

# Appendix D. Errors

All numbers from the American Housing Survey (AHS), except for sample size, are estimates. As in other surveys, errors come primarily from the following:

- Incomplete data (Incomplete data are adjusted by assuming that the respondents are similar to those not answering, and the size of these errors is estimated.)
- Wrong answers (The U.S. Census Bureau does not adjust for wrong answers and does not estimate the size of the errors.)
- Sampling (Sampling errors are not adjusted and the size of the error is estimated.)

Incomplete data and wrong answers are usually the largest source of errors, larger than sampling errors. For example, in the American Housing Survey–National Sample (AHS-N), the changes in weighting in 1981 and 1991 (see Appendix C) corrected some of the error due to incomplete data. That one correction averaged 2.5 percent in 1991. Worse errors from incomplete data and from wrong answers apply to some items, discussed below.

Additional information on the quality of AHS data can be obtained from the U.S. Census Bureau’s, *American Housing Survey: A Quality Profile*, Series H121/95-1.

## INCOMPLETE DATA

**Coverage errors.** Because of deficiencies with our sampling lists, the homes in the survey do not represent all homes in the country. The Census Bureau attempts to adjust for the deficiencies by raising the raw numbers from the survey proportionally, so that the numbers published here match independent estimates of the total number of homes. Housing unit undercoverage is about 2.2 percent for the 2003 AHS-N.

The independent estimates changed around 2.5 percent in both 1981 and 1991 (after the 1980 and 1990 censuses, respectively), which implies that some error existed in the years just before the adjustment. By comparison, the independent estimates changed by 0.8 percent in 2003 (after the 2000 census). Before adjustments, undercoverage varies from 2 percent to 20 percent for major categories of units (see Table 2 in Appendix D of *American Housing Survey for the U.S. in 1995*), but is usually less than 2 percent, on average. Table W lists units that have known coverage deficiencies.

Table W. **Poorly Covered Units**

Type of unit	Type of deficiency
Manufactured/mobile homes, boats, and recreational vehicles (RVs)	No coverage of new manufactured/mobile home parks, new marinas, and new RV parks since April 1980 in areas where addresses are complete and permits are required for new construction.
Conventional new construction	No coverage of permits issued fewer than 8 months before interviewing or homes built without permits where permits are required. In addition, eligible units could be missed and ineligible units included because of incorrect answers to questions used to screen out ineligible units.
New construction in special places (for example, college campuses, prisons, etc.)	Not covered in either permit-issuing or nonpermit-issuing areas.
Group quarters and houses moved in	Eligible units could be missed because of incorrect answers to questions used to screen out ineligible units.
Conversions from nonresidential units	Minimal coverage of nonresidential units in buildings with no living quarters at the time of the 1980 census that converted to housing units by 1991 (and no coverage since 1991) in areas where addresses are complete and permits are required for new construction.
Within-structure additions	Some extra apartments created illegally or occupied by fugitives are probably missed because people do not report them for fear of penalties.
Whole structure additions not covered by permit sampling	These units are chosen with the aid of screening questions. Eligible units could be missed and ineligible units included because of incorrect answers to the screening questions.

**Missing data.** Some people refuse the interview or some of the questions, or do not know the answers. When the entire interview is missing, other similar interviews represent the missing ones (see Appendix B). For most missing answers, an answer from a similar household is copied.<sup>1</sup> The Census Bureau does not know how close the imputed values are to the actual values. For other items, “not reported” is used as an answer category. The items with the most missing data are primarily those that people forget or consider personal: mortgages, other housing costs, and income.

<sup>1</sup>Hot deck allocation is used: an answer is copied from the most recently processed similar household before the household with the missing item.

Incompleteness can cause large errors since, when even 10 percent of homes are missed by a particular question, they represent about 10 million homes that have to be estimated *on little or no basis* (there are about 100 million homes in the U.S.). The survey estimates them by assuming that they are like some group of homes that did give data, an assumption that is *never exactly true* although it is usually better than ignoring the homes with the missing data. Thus, it is not surprising that large biases, as shown in Tables X and Y in the tables section, are possible when the survey has data for only 50 to 90 percent of homes for particular items. Again, readers should be wary of items with highly incomplete data.<sup>2</sup>

Rates of completeness were not computed for 2003. Table 2 in Appendix D of *American Housing Survey for the United States in 1995* gives the completeness rates for 1995. Because of the change in data collection methodology, the rates for 2003 may be higher or lower than in the past. However, the items that were most incomplete in 1995 are probably still the most incomplete for 2003.

**Effect on income.** The nonsampling errors interact particularly badly for income. Income questions are inconsistently answered (Table Z), incompletely answered, and the totals fall short of totals known from the National Income Accounts, especially for the elderly.<sup>3</sup>

**Change over time.** Several aspects of the AHS make estimates of change from previous data unreliable. These changes may elicit different answers from the past, even if nothing changed in the housing unit. Wording and question order for most questions changed. Also, the questionnaire now runs on interviewers' portable computers (as described in Appendix C), resulting in the following possible changes:

- The correct questions should be asked. Skip patterns will be followed more accurately.
- Inconsistent answers (such as reporting a move-in date before the date built) are probed during the interview, rather than just being changed in later computer processing, so these problems should be resolved more accurately.

<sup>2</sup>Statistical note: The November 1990 paper, *How Response Error, Missing Data and Undercoverage Bias Survey Data*, estimates that 90 percent of errors from incomplete data are less than:  $1.645 \times (.0012 \times U + .0363 \times (\text{lesser of } A \text{ or } U - A))$  where A is any count from the AHS and U is the total number of housing units in the U.S. or metropolitan area (both in thousands, result also in thousands). Weights are adjusted to reduce these errors, but it is not known how much error remains. *How Response Error, Missing Data and Undercoverage Bias Survey Data*, order number HUD-6458, is available from HUD USER (see "Where to Get AHS Data").

<sup>3</sup>Data are in the *Codebook for the American Housing Survey Volume 1*, available from HUD USER. Newer comparisons, though for a different survey, are in *Money Income of Households Families, and Persons in the United States: 1992*, Series P60-184, pages C12-C14, available from the Superintendent of Documents (see "Where to get AHS data").

- For some questions, large changes from prior year data are probed during the interview to reduce mistaken measurements of large change.
- Some respondents may dislike the presence of the computer, though interviewers do not report many problems.
- It is now a little harder for interviewers to go back to a question much earlier in the questionnaire if a respondent suddenly remembers something.

## WRONG ANSWERS

Wrong answers happen because people misunderstand questions, cannot recall the correct answer, or do not want to give the right answer. Table Y shows which items have been measured for inconsistency when people are reinterviewed after a few weeks. The actual survey did not catch and reconcile these inconsistencies and continuously occurring errors are not measured at all. Thus, a high rate of wrong answers remains for some items. The Census Bureau categorizes these levels of inconsistency into three ranges:

1. Less than 20 is considered a low level of inconsistency.
2. Between 20 and 50 is considered a moderate level of inconsistency.
3. Greater than 50 is considered a high level of inconsistency indicating that responses are not reliable.

Not all questions have been checked for inconsistencies; the ones checked were the questions where inconsistencies seemed likely. Questions measuring opinions were likely to have high inconsistencies. For the 2003 AHS-N, the wording for some questions changed. This change is expected to lower the level of inconsistency for the changed items. The numbers in Table Z are percents. They are nearly the same as 100 minus the correlation between answers in the original interview and the reinterview. For example, an inconsistency of 15 means a correlation of 85 percent, which is good. This is the correlation between answers to the same question, usually from the same respondents, a month apart. Wrong answers create wrong results and mean that data about groups (for example, income groups) are infected with data from people who really are not like the group at all. Errors are especially troublesome for rare items for which even small errors overwhelm the true data. Readers should be wary of drawing firm conclusions from items with high levels of inconsistency or from categories smaller than a few million homes.

## SAMPLING ERRORS

**Definition.** Errors from sampling reflect how estimates from a sample vary from the actual value. (Note: “actual value” means the value derived if all housing units had been interviewed under the same conditions, rather than only a sample.) A confidence interval is a range that contains the actual value with a specified probability. The range of nonsampling error is usually larger than this confidence interval.

**Counts.** Most numbers from the AHS are counts of housing units (for example, units with basements or units with elderly persons). These counts have error from sampling. As with the other types of errors, readers should be wary of numbers with large errors from sampling.

Table AA gives a convenient list of errors for a range of numbers for 2003 AHS-N. The error from sampling cannot be known exactly. For numbers not in Table AA, the error from sampling is approximated using the following formula for constructing a 90-percent confidence interval:

$$1.645 \times \sqrt{4.74 \times A - .000039 \times A^2}$$

where A is a number (a count of units in thousands) from the AHS. This formula is an overestimate for most items. For more accurate estimates, use the formula in Table Y.

For example if A is 200:

$$1.645 \times \sqrt{4.74 \times 200 - .000039 \times 200 \times 200} = 51$$

The 90-percent confidence interval can then be formed by adding and subtracting this error to the survey estimate of 200 (that is, 200 plus or minus 51). Statements such as “the actual value is in the range 200 plus or minus 51 (149 to 251),” are right 90 percent of the time and wrong 10 percent of the time.<sup>4</sup>

Numbers in the book are printed in thousands, so 200 means 200,000. The formulas are designed to use numbers directly from the book; do not add zeros. The result is also in thousands, so 51 means 51,000.

Tables BB gives a list of errors for a range of numbers for the 2003 AHS-National Metropolitan Areas. For numbers not found in this table, interpolate between the numbers in the table or use the appropriate formula from Table EE for the 2003 AHS-National Metropolitan Areas. Remember, in any case, that the total error is larger than the sampling error.

**Percents.** Any subgroup can be shown as a percent of a larger group. For AHS-N, the error from sampling for a 90-percent confidence interval for this percent is:

$$1.645 \times \sqrt{4.74 \times p \times (100 - p) / A}$$

<sup>4</sup>The formula in the text is based on 1.645 times the standard error from sampling. This formula gives “90-percent confidence interval errors.” For 95-percent confidence interval errors, multiply by 1.960 instead of 1.645; for 99-percent confidence, multiply by 2.576 instead of 1.645.

where p is the percent; A is the denominator, or base of the percent in thousands.

For example, the error from sampling for a 90-percent confidence interval for 40 percent of 200 (meaning 200,000) is:

$$1.645 \times \sqrt{4.74 \times 40 \times 60 / 200} = 12.4$$

Statements such as “the actual percent is in the range 27.6 percent to 52.4 percent” are right 90 percent of the time.

This formula is an overestimate for most items. To get a more accurate estimate for AHS-N, replace the first number under the square root sign above with the first number under the square root sign of the formula for the appropriate universe in Table CC.<sup>5</sup>

For the AHS-National Metropolitan areas, use the appropriate formula in Table FF.

Note that when a ratio C/D is computed when C is not a subgroup of D (for example, the number of Hispanics as a ratio of the number of Blacks), the error from sampling is different.<sup>6</sup>

**Medians.** The steps in Table DD calculate the error from sampling for a 90-percent confidence interval for a median. This is an approximation to the error.

For small bases the confidence interval on medians can not be estimated reliably. To estimate a median’s sampling error more accurately, find the sampling error on 50 percent as described in Table GG and compute the 90-percent confidence interval.

**Differences.** Two numbers from the AHS, like 34 and 40, or 40 percent have a “statistically significant difference” if their ranges of error from sampling for a 90-percent confidence interval do not overlap.<sup>7</sup>

**Formulas for Error From Sampling.** The letter “A” in the formulas in Table DD represents a number (a count of units in thousands) from AHS, (see the “Counts” section for an

<sup>5</sup>This formula is actually  $1.645 \times \sqrt{(p(100-p)/n)}$ , since 4.74/A adjusts the data to the effective sample size.

<sup>6</sup>The error from sampling for a 90-percent confidence interval for a ratio C/D is

$$C/D \sqrt{(\text{error for } C/C)^2 + (\text{error for } D/D)^2}$$

when the error for C should be interpreted as the error for a 90-percent confidence interval for C. Likewise, the error for D should be interpreted as the error for a 90-percent confidence interval for D.

<sup>7</sup>When ranges of error from sampling for a 90-percent confidence interval do overlap, numbers are still statistically different if the result of subtracting one from the other is more than

$$\sqrt{(\text{error for first number})^2 + (\text{error for second number})^2}$$

The error for the first and second numbers should be interpreted as the error for a 90-percent confidence interval for the first and second numbers, respectively.

example of how “A” is used). For AHS-N, the minimum error from sampling is  $\pm 10$  (meaning  $\pm 10$  thousand).<sup>8</sup> If a formula gives an error smaller than 10, use 10.

For a confidence interval on zero for the 2003 AHS-National Metropolitan Areas, refer to Table Y. If a formula gives an error smaller than the error of zero, use the error for zero.

The formulas give the errors for a 90-percent confidence interval. For a 95-percent confidence interval, multiply by 1.960 instead of 1.645; for a 99-percent confidence interval, multiply by 2.576 instead of 1.645.

For AHS-N, if an item falls into two different categories in Table CC, use the formula that gives the largest error. For example, for Hispanics’ income in the South, use the formulas for the South (since there is no specific formula for income, and errors for the South will be bigger than those for Hispanics). For the following neighborhood characteristics, use the neighborhood formulas:

- Opinion of neighborhood
- Street noise or traffic
- Neighborhood crime
- Odors
- Other bothersome neighborhood conditions
- Public elementary school
- Public transportation
- Neighborhood shopping
- Police protection
- Parking lots
- Description of area (except open space, park, farm, or ranch) within 300 feet
- Age of other residential buildings within 300 feet
- Other buildings vandalized or with interior exposed within 300 feet
- Bars on windows of buildings within 300 feet

<sup>8</sup>This minimum formula is based on the binomial 90-percent confidence interval on zero  $U \times (1 - .1^{4.74/U}) = 11$  (where U is the total number of homes from the AHS). For a 95-percent confidence interval, substitute .05 for .1 in the above formula. For a 99-percent confidence interval, substitute .01 for .1. More discussion and other approximations are in the paper “Sampling Errors for Small Groups” available from HUD USER (see “Where to Get AHS Data”).

- Conditions of streets within 300 feet
- Trash, litter, or junk on streets or any properties within 300 feet
- Manufactured/mobile homes in group

For the following items, which have larger standard errors, use the special characteristics formulas:

- Cooperatives or condominiums
- No complete bathroom
- Less than 1,500 square feet of detached one-family or mobile homes
- Well serving 1 to 5 units
- Mobile homes in a group
- Area within 300 feet includes open space, park, farm, or ranch
- Septic tank, cesspool, chemical toilet
- Five or more acres in lot size
- No bedroom
- Lacking complete kitchen facilities
- Lacking some plumbing facilities
- No flush toilet
- Major street repairs needed

**Table X. Errors for Incomplete Data Bias: 2003 AHS-N**

[Numbers in thousands]

When the AHS gives one of the following numbers—	The chances are 90 percent that the complete value <sup>1</sup> is inside the range of plus or minus
0 .....	238
10 .....	239
100 .....	244
1,000 .....	298
2,500 .....	388
5,000 .....	537
10,000 .....	835
25,000 .....	1,731
50,000 .....	3,224
75,000 .....	2,967
100,000 .....	1,474
110,000 .....	887
120,000 .....	280

<sup>1</sup>“Complete value” means the value derived if there were no missing data.

**Table Y. Errors for Incomplete Data Bias: 2003 AHS-National Metropolitan Areas**

[Errors in thousands]

Size of estimate	Chicago, IL	Detroit, MI	Los Angeles- Long Beach, CA	New York- Nassau-Suffolk- Orange, NY	Northern New Jersey	Philadelphia, PA-NJ
0 .....	6.3	3.7	6.5	9.6	5.1	4.1
10 .....	6.9	4.3	7.1	10.2	5.7	4.7
25 .....	7.8	5.2	8.0	11.1	6.6	5.6
50 .....	9.3	6.7	9.5	12.6	8.1	7.1
100 .....	12.3	9.6	12.5	15.5	11.1	10.0
300 .....	24.2	21.6	24.5	27.5	23.0	22.0
500 .....	36.2	33.5	36.4	39.4	35.0	33.9
700 .....	48.1	45.4	48.3	51.4	46.9	45.9
1,100 .....	72.0	49.9	72.2	75.3	70.8	61.5
1,200 .....	78.0	44.0	78.2	81.2	76.8	55.5
1,400 .....	89.9	32.1	90.1	93.2	76.1	43.6
2,000 .....	77.9	NA	85.0	129.0	40.3	7.7
2,500 .....	48.0	NA	55.1	149.6	10.4	NA
3,000 .....	18.1	NA	25.3	119.8	NA	NA
3,500 .....	NA	NA	NA	89.9	NA	NA
4,000 .....	NA	NA	NA	60.1	NA	NA
4,500 .....	NA	NA	NA	30.2	NA	NA

NA means no error estimates are provided because the estimate is larger than the estimated total number of housing units in the MSA.

Table Z. **Different Answers a Month Apart**

Item	Level of inconsistency <sup>1</sup>	Confidence interval <sup>2</sup>	When measured <sup>3</sup>
<b>HIGH LEVEL OF INCONSISTENCY</b>			
Other kinds of heating equipment (central warm-air) . . . . .	91	(73 - 100)	89-MS
Mortgage payment includes anything else (first mortgage) . . . . .	90	(72 - 111)	90-MS
Water came in from other places . . . . .	81	(64 - 100)	89-MS
Moved for other, financial/employment . . . . .	80	(62 - 104)	85-MS
Moved for other, housing related . . . . .	79	(65 - 97)	85-MS
Poor city/county service in neighborhood . . . . .	78	(63 - 95)	89-MS
Police protection problem in neighborhood . . . . .	78	(63 - 95)	89-MS
Number of business rooms with direct access to outside . . . . .	76	(63 - 91)	95-N
Moved for other reasons . . . . .	73	(64 - 85)	85-MS
Number of other rooms . . . . .	73	(64 - 83)	95-N
Difficulty hearing with or without a hearing aid . . . . .	72	(59 - 88)	95-N
Rooms used both as business space and for something else . . . . .	70	(62 - 80)	95-N
Cost for routine repairs and maintenance . . . . .	70	(65 - 75)	95-N
Moved for better quality house . . . . .	69	(58 - 82)	85-MS
Moved for other family/personal related . . . . .	68	(54 - 86)	85-MS
Cost for water supply and sewage disposal . . . . .	68	(61 - 76)	81-N
Lower cost state or local mortgage . . . . .	67	(54 - 83)	95-N
Other problem in neighborhood . . . . .	67	(61 - 74)	89-MS
Number of living rooms . . . . .	66	(53 - 82)	95-N
Shed, detached garage, or other building added or replaced in last 2 years . . . . .	66	(49 - 88)	95-N
Water safe for drinking . . . . .	66	(56 - 77)	95-N
Undesirable industries/businesses in neighborhood . . . . .	66	(54 - 82)	89-MS
Difficulty reaching kitchen facilities . . . . .	65	(49 - 87)	95-N
Number of family rooms, dens, recreation rooms and/or libraries . . . . .	65	(57 - 75)	95-N
Rats . . . . .	65	(54 - 69)	89-MS
Difficulty opening, closing, or going through any doors of home . . . . .	64	(46 - 87)	95-N
Noise in neighborhood . . . . .	64	(57 - 72)	89-MS
Difficulty moving between rooms . . . . .	64	(49 - 84)	95-N
Number of business rooms without direct access to outside . . . . .	64	(54 - 76)	95-N
Peeling paint on the ceiling . . . . .	63	(49 - 80)	81-N
Other kinds of heating equipment (none) . . . . .	63	(60 - 67)	89-MS
How LIKELY to move to place prefer to live in 5 years . . . . .	62	(54 - 71)	85-MS
Difficulty reaching bathroom facilities . . . . .	62	(47 - 82)	95-N
Other kinds of heating equipment (unvented room) . . . . .	62	(45 - 86)	89-MS
Difficulty seeing with or without glasses or contact lenses . . . . .	60	(49 - 72)	95-N
How LIKELY to still be living in this unit in 5 years . . . . .	60	(49 - 74)	85-MS
Gross income . . . . .	59	Not available	82-MS
Number of days worked at home . . . . .	59	(49 - 72)	95-N
Patio, terrace, or detached deck added or replaced in last 2 years . . . . .	58	(42 - 81)	95-N
Electric fuses or breaker switches blown . . . . .	58	(50 - 68)	81-N
Open cracks or holes in building . . . . .	58	(47 - 72)	81-N
People in neighborhood . . . . .	57	(52 - 62)	89-MS
Other major repairs over \$500 each—repair done . . . . .	57	(50 - 64)	85-MS
Work done in last 2 years to attic, basement, garage, or unfinished area of home . . . . .	56	(44 - 71)	95-N
Difficulty going up and down steps . . . . .	56	(46 - 69)	95-N
Central air conditioning/dehumidifier . . . . .	56	Not available	80-N
Satisfactory police protection . . . . .	55	(49 - 62)	77-N
Moved for lower rent or less expensive house to maintain . . . . .	55	(43 - 70)	85-MS
Broken plaster or peeling paint . . . . .	55	(46 - 65)	89-MS
Water came in from walls, doors, windows . . . . .	55	(45 - 67)	89-MS
A working electric wall outlet . . . . .	55	(42 - 71)	77-N
Home equity loans . . . . .	55	(48 - 64)	95-N
Other kinds of heating equipment (fireplace with no insert) . . . . .	54	(49 - 59)	89-MS
Shopping . . . . .	54	(47 - 61)	77-N
Special modifications, equipment, or assistance needed because of physical limitation . . . . .	54	(44 - 66)	95-N

Table Z. **Different Answers a Month Apart—Con.**

Item	Level of inconsistency <sup>1</sup>	Confidence interval <sup>2</sup>	When measured <sup>3</sup>
<b>HIGH LEVEL OF INCONSISTENCY—Con.</b>			
Difficulty entering and exiting home .....	54	(43 - 67)	95-N
Broken plaster on the ceiling .....	53	(40 - 70)	81-N
Water came in from roof .....	53	(46 - 60)	89-MS
Driveways or walkways added or replaced in last 2 years .....	53	(42 - 67)	95-N
Difficulty with personal activities—bathing/showering .....	53	(42 - 66)	95-N
Payments the same during whole length of the mortgage .....	52	(46 - 59)	85-MS
Difficulty with personal activities—cooking and preparing food .....	52	(41 - 66)	95-N
Other major repairs over \$500 each—someone in household did the work .....	51	(36 - 72)	85-MS
Number of hours worked at home as self-employed, contract worker, or business owner .....	51	(43 - 61)	95-N
Litter in neighborhood .....	51	(44 - 60)	89-MS
Which best describes place at that time .....	51	(46 - 55)	85-MS
Rate the place (10 categories) .....	51	(49 - 53)	89-MS
Main reason moved .....	51	(47 - 55)	85-MS
Yearly cost for garbage .....	51	(43 - 62)	81-N
<b>MODERATE LEVEL OF INCONSISTENCY</b>			
Holes in the floors .....	50	(33 - 74)	81-N
Type of vacant unit .....	50	(38 - 65)	81-N
Cookstove or range with oven .....	50	(39 - 64)	85-N
Public transportation .....	50	(44 - 56)	77-N
Oil, coal, kerosene, wood, and any other fuel cost .....	50	(40 - 64)	81-N
Other kinds of heating equipment (other built-in electric) .....	50	(38 - 66)	89-MS
Central air fuel .....	50	(40 - 63)	85-N
At age 16, live in this area/different place .....	50	(44 - 57)	85-MS
Difficulty with personal activities—housework/laundry .....	50	(41 - 61)	95-N
Do work at home .....	50	(43 - 58)	95-N
Traffic in neighborhood .....	49	(43 - 54)	89-MS
Moved to establish own household .....	48	(38 - 59)	85-MS
Rate the place (categories 1–6 combined) .....	48	(46 - 51)	89-MS
Fencing or walls added or replaced in last 2 years .....	48	(37 - 61)	95-N
Drive to work alone or with others .....	48	(38 - 59)	95-N
Real estate taxes .....	47	(33 - 67)	81-N
Other kinds of heating equipment (portable electric) .....	47	(41 - 54)	89-MS
Central air conditioning/none .....	47	Not available	80-N
Crime in neighborhood .....	47	(41 - 53)	89-MS
Bathroom or kitchen remodeled in last 2 years .....	46	(39 - 54)	95-N
Fixed place of work .....	46	(37 - 57)	95-N
Any additions built—repair done .....	46	(35 - 61)	85-MS
Water came in from basement .....	45	(38 - 55)	89-MS
Any other rooms .....	45	(42 - 49)	95-N
Moved to change from owner to renter/renter to owner .....	44	(36 - 55)	85-MS
Five years from now, would you prefer living in this area or someplace else .....	44	(32 - 60)	80-N
Major equipment, such as furnace or central air replaced or added—repair done .....	44	(35 - 55)	85-MS
Major disaster in last 2 years required repairs .....	44	(31 - 60)	95-N
Water leaked into home from outdoors .....	43	(39 - 47)	89-MS
Concealed wiring .....	43	(33 - 57)	89-MS
Other kinds of heating equipment (fireplace with insert) .....	43	(35 - 52)	89-MS
Rate the place (4 combined categories) .....	43	(41 - 46)	89-MS
Difficulty with personal activities—grooming/dressing .....	43	(30 - 60)	95-N
Siding replaced or added in last 2 years—repair done .....	42	(32 - 56)	85-MS
Moved to be closer to school/work .....	41	(32 - 53)	85-MS
Yearly cost of insurance (reported in \$100 increments to \$1,000) .....	41	(38 - 44)	89-MS
Heat breakdown .....	41	(30 - 56)	89-MS
Heating equipment broke down for 6 hours or more .....	41	(30 - 56)	89-MS
Public elementary school satisfactory .....	40	(34 - 47)	89-MS
Cost for real estate taxes .....	40	(35 - 46)	81-N

Table Z. **Different Answers a Month Apart—Con.**

Item	Level of inconsistency <sup>1</sup>	Confidence interval <sup>2</sup>	When measured <sup>3</sup>
<b>MODERATE LEVEL OF INCONSISTENCY—Con.</b>			
Mice or rats or signs of .....	40	Not available	76-N
House/apartment cold for 24 hours .....	40	(36 - 45)	89-MS
Central air conditioning/portable fan .....	40	Not available	80-N
Current mortgage same year as bought home .....	39	(27 - 56)	85-MS
Mode of transportation to work last week .....	38	(31 - 46)	95-N
Anything about the neighborhood that bothers you .....	38	(35 - 41)	89-MS
Prefer to be living in another home in this area in 5 years .....	38	(31 - 48)	85-MS
Change in taxes/insurance/principal balance .....	37	(28 - 51)	85-MS
Number of mortgages on home/property .....	36	(28 - 47)	95-N
Other kinds of heating equipment (stove) .....	36	(28 - 47)	89-MS
Costs for gas for the month of August .....	35	(24 - 54)	89-N
Bathrooms remodeled or added—repair done .....	35	(28 - 45)	85-MS
All or part of roof replaced in last 2 years—repair done .....	35	(29 - 42)	85-MS
Married, widowed, divorced, or separated .....	35	Not available	85-MS
Number of dining rooms .....	35	(32 - 38)	95-N
Highest level of school/degree .....	34	(32 - 35)	95-N
New storm doors or storm windows bought and installed—repair done .....	33	(27 - 41)	85-MS
Moved because needed larger house or apartment .....	33	(26 - 41)	85-MS
Number of homes source of water serving .....	33	(22 - 49)	95-N
Insulation added—repair done .....	32	(25 - 44)	85-MS
Kitchen remodeled or added—repair done .....	32	(25 - 41)	85-MS
House and lot sell on today's market .....	31	(29 - 34)	90-MS
Moved for new job or job transfer .....	30	(22 - 39)	85-MS
Average monthly cost for gas .....	29	(23 - 37)	89-N
Average monthly cost for electricity .....	28	(24 - 34)	89-N
Type of mortgage (for the first mortgage/loan) (non-CATI) <sup>4</sup> .....	27	(21 - 36)	89-N
Change based on interest rates .....	26	(18 - 38)	85-MS
Year the building was built .....	25	Not available	85-MS
All or part of roof replaced in last 2 years—someone in household did the work .....	25	(15 - 44)	85-MS
Number of family rooms .....	25	(21 - 30)	85-N
Mortgage payment included homeowner's insurance (first mortgage) .....	24	(21 - 27)	90-MS
Prefer to be living in this house/apartment/someplace else .....	24	(20 - 29)	85-MS
Number of half bathrooms .....	24	(20 - 27)	95-N
Clothes washer age .....	22	(19 - 25)	85-N
How many years for mortgage .....	22	(17 - 29)	85-MS
<b>LOW LEVEL OF INCONSISTENCY</b>			
Attend a public school or a private school .....	19	(15 - 25)	89-MS
New storm doors or storm windows bought and installed—someone in household did the work .....	19	(11 - 35)	85-MS
Garbage disposal age .....	18	(15 - 22)	85-N
Refrigerator age .....	18	(16 - 20)	85-N
Heating equipment broke .....	18	(9 - 34)	89-MS
Clothes dryer age .....	18	(15 - 21)	85-N
Oven/cooking burner age .....	18	(16 - 21)	85-N
Monthly payment (first mortgage) .....	16	(14 - 18)	90-MS
Insulation added—someone in household did the work .....	16	(8 - 33)	85-MS
New storm doors or storm windows bought and installed—job cost .....	15	(8 - 32)	85-MS
Mortgage payment included property tax (first mortgage) .....	15	(12 - 18)	90-MS
New/assumed mortgage .....	15	(11 - 22)	85-MS
How much was borrowed .....	14	(11 - 18)	85-MS
Monthly payment (for first mortgage/loan) (non-CATI) <sup>4</sup> .....	14	(11 - 19)	89-N
Mortgage, home equity loan, or other loan on this house/ apartment .....	14	(11 - 17)	95-N
Dishwasher age .....	14	(11 - 17)	85-N
Number of full bathrooms .....	13	(11 - 15)	95-N
Where was mortgage borrowed (non-CATI) <sup>4</sup> .....	13	(7 - 28)	89-N
How much was borrowed (for the first mortgage/loan) (non-CATI) <sup>4</sup> .....	13	(10 - 17)	89-N
Number of bedrooms .....	12	(11 - 14)	95-N

Table Z. **Different Answers a Month Apart—Con.**

Item	Level of inconsistency <sup>1</sup>	Confidence interval <sup>2</sup>	When measured <sup>3</sup>
<b>LOW LEVEL OF INCONSISTENCY—Con.</b>			
Clothes dryer fuel .....	12	(9 - 14)	85-N
Have property insurance .....	12	(10 - 14)	89-MS
Number of room air conditioners .....	11	(9 - 15)	85-N
Room air conditioners .....	10	(8 - 12)	85-N
Interest rate on the mortgage (for the first mortgage/loan) (non-CATI) <sup>4</sup> ...	10	(7 - 15)	89-N
Source of water serving 15 or more homes .....	10	(8 - 13)	95-N
Kitchen remodeled or added—someone in household did the work .....	9	(3 - 26)	85-MS
Number of units in building .....	8	(6 - 9)	85-N
Clothes washer .....	8	(6 - 9)	85-N
Living quarters .....	8	(6 - 9)	85-N
Source of water .....	8	(6 - 11)	95-N
Dishwasher .....	6	(5 - 7)	85-N
Garbage disposal .....	5	(4 - 7)	85-N
Number of apartments .....	5	(4 - 8)	85-N
Central air conditioning .....	5	(4 - 6)	85-N
Clothes dryer .....	5	(4 - 7)	85-N
Cooking fuel .....	5	(4 - 6)	85-N

<sup>1</sup>Levels are in percents. They are nearly the same as 100 minus the correlation between answers in the original interview and the reinterview a month later. For example, an inconsistency of 80 means a correlation of 20 percent, which is not good.

<sup>2</sup>Square brackets show 90-percent confidence intervals. Parentheses show 95-percent confidence intervals (used in 1988 and before).

<sup>3</sup>Measured in national surveys (N) or metropolitan surveys (MS).

<sup>4</sup>CATI is computer-assisted telephone interviewing; where shown, inconsistency was measured separately for CATI and non-CATI interviews.

Table AA. **Errors From Sampling: 2003 AHS-National**

[Numbers in thousands]

When the AHS gives one of the following numbers—	The chances are 90 percent that the actual value is inside the range of plus or minus
0 .....	11
10 .....	11
100 .....	36
1,000 .....	113
2,500 .....	177
5,000 .....	248
10,000 .....	343
25,000 .....	505
50,000 .....	614
75,000 .....	607
100,000 .....	477
110,000 .....	366
120,000 .....	140

Source: These errors were computed based on a formula with high sampling error Table Y. This table represents a conservative example.

**Table BB. Errors From Sampling to Compute a 90-Percent Confidence Interval: 2003 AHS-National Metropolitan Areas**

(Numbers in thousands)

Size of estimate	Chicago, IL	Detroit, MI	Los Angeles-Long Beach, CA	New York-Nassau-Suffolk-Orange, NY	Northern New Jersey	Philadelphia, PA-NJ
0	2.5	2.5	2.5	5.4	5.3	2.5
10	5.4	5.4	5.4	8.0	7.9	5.4
25	8.6	8.6	8.6	12.6	12.4	8.6
50	12.1	12.0	12.1	17.7	17.5	12.1
100	17.0	16.8	17.0	25.0	24.5	16.8
300	28.4	27.4	28.5	42.3	40.6	27.6
500	35.4	33.1	35.5	53.4	50.1	33.6
700	40.3	36.3	40.5	61.7	56.4	37.1
1,100	46.3	37.1	46.8	73.5	62.8	39.1
1,200	47.2	36.3	47.7	75.8	63.3	38.6
1,400	48.4	33.1	49.1	79.6	63.3	36.5
2,000	47.2	NA	48.6	86.4	53.3	13.2
2,500	40.3	NA	42.7	87.7	23.3	NA
3,000	23.5	NA	29.1	85.2	NA	NA
3,500	NA	NA	NA	78.6	NA	NA
4,000	NA	NA	NA	66.6	NA	NA
4,500	NA	NA	NA	45.2	NA	NA

NA means no error estimates are provided because the estimate is larger than the estimated total number of housing units in the MSA.

**Table CC. Formulas for 90-Percent Confidence Intervals: 2003 AHS-National**

Characteristics	General formulas —	
	All characteristics except those listed under other formulas	Fuels, heating/cooling equipment and neighborhood characteristics
Total units, elderly, new construction, vacants, Northeast, Midwest, West	$1.645 \times \sqrt{2.48 \times A - 0.000021 \times A^2}$	$1.645 \times \sqrt{4.74 \times A - 0.000039 \times A^2}$
Central city, mobile homes, Hispanic, urban, suburbs, Black	$1.645 \times \sqrt{2.48 \times A - 0.000021 \times A^2}$	$1.645 \times \sqrt{2.48 \times A - 0.000021 \times A^2}$
Rural, South, outside (P)MSAs	$1.645 \times \sqrt{3.12 \times A - 0.000026 \times A^2}$	$1.645 \times \sqrt{4.74 \times A - 0.000039 \times A^2}$

**Table DD. How to Compute a 90-Percent Confidence Interval for a Median**

Steps for calculations	The formula	An example	Your data
How many total units is the median based on (in thousands, exclude "not reported" and "don't know")? . . . .	A	200	_____
What are the end-points of the category the median is in? . . . . .	X - Y	\$50-75	_____
What is the width of this category (in dollars, rooms, or whatever the item measures)? . . . . .	W	\$25	_____
How many housing units are in this median category (in thousands)? . . . . .	B	30	_____
Then the error from sampling for the median is approximately: <sup>1</sup> . . . . .	$\frac{K \times W \times \sqrt{A}}{B}$	$\frac{1.8 \times 25 \times \sqrt{200}}{30.0} = \$21$	_____
The 90-percent confidence interval for the median is: . . .	median $\pm \frac{K \times W \times \sqrt{A}}{B}$	median $\pm$ \$21	_____

<sup>1</sup>Note: To obtain an appropriate value for K, multiply the **numerator** of the formula for computing the error from sampling for 50 percent by a factor of .01. Refer to the "Percents" section of this appendix for the appropriate formula for AHS-National. Refer to Table FF for the appropriate formula for the AHS-National Metropolitan areas. For example, estimates in the Chicago, IL, metropolitan area  $K = .01 \times (1.645 \times \sqrt{1.100 \times 50 \times 50}) = .863$ .

Table EE. **Formulas for 90-percent Confidence Intervals: 2003 AHS-National Metropolitan Areas**

MSA	The formula is:
Chicago, IL .....	$1.645 \times \sqrt{(1.100 \times A) - (.000344 \times A^2)}$
Detroit, MI .....	$1.645 \times \sqrt{(1.100 \times A) - (.000579 \times A^2)}$
Los Angeles-Long Beach, CA .....	$1.645 \times \sqrt{(1.100 \times A) - (.000332 \times A^2)}$
New York-Nassau-Suffolk-Orange, NY .....	$1.645 \times \sqrt{(2.350 \times A) - (.000485 \times A^2)}$
Northern New Jersey .....	$1.645 \times \sqrt{(2.300 \times A) - (.000888 \times A^2)}$
Philadelphia, PA-NJ .....	$1.645 \times \sqrt{(1.100 \times A) - (.000534 \times A^2)}$

<sup>1</sup>The formulas in the text are based on 1.645 times the errors from sampling. This formula gives 90-percent confidence interval errors. For 95-percent confidence interval errors, multiply by 1.960 instead of 1.645; for 99-percent confidence, multiply by 2.576 instead of 1.645.

Table FF. **Formulas for 90-percent Confidence Intervals Associated with a Percentage: 2003 AHS-National Metropolitan Areas**

MSA	The formula is:
Chicago, IL .....	$1.645 \times \sqrt{(1.100 \times p \times (100-p))/A}$
Detroit, MI .....	$1.645 \times \sqrt{(1.100 \times p \times (100-p))/A}$
Los Angeles-Long Beach, CA .....	$1.645 \times \sqrt{(1.100 \times p \times (100-p))/A}$
New York-Nassau-Suffolk-Orange, NY .....	$1.645 \times \sqrt{(2.350 \times p \times (100-p))/A}$
Northern New Jersey .....	$1.645 \times \sqrt{(2.300 \times p \times (100-p))/A}$
Philadelphia, PA-NJ .....	$1.645 \times \sqrt{(1.100 \times p \times (100-p))/A}$

<sup>1</sup>The formulas in the text are based on 1.645 times the errors from sampling. This formula gives 90-percent confidence interval errors. For 95-percent confidence interval errors, multiply by 1.960 instead of 1.645; for 99-percent confidence, multiply by 2.576 instead of 1.645.

**Table GG. Calculation of the 90-Percent Confidence Interval for Medians**

In the following example, cost data are used to calculate the 90-percent confidence interval for medians (all numbers are in thousands):

		Cumulative number of housing units
Total housing units	209	
Less than \$25	50	50
\$25 to \$49	45	95
\$50 to \$74	30	125
\$75 to \$99	20	145
\$100 or more	55	200
Not reported	9	
Median	\$54	

Item	Formula	Bottom limit		Top limit	
		Example	Your data	Example	Your data
How many total units is the median based on (in thousands, exclude "not reported" and "no cash rent")? .....	A	200	_____		
Half the total, for the median (in thousands).....	A/2	100	_____		
Error from sampling for 50 percent of the base of this median (first line) <sup>1</sup> .....	$1.79/\sqrt{A}$	12.7	_____		
Multiply this percentage error by .01 to turn it into a fraction and by total units to give the error in housing units.....	$1.79\sqrt{A}$	25.3	_____		
Bottom of error range (second line minus fourth line, in thousands).....	B <sub>bottom</sub>	*74.7	_____		
Top of error range (second line plus fourth line, in thousands) .....	B <sub>top</sub>			*125.3	_____
* Start adding up the housing units in the table, category by category, cumulatively from the beginning of the table until you exceed the starred number above. What interval does the starred number fall in? .....		\$25-49	_____	\$75-99	_____
How many housing units are in all the categories before this one (in thousands)?.....	C	50	_____	125	_____
How many housing units are in this category (in thousands)? .....	D	45	_____	20	_____
What is the bottom limit of this category (in dollars, rooms, or whatever the item measures)? .....	E	\$25	_____	\$75	_____
What is the bottom limit of the next category (in dollars, rooms, etc)? .....	F	\$50	_____	\$100	_____
Formula to calculate limits of confidence interval .....	$\frac{(B-C)}{D}(F-E)+E$	$\frac{(74.7 - 50)}{45}(25)+25$		$\frac{(125.3 - 125)}{20}(25)+75$	
Limits of confidence interval (in dollars, rooms, etc.) .....		\$39		\$75	

\* Starting with the starred step, this worksheet is equivalent to interpolation, for those who are familiar with this term.

<sup>1</sup>Statistical note: This formula is based on the error from sampling for 50 percent (using the appropriate formula,  $1.645 \times \sqrt{4.74 \times 50 \times (100 - 50)/A} = 179/\sqrt{A}$ ).