,000 to $149,999			
$150,000 or more			
Not available			

The group quarters housing units were divided into two strata: (1) institutionalized group quarters; and (2) noninstitutionalized group quarters.

The following sample selection procedures were then implemented separately within the 1980 central city and balance of the metropolitan area. For the Birmingham, AL MSA; Memphis, TN-AR-MS MSA; Norfolk-Virginia Beach-Newport News, VA MSA; and the Providence-Pawtucket-Warwick, RI-MA PMSA's, the sample selections were implemented separately by geographic zone. First, all units were sorted by the 1980 central city and balance, stratum, State, district office, ED, and census serial number. The

sample selection procedure was then implemented separately for noninstitutionalized group quarters and for institutionalized group quarters and nongroup quarters housing units. For the institutionalized group quarters and nongroup quarters housing units, the sample selection was done across the 61 strata. Individual housing units were selected for the nongroup quarters while each institutionalized group quarters had one chance of selection. Before the sample selection for the noninstitutionalized group quarters was implemented, the following measure of size was calculated for each record:

$$(1/4) \times \frac{\text{(Total group quarters population)}}{2.75}$$

The noninstitutionalized group quarters were then selected proportionate to the measure of size.

The second frame from which the metropolitan area was selected was a list of new construction building permits issued since 1980 (i.e., the new construction universe). The sample selection from the list of new construction building permits was an independent operation within this metropolitan area. This operation was described previously in the discussion on computerized building permit sampling in the 1970-based areas.

For those metropolitan areas which were not 100-percent permit-issuing, the remainder of the sample was selected from a frame consisting of areas not under the jurisdiction of permit-issuing offices (i.e., the 1980-based nonpermit universe). The first step in the sampling operation for the nonpermit universe was the selection of a sample of census ED's within these areas (using the overall sampling rate). Prior to this sample selection, the ED's were sorted by State, district office, and enumeration district number. The probability of selection of an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1980 census ED} + \dfrac{\text{Noninstitutionalized group quarters population in 1980 census ED}}{2.75}}{4}$$

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 housing units were further subdivided to produce an expected four sample housing units. Following the division, a segment from each sample ED was selected. All housing units in existence at the time of interview in these selected segments were eligible for sample. Thus, housing units enumerated in the 1980 census as well as housing units built since the 1980 census are included.

ESTIMATION

The 1988 AHS-MS produced estimates pertaining to characteristics of the housing inventory at the time of the interview (i.e., the 1988 housing inventory).

1988 housing inventory. The AHS estimates of characteristics of the 1988 housing inventory were produced using a multistage ratio estimation procedure. Prior to the implementation of the ratio estimation procedures, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for Type M and Type A noninterviews.

Type M noninterview adjustment. The Type M noninterviews are sample units which were dropped due to selection by another survey or because of permit unavailability. These noninterviews occur only in the 1980-based permit-issuing area universe, the 1980-based nonpermit-issuing area universe, and the 1980-based new construction universe. This adjustment was done separately for the 1980 central city and balance of each metropolitan area for housing units in the 1980-based permit-issuing universe, in group quarters, in the 1980-based nonpermit issuing area universe, and in the 1980-based new construction universe. The Type M noninterview adjustment was computed separately for each cell and was equal to the following:

$$\frac{\text{AHS-MS sample estimate of 1980 housing units in the cell} + \text{Weighted count of Type M noninterviewed housing units}}{\text{AHS-MS sample estimate of 1980 housing units in the cell}}$$

Type A Noninterview Adjustment. The next adjustment was the Type A noninterview adjustment. This adjustment was done on occupied housing units and was computed separately for units in the 1980-based permit-issuing area universe, for new construction, and for all other housing units (this includes the 1970-based permit-issuing universe, the 1970-based and 1980-based nonpermit-issuing universes, and the 1970-based new construction housing units built prior to the last survey). For units in the 1980-based permit-issuing universe, a Type A noninterview adjustment factor was computed separately for each of the 62 strata used in the sample selection process by 1980 central city and balance. For new construction units, a Type A noninterview adjustment factor was computed separately by tenure for each of the central city and balance. For all other units, a Type A noninterview adjustment factor was calculated separately by tenure and 1970 central city and balance for each of the following: (1) 24 noninterview cells for sample housing units from the permit-issuing universe (where the cell consisted of one or more of the 50 different strata used in the 1970-based permit-issuing universe as previously described); (2) 1 noninterview cell for new construction housing units; (3) 1

noninterview cell for mobile homes or trailers from the nonpermit-issuing universe; (4) 1 noninterview cell for units that were not mobile homes or trailers from the nonpermit-issuing universe; (5) 3 noninterview cells for units from the coverage improvement universe; (6) 1 noninterview cell for units classified as vacant at the time of the 1970 census; and (7) 1 noninterview cell for units classified as group quarters at the time of the 1970 census. Within a given cell, the Type A noninterview adjustment factor was equal to the following ratio using the basic weight times the Type M noninterview adjustment factor for the sample weight:

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

Ratio estimation procedures. 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 1970-based permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor for each cell was equal to the following:

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

For each metropolitan area, the numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of housing units enumerated in areas under the jurisdiction of permit-issuing offices.

The denominators of the ratio estimation factors were then computed and were obtained from weighted estimates of all the AHS-MS housing units from the 1970-based permit-issuing universe within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the Type A noninterview adjustment). The computed ratio estimation factor was then applied to the existing weight for each sample housing unit within the corresponding ratio estimation cells. This ratio estimation procedure was introduced to correct the probabilities of selection for samples in each of the strata used in the sample selection of the 1970-based permit-issuing universe. Prior to the AHS-MS sample selection within each metropolitan area, housing units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. The same probability of selection was then applied to the remaining units to select the AHS sample. Since the number of housing units deleted from the AHS universe frame was not necessarily proportional among all strata, some variation in the actual probability of selection between strata was introduced during the AHS-MS sample selection process.

The following ratio estimation procedure was employed for all sample units from the 1980-based permit-issuing universe. This factor was computed separately for all metropolitan areas excluding the Buffalo, NY CMSA; Cleveland, OH PMSA; Indianapolis, IN MSA; and the Milwaukee, WI PMSA within each 1980-based permit-issuing universe noninterview cell mentioned previously. The ratio estimation factor was equal to the following:

$$\frac{\text{1980 census count of housing units from the 1980-based permit-issuing universe in the corresponding cell}}{\text{AHS sample estimate of 1980-based housing units from the 1980 permit-issuing universe in the corresponding cell}}$$

For each metropolitan area, the numerator of the ratio was obtained from the 1980 Census of Population and Housing 100-percent file of housing units enumerated in areas under the jurisdiction of permit-issuing offices. The denominator of the ratio was obtained from weighted estimates of all the AHS sample housing units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the Type M noninterview adjustment factor times the Type A noninterview adjustment factor).

The computed ratio estimation factor was then applied to the existing weight for each sample housing unit within the corresponding ratio estimation categories.

This ratio estimation procedure was introduced to adjust the sample estimate in each of the strata used in the sample selection of the 1980-based permit-issuing universe to an independent estimate (1980 census count) for the strata. This adjustment was necessary since some sample units were dropped during the processing procedures.

The final ratio estimation procedure was applied in all metropolitan areas. The metropolitan areas were subdivided into geographic areas consisting of a combination of counties. This procedure involved the ratio estimation of the AHS-MS weighted sample estimate of the September 1, 1988, housing inventory in each geographic area for each metropolitan area to an independent estimate of total housing units for the corresponding cell. This ratio estimation factor equalled the following:

$$\frac{\text{Independent estimate of the September 1, 1988, housing unit inventory for the corresponding geographic area of the metropolitan area}}{\text{AHS-Metropolitan Area sample estimate of the housing inventory for the corresponding geographic area of the metropolitan area}}$$

The independent estimates of total housing units that were used as the numerator of this ratio are described below. The denominator of this ratio was obtained from the weighted estimate of the AHS Metropolitan Area sample housing units using the existing weight.

Independent estimates were derived for the September 1, 1988, occupied housing inventory for each geographic area within each metropolitan area. For all metropolitan areas excluding the San Jose, CA PMSA, the estimates were based on the following ratio:

$$\frac{\text{1988 estimate of total population excluding group quarters in the county}}{\text{1988 estimate of total population per household excluding group quarters in the county}}$$

The methodology used to derive the independent estimates was based on the population-per-household method as described in the *Proceedings of the Bureau of the Census Second Annual Research Conference*, March 23-26, 1986, pages 83-110. This method is based on the national trend of the adult population per household and assumes that this trend is uniform throughout the country.

For the San Jose, CA PMSA, the independent estimates were obtained from the State of California, Department of Finance. Here, the population per-household method could not be applied since the national trend underestimated the true population per household.

The AHS-MS sample estimate of the housing inventory for the corresponding geographic area was obtained using the existing weight. The computed ratio estimation factors were then applied to all housing units in the corresponding geographic area of each metropolitan area, and the resulting product was used as the final weight for tabulation purposes.

The effect of this ratio estimation procedure, as well as the overall estimation procedures, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat, by chance, from the metropolitan area as a whole, it can be expected that the sample estimates will be improved when the sample housing population, or different portions of it, is brought into agreement with known good estimates of the metropolitan area housing population.

ACCURACY OF THE DATA

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-MS.

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 of respondents to provide correct information; mistakes in recording or coding the data; other errors of collection, response, processing, and 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 1988 AHS-MS.

AHS-MS content errors. A Content Reinterview Program was done for the 1988 AHS-MS units. A sample of these units was revisited and answers to some of the questions on the questionnaire were obtained again. The original interview and reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the data collected from interviewed units.

The 1988 Content Reinterview Program served solely as an interviewer quality check. All interviewers were selected for the quality check; this reviewed their proficiency in properly evaluating the following items: (1) correct unit visited; (2) area segment coverage; (3) living quarters classification; (4) tenure (i.e., owner/renter); (5) interview status; and (6) household composition. The results of this reinterview program, however, are not available.

Although a quality analysis of particular survey questions was not done as part of the 1988 Content Reinterview Program, past reinterview programs have shown that certain items are likely to produce moderate or high response variability. Response variability is defined as a measure of consistency between the original survey response to an item and the reinterview response to that item.

Moderate levels of variability indicate that the response error is not insignificant in comparison to the sampling error. High variability indicates that the response errors are very significant in relation to the sampling errors with which they are associated; therefore, caution should be used when considering estimates of these characteristics. A 1984 Content Reinterview Program was done for each 1988 metropolitan area. Both the response variance and bias components of response error were estimated for selected questionnaire items. The items reinterviewed fell into three groups: units in structure, number of rooms, and appliances. The results of this study are presented in the Census Bureau memorandum, "1984 AHS-MS Reinterview Results." Some of these results are presented below (note that the results are based on interviews across all 1984 metropolitan areas and not for any specific metropolitan area):

a. The rarity of responses to a majority of the items in the units-in-structure group resulted in valid measures for only the living-quarters, number-of-apartments, and units-in-building items. All showed low response variability. The number-of-rooms group showed low variability for bedrooms and bathrooms. Moderate levels existed for other rooms, except kitchens and living rooms for which measures could not be computed. The appliance group generally exhibited low variability for the existence of appliances and moderate variability for appliance age. Only central air conditioning fuel showed high variability.

b. The level of response bias present in the data did not appear to be a significant problem. The few categories which exhibited a significant level of bias were mostly in the number-of-rooms group. Several of the categories concerned with air conditioning also showed some bias.

c. The square footage question was analyzed only for the response variance interview. The results showed that individuals did not know the square footage or floor dimensions of their house or apartment. However, the individuals who did estimate their square footage in both interviews rather than opting for "Don't Know" were within 100 square feet of their original estimate two-thirds of the time.

The results of this study were based on sample data. Sampling error associated with the corresponding estimates of nonsampling error must be taken into account when considering the results of this study.

Other reinterview studies were also conducted in conjunction with previous AHS-National Sample and AHS-MS enumerations. These studies included items dealing primarily with poor housing quality, attitudes about the neighborhood, and certain housing costs. The following table shows the items which had relatively high levels of variability. While these reinterview studies were not done in 1988, questions from previous enumerations were not altered enough to lead one to believe that the level of inconsistent responses would change.

Survey items	Level of variability
Open cracks or holes on inside of building	Moderate to High
Holes in floors	Moderate to High
Broken plaster or peeling paint on ceiling and walls	High
Mice or rats	Moderate
Working electrical outlet in all rooms	High
Concealed wiring	High
Blown fuses/tripped circuit breakers	Moderate to High
Neighborhood conditions: street noise; roads in need of repair; crime; trash,litter, junk in streets or on properties;boarded up/abandoned structures; non-residential activities; odors, smoke, gas	Moderate to High
Satisfactory neighborhood services: police protection; hospitals/health clinics; public transportation; shopping; elementary schools	Moderate to High
Electricity cost	Moderate to High
Gas cost	High
Oil, coal, kerosene, wood or other fuel cost	Moderate to High
Fire/hazard insurance	Moderate to High
Real estate taxes	Moderate
Cost of real estate taxes	Moderate
Cost of water supply and sewage disposal	High
Cost of garbage collection	Moderate to High
Gross income	High
Prefer to live in same area or somewhere else	Moderate

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

Coverage errors. In errors of coverage and estimation for missing data, the AHS new construction sample had deficiencies in the representation of conventional (nonmobile home or trailer) new construction. Due to time constraints, only those building permits issued more than 7 months before the survey ended were eligible to be sampled to represent conventional new construction in permit-issuing areas for each metropolitan area. However, these permits issued during the last 7 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 housing 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, is not adequately represented.

AHS-MS misses a significant portion of new mobile homes. It is believed that most of the difference is due to poor coverage of mobile home parks in address ED's. Undercoverage exists for those mobile homes built between the time of the last coverage improvement procedure and the 1980 census. It has been estimated that on a national level as much as 25 percent of those mobile homes built after January 1, 1980, may be missed.

Deficiencies also exist in ED's where area sampling methods are used. It had been assumed that all housing units located inside these ED's would be represented in the sample. However, it has been established that the AHS-MS missed as much as 2 percent of all housing units in these ED's because they were not listed during the initial canvassing. It should be noted that since these ED's were recanvassed each time this metropolitan area was surveyed, the number of missed housing units may be considerably less for the 1988 survey than for the initial survey.

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

Rounding errors. For errors associated with processing, 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-MS. 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 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:

> 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.

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

The figures presented in the tables that follow (page App-64) are approximations to the standard errors of various estimates shown in this report for this metropolitan area. 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.

Table I presents the standard errors applicable to estimates of characteristics of the 1988 housing inventory. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in this table.

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 for the 1988 housing inventory. 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 estimates of zero are 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:

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

where: S_x = the standard error of x
S_y = the standard error of y

Illustration of the use of the standard error tables. Table 2-1 of this report shows that in the Birmingham, AL, metropolitan area, there were 243,700 owner-occupied housing units. Interpolation using table I of this appendix shows that the standard error of an estimate of this size is approximately 3,410. 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	3,550
243,700	x
250,000	3,390

The entry of "x" is determined as follows by vertically interpolating between 3,550 and 3,390.

$$243{,}700 - 200{,}000 = 43{,}700$$
$$250{,}000 - 200{,}000 = 50{,}000$$

$$3{,}550 + \frac{43{,}700}{50{,}000}(3{,}390 - 3{,}550) = 3{,}410$$

Consequently, the 90-percent confidence interval, as shown by these data, is from 238,240 to 249,160 housing units. Therefore, a conclusion that the average estimate,

derived from all possible samples, of 1988 owner-occupied housing units lies within a range computed in this way would be correct for roughly 90 percent of all possible samples.

Table 2-3 also shows that of the 243,700 owner-occupied housing units, 62,100 or 25.5 percent had two bedrooms. Interpolation using table II of this appendix (i.e., interpolation on both the base and percent) and applying the factor given in the footnote shows that the standard error of the 25.5 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 (see table II footnotes).

Base of percentage	Estimated percentage		
	25 or 75	25.5	50
200,000	1.0	a	1.1
243,700		p	
250,000	0.9	b	1.0

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

$$25.5 - 25.0 = 0.5$$
$$50.0 - 25.0 = 25.0$$

$$1.0 + \frac{0.5}{25.0}(1.1 - 1.0) = 1.0$$

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

$$25.5 - 25.0 = 0.5$$
$$50.0 - 25.0 = 25.0$$

$$0.9 + \frac{0.5}{25.0}(1.0 - 0.9) = 0.9$$

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

$$243{,}700 - 200{,}000 = 43{,}700$$
$$250{,}000 - 200{,}000 = 50{,}000$$

$$1.0 + \frac{43{,}700}{50{,}000}(0.9 - 1.0) = 0.9$$

Applying a factor of 1.1 according to the footnote from table II gives a standard error of 1.0 percentage points. Consequently, the 90-percent confidence interval, as shown by these data, is from 23.9 to 27.1 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 metropolitan areas or the difference between separate and uncorrelated characteristics in the same metropolitan area. If there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error, but if there is a high negative correlation, the formula will underestimate the true standard error.

Illustration of the computation of the standard error of a difference. Table 2-3 of this report shows that in the Birmingham, AL, metropolitan area, there were 141,100 owner-occupied housing units with three bedrooms.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 79,000. Table I shows that the standard error of 141,100 is approximately 3,400 and the standard error of 62,100 is approximately 2,590. Therefore, the standard error of the estimated difference of 79,000 is about 4,270.

$$4{,}270 = \sqrt{(3{,}400)^2 + (2{,}590)^2}$$

Consequently, the 90-percent confidence interval for the 79,000 difference is from 72,170 to 85,830 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 90 percent of all possible samples. Thus, we can conclude with 90-percent confidence that the number of owner-occupied housing units with three bedrooms is greater than the number of owner-occupied units with two bedrooms since the 90-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, determine the confidence interval corresponding to the two points established in step 2. To find the lower endpoint of the confidence interval, it is necessary to know into which interval of the distribution the lower percentage limit falls. Similarly, to find the upper endpoint of the confidence interval, it is necessary to know into which interval of the distribution the upper percentage limit falls. Note that these distribution intervals could be different, although this will not happen very often.

A 1.6 standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus 1.6 times the standard error determined in step 1. For about 90 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the 90-percent confidence interval of a median. Table 3-13 of this report shows the median monthly housing cost for owner-occupied housing units is $268. The base of this distribution is obtained by subtracting "mortgage payment not reported" from the total number of occupied units. Using this definition, the base is equal to 224,900.

1. Interpolation using table II and applying the appropriate footnote shows that the standard error of 50 percent on a base of 224,900 is approximately 1.2 percentage points.
2. To obtain a 90-percent confidence interval on the estimated median, initially add to and subtract from 50 percent 1.6 times the standard error determined in step 1.This yields percentage limits of 48.1 and 51.9.
3. From the distribution for monthly housing cost in table 3-13, the $250 to $300 interval for owner-occupied housing units contains the 48.1 percent derived in step 2. About 106,100 housing units or 47.2 percent fall below this interval, and 17,600 housing units or 7.8 percent fall within this interval. By linear interpolation, the lower limit of the 90-percent confidence interval is found to be about $256.

$$250 + (300 - 250)\frac{48.1 - 47.2}{7.8} = 256$$

Similarly, the $250 to $300 interval for owner-occupied housing units contains the 51.9 percent derived in step 2. About 106,100 housing units or 47.2 percent fall below this interval, and 17,600 housing units or 7.8 percent fall within this interval. The upper limit of the 90-percent confidence interval is found to be about $280.

$$250 + (300 - 250)\frac{51.9 - 47.2}{7.8} = 280$$

Thus, the 90-percent confidence interval ranges from $256 to $280.

Table I. Standard Errors for Estimated Number of Housing Units in the 1988 Birmingham, AL MSA

Size of estimate	Standard error[1]		
	Combined owner and renter housing units[2]	Owner housing units[3]	Renter housing units[4]
0	130	130	100
100	130	130	100
300	200	200	180
500	260	260	230
700	310	300	270
1,000	370	360	320
2,500	580	570	510
5,000	810	800	720
10,000	1,140	1,130	1,010
25,000	1,770	1,750	1,570
50,000	2,420	2,390	2,140
75,000	2,850	2,810	2,520
100,000	3,160	3,120	2,790
150,000	3,510	3,460	3,100
200,000	3,590	3,550	-
250,000	3,430	3,390	-
300,000	2,980	2,940	-
350,000	2,050	-	-
384,400	-	-	-

[1]To compute standard errors for new construction estimates, the standard errors in the table should be multiplied by a factor of 1.2 for owner housing units, 1.4 for renter housing units, and 1.2 for the combined owner and renter housing units.

[2]Some examples that pertain to both owner and renter housing units are: total housing units; all occupied housing units; all year-round housing units; mobile homes or trailers; and total vacant housing units.

[3]The owner housing units pertain to owner-occupied housing units and vacant housing units excluding vacant-for-rent housing units.

[4]The renter housing units pertain to renter-occupied housing units and vacant-for-rent housing units.

Table II. Standard Errors for Estimated Percentages of Housing Units in the 1988 Housing Inventory of the Birmingham, AL MSA

Base of percentage	Estimated percentage[1]					
	0 or 100	1 or 99	5 or 95	10 or 90	25 or 75	50
100	51.2	51.2	51.2	51.2	51.2	51.2
300	25.9	25.9	25.9	25.9	25.9	29.6
500	17.4	17.4	17.4	17.4	19.8	22.9
700	13.0	13.0	13.0	13.0	16.8	19.4
1,000	9.5	9.5	9.5	9.7	14.0	16.2
2,500	4.0	4.0	4.5	6.1	8.9	10.2
5,000	2.1	2.1	3.2	4.3	6.3	7.2
10,000	1.0	1.0	2.2	3.1	4.4	5.1
25,000	0.4	0.6	1.4	1.9	2.8	3.2
50,000	0.2	0.5	1.0	1.4	2.0	2.3
75,000	0.14	0.4	0.8	1.1	1.6	1.9
100,000	0.10	0.3	0.7	1.0	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	0.6	0.8	0.9
350,000	0.03	0.2	0.4	0.5	0.7	0.9
384,400	0.03	0.2	0.4	0.5	0.7	0.8

[1]Standard errors are presented to the nearest one-tenth of one percentage point except when the standard error is less than or equal to 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.

The following factors should be applied to estimates that do not pertain strictly to new construction. For estimates pertaining to both owners and renters, apply a factor of 1.1. For estimates pertaining to owner housing units, apply a factor of 1.1. For estimates pertaining to renter housing units, apply a factor of 1.0.