Chapter 1

ERRORS IN AHS DATA

As you may know a sample survey usually differs from the rest of the country in small random ways. These random differences are called sampling errors, and will be discussed in Section 2.0. However, first we will discuss nonsampling errors in the AHS, which are often larger.

To summarize both kinds of errors, you should probably assume that any percentage may be off by at least plus or minus 2 percentage points (more on some variables). If the percentage is calculated from less than 2,000 cases, the error may be even larger, and you should use Tables 1-4 and 1-5 to find it.

When the AHS shows an actual number of households, not a percent, it may be off by plus or minus 2 percent of the whole sample. For example, the 1.3 million U.S. homes without kitchens in 1983 may be off by 1.8 million (2 percent of the U.S housing stock), and the 1,900 homes without kitchens in Washington, D.C. may be off by 5,600 (2 percent of the DC housing stock). These are very rough judgements, and other researchers may assert more or less overall error.

1.0 NONSAMPLING ERRORS

Nonsampling errors come from four sources.

Interviewer and Respondent Errors. These errors are due to misunderstandings, people not knowing the answers, or not wanting to give them. Re-interviews and comparisons with other data measure this type of error. Reinterviews show that from a few percent to over half the people change their answers when re-asked the same question a few weeks later. These measurements are discussed in Sections 1.1 and 1.2.

Omissions. Omissions from the sample design are hard to measure but are thought to equal about 6 percent of the country's homes in 1980. Omissions are discussed in Section 1.3. Occupants of the omitted homes tend to have lower income than average. Weights are adjusted to account for about two thirds of these omissions, but this is an imperfect solution. Section 1.3 describes the omissions.

Unanswered Questions. When people in the sample are not interviewed or omit some answers, the AHS assumes they are like people who do answer. This is probably wrong, and the error is unmeasured. About 5 percent of the sample each year is not interviewed (besides the 6 percent omitted from the sample, mentioned above). Chapter 2 discusses how the AHS adjusts weights for them. Further refusals, on individual questions, range up to about 15 percent on wages, and even 30 percent on business and farm income when those were separate question in 1974-83. The

Sample Status section describes allocations for these answered questions, and the Directory reports the missing answers on every question in every survey.

> Processing Errors. Processing errors can occur at places that use the data, and we have no way to measure these. Processing errors at Census include programming mistakes, and mistakes in tracking or keying the data. Programming mistakes are removed when they are found, and appear to be rare. Keying errors are measured regularly and average one in 10,000 numbers.

1.1 Discrepancies in People's Answers in the AHS.

People who do not understand questions, answer wrongly, or are recorded inaccurately by the interviewer, can lead to errors in the data. Therefore, Census conducts periodic studies to determine the extent of the problem. Recently, the answers to selected questions provided by a sample of 6,268 households to the 1987 National Survey were compared to the answers provided by the same respondents in 1985. Households with different answers in both years were asked during the 1987 interview to explain the discrepancies. Table 1.1A below presents the results of the study.

Besides that two year comparison of 1985 and 1987, Census conducts a short second interview within four weeks off the first, at two thousand or so units in each survey. By telephone, an experienced interviewer tries to talk to the same respondent who talked to the first interviewer. Different answers imply that someone made a mistake in at least one of the interviews. However, people who give the wrong answer both times cannot be measured. Table 1-1B shows the reason for discrepancies found in each reinterview for the 1985 and 1986 metro surveys. Table 1-1C shows the rate of discrepancies, though not the reasons, for a much longer list of variables, for many different years, so their importance for your work can be judged. For example, the first line shows different reporting of tenure between the original interview and the reinterview: One percent of all households changed tenure. In particular, one percent of the owners were re-classified as renters, and two percent of the renters were re-classified as owners. The two interviews asked about tenure within four weeks of each other, so an actual change in tenure would be rare. The differences may be simple misunderstandings. They may also be ambiguous cases (such as a property loaned by a relative, which should be called rental).

The reinterviews measure some of the error, but they do not catch people who answer both questions wrong. Also, it is unlikely that errors in different directions cancel. For example, with a variable like kitchens, 99 percent of households have kitchens, so even a very small proportion who misunderstand the question, or give a wrong answer can greatly increase the number who appear to lack kitchens. For any rare items, like kitchens, even a small error can create substantial over-estimates, and the following errors are not always small.

Table 1.0
UNITS OMITTED FROM SAMPLE OR NOT ANSWERING 85N AHS
All These Are Adjusted For In Weight

| | OCCUPIED | | | VACANT | |
	Total	Blacks & Hispanics	For Rent*	Other Year Round	Seasonal
All Units	8%	16%	-30%	3%	30%
New Construction (81-85)	22%	31%	-13%	13%	41%
New Mobile Homes (81-85)	33%	36%	NA	24%	52%

* In this column, AHS has over-coverage, more units than the control total.
  Source: Comparison of variables PWT and weight.

To account for the incomplete response rate, the confidence interval is about $\pm$ .01xPxQx (100/R-1) where R is the response rate, P is the percent in a category, and Q is (100-P). For example, for a trait that occurs in 20% of all units: $\pm$ .01x20x80x (100/92-1) = 1.4%.

Table 1.1A

Discrepancies Found Between 1985 and 1987 Out of 6268 Households Examined

TENURE

	Reason
Purchased since 1985	21
SOLD, now renting	4
Began charging rent since 1985	1
Stopped charging rent since 1985	2
1985 answer wrong	42
1987 answer wrong	41
Other	38
	149

BASEMENT

	Reason
Built under house	3
Old basement filled in	1
House is split-level, don't know what to call it	17
Have a partial basement, don't know what to call it	18
Walkout basement, don't know what to call it	0
Shallow basement, don't know what to call it	2
1985 answer wrong	305
1987 answer wrong	349
Other	60
	755

BEDROOM

	Reason
Another room converted	144
Addition added	34
Bedroom now used for something else	219
Part of house/apt. merged	4
Attic or basement finished	19
1985 answer wrong	127
1987 answer wrong	164
Other	61
	772

BATHROOM

	1st Reason	2nd Reason
Half converted to full	15	0
Added in addition	52	0
Space converted	7	0
Some/all fixtures removed	5	0
Destroyed in merger	0	0
1985 answer included half bathrooms	6	1
1987 answer included half bathrooms	6	0
1985 answer wrong	253	4
1987 answer wrong	152	1
Other	29	2
Refused	1	
	526	8

FUEL

	Reason
Fuel used less often in 1985, now more	152
New/converted equipment uses other fuel	87
1985 answer wrong	133
1987 answer wrong	155
Other	83
Refused	4
	614

HEATING EQUIPMENT

	1st Reason	2nd Reason
Old equipment replaces	80	0
Types used less 1985, now more	150	3
Installed since 1985	36	1
1985 answer wrong	359	2
1987 answer wrong	480	2
Other	80	5
Refused	11	
	1,196	13

RENT

	Pd Monthly		Pd Yearly	
	1st	2nd	1st	2nd
Major alterations/improvements	6	0	1	0
Conversion or merger changed size of unit	0	0	0	0
Disaster/partial demolition changed size of unit	0	0	0	0
No longer rent controlled	1	0	0	0
Now rent controlled	1	0	0	0
No longer subsidized	1	0	0	0
Now subsidized	6	0	0	0
Owner raised/lowered rent	76	0	5	0
1985 answer wrong	12	5	4	1
1987 answer wrong	10	0	4	1
Other	33	5	3	2
Refused	1	1	1	0
	147	11	18	4

VALUE

	1st Reason	2nd Reason
Major alterations/improvemts	89	13
Disaster/demolition	0	1
Sold/purchased land	3	0
Area more developed	68	23
Area had major disaster	3	1
Changes in the economy	253	54
Rezoning	4	1
1985 answer wrong	296	7
1987 answer wrong	77	4
Other	190	25
Refused	8	1
	991	130

Table 1-1B
Discrepancies Found During Reinterviews for the 1985 and 1986 Metropolitan Survey
(Holes/Cracks)

	1985		1986	
	Same Respondent	Different Respondent	Same Respondent	Different Respondent
Occupied Units				
Number of Interviews	1,194	160	1,801	343
Numbers of Discrepancies	74	14	65	10
Percent	6.2%	8.8%	3.6%	2.9%
Reasons Stated by Respondent				
Hole/Crack fixed/appeared since original interview	14		11	2
Original interviewer marked wrong answer	14	2	7	2
Original response was correct	5			
Respondent changed mind	10		8	
Respondent changed mind on size of hole/crack	6			
Respondent confused about question	6			
Qualifying statement in ( ) not read originally	4		6	
Other reasons	11	6	8	2
No reason given	4		19	2
Original respondent forgot		4		
Different respondent		2		2
Respondent does not remember			6	
Vacant Units				
Number of reinterviews	128	55	177	71
Number of discrepancies	8	4	10	3
Percent	6.3%	7.3%	5.6%	4.2%
Reasons Stated by Respondent				
Holes/Cracks fixed/approved since original interview		1		
Respondent not sure of situation			2	
Original response correct	2			
No reason given	2		5	2
Other reasons	4		3	
Different respondent		3		1

Table 1-1C
Differences Found During Re-Interview, by Original Answer

	All Units	Owners	Renters	Vacant	Survey	Variable
Different tenure	1%	1%	2%	NA	81N	TENURE
Different occupied/vacant status	3	2	4	4	81N	ISTATUS
Different unit visited	.4				81N	NA
Different unit visited	.2				78N	NA
Different household composition	1.0				81N	NA
Different household composition	1.5				78N	NA
Different birthdate	6				78N	ZAGE
Different age	5				78N	ZAGE
Different move date	3				78N	MOVED

	All*	Yes	No	Don't Know	Survey	Variable
Air conditioned	6%	7%	6%		80N	AIR
To reduce central air use:						
Room unit	1	50	1		80N	RARUNT
Awnings	4	50	3		80N	RAAWNS
Dehumidifier	9	50	5		80N	RADHMD
Ceiling fan	5	29	3		80N	RACFAN
Attic fan	6	24	5		80N	RAATFN
Window fan	4	44	3		80N	RAWNFN
Portable fan	15	25	12		80N	RAPOFN
Nothing	23	24	23		80N	RANONE
Added wood/coal stove	3	61	1		80N	HA02WS
Added fireplace	1	67	1		80N	HA03FP
Added portable electric heater	5	59	3		80N	HA04EH
Added unventilated kerosene heater	1	86	.3		80N	HA05UV
Added other heater	1	69	1		80N	HA06OT
Added no heater	10	5	58		80N	HA07NO
Have fireplace/stove	6	9	5		80N	FRPL
Fire/stove works	3	2	38		80N	FPLWK
All wood bought	14	26	9		80N	BUYFUL
Had job last week	7	6	7		80N	IFJ1
Public transportation besides car	1	55	1		80N	PUBTR1
Car besides public transportation	7	43	2		80N	CARTO1
Same work place daily	5	3	30		80N	HJOB1
Garage or carport	5	5	6		78N	GARAGE
Piped water in building	40	0	54		77N	WPIPED
Had to use extra heat sources	8%	44%	5%		77N	HADDL
Had to use extra heat sources	9	61	.5		76N	HADDL
Heating breakdown	6	54	4		77N	IFCOLD
Heating breakdown	5	40	2		76N	IFCOLD
Closed unheatable rooms	5	47	3		77N	IFCLSD
Closed unheatable rooms	4	60	2		76N	IFCLSD
Interior open cracks/holes	5	49	2		77N	CRACKS
Interior open cracks/holes	5	51	3		76N	CRACKS
Holes in floors	2	35	1		77N	HOLES
Holes in floors	2	58	1		76N	HOLES
Seen mice or rats	9	40	4		76N	RATS
Basement	5	5	4		76N	CELLAR
Basement leak	15	27	10	38	76N	BLEAK
Electric plug in every room	3	2	49		76N	PLUGS

* "All" means applicable households. For example, piped water was only asked at occupied homes, not vacant.

Table 1-1C (continued)
Differences Found During Re-Interview, by Orginal Answer

	All*	Yes	No	Don't Know	Survey	Variable
All wiring concealed	3	2	75		76N	NOWIRE
Attic or roof insulation	28	11	40	55	76N	INSUL
Thru other bedroom to bath	10	32	5		76N	PRIVB
Thru bedroom to other room	6	50	2		76N	PRIVN
13+ shares bedroom with 2 others	19	14	29		76N	NOPRIV
Blown fuses	10	51	5	100	76N	IFBLOW
Garbage collection	7	4	14	100	76N	TRASH
Mobile home loans	22	17	27		75N	MLOAN
Mortgage	1	4	2		75N	MORT
Water stopped 6+ hours	13	11	5	75	75N	BADDRY
Roof leaked in last 3 months	5	29	2	42	74N	RLEAK
Roof leaked in last 3 months	5	28	2	51	73N	RLEAK
Main reason for move	15	NA	NA		73N	WHYMOVE

	All	One	Two	Three	Four or More	Survey	Variable
Number of carpool	17	NA	11	37	46	80N	PASS1
Number of rooms	3	22	30	14	1	78N	ROOMS
Number of bedrooms**	6	4	5	6	8	78N	BEDRMS
Number of bedrooms**	5	6	5	4	7	77N	BEDRMS
Heating breakdowns	22	15	40	0	50	77N	NUMCOLD
Heating breakdowns	26	20	50	25	40	76N	NUMCOLD

	All	None	Two	Three	Four or More	Survey	Variable
Cars owned or used	14	13	10	19	26	80N	CARS
Cars owned or used	8	8	5	9	13	77N	CARS
Cars owned or used	6	6	4	8	5	73N	CARS
Trucks owned or used	9	4	15	37	18	80N	TRUCKS
Trucks owned or used	5	3	8	21 (2+)		77N	TRUCKS
Rooms without heating ducts	11	5	57	52	29	77N	NUMND
Rooms without heating ducts	85	6	57	54	34	76N	NUMND
Blown fuses	17	NA	16	30	9	76N	NUMBLOW

	All	Exclusive Use	Shared	No	Survey	Variable
Complete kitchen	1	.3	88	14	78N	KITCHEN
Complete kitchen	1	.2	NA	26	77N	KITCHEN
Complete kitchen	1	.3	89	11	75N	KITCHEN
Complete plumbing	1	.2	33	19	77N	PLUMB
Complete plumbing	1	1	46	23	74N	PLUMB

	All*	Ex.	Good	Fair	Poor	Survey	Variable
House rating 2+ points difference	2%	2%	.3%	4%	8%	77N	HOWH
House rating 2+ points difference	2	2	.4	5	10	76N	HOWH
House rating 2+ points difference	1	1	.2	3	10	75N	HOWH
House rating 2+ points difference	1	1	.4	2	9	74N	HOWH
Neighborhood rating 2+ points difference	2	2	.1	3	39	77N	HOWN
Neighborhood rating 2+ points difference	2	2	.4	4	16	76N	HOWN
Neighborhood rating 2+ points difference	2	3	0	8	19	75N	HOWN
Neighborhood rating 2+ points difference	1	1	.1	2	11	74N	HOWN
Neighborhood rating 2+ points difference	1	1	.8	3	1	73N	HOWN

* "All" means applicable households. For example, piped water was only asked at occupied homes, not vacant.
** Not clear what efficiencies are.

Table 1-1C (continued)
Differences Found During Re-Interview, by Original Answer

	All	Have Cond.	Do Not	Don't Know	All** with Cond.	No Bother	Little Bother	Much Bother	Want Move	Survey	Variable
Street noise	19	32	14		5	5	3	11	10	77N	STRN
Heavy traffic	16	27	12							77N	TRAF
Streets need repair	15	44	8							77N	ROAD
Snow blocks road	12	48	7							77N	SNOW
Poor street lighting	17	29	13							77N	STRL
Neighborhood crime	12	41	6							77N	CRIME
Littered streets/lots	13	48	6							77N	JUNK
Boarded/abndned bldgs	5	31	3							77N	ABAN
Rundwn occup. homes	8	45	5							77N	DUMP
Non-residential act.	18	39	14							77N	NONRES
Odors	8	49	4							77N	ODOR
Plane noise	13	29	10							77N	AIRN
Unsatis. public trans.	28	31	20	61						74N	TRN
Unsatisfactory schools	14	42	7	50						77N	SCH
Neighbrhood shopping	13	43	8	100						77N	SHP
Police protection	85	50	6	68						77N	FUZZ
Recreation facility	24	43	14	65						77N	RECR
Hospitals/clinics	18	48	11	61						77N	HOSP

	All Renters	Utility paid by Household	Included in Rent	Not Used	Survey	Variable
Different payee for:						
Electricity	2	2	8	0	81N	PAYE
Gas	13	3	26	20	81N	PAYG
Other fuels	17	17	47	11	81N	PAYO
Water	3	10	2	NA	81N	PAYW
Garbage	3	19	1	NA	81N	PAYT

	All Owners	Utility paid by Household	Not Used	Survey	Variable
Electricity	.2	0	40	77N	BUYE
Gas	1	.5	2	77N	BUYG

	All	Ducts	Heat Pump	Radia-tors	Built in Electric	Floor or Wall Furnace	Room Heaters Vented	Room Heaters Unvented	Fireplace Stove or Portable	None	Survey	Variable
Main htg.	16	11	27	15	13	26	38	21	33	40	80N	HEQUIP
Main htg.	13	6	53	9	18	26	43	21	28	46	77N	HEQUIP
Main htg.	7	3	NA	4	8	10	19	19	14	18	75N	HEQUIP
Main htg.	3	4	NA	7	8	15	18	14	30	0	74N	HEQUIP

	All*	None	Gas Piped	Bottled	Oil	Kero	Elec	Coal Coke	Wood	Solar	Other	Survey	Variable
Main htg. fuel	7%	18%	5%	9%	6%	27%	14%	0%	17%	NA	25%	78N	HFUEL
Main htg. fuel	5	NA	3	19	6	50	5	15	16	NA	100	77N	HFUEL

	All	Wood	Coal	Other	None	Survey	Variable
Fire/stove fuel	9	3	17	25	44	80N	FPLFUL

	All	Central	Room Units	Survey	Variable
Type of air conditioning	3	2	4	80N	AIRSYS

"All" means applicable households. For example, piped water was only asked at occupied homes, not vacant.
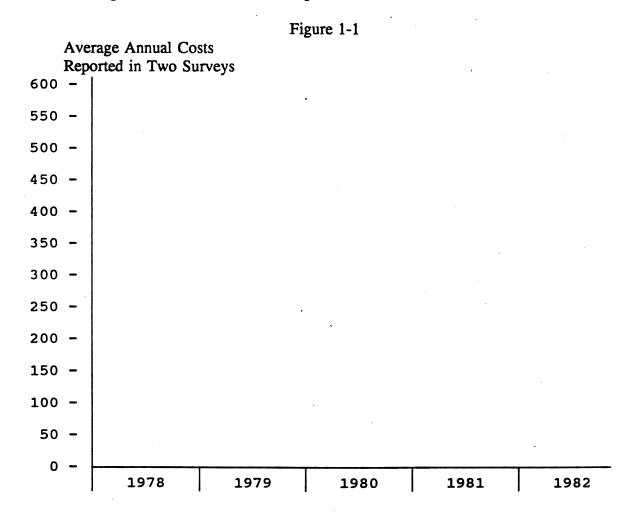** Different by two or more points.

Table 1-1C (continued)
Differences Found During Re-Interview, by Original Answer

Commute in	All	Car	Truck	Van	Survey	Variable
	6	3	14	48	80N	VEHCL1

Commute by	All	Drive alone	Pool	Bus	Sub-way	RR	Taxi	Mcyc	Bike	Walks	Horse	Other	Survey	Variable
	10	5	18	23	19	44	0	27	18	21	26	100	80N	TRAN1

Why drive alone	All	Irreg hours	Irreg place	Need car	Know no one	Like privacy	No detour	No waiting	Emer/ o'time	Mistrust Other drivers	Other	Survey	Variables
	52	41	67	54	39	71	84	86	94	75	77	80N	ALONE1

Why not pub trans	All	Prefer car	Far fm. work	Slow	Bad Schedule	No pub. trans	Far fm. home	Cost	Need car	Handi-cap	Other	Survey	Variables
	43	65	80	82	76	22	89	67	53	0	77	80N	NOPUB1

Why use pub trans	All	No license	No car	Cheap	Parking	Driving strain	Fast	Other	Survey	Variables
	33	0	20	21	75	100	71	50	80N	WMEAN1

Number of bathrooms	All	Diff. rooms	1 Bathrm	1 1/2 w/o Tlt	1 1/2 w/Tlt	2	Over 2	Survey	Variables
	11	7	3	83	20	15	20	77N	BATHS

Water source	All	System	Well	Other	Survey	Variable
	2	1	3	13	78N	WATER

Sewage disposal	All	System	Septic	Privy	Other	Survey	Variable
	2	1	4	4	100	78N	SEWDIS

	All*	Everywhere	Some	None	Survey	Variable
Storm windows	12%	10%	37%	9%	76N	STORM
Storm doors	15	11	41	9	76N	STORM

3 people/bedroom	All	In 1 Bedrm	In 2 Bedrms	No	Survey	Variable
	3	21	67	1	76N	IF3BED

Different seasonality	All	Year round	Migratory	Summer only	Winter only	Other Seasonal	Survey	Variables
	12	6	67	33	NA	0	80N	VACANCY

Reason for Vacancy	All	For rent	Reg. sale	Sale condo	Sale coop	Not occupied rented	Not occupied sold	Occas. use	Other	Survey	Variables
	31	29	6	33	NA	67	50	35	25	80N	VACANCY

Preferred area in 5 years	All	Here	Elsewhere	Survey	Variables
	20	16	27	80N	WHRINS

"All" means applicable households. For example, piped water was only asked at occupied homes, not vacant.
** Different by two or more points.

Table 1-1C (continued)
Differences Found During Re-Interview, by Original Answer

	All households with any condition on orig. or re-interview	Have this condition	Don't	Survey	Variable
Asthma	8	65	3	78N	ASTHMA
Tuberculosis	0	50	0	78N	TUBERC
Chronic bronchitis	4	58	2	78N	BRONCH
Emphysema	2	47	1	78N	EMPHY
Other lung problem	3	74	1	78N	LUNGO
Heart attack	5	66	2	78N	ATTACK
Other heart trouble	10	69	4	78N	HEARTO
Stroke	1	42	1	78N	STROKE
Arthritis or rheumatism	26	55	16	78N	ARTH
Convulsions or epileptic seizures	1	64	1	78N	EPILEP
Cerebral palsy	0	100	0	78N	PALSY
Deaf or serious trouble hearing	6	68	3	78N	DEAF
Blind or serious trouble seeing	6	80	2	78N	BLIND
Missing legs, feet or toes	0	50	0	78N	LEGSM
Missing arms, hands or fingers	1	67	0	78N	ARMSM
Chronic limb stiffness or deformity	4	76	2	78N	LIMBST
Chronic back stiffness or deformity	4	88	2	78N	BACKST
Other trouble with back or spine	11	85	4	78N	BACKO
Paralysis	1	71	0	78N	PARAL
Other conditions	17	76	9	78N	CONDOR

	1 problem	2 problems	3+ problems	Survey	Variable
Number of conditions 73	60	79	93	78N ·	NA
Number of difficulties 89	81	100		78N	NA

	All households with any condition on orig. or re-interview	Have this difficulty	Don't	Survey	Variable
Hard to go in/out of house	31	65	21	78N	INOUT
Hard to get around inside	24	92	18	78N	AROUND
Hard to get up/down stairs	60	79	49	78N	UPDOWN
Hard to use bathroom/kitchen/etc.	14	100	11	78N	USEQUI

1.2 Comparison with Other Data

In addition to re-interviews, which repeat the survey's own questions in order to find errors, we can compare some AHS items to more accurate data from other sources. Three comparisons are shown here, covering utility costs and income. We are not aware of comparisons of other subjects. The data below suggest that AHS utility costs are a little high, and incomes a little low. Therefore, rent to income ratios are also too high in the AHS, on average. The details of these comparisons are discussed below.

Figure 1-1

Average Annual Costs
Reported in Two Surveys

| | 1978 | 1979 | 1980 | 1981 | 1982 |

Source: Energy Information Administration, Consumption Expenditures, April 1981 through March 1982, Part 1: National Data, Washington, Government Printing Office, 1983 (and earlier editions), and HUD special tabulations.

The discrepancy is fairly consistent over time, and data not presented here show it is also consistent for single family detached homes. Studies do show that 1980 Census data are even more over-stated, and vary greatly from area to area.

Independent estimates of income from GNP accounts, the Social Security Administration, the Veterans Administration, and so forth are shown in Table 1-2 AHS figures are lower than the independent estimates for total income and for every category other than self-employment income. The Current Population Survey (CPS) is done by the Census Bureau for the Labor Department. It is also low but comes closer to the independent estimates. When income is asked in CPS it is a major part of the questionnaire, while it is only a small part of the AHS questionnaire.

Table 1-2
Money Income of All U.S. Households
Billions of Dollars

	Independent Estimate	CPS	AHS	AHS as % of Ind. Est.
Total money income	$2,403*	$2,201	$2,073	<86
Wages or salaries	1,632	1,161	1,505	92
Interest	221	99	67	30
Soc. Sec., RR ret.	155	142	139	90
Nonfarm self-employment	104	120	142	137
Dividends	60	27		
Estates & trusts	NA	7	38**	<63
Fed. & mil. retirement	35	32		
State & local govt ret.	21	13	33	59
Private pensions & anns.	55	35	27	49
Net rent & royalties	34	17	23**	68
Unemployment comp	26	20	18	69
AFDC	14	11		
SSI	9	8		
Other public assistance	NA	2	17	<74
Workers' comp	14	7	5	36
Veterans' payments	14	9	13**	93
Farm self-employment	9	10	25	278
Alimony & child support	NA	8	8	
Reg. contrib. from people	NA	5	5	
Other money income	NA	14	9	
12 months ending	12/83	12/83	10/83	

* Excludes 5 categories, shown as NA. There are other differences such as the exclusion of children's income (0-14) from CPS and AHS, military households from CPS, and group quarters from AHS.

** AHS comes closer to independent estimate than CPS does. This is considered desirable, but even the independent estimates contain unknown amounts of errors.

Source: Census Series P-60 No. 151, p. 170 and HUD special tabulation. (Since the AHS public use tape does not distinguish among amounts of $50,000 or more, they have each been treated as $60,000.)
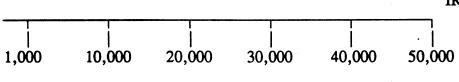
In addition, Census Bureau staff have compared the 1973 CPS to individual tax returns from the same people, using social security numbers to match the data. Some people of course did not submit tax returns, usually because their income was too low. Figure 1-2 shows people who appeared in both CPS and IRS records.

Figure 1-2
Distribution of Tax Filers, at Each Income, by
Difference between IRS and CPS wages

IRS over 50% above CPS
IRS 15-50% above CPS
IRS 5-15% above CPS

IRS = CPS ± 5%

IRS 5-15% below CPS
IRS 15-50% below CPS
IRS 50-100% below CPS

1,000   10,000   20,000   30,000   40,000   50,000

CPS Wages (at 1986 Prices)

Source: Drawn from data in: Roger A. Herriot and Emmett F. Spiers, "Measuring the Impact on Income Statistics of Reporting Differences between the Current Population Survey and Administrative Sources" in Social Security Administration, Studies from Interagency Data Linkages Report 11, SSA, 1980, Table 6.

The large area in the middle of the graph represents families where IRS data are within 5 percent of CPS. This area only covers about half the families. For other families, IRS information is progessively farther from CPS. The bands at the top are wider than those at the bottom, reflecting more families where the IRS is above the CPS than below. There is reason to believe that IRS data are more accurate than CPS, because taxpayers used their W-2 forms when subitting to IRS, and there are penalties for mistakes. Therefore, each income category in CPS (or AHS) data includes families whose incomes are really higher or lower than that category.

The study was done with 1972 income; current data would be expected to be similar. AHS is similar to CPS, so the findings probably apply to AHS also. The graph covers wages, salaries, tips, and overtime reported to CPS. Some people do not tell CPS (or AHS) what their wages were. Their wages are allocated on the tape, and would be farther from IRS than most of the cases shown here.

1.3 Omissions from the AHS Sample Design

The sample design is fully discussed in Chapter 3. This section just summarizes the main gaps, so if you are studying one of these areas, you can be aware of the AHS' weakness.

Some new construction units are missed. The sample of building permits is selected several months before the end of the survey, so the most recent spring and summer of building permits are missed in each national survey. Summer and fall are missed in each SMSA survey. The total is adjusted by weighting, but detailed characteristics for the most recent year may well be wrong. New constructions occuring just before the most recent Census may also be missed, since they can be accidentally screened out of the sample.

Other new housing is also hard to sample: conversions from non-residential buildings, new locations for mobile homes, etc. Finally there are three types of illegal housing listed at the bottom of the table. Some people hide from Census interviewers if their situation is illegal, even though the survey is confidential.

Table 1-3
Types of Homes Covered poorly by AHS

	Surveys with Problem	
	National	SMSA
New homes built just before survey was over or just before Census was taken	All	All
Homes created in non-resdntial bldgs after 3/70	78N-83N	All
Homes created in already-residential buildings after 3/70		All but 76S-81S
Mobile home parks founded after 3/70	73N-75N 77N-80N	All but 76S-78S
Mobile homes outside parks and installed after 3/70 or vacant then	73N-83N	All
Conventional homes moved onto new sites after Census	All	All
Homes created on Federal land (e.g., military) after 3/70	All	All
Homes on boats, RVs, etc.	All	All
Extra apartments in houses zoned single family	All	All
Homes built w/o permits where bldg permits req.	All	All
Households of illegal aliens	All	All

2.0 SAMPLING ERROR

The AHS does not cover every home in the United States. A large random sample was selected, and it should be representative, but there is always a chance that the people in the sample could be a little different from the average.

2.1 Error Tables

Tables 1-4 and 1-5 show by how much the AHS sample may differ from the rest of the country.

Table 1-4

When the national AHS shows one of these numbers:	Then the odds are 95 out of 100 that the sampling error in the whole country is within plus or minus:	
0	5,000	4,000
25,000	16,000	13,000
100,000	31,000	27,000
250,000	49,000	42,000
500,000	70,000	60,000
1,000,000	98,000	85,000
5,000,000	215,000	185,000
10,000,000	295,000	254,000
25,000,000	423,000	364,000
50,000,000	477,000	410,000
	Use this column for numbers on: Hispanics Mobile homes Built since 1970 Incomplete kitchen or plumbing Zero bedrooms or bathrooms	Use this column for all other topics.

Note: For mobile homes in the South Region, multiply these national errors on mobile homes by 1.3. For mobile homes in the West, multiply by 1.8. For other regional numbers, use the figures in the table without adjustment.

For example, if the AHS shows 100,000 units of a certain type, then this table shows that the sampleing error can be as large as 31,000. Therefore, the true number of units of this type can be anywhere between 69,000 and 131,000. It should be noted that the right hand column applies to most topics. Also, remember that errors caused by the sample are in addition to other errors such as the errors found by re-interviews, as discussed above.

Table 1-5 shows sampling errors when a percentage is calcualted from the national AHS:

Table 1-5

When the base of the percent is:	and AHS shows that a category is....			
	2% or 98%	10% or 90%	25% or 75%	50%
	...of the base, then the odds are 95 out of 100 that the sampling error is within plus or minus (in percentage points):			
5,000	53.6	53.6	53.6	49.9
10,000	31.1	31.1	36.8	42.5
25,000	13.7	16.1	23.3	26.9
50,000	7.1	11.4	16.5	19.0
100,000	3.8	8.1	11.6	13.4
250,000	2.4	5.1	7.4	8.5
500,000	1.7	3.6	5.2	6.0
1,000,000	1.2	2.6	3.7	4.3
2,500,000	.8	1.6	2.3	2.7
5,000,000	.5	1.1	1.6	1.9
7,500,000	.4	.9	1.3	1.6
10,000,000	.4	.8	1.2	1.3
25,000,000	.2	.5	.7	.9
50,000,000	.2	.4	.5	.6
75,000,000	.1	.3	.4	.5
90,000,000	.1	.3	.4	.4

Note: For the special topics listed in the last table, multiply these errors by 1.2. For mobile homes in the South Region, multiply by 1.5. For mobile homes in the West, multiply by 2.

For example, suppose the AHS shows 9,000,000 black households, and 45.0 percent of them own their own homes. To measure the error in this percent, we go to '10,000,000' and '50%' in the table, and find an error of '1.3'. This means the true percent may be 45.0 plus or minus 1.3, so black ownership is likely to be between 43.7 and 46.3 percent.

The tables above measure sampling errors in national AHS surveys. The sampling errors in metropolitan surveys are smaller, but vary widely from place to place. When you obtain a number from a metropolitan survey, find the square root of that number, mulitply this square root by the factor in Table 1-6, and you will have the sampling error. (Error = (factor from Table 1-6) x number from AHS). For example, the factor for Albany in 1980 is 16. If the Albany area had 40,000 homes of a particular type in 1980, the square root is 200, multiplied by 16 is 3,200, so the chances are 95 out of 100 that the true number is between 36,800 and 43,200.

When you obtain a percent (P) from a metropolitan survey, use the following formula:

$$\text{Error} = (\text{factor from Table 1-6}) \times \frac{P(100 - P)}{\text{base of percent from AHS}}$$

Table 1-6
METROPOLITAN ERROR FACTORS

Name	74	75	76	77	78	79	80	81	82	83	84	85	86	87
Albany-Schenectady-Troy	16			16			16							
Allentown-Bethlehem-Easton			15				14							
Anaheim-Santa Ana-Garden Grove	24			24				26						
Atlanta		16			17				27					
Baltimore			27			28				32				
Birmingham			16				17							
Boston-Lawrence-Lowell (1)	24			24				34						
Buffalo			21			20								
Chicago		28				29				43				
Cincinnati		21			22				24					
Cleveland			25			26								
Colorado Springs		10			10									
Columbus		17			18				21					
Dallas-Fort Worth (2)	24			24				28						
Denver			24			23				29				
Detroit	24			24				38						
Fort Worth (2)	17			17				19						
Grand Rapids			14				14							
Hartford		14				15				18				
Honolulu			15			15				18				
Houston (7)			18			20				42				
Indianapolis			20				20							
Kansas City		20			21				25					
Las Vegas			11			12								
Los Angeles-Long Beach	29			29			14							
Louisville			18				17			33				
Madison		10		10				10						
Memphis	18			18			17							
Miami-Fort Lauderdale (3)		24				24				29				
Milwaukee		21				21								
Minneapolis-St. Paul	25			25				27						
New Orleans		19			21				23					
New York			44				44			51				
Newark-Northeastern NJ (4)	24			24				29						
Nwprt News-Hampton-Norfolk (5)		19			11		17							
Oklahoma City			18				17							
Omaha			14			14								
Orlando	15			15				17						
Paterson-Clifton-Passaic (4)		19			21				23					
Philadelphia		21			26				44					
Phoenix	20			20				24						
Pittsburgh	28			28				31						
Portland, Oregon		19				20				26				
Providence-Pawtucket-Warwick			18				18							
Raleigh			9			10								
Rochester		17			17				19					
Sacramento			19				20			24				
Saginaw	**			8			8							
St. Louis			20				21			35				
Salt Lake City	13			13			14							
San Antonio		17			18				20					
San Bernardino-Riverside-Ontario		21			6				29					
San Diego		24			6				29					
San Francisco-Oakland					21				40					
San Jose														
Seattle-Everett-Tacoma (6)			14			14				33				
Spokane	10			10				12						
Springfield-Chicopee-Holyoke		12			13									
Tacoma (6)	12			12				14						
Tampa-St. Petersburg														
Washington, DC	23			23				23						
Wichita	11			11				12						

Note: Each factor is approximately 1.96 times the square root of the average weight.

For example, if the AHS shows 96,000 renters in the Albany area, and 14.1 percent of them rent single family homes, the error is:

$$16 \times \frac{14.1 \times (100 - 14.1)}{96000}$$

or 1.8. The true percent is likely to be between 12.3 and 15.9.

This approach will also give more accuracy for national errors than Tables 1-4 and 1-5 above. Therefore, factors for various groups are provided in Table 1-7.

Table 1-7
National Error Factors for Selected Groups

Survey Year	73	74	75	76	(77-80)	81	83
National Tapes							
U.S.: Mobile Homes, NCPK*, Hispanics	76	87	90	92	95	104	99
U.S.: Other	87	86	75	76	82	89	85
South: Mobile Homes					126	138	131
West: Mobile Homes					166	182	173
NE, MW: Mobile Homes, NE, MW, West: NCPK					100	110	104
Regions: Other	NA	NA	NA	NA	85	92	88

* NCPK means New Construction, or incomplete Plumbing or Kitchens, except in 73-76, when it omits new construction. In 73 it also includes Blacks.

Note: Each metropolitan factor is about 1.96 times the square root of the average weight. National factors are larger than that, because of the clustered sample.

These tables of errors may meet the needs of most users. Sections 2.2 and 2.3 present two other methods for statisticians and econometricians, if they need more refined estimates of the standard error and confidence intervals.

2.2 Statistical Test From Computer Programs

Many computer packages calculate statistical tests and confidence intervals for you. First, it is important to be sure your computer package uses the correct sample size in calculating statistics. Some packages, including BMDP, SAS, and SPSS, wrongly treat the weighted count as a sample size; for example the package might assume there were 85 million households in the 1983 National AHS, rather than the actual count of 61,000 occupied homes. To tell if your package makes this mistake, you should divide all weights by the average weight. If this changes the standard errors, there is a mistake, and the ones after the division should be used.

Second, the packages usually assume simple random sampling. Due to the stratification and clustering of the AHS design, confidence intervals from statistical packages are too small. Fortunately there is also a simple adjustment for the sample design.

The "design effect" is number which is different for different variables, and is shown in Table 1-8. Any time you obtain a standard error from the computer, multiply it by the square root of the design effect. Alternatively you can correct standard errors: Find the largest design effect that applied to your analysis, and divide all weights by this number (i.e., take the original weight, divided by the average weight, then divided again by the design effect). This computes an effective sample size that adjusts for the complex AHS design. This calculation will give good values although it will slightly over-estimate significance tests with regressions. The package will assume degrees of freedom equal to the effective sample size minus the number of variables, while the actual degrees of freedom are about 50, base on how the design effect was calculated. (This discussion is indebted to Appendix 4 of Procedural Handbook: 1981-82 Mathematics and Citizenship/Social Studies Assessments, undated, from the National Assessment of Educational Progress, ETC, Princeton, NJ 08541-6710, (800) 223-0267 that cites several articles, including Kish and Frankel, "Inference from Complex Samples," J. of the Roy. Stat. Soc. Series B v.36, 1974, and other articles.)

The following rules may also help you:

The 95 percent confidence interval for the proportion P is:

$$P \pm \left[1.96 \times \left[\frac{P\,(1-P)}{\text{sample size}-1} \right]^{1/2} \times (\text{design effect})^{1/2} \right]$$

The 95 percent confidence interval for the difference between two proportions P and R is:

$$P - R \pm \left[1.96 \times \left[\frac{P(1-P) \times \text{design effect for P}}{P \text{ sample size} - 1} + \frac{R(1-R) \times \text{design effect for}}{R \text{ sample size} - 1} \right] \right]^{1/2}$$

The 95 percent confidence interval for a mean M is:

$$M \pm \left[1.96 \times \text{standard deviation} \times \left[\frac{\text{design effect}}{\text{sample size}} \right]^{1/2} \right]$$

The 95 percent confidence interval for the difference between two means L and M is:

$$L-M \pm \left[1.96 \times \left[L \text{ std. dev.}^2 \times \frac{L \text{ design effect}}{L \text{ sample size}} + M \text{ std. dev.}^2 \times \frac{M \text{ design effect}}{M \text{ sample size}} \right]^{1/2} \right]$$

For a Chi squared test of independence you should divide the usual chi squared statistic by the design effect.

Table 1-8
Design Effects for National and Regional Standard Errors

	US	NE	MW	S	W
Incomplete Kitchen,					
Plumbing or bathroom	2.61	2.18	2.61	2.18	2.61
New construction	2.30	2.58	2.58	1.88	2.30
Mobile Home, Individual Well	2.70	3.26	3.26	5.18	8.29
Other Topics, Depending Whether Cases Are:					
0 - 9% Rural	1.36	1.25	1.25	1.25	1.58
10 - 29% Rural (Metro Areas)	1.63	1.50	1.50	1.50	1.89
30 - 49% Rural (National Averages)	1.90	1.75	1.75	1.75	2.21
50 - 69% Rural	2.17	2.01	2.01	2.01	2.52
70 - 89% Rural (Non-Metro Areas)	2.45	2.26	2.26	2.26	2.87
90 - 100% Rural	2.72	2.51	2.51	2.51	3.15

Note: Design effect is the ratio of estimated variance in the AHS to variance in a simple random sample of the same size.

2.3 Formal Calculation of Variances

The confidence intervals in Sections 2.1 and 2.2 are approximately right, but actual variances depend a lot on individual questions. The Census Bureau uses detailed calculations to prepare the summary tables above, and you can use these calculations yourself if you wish. The following is a modification of what Census actually does on national AHS data.

This modification is necessary to reflect the information that is available in the AHS public use file. The calculation can be done weighting the cases with either **WEIGHT** or **PWT**. Use of **WEIGHT** will result in a very slight overestimate of variance, but the difference is negligible. Variance calculations for the MSA surveys are not presented here; the tables of variances shown in Appendix B of each published MSA report can be used.

The variance for an estimate is equal to the sum of the variance for the estimate from self-representing (SR) PSUs plus the variance for the estimate from non-self-representing (NSR) PSUs. The SR component reflects the effect of the sampling of clusters within the SR PSUs. The NSR component reflects the effects of the sampling of PSUs within the NSR strata as well as the sampling of clusters within the NSR PSUs.

Self-Representing PSUs

The technique for the SR PSUs is to group these PSUs into 46 relatively homogeneous groupings, then divide the cases in each group in half at random, prepare the estimate in each half, square the difference of these two estimates, and add up these squared differences across all 46 groupings. The process is carried out ten times and the results are averaged to get the final estimate of variance in SR PSUs. The repetitions and the averaging are not strictly required. In effect they reduce the variance of the estimate of variance. The SR variance component of an estimate of characteristic X is calculated using the following formula:

$$Var(X')_{SR} = \frac{\sum\limits_{r=1}^{10} \sum\limits_{s=1}^{46} (X'_{rs1} - X'_{rs2})^2}{10}$$

where: X' is the estimate of characteristic X,

E = Summation

r is a subscript identifying the half-sample replications used in this variance estimation. For AHS, these half-sample replications are formed by utilizing the panel numbers assigned to the AHS cases. Panel number can be found in the 4th digit (counting from the left) of the control number,

s is the subscript identifying the groupings of socio-economically similar SR PSUs used in this variance estimation. The general class of SR PSUs can be distinguished by a code 1.0000 in AWT. Then the cases in each socio-economically similar grouping are identified by unique codes in NCLUS (codes are scattered between 1001 and 2060),

X'_{rs1} is the estimate of characteristic X based on the AHS sample cases in the first half-sample of the rth replication in the sth group of PSUs,

and

X'_{rs2} is the estimate of characteristic X based on the AHS sample cases in the second half of the rth replication in the sth group of PSUs.

For AHS, the half-samples within each of the replications are defined as follows:

Replication	Panel Numbers Included In First Half-Sample	Panel Numbers Included In Second Half-Sample
1	1,2,3	4,5,6
2	1,2,4	3,5,6
3	1,2,5	3,4,6
4	1,2,6	3,4,5
5	2,3,4	1,5,6
6	2,3,5	1,4,6
7	2,3,6	1,4,5
8	1,3,4	2,5,6
9	1,3,5	2,4,6
10	1,3,6	2,4,5

Non-Self-Representing PSUs

The NSR variance component of an estimate of characteristic X is calculated using the following formula:

$$VAR(X')_{NSR} = \sum_{s=101}^{210} \left[A_{s3} \frac{X_{s1} + X'_{s2}}{2} - A_{s1}X'_{s3} \right]$$

$$+ \frac{21}{4} \sum_{s=101}^{210} (A_{s3})^2 [X'_{s1} - X'_{s2}]^2$$

where: X' is the estimate of characteristic X,

E = Summation

s is a subscript identifying the pairs of strata used in the sample selection. For AHS, s goes from 101 to 210 and it can be identified by the third through fifth digits (counting from the right) of RCLUS.

s1, s2, s3--These are subscripts identifying the individual PSUs within the sth pair of strata. These subscripts can be calculated by analyzing the second digit, counting from the right, of RCLUS, in combination with AWT. The cases with "1" in this digit are in PSU s1. Other cases with the same value of s and the same value of AWT as PSU s1 are in PSU s2. Remaining cases with the same value of s but a different value of AWT are in PSU s3. Note that this calculation depends on there always being three PSUs in each pair of strata. In instances where the third PSU sampled from the pair of strata is identical to one of the other PSUs (possible, because it was chosen independently), some cases from the PSU that was selected twice are identified as being in the "third" PSU.

X'_{s0} is the estimate of characteristic X based on the AHS sample cases in PSU 0 of the sth NSR Random Cluster.

X'_{s1} is the estimate of characteristic X based on the AHS sample cases in PSU 1 of the sth NSR Random Cluster.

X'_{s2} is the estimate of the characteristic X based on the AHS sample cases in PSU 2 of the sth NSR Random Cluster.

A_{s0} is the value of AWT associated with AHS sample cases in PSU 0 of the sth NSR Random Cluster.

A_{s2} is the value of AWT associated with AHS sample cases in PSU 2 of the sth NSR Random Cluster.

Special PSUs

Five cases on the 1981-83 files are outside the normal PSUs chosen for the AHS. These cases can be used for normal estimates, but must be ignored for variance calculations. They have codes of 9999 in NCLUS, 99999 in RCLUS, and 999999 in AWT. All other cases derived from the business sample fall in normal AHS PSUs, so they will be included in the calculations described above.

Chapter 2

WEIGHTS

1.0 WEIGHTS

Weights to prepare national or MSA estimates are provided on the AHS tapes and are used in preparing the numbers presented in the Census publications and custom analyses prepared by the AHSDP Project. These weights and the various adjustments used to compute the weights are discussed in detail below for the 1973 National Sample and MSA samples from 1974 through 1983. Since the Census Bureau had not finalized its definitions and computations of weights for the 1985 National Survey and upcoming MSA surveys, it was not possible to include the documentation of the new weights in this version of the Codebook. A section will be added to this Chapter and made available to users as soon as the documentation becomes available.

The variable WEIGHT is not simply the inverse of the probability of selection for each unit, but includes several adjustments, designed to adjust for random variation in the original selection of the NSR PSUs, to account for refusals and other missed interviews, and to make AHS estimates conform to Survey of Construction (SOC) estimates, Housing Vacancy Survey (HVS) estimates, and Current Population (CPS) estimates. CPS estimates themselves are adjusted to independent estimates of total population, based on census counts, National Center for Health Statistics data on births and deaths, and Census Bureau estimates of net migration. These adjustments change each year, so the variable WEIGHT also changes.

The variable "pure weight," (PWT) is more appropriate than WEIGHT for longitudinal analyses. It is the inverse of the probability of selection and incorporates none of the adjustments described below. It changes only because of formal sample reductions. For example, the 7/97 reduction in 1977 increased the weight of remaining units by 97/90ths. Otherwise it is invariant over time.

1.1 Adjustments

1.1.A Non-Interview Adjustment

Type A non-interviews include refusals and other situations where data should have been but were not collected. The weights of these units are set to zero, and weights of responding units are correspondingly increased by the following ratio to represent the type A non-interviews:

$$\frac{\text{Interviewed units + Type A non-interviews}}{\text{Interviewed units}}$$

This approach assumes that non-responding units (about 5 percent of the total) would have responded like the others. This ratio is calculated by weighting each unit at the inverse of its probability of selection. It is done separately for each of the cells in Table 1-1, in each

Census region. However, if there are less than 30 cases in a cell, or if the ratio in the cell would be 1.5 or more, the cell is combined with the cell having the closest scale value, as shown in Table 1-1. Such combinations continue until both conditions are met.

Table 1-1

NON-INTERVIEW ADJUSTMENT CELLS AND SCALE VALUES,
NORTHEAST REGION (repeated for each region)

	MSA			Non-MSA	
	Central City of MSA	Balance Urban	Balance Rural	Urban	Rural
Occupied					
Permit Segments	1	3	4	101	103
Area Segments					
Mobile Homes	51	61	64	151	161
Non-Mobile Homes	21	31	34	121	131
Address & Other Segments					
Mobile Homes	50	60	63	150	160
Bon-Mobile Homes	20	30	33	120	130
Vacant & URE					
Permit Segments	301	303	304	501	503
Area Segments	321	341	351	521	541
Address & Other Segments	320	340	350	520	540

1.1.B PSU Adjustment

This adjustment is used only in NSR PSUs. It takes into account the differences that existed at the time of the 1970 Census between the sampled NSR PSUs and all other NSR PSUs. This difference arises purely from the natural variability of samples, and is adjusted by multiplying the following fraction times the inverse of the probability of selection for each unit:

$$\frac{305}{\displaystyle\mathop{\text{E}}_{i=1}}\quad\frac{\text{1970 Census count of housing units in ith NSR PSU in sample}}{\text{Probability of selection of ith NSR PSU}}$$

It is done separately in each cell listed in Table 1-2. In 1983, the factors were recalculated using 1980 Census counts.

It might be though that this ratio could be multiplied by the pure weight to yield an adjusted weight that would be invariant over time and would be useful for longitudinal analysis. However, the appropriate ratio for a case depends on its tenure, which is not invariant, so the adjusted weight would not be invariant and would not be useful for longitudinal analysis.

Table 1-2

PSU ADJUSTMENT CELLS

	Region			
	Northeast	North Central (Midwest)	South	West
Occupied HUs				
MSAs				
Central City: Owner				
Renter				
Balance Urban: Owner				
Renter				
Balance Rural: Owner				
Renter				
Vacant HUs (including UREs)				
MSAs				
Central City				
Balance				
Non-MSAs				
Urban				
Rural				

1.1.C New Construction Adjustment

This adjustment is used first for units where the value of BUILT is April 1, 1970 or later. It should be noted that units derived from the sample of building permits are forced to have a date of April 1, 1970 or later. If they have an earlier date, it is changed to match the date when the current reference person moved in, unless that date is also before April 1, 1970, in which case both are arbitrarily changed to January 1974. After these edits, the weights for all units built after April 1, 1970, whether from permit samples or area samples, are adjusted by the following ratio:

$$\frac{\text{Survey of Construction estimates of units built 4/1/70 or later}}{\text{AHS estimate of units built 4/1/70 or later}}$$

This ratio is calculated using the inverse of the probability of selection of each unit, adjusted by the non-interview adjustment and the PSU adjustment.

Note that the denominator includes units in the current AHS sample and also cases dropped from the sample, because of demolitions, mergers, etc., as long as they were built April 1, 1970 or later. The SOC data used are total construction as of the end of the interviewing period. The ratio is calculated separately in each of the cells in Table 1-3.

Table 1-3

NEW CONSTRUCTION ADJUSTMENT CELLS

Date Built	Non-Mobile Homes	
	1 Unit	2+ Units
April 1970 - October 1973		
November 1973 - October 1974		
November 1974 - December 1975		
January 1976 - December 1976		
January 1977 - January 1978		
February 1978 - January 1979		
February 1979 - December 1979		
January 1980 - December 1980		
January 1981 - December 1981		

The above adjustment makes AHS estimate of total cumulative construction match SOC estimates. The occupied new construction units in the AHS are then counted, using these new adjusted weights. The resulting estimate of occupied new construction is subtracted from the Current Population Survey (CPS) estimate of total occupied units, to produce an

estimate of occupied old units. The following ratio is then calculated:

$$\frac{\text{Estimate of occupied old units}}{\text{AHS estimate of occupied units built before 4/1/70}}$$

This ratio is calculated using weights as adjusted by all previous calculations. It is done for one cell consisting of all occupied units. In 1983 the denominator of the ratio was changed to occupied units built before 4/1/80.

The result of this procedure is that if SOC misses some newly built units (which is possible in any survey), but CPS picks them up (which is possible because CPS totals are forced to match control totals based on vital statistics registrations and migrations), then AHS considers these extra households to be living in old construction, since new construction is matched to SOC, while total units are matched to CPS.

The adjustment process so far makes total occupied units match CPS counts, but does not necessarily make types of units match. That is done by the occupied units adjustment explained below.

1.1.D Vacant and URE Units Adjustment

This adjustment is used only for vacant and URE units. The effect is to match Housing Vacancy Survey (HVS) data on the percent distribution of types of vacancies while preserving AHS data on the total number of vacancies. The weights resulting from the previous steps are adjusted by the following ratio:

$$\frac{\dfrac{\text{HVS estimate of vacant and URE units in a cell}}{\text{HVS estimate of vacant and URE units in all cells}}}{\dfrac{\text{AHS estimate of vacant and URE units in the cell}}{\text{AHS estimates of vacant and URE units in all cells}}}$$

This ratio is calculated using weights adjusted by all the previous calculations. It is calculated separately in each of the cells in Table 1-4. However, if there are less than 50 cases in a cell, or if the ratio for the cell would be 2.0 or greater or 0.2 or less, the cell is combined with the cell having the closest scale value, as shown in the table. Such combinations continue until both conditions are met. The HVS data used are an average of the third and fourth quarters of the calendar year.

Table 1-4

VACANT AND URE UNITS ADJUSTMENT CELLS & SCALE VALUES

Year-round Vacant + UREs
  For Rent   900
  For Sale Only   901
  Other   903

Seasonal and Migratory   (This cell should not be
  Vacant + UREs   collapsed with any other cell)

1.1.E Occupied Units Adjustment

This adjustment is used only for occupied units. The weights resulting from the previous steps are adjusted by the following ratio:

$$\frac{\text{Estimate of Occupied Units Based on Current Population Survey Data}}{\text{AHS estimate of occupied units}}$$

This ratio is calculated using weights adjusted by all the previous calculations. It is calculated separately in each of the cells in Table 1-5. However, cells can be combined as described for the Vacant and URE Units Adjustment, if the conditions mentioned there are not met. The CPS data used are estimated for October, being estimated by a regression on 35 months of CPS data, ending six months after the survey.

Table 1-5

OCCUPIED UNITS ADJUSTMENT CELLS & SCALE VALUES

		Inside MSAs		Outside
		In Central Cities	Not In Central Cities	MSAs
Owner-Occupied				
White & Other:	Male	10	40	90
	Female	20	50	100
Black:	Male	210	240	290
	Female	220	250	300
Renter-Occupied				
White & Other:	Male	510	540	590
	Female	520	550	600
Black:	Male	710	740	790
	Female	720	750	800

1.2 Iteration

After all the above adjustments are done, the New Construction Adjustment, the Vacant and URE Units Adjustment, and the Occupied Units Adjustment are repeated, using the same numerators as before, but using denominators recalculated with weights adjusted by the results of the first iteration. The end result of this process is the weight factor, **WEIGHT**.

1.3 Journey to Work Supplement

The process above produces a household weight. The composition of households in the AHS sample differs very slightly form the CPS sample, due to normal sampling variability and perhaps procedural differences. Therefore, analysis of workers in the AHS Journey to Work Supplement might differ from analysis of workers in CPS. Another step of ratio estimation was therefore used to make AHS population characteristics match CPS population characteristics. The ratio was multiplied by the household weight to obtain a personal weight, which is stored for each worker as WWT1-8. For most purposes, and certainly for any research on households, these individual weights can be ignored, but they do produce estimates of employed workers more similar to CPS estimates.

1.4 1980 Census Adjustments

The weights of the national AHS are not directly benchmarked to the 1970 or 1980 Census. They are benchmarked to CPS and HVS which in turn were benchmarked to the 1970 Census until 1980, when they were adjusted to match the 1980 Census. The AHS national data from 1973-1980 are ultimately based on the 1970 Census, and from 1981 on are based on the 1980 Census. This introduces a discontinuity in time series. The discontinuity was extensively discussed in U.S. Bureau of the Census, Current Population Reports, Series P-60, No. 127, Money Income and Poverty Status of Families and Persons in the United States: 1980. The publication provides data comparing the old 1970 base and the 1980 base. In order for users of AHS tapes to test the effect of the discontinuity, the variable OLDWT is present on the 1981 National file. It is adjusted to the old series of CPS and HVS estimates, based on the 1980 Census. None of these considerations applies to PWT, which remains the inverse of the probability of selection and is most appropriate for longitudinal analysis.

1.5 Rural Weights

All of these adjustments are calculated in exactly the same way for rural and urban units, the only difference being that the initial probability of selection, as noted above, is twice as high for rural units, so their weights end up half as much as the weights of urban units.

On the public use tapes, as discussed in the section on geographic codes, rural units are normally identified as rural, and the interested researcher can verify that their weights are half as much as for urban units.

In the 125 MSAs identified on the public use tapes (in the variable "MSA"), rural-urban codes are suppressed to preserve confidentiality of families in the small rural portions of these MSAs. If the weights had been left alone, it would have been possible to identify rural units by their small weights. On the other hand, if the weights had been doubled, then any results in these MSAs would have been over-estimated. Therefore, half of the rural units in these 125 MSAs were deleted from the basic file, and the weights of the other half of the rural units were doubled. The interested researcher can verify that no low-weighted units appear in these MSAs.

The half of the rural units in these MSAs that were initially deleted were also included on the data tapes and can be identified with a special variable, RURREC (they have a code 2 in RURREC, all other cases have code 1). In order still to protect the confidentiality of · families in rural areas of these 125 MSAs, the MSA code was suppressed. Therefore, the researcher can be sure that any cases with 2 in RURREC are rural, and are somewhere in the 125 MSAs, but will not know which MSA. Since this is half the original sample of rual cases in the 125 MSAs, their weights have been doubled to provide and estimate of these rural areas. The interested researcher can use these RURREC=2 cases for estimates of rural portions of the 125 MSAs, or can combine them with other rural cases, to obtain estimates for all rural areas in the country.

Table 1-6
Non-Interview Adjustment Cells and Scale Values by MSA

		Units Occupied in 1970 in Address Segments										
		Central City 1										
Income in 1970		Race of Head in 1970										
		Nonblack						Black				
		Owner in 1970			Renter in 1970			Owner in 1970			Renter in 1970		
		Family Size in 1970			Family Size in 1970			Family Size in 1970			Family Size in 1970		
		1	2	3+	1	2	3+	1	2	3+	1	2	3+
Under $3,000		1	4	8	61	64	68	201	204	208	261	264	268
$3,000-$9,999		2	5	9	62	65	69	202	205	209	262	265	269
$10,000-$14,999		18	23	29	78	83	89	218	223	229	278	283	289
$15,000 and over		19	24	30	79	84	90	219	224	230	279	284	290

		Units Occupied in 1970 in Address Segments										
		Central City 2										
Income in 1970		Race of Head in 1970										
		Nonblack						Black				
		Owner in 1970			Renter in 1970			Owner in 1970			Renter in 1970		
		Family Size in 1970			Family Size in 1970			Family Size in 1970			Family Size in 1970		
		1	2	3+	1	2	3+	1	2	3+	1	2	3+
Under $3,000		501	504	508	561	564	568	701	704	708	761	764	768
$3,000-$9,999		502	505	509	562	565	569	702	705	709	762	765	769
$10,000-$14,999		518	523	529	578	583	589	718	723	729	778	783	789
$15,000 and over		519	524	530	579	584	590	719	724	730	779	784	790

		Units Occupied in 1970 in Address Segments										
		Central City 3										
Income in 1970		Race of Head in 1970										
		Nonblack						Black				
		Owner in 1970			Renter in 1970			Owner in 1970			Renter in 1970		
		Family Size in 1970			Family Size in 1970			Family Size in 1970			Family Size in 1970		
		1	2	3+	1	2	3+	1	2	3+	1	2	3+
Under $3,000		1501	1504	1508	1561	1564	1566	1701	1704	1708	1761	1764	1768
$3,000-$9,999		1502	1505	1509	1562	1565	1569	1702	1705	1709	1762	1765	1769
$10,000-$14,999		1518	1523	1529	1578	1583	1589	1718	1723	1729	1778	1783	1789
$15,000 and over		1519	1524	1530	1579	1584	1590	1719	1724	1730	1779	1784	1790

Vacant Units and Units in Other Segments Not Included Above			
	Central City 1	Central City 2	Balance of MSA
Units vacant in 1970 in address segments	105	605	1605
Units in special places in address segments	53	553	1533
New construction in permit segments and coverage improvement samples	96	596	1596
Mobile homes in area segments and coverage improvement samples	98	598	1598
Other units	99	599	1599

2.0 SMSA SURVEYS

Weighting for the MSA surveys is similar to that described for the national surveys. There are five adjustments in **WEIGHT**: the non-interview adjustment, a special stratification adjustment, a new construction adjustment, a special Houston adjustment, and a Decennial Census adjustment.

The non-interview adjustment (see Section 1.1.A) is calculated separately for each cell listed in Table 1-6. However, if there are less than 30 cases in a cell, or if the ratio in the cell would be 1.5 or more, the cell is combined with the cell having the closest scale value, as shown in Table 1-6.

The stratification adjustment is calculated separately for each cell listed in Table 1-7. It is calculated only for units taken from the 1970 Census long form questionnaries that are not group quarters and special places, i.e., for old housing units in permit-issuing areas. The weights resulting from the non-interview adjustment are adjusted by the following ratio:

$$\frac{\text{1970 Census count of housing units in permit-issuing areas}}{\text{AHS estimate of 1970 housing units in premit-issuing areas}}$$

This ratio is calculated using the inverse of the probability of selection of each unit, adjusted by the non-interview adjustment. Note that the denominator includes units in the current AHS sample and also cases dropped from the sample, because of demolitions, mergers, etc. All units are classified by their 1970 characteristics, not the current characteristics. The ratio is calculated separately in each of the cells shown in Table 1-7. If there are less than 20 cases in a cell or if the ratio in a cell would be 2.0 or greater, or 0.2 or less, the cell is combined with the cell having the closest scale value.

The effect of this adjustment procedure in MSAs is to reduce somewhat the variance due to variation in sampling rates for different strata in the address-listing areas. In principle there should be no difference in sampling rates for different strata. However, before the AHS sample selection in each MSA, units already selected for other Census Bureau surveys were deleted from the lists. Thus, some variation in effective sampling rates was introduced during the AHS sample selection process.

Table 1-7
Stratification Adjudication Cells and Scales Values by MSA

Units Occupied in 1970 in Address Segments

Central City 1

Race of Head in 1970

Income in 1970	Nonblack								Black							
	Owner in 1970				Renter in 1970				Owner in 1970				Renter in 1970			
	Family Size in 1970				Family Size in 1970				Family Size in 1970				Family Size in 1970			
	1	2	3-4	5+	1	2	3-4	5+	1	2	3-4	5+	1	2	3-4	5+
Under $3,000	1	4	9	15	30	34	39	45	401	404	409	415	430	434	439	445
$3,000-$9,999	2	5	10	16	31	35	40	46	402	405	410	416	431	435	440	446
$10,000-$14,999			70				100				470				500	
$15,000 and over	80	81	71	76	110	111	101	106	480	481	471	476	510	511	501	506

Units Occupied in 1970 in Address Segments

Central City 2

Race of Head in 1970

Income in 1970	Nonblack								Black							
	Owner in 1970				Renter in 1970				Owner in 1970				Renter in 1970			
	Family Size in 1970				Family Size in 1970				Family Size in 1970				Family Size in 1970			
	1	2	3-4	5+	1	2	3-4	5+	1	2	3-4	5+	1	2	3-4	5+
Under $3,000	1001	1004	1009	1015	1030	1034	1039	1045	1401	1404	1409	1415	1430	1434	1439	1445
$3,000-$9,999	1002	1005	1010	1016	1031	1035	1040	1046	1402	1005	1410	1416	1431	1435	1440	1446
$10,000-$14,999			1070				1100				1470				1500	
$15,000 and over	1080	1081	1071	1076	1110	1111	1101	1106	1480	1481	1471	1476	1510	1511	1501	1506

Units Occupied in 1970 in Address Segments

Balance

Race of Head in 1970

Income in 1970	Nonblack								Black							
	Owner in 1970				Renter in 1970				Owner in 1970				Renter in 1970			
	Family Size in 1970				Family Size in 1970				Family Size in 1970				Family Size in 1970			
	1	2	3-4	5+	1	2	3-4	5+	1	2	3-4	5+	1	2	3-4	5+
Under $3,000	3001	3004	3009	3015	3030	3034	3039	3045	3401	3404	3409	3415	3430	3434	3439	3445
$3,000-$9,999	3002	3005	3010	3016	3031	3035	3040	3046	3402	3405	3410	3416	3431	3435	3440	3446
$10,000-$14,999			3070				3100				3470				3500	
$15,000 and over	3080	3081	3071	3076	3110	3111	3101	3106	3480	3481	3471	3476	3510	3511	3501	3506

Vacant Units and Units in Other Segments Not Included Above

Rent or Value in 1970	Central City 1	Central City 2	Balance of MSA
Rent under $80 or value under $15,000	199	1199	3119
Rent of $80-$119 or value of $15,000-$24,999	201	1201	3201
Rent of $120 and over or value of $25,000 and over	202	1202	3202
Remaining vacants	205	1205	3205

The new construction adjustment is calculated only in the 1979-83 surveys, and only for sample units resulting from building permits issued since the previous survey in the MSA. It is used in 35 of 60 MSAs. Where this adjustment was used, its effect is to match AHS estimates to estimates from the Survey of Construction (SOC) on the proportion of new construction done in central cities and suburbs, while preserving AHS estimates of total new construction. The weights resulting from the previous steps are adjusted by the following ratio:

$$\frac{\dfrac{\text{SOC estimate of permits issued in the cell since five months from end of enumeration}}{\text{SOC estimate of permits issued in the MSA since five months from end of enumeration}}}{\dfrac{\text{AHS estimate of permits issued in the cell since five months from end of enumeration}}{\text{AHS estimate of permits issued in the MSA since five months from end of enumeration}}}$$

This ratio is calculated weighting each AHS sample case by the inverse of its probability of selection. It is calculated separately in only two cells in each MSA: central city(ies) and balance of MSA.

The special Houston adjustment is calculated ony in the 1979 survey, and only for housing units built since April 1, 1970 in the Houston MSA, outside the central city. The effect is to obtain a better estimate of the great growth the Houston area experienced during the seventies in non-permit-issuing areas. The weights resulting from the previous steps are adjusted by the following ratio:

$$\frac{\text{AHS estimate + Census Bureau estimate of the undercount}}{\text{AHS estimate}}$$

This ratio is calculated using the weights adjusted by the non-interview adjustment and the stratification adjustment. It is calculated separately in four cells:

1. Conventional housing units (non-mobile home or trailer) outside the central city, which were built between the 1970 census and the 1976 survey.

2. Mobile homes and trailers outside the central city with a model year between the 1970 census and 1976 survey.

3. Conventional housing units outside the central city built between the 1976 and 1979 surveys.

4. Mobile homes and trailers outside the central city with a model year between the 1976 and 1979 surveys.

A total housing unit ratio estimation was done for the 1974, 1975, 1977, and 1979-1983 surveys for some MSAs. For years other than 1979 and 1980, the independent estimates were based on census counts plus estimates of change. In 1979 and 1980 the independent estimates were based solely on 1970 and 1980 census counts:

in 1979:
$$\frac{\underline{1980 \text{ Census count x } 19 + 1970 \text{ Census count}}}{20}$$

AHS estimate

in 1980:
$$\frac{\underline{1980 \text{ Census count x } 21 - 1970 \text{ Census count}}}{20}$$

AHS estimate

This ratio is calculated using weights adjusted by all the previous steps. It is calculated separtely for two cells in each MSA: central city(ies) and balance of MSA. At the time this is written, it has not been decided how 1981 and later MSA surveys will be adjusted to 1980 Census counts. The above extrapolation technique may be used, or some other Census Bureau estimate of the total number of housing units may be used.

The weight resulting from all these steps is stored in the variable **WEIGHT**. The reader will note that up through the 1978 survey, only two adjustments are used: the non-interview adjustment and the stratification adjustment. All the other adjustments began in 1979. For comparability, a special variable called "OLDWT" is available on the 1979 and later surveys, including only the non-interview adjustment and the statification adjustment.

Chapter 3

SAMPLING DESIGN

The original samples for the National and Metropolitan Statistical Area (MSA) Surveys were drawn from the 1970 Census. The National sample was used from 1973 to 1983. In 1985, the National Survey was administered to a completely new sample drawn from the 1980 Census. This sample will be reinterviewed every two years until 1993. The procedures used to draw the three basic samples--the original National Sample, the new National sample, and the MSA samples--are similar, although differences are important enough to be documented. Sections 1.0 and 2.0 discuss in detail the selection of the original National Survey and the changes which occurred over the years. Section 3.0 presents the sample design for the new National Survey with emphasis on the differences between the two National samples. Section 4.0 describes the MSA samples.

1.0 THE 1973 NATIONAL SAMPLE

1.1 Selection of Sample Areas

The United States was divided into areas made up of counties and independent cities referred to as primary sampling units (PSUs). These PSUs were then grouped into 376 strata, 156 of which consisted of only 1 PSU each, which were therefore in the sample with certainty. These 156 strata were mostly the larger MSAs and were called self-representing (SR), since the sample from each area represented just that PSU. Each of the other 220 strata consisted of a group of PSUs and were referred to as non-self-representing (NSR), since the sample of housing units from the sample PSU in a stratum represented the other PSUs in the stratum as well as the sample PSU.

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

Sampling Plan A and Sampling Plan B are both representative of the NSR areas of the country. Therefore, if one added up the selected PSUs, weighting each by the inverse of its probability of selection, one would double-count the NSR areas. This result is avoided by weighting all data from Plan A by two-thirds, and data from Plan B by one-third, so their total adds up to a single count of the NSR areas.

1.2 Sampling Rates Within Sample PSUs

The national average sampling rate was determined by dividing the number of housing units in the country by the original desired sample size. This national average sampling rate was 1 in 1,366. In each of the 461 PSUs, this rate was adjusted so that the overall probability of selection for each sample housing unit was the same. For example, if the probability of selecting an NSR PSU was 1 in 10, then the within-PSU sampling rate would be 1 in 136.6. In the 25 PSUs which were selected twice, under Sampling Plans A and B, the appropriate within-PSU sampling rate was calculated twice, once under each sampling plan, and the rates were added to obtain an overall within-PSU sampling rate.

In order to save interviewer travel costs, it was decided to sample units in clusters, so that instead of counting off 1,366 units and then taking one unit into the sample, every 2,732 units were counted off (this is the national average; as just noted, the sampling rate and hence the count-off rate varied in each PSU) and then 4 units were taken into the sample. In some areas, half the clusters of four units were surveyed, with the other clusters held in reserve. (This was the pattern followed in so-called "area segments" which are defined below.) In other areas, two of the four units in each cluster were surveyed, with the other half of each cluster held in reserve. (This was the pattern for all sample cases other than "area segments.")

Starting in 1974, all of the units in rural clusters were surveyed, (i.e., the reserve sample was activated in rural areas.) This resulted in a probability of selection in urban areas of 1 in 1,366 and in rural areas of 1 in 783. The purpose of the double-sampling in rural areas was to obtain more accurate data on rural housing, to compare with urban housing.

1.3 Use of Enumeration Districts

Each of the 461 PSUs was divided into enumeration districts (EDs) of type A, B, or C. Each Ed within a sample PSU could be classified into one of the following four ED geographic strata: (1) central city, (2) urbanized area outside of the central city, (3) urban place outside of urbanized areas, and (4) rural. For each ED the following number was calculated: the sum of the 1970 count of housing units, plus one-third of the 1970 count of persons in group quarters divided by four. This number was the ED measure of size which was used in the ED sample selection. Enumeration districts were stratified according to ED geographical code, place size code, and ED code. Within each ED geographical category, the Ed measures of size were cumulated. For each ED in the established ordering of EDs within an ED geographic stratum, there was an associated cumulative total.
For the i_{th} ED in the stratum ordering the cumulative total, denoted T_i, was

$$\sum_{k=1}^{i} M_k$$

where M_k is the measure of size for the k_{th} ED.

$\sum$ = Summation

A cumulative total of the measure of size was computed for each ED geographic stratum and divided by s, the expected sample size of EDs for the stratum, to obtain a systematic probability proportional to size sampling interval, TE. A random start R was designated and the following set of numbers was determined:

$$R + TE, R + 2 TE, \cdots, R + sTE.$$

Under this selection procedure the ED sample size, s, was the greatest integer such that the quantity $R + sTE$ did not exceed the cumulative total for the last ED in the Ed geographic stratum. The ith ED was included in the ED sample if

$$T_{i-1} < R + m TE \leq T_i$$

for one of the values of m in the set $\{0,1,2,\cdots,s\}$.

The selected EDs were then divided into address EDs and area EDs. An ED was classified as address if 90 percent or more of the 1970 census addresses recorded in the Census Ed Address Register had a complete house number and Street name, and furthermore the ED was geographically located in a jurisdiction issuing permits for new construction. An ED could have been classified as an area ED either because it was not within a jurisdiction issuing permits for new construction and/or more than 10 percent of the addresses in the census adress registers did not have a complete house number and street name. The sample selection procedures for address EDs are discussed in Section 1.3.A while the procedures for selecting the sample in area EDs are discussed in Section 1.3.B.

1.3.A List or Address EDs

Enumeration districts within jurisdictions which issued building permits for new construction and which had a proportion of complete addresses of 90 percent or more were classified as list or address EDs. Addresses recorded in the 1970 Census ED Address Register were then used to form clusters having an expected four units. Adjacent address listings on the Ed Address Register corresponded to housing units which were physically adjacent, so that the clusters formed would be compact clusters. Addresses having all their units in a single cluster were designated TA addresses while all other addresses were defined as NTA addresses.

After clusters in sample address EDs were formed using all addresses within the ED address register, sample clusters were determined. For TA addresses, where all units for an address fall in the same cluster, the interviewer listed and interviewed all units found at the TA address. For an NTA address, only a proportion of units at the address was part of the sample cluster. The AHS interviewer had to list all units found at the address using established listing procedures, and units located on listing lines which had AHS sample unit identifiers predesignated in the regional office. Suppose that the cluster fomation had formed the following two clusters from units at 103 Maple Street where a two-unit structure is located and 106 Maple Street where a six-unit structure is located:

103 Maple	Unit 1	
	Unit 2	Cluster A
	Unit 1	
	Unit 2	

	Unit 3	
	Unit 4	Cluster B
106 Maple	Unit 5	
	Unit 6	

The address 103 Maple was a TA address while 106 Maple was a NTA address. If Cluster A was determined to be a sample cluster, then an interviewer would use the listing sheet for 106 Maple Street to transcribe all units found at that address. There would have been at least two listing lines having AHS sample designation codes, and the units falling on those listing lines would have been part of sample Cluster A.

Two of the housing units within a sample cluster were assigned to the primary sample while the remaining two units were assigned to the reserve sample.

Each mobile home park is treated as one "structure". The list is treated much as described for the list of building permits: each building or mobile home park is listed, with the number of units found at the 1970 Census; the count-off rate is applied, and clusters are selected. These clusters in effect determine a sampling rate within the building or park, e.g., 2 units may be selected out of a 20-unit building (10 percent sampling within the building) or 1 unit out of a 1-unit building (100 percent). In rural areas the entire cluster of 4 is included in the sample, while in urban areas two of the four units in each cluster are held in reserve. The interviewer who goes to the building or park makes a list of all units actually there, and samples them at the appropriate sampling rate. In later surveys, any additional units in the building or park are added at the bottom of the list, and the sampling rate continues to apply to them, systematically. In fact, a preprinted listing sheet, with certain lines earmarked for the AHS sample, is used to record and sample additional units.

1.3.B Area EDs

EDs where permits are not required for new construction and EDs where permits are required for new construction but less than 90 percent of the addresses within the ED are complete, use "area samples." These are primarily rural, but include some large cities, such as Houston. The boundaries of each such ED were subdivided into small land areas having recognizable features such as county roads, rivers, railroad tracks, etc., as boundaries. These small land areas are called "area segments". Wherever possible, an area segment was formed so that it contained between 7 and 20 housing units which were enumerated in the 1970 Census. The housing unit count within each area segment within a sample ED was converted to a cluster count by dividing the housing unit count of the area segment by four and rounding the result. Within each area segment cluster identifiers were listed. From the total ED list of cluster identifiers, sample clusters were determined and every other cluster that was drawn into sample was designated for the reserve sample. Area segments

containing sample clusters were the sample area segments. The sample area segments that had to be listed by field enumerators as sample clusters at this stage were undefined for the AHS interviewers. A few months before interviewing started, interviewers went to the defined area and listed every unit there. When field enumerators listed sample area segments within permit-issuing area EDs, they determined for each housing unit within the area segment if the unit was built before or after April 1, 1970.

Housing units built after April 1, 1970 were ineligible for the sampling of the primary and reserve noncompact clusters. These housing units built after April 1, 1970, would have a chance of being selected in the building permit frame. Sampling of new construction in this way in permit issuing area EDs led to a more dispersed and heterogeneous sample. If the area, according to the 1970 Census, was expected to have four units, then all units actually found were interviewed. If the area was expected to have more than four units, e.g., 20, then the sampling rate was determined based on the number expected, e.g., 4 out of 20, and was applied to all units found. For example, if 30 were found, 4/20 of 30, or 6, would be interviewed. This sampling rate assured that all units, even those missed by the 1970 Census, had a predetermined probability of selection.

The only exception was when a very large number of units was found, as in a new subdivision or a large, new apartment building, in which case the area was subsampled and the new probability of selection was reported back, so that weights could be adjusted to make the subsample represent all of the original sample. The list is updated by interviewers before every survey, and the sampling rate is applied by Census Bureau regional office staff to the additional units also, so that new units are included in the survey at the same rate as old units.

1.4 Building Permit Sampling

Housing units built after April 1, 1970, were classified as new construction housing units in the AHS, and new construction housing units located in address EDs and permit-issuing area EDs would be represented in the building permit frame. The building permit sample selection procedures parallel the procedures for the address EDs.

Analogous to the sampling of enumeration districts within a sample PSU was the monthly sampling of building permit offices. For each permit office, the total number of housing units associated with permits issued in the period April 1, 1970 to five months prior to the first AHS enumeration was obtained. This was the building permit office's measure of size for each month. Building permit offices within a sample PSU were then grouped into the following three geographic strata: (1) central city of MSA, (2) balance of an MSA, and (3) non MSA. After the stratification of building permit offices, permit offices were sampled monthly using systematic probability proportional to size sampling.

Building permits in sample permit offices had to be listed by field enumerators. From the permit office listing sheets, new construction units were assigned to map grid coordinates which grouped together new construction units located within a small land area within the jurisdiction of the permit office. Clusters of four new construction units were formed from

units within each set of map coordinates. Clusters were then systematically sampled from the total permit office cluster universe.

Normally, a permit represents one unit, so this would be a cluster of 4 permits, but in multi-family construction it might be a cluster of the 12th-15th units in a planned 20-unit structure, under the same building permit. Originally, two units of each cluster were held in reserve, but since the reserve sample has been activated in rural areas, all units of every cluster are used for the sample in rural areas. In urban ares, two units from each cluster are kept in the sample with the other two still held in reserve. Then, each permit was followed up, and if the unit had not been built by April 1, 1970, but had been built by the time of the interview, it was included in the sample. Where the permit was for a multi-unit building, the count-off process as already discussed took this into account, and indicated what fraction of the units was to be sampled.

When interviewing began, all units in the building were listed and the appropriate fraction was sampled. The count-off process indicated what fraction of the units in the structure was to be sampled. For example, if the count-off process indicated 4 units were to be taken from a 20-unit structure, then 4/20 of the units actually found in the structure were sampled. (In a sense, each building permit was treated like an area segment as described earlier.) The permit sampling is repeated every year, so new units are continuously sampled. Since permits are usually not required for mobile homes, new mobile homes are not sampled by these building permits, but are covered by the area sampling, like all other kinds of units other than new construction.

The cut-off of building permits 5 months before the survey means that some units are missed each year, because they are built before the survey, with permits issued after the cut-off. It is estimated that the 1980 National AHS sample missed about 1.4 percent (i.e., about 251,000 units) of conventional housing units built after April 1, 1970 and before September 1980 because of the permits for these units were issued less than 5 months in advance of the survey. These missed units would be even more serious as a percentage of 1980 construction. The new construction adjustment described in Section 1.1.C of Chapter 3 below is designed to reduce the effect of this deficiency, although some bias probably still exists. Review of the adjustment indicates that there has been a consistent overcompensation for this deficiency in every year since 1975 by adjusting to counts of new construction for the end of the interview period, which has been December or January, instead of October. This overcompensation may inflate the new construction counts by 100,000 to 300,000 units. Researchers needing precise estimates of new construction should refer to the Census Bureau's monthly "Survey of Construction" (SOC). The AHS is designed to show the characteristics of people who live in new construction, which SOC does not show, but is not designed to replace basic statistics provided by SOC.

1.5 Coverage of Units in Different Types of EDs

In area EDs of Types A and B, the area sampling methods successfully reach all kinds of housing, but they are very expensive because the job of listing all housing units within a geographic area is time-consuming for interviewers. Therefore, whenever possible, the

Census Bureau uses address lists as described in Section 1.3.A above. Address EDs cover about 75 percent of housing in the country. (They only cover about 60 percent of the sample, because these EDs are largely urban, which are sampled half as much as rural EDs.) The weakness of these address lists is that they do not cover units missed by the 1970 Census, structures that were non-residential in 1970 but now have housing units in them, or mobile homes placed into a new park or onto an individual site after the 1970 Census.

Furthermore, the permit sampling, which is used in address as well as permit-issuing area EDs, can miss certain new construction units. Special efforts have been made to fill these gaps:

CEN-SUP--This was an evaluation study to estimate units missed by the 1970 Census, which identified a sample of such units. This sample was divided among the various current surveys of the Census Bureau, including the AHS, starting with the 1973 and 1974 surveys. The overall probability of selection for these units was quite variable but averaged about 1 in 1,900. Unfortunately, this evaluation study did not cover mobile home parks.

New construction from old permits--A sample of new construction units whose permits were issued before January 1970 was selected as follows. Units whose permits were issued before January 1970, but which were completed after the census, were identified from the Survey of Construction (SOC), a survey of building permits conducted monthly by the Bureau of the Census. These units were then sampled so that the overall probability of selection was about 1 in 1,320. They were added to the AHS starting with the 1976 survey.

Mobile home parks--A sample of mobile homes placed in parks missed by the Census or established after the Census was selected as follows. A list of mobile home parks was obtained from commercial listings. This list was then supplemented by additional parks identified by a canvassing operation similar to that performed in EDs where area sampling methods are used. Unfortunately, however, this canvassing operation only represented about 92 percent of all address EDs. The parks were divided into clusters of four sites. These clusters were sampled so that the overall probability of selection of a unit was about 1 in 1,366. These were added to the AHS sample starting in 1976. The canvassing operation was repeated in 1980, and the resulting mobile homes were added to the AHS in 1981.

Residual problems--The efforts described above leave a residual of hard-to-find units. The following list summarizes the units covered by all these methods, and the other types of units where a special residual effort was needed in address EDs:

1. Units in structures or mobile home parks covered in the 1970 Census (address lists)

2. Group quarters covered in the 1970 Census and converted to housing units (address lists)

3. Units erroneously missed by the 1970 Census (CEN-SUP units)

4.   Building permits issued from January 1970 on (building permit sample)

5.   Building permits issued up through December 1969 (SOC sample)

6.   Mobile homes outside parks and vacant at the time of the 1970 Census (residual effort)

7.   Mobile homes in parks missed by the 1970 Census or established after the Census (mobile home park sample)

8.   Mobile homes placed outside parks after the 1970 Census (residual effort)

9.   Units in structures that were totally non-residential at the time of the 1970 Census and were later converted to residential use (residual effort)

10.   Houses moved onto a new site after 1970 Census (residual effort)

The residual effort was done in three stages. First, a subsample of the regular AHS sample units from the Census address list was selected. Second, succeeding structures that had been eligible to be selected from the Census address list were then listed until eight such additional structures (including mobile home parks) were found. Third, any structure between these nine was assessed. If it was of a type not covered by any of the other coverage improvement efforts, it was added to the sample and all or a subsample of units in it were interviewed, starting in 1976. The overall probability of selection for these units was highly variable, but averaged about 1 in 2,400. This procedure was not considered very efficient for finding non-residential conversions (which might be primarily in business districts), since the listing procedure started from a residential unit.

1.6   Business Sample

Non-residential conversions were surveyed again in 1980. Interviewers went to a sampling frame of businesses, originally prepared for a Department of Justice survey. Any structures which did not have housing units in them in the 1970 Census, but did by 1980, were considered non-residential conversions, and were added to the AHS sample in 1981.

2.0   FATE OF SAMPLE UNITS OVER TIME

The previous discussion describes how units join the AHS sample. They remain in it every year as long as possible. Even a unit which is converted to commercial purposes, or is boarded up and uninhabitable, is kept on the list as a Type B non-interview (a "recoverable loss"). It is revisited each survey in case it has been fixed up for residential use again. It drops out when it becomes a Type C non-interview ("unrecoverable loss," i.e., when it is demolished, totally burned down, etc.). Some units have also been dropped because of sample reductions, as explained below.

2.1 Splits and Mergers

When two or more units are merged into one unit, the order of the units on the listing sheet determines whether the unit stays in the sample. If the first unit was in the AHS sample, the new unit stays in the sample with the same control number the old unit had. Otherwise, it drops from the sample. For example, if Apartment 701 is in the AHS and Apartment 702 is in the Current Population Survey (or not in any Census Bureau survey), and they are merged into Apartment 701, AHS keeps the new unit. If the new unit is called Apartment 702, however, AHS loses the unit. It becomes a Type C non-interview ("merged, not in current sample").

When a unit is split into two or more units, both units stay in the sample. One unit retains the old control number, while the other will have a new control number, while the other will have a new control number, which cannot be linked to the control number of the unit from which it split.

This sampling plan makes longitudinal study of splits and mergers difficult. It was modified when the new sample was drawn in 1985, by having AHS units as distant as possible from units in other Census Bureau samples, so that split and merged units can almost always be retained in the sample. (See Section 3.4 below.) The data base was also modified to identify which units split from and which merged with each other.

2.2 Sample Reductions

By 1977, the additions to the sample from new construction and other additions to the housing stock had increased the total sample size (interviews plus non-interviews) to about 81,000. The sample was reduced by 7/97 to approximately 75,000 in 1977. However, this reduction did not include any CEN-SUP units or units which were selected as part of the 1976 Coverage Improvement Program. Thus, the overall probability of selection for these latter units remained unchanged; the probability of selection for the rest of the units was changed to about 1 in 1,472 if they were urban and about 1 in 736 if they were rural.

In 1981, there was a similar cut-back of 5/90, so the probability of selection for the basic sample of urban units was 1 in 1,559. At the same time, the rural sample was cut back further by dropping half of the rural sample outside the 125 biggest MSAs, and a quarter of the rural sample inside these 125 MSAs. Thus, outside the 125 MSAs, the probability of selection of rural units become equal to the probability for urban units; and inside the 125 MSAs, the probability was about 1 in 1,039. Again, CEN-SUP and Coverage Improvement Program units were omitted from the cut-backs. In 1983, however, the portion of the rural sample that had been cut back was reinstated.