

3. Coverage

The specification of goals for coverage is an important element of the design of any survey. What is the target population for which estimates are needed? Given a clear and precise definition of the target population, one can ask how well the sampling frame developed for the survey and the sample selection procedures that are used have succeeded in providing an unbiased sample for that population. If some units in the target population have no chance of selection in the sample (undercoverage), the resulting estimates may be biased. Multiple chances of selection leading to overcoverage are also possible, but much less likely in a survey like RECS. The effects of coverage bias can be reduced, but seldom fully eliminated, by benchmarking survey estimates to data for the target population available from sources external to the survey.

The first section of this chapter describes the target populations for RECS. The second section describes those parts of the frame development and sample selection procedures that involve individual housing units and are, therefore, a potential source of coverage error. The final section examines the relationships between sample-weighted survey estimates and the external data to which they are benchmarked.

RECS Target Populations

In broad terms, the goal of RECS is to provide data for the United States residential energy sector. The sample for the 1978 and 1979 surveys did not include any representation of Alaska and Hawaii. From 1980 on, the sample for RECS has covered all 50 States and the District of Columbia.

For both conceptual and operational reasons, survey target populations must be defined as precisely as possible. For RECS, some compromises have been required between coverage goals that would be ideal if cost were no object and those judged to be achievable with the resources available.

In this section, two kinds of target "populations" are identified. The first consists of the housing units and households that are meant to be included in the surveys; the second consists of the energy consumption associated with those units. Consumption could be treated as a survey variable and discussion of goals for its measurement deferred to Chapter 5 (Measurement Error), but because of its basic importance to RECS, it seems desirable to introduce the subject in this chapter.

Housing Units and Households

The target population for RECS consists of housing units occupied as primary residences in the United States. The housing unit and household definitions for RECS are the same as those used by the Census Bureau in the decennial censuses of population and housing and in the Current Population Survey (CPS). Mobile homes are included. Housing units on military installations were not included in the 1978 and 1979 surveys, but have been included in RECS since 1980

RECS Target Populations

- Housing Units/Households

Housing unit - Census definition

Occupied - Excludes vacant units

Primary - Excludes seasonal and occasional residences

Reference date

Through 1990 - November of survey year

1993 - July of of survey year

- Energy Consumption and Expenditures

For housing units in target population

For major energy sources:

Electricity, natural gas, fuel oil, kerosene, LPG

Rough estimate for wood

Solar energy not included

At site - Excludes primary fuels used to produce electricity

Exclusions

Business uses

Some outdoor uses

Common areas in multi-unit facilities

Reference period

Through 1984 - 12 months starting in April of survey year

1987 on - calendar year

(EIA 1982a). Group quarters, such as military barracks and nursing homes, occupied by 10 or more unrelated persons are excluded.

The restriction of the target population to primary residences means that housing units that are vacant or occupied only on a seasonal or occasional basis are excluded. Although it would have been desirable to collect information about housing characteristics and energy consumption for secondary residences and vacant housing units, attempting to do so would have substantially increased the cost of the survey and the complexity of the procedures. An advantage of the

exclusion is that it creates a one-to-one equivalence between housing units and households, so that estimates of households from the Census Bureau's Current Population Survey (CPS) can be used directly to benchmark the RECS estimates.

Part of the definition of a target population is the date or time period to which it refers. For each of the Residential Energy Consumption Surveys, the data collection for the Household Survey has taken place from the fall of the survey year through the winter and, in some instances, to the early spring of the following year. For all surveys through 1990, November of the survey year was taken as a rough midpoint of the data collection period and the survey estimates of housing units were benchmarked, or adjusted to agree with, estimates of households (equivalent to housing units) for that month based on data from the CPS. For the 1993 RECS, the data collection period for the Household Survey was essentially the same as for previous surveys--that is, it took place from the fall of 1993 through the spring of 1994--but the survey estimates were benchmarked to CPS estimates of households for July 1993.

Ideally, each RECS Household Survey should include a probability sample of all housing units occupied as primary residences in the month used as a benchmark for the survey estimates--that is, November of the survey year for all surveys through 1990 and July for the 1993 survey year. In actual practice, the decision on whether to include a housing unit in RECS is based on its occupancy status at the time that a survey interviewer succeeds in contacting the unit and determining its status. Thus the 1993 survey sample includes some housing units that were not occupied as primary residences in July 1993 but were occupied when they were contacted by survey interviewers during the data collection period that started in October 1993.

Energy Consumption

For energy consumption and expenditures, the goal of the surveys is to collect data on energy consumption for residential purposes, during a specified time period, by the housing units and households in the target population. Energy consumed for businesses located in or closely associated with a residence is meant to be excluded. The surveys attempt to measure consumption *at the site*, i.e., at the point it enters the residence. For electricity, this measure of consumption does not reflect the total amount of energy used to generate it (primary energy consumption). Site values for residential consumption of electricity can be multiplied by a factor of three to provide a rough estimate of total energy consumed in the production of electricity for residential use (EIA 1993a, p. 13). The *Household Energy Consumption and Expenditures 1993* report includes two tables--5.2 and 5.4--that show primary consumption of electricity for residential purposes.

For the first six surveys, through 1984, the goal was to measure residential energy consumption by households in the target population for the 1-year period from April of the survey year through March of the following year. Thus, for the 1984 RECS, estimates of consumption were for the 12 months from April 1984 through March 1985. Complete data cannot be collected from suppliers until after the end of the consumption reference period. For the 1987 RECS and in subsequent surveys, estimates of consumption have been for the calendar year corresponding to the survey year, so that collection of data from suppliers could begin at the start of the calendar

year following the survey year. Operational problems in using billing data from suppliers to estimate consumption and expenditures for a specific 12-month reference period are discussed in Chapter 6.

Some Implications of the Definitions

The seasonal and year-round vacant housing units that are excluded from RECS do consume some energy. Some of them are occupied for part of the reference year and even those that are not occupied at all may consume moderate amounts of energy for various purposes. Table 3.1 provides information from two sources, RECS and the American Housing Survey, about the portion of all housing units that are vacant.

For sampling purposes, the RECS procedures call for a listing of all households in a designated sample of areas, whether or not they are eligible for interviews. Therefore, a rough indication of trends in the proportion of housing units excluded by definition can be derived from examination of the data on the proportion of assigned sample housing units in each survey year that turned out to be in the excluded group. The RECS data shown in Table 3.1 are *unweighted* counts of sample units; the estimates of percent ineligible (seasonal and year-round vacant) would be somewhat different if the appropriate sampling weights were applied to each unit. A more precise indication of the proportions of excluded units is provided by the data from the American Housing Survey shown in the last column of Table 3.1; these are weighted sample estimates.

The biennial American Housing Survey estimates of vacant units have varied within a narrow range, 9.2 to 11.5 percent, between 1981 and 1993. The mean energy consumption of these vacant units was almost certainly substantially lower than the mean consumption of occupied units, but no reliable estimates are available. Both the level and the trends in the proportion of RECS sample housing units excluded are quite similar to the corresponding American Housing Survey estimates.

The RECS estimates of consumption do not cover all kinds of residential energy consumption. Although data are collected on the numbers of housing units using solar collectors for main and auxiliary space and water heating, no estimates are developed for fuel equivalents of solar energy. Some, but not all, outdoor uses of fuels--such as for lawn mowers and outdoor grills--are not included. Fuel consumption for common areas, such as lobbies and meeting rooms, in multiunit apartments and other residential facilities is also excluded.

The consumption of wood as a fuel is not included in the basic RECS estimates of total consumption of major fuels, but sufficient data on wood used are collected to permit a rough estimate. For 1993, it was estimated that inclusion of wood would have added 5 percent to the RECS estimate of total fuel consumption (EIA 1995d, Tables 5.2 and 5.9). Since 1984, residential consumption of wood as a fuel has declined substantially.

Table 3.1. Ineligible Housing Units in Sample by Survey Year: 1978-1993

Year	Housing Units in RECS Sample			American Housing Survey: Percent Seasonal or Year-Round Vacant
	Total	Ineligible		
		Number	Percent	
1978.....	4,849	342	7.1	NS
1979.....	NA	NA	NA	NS
1980.....	7,232	598	8.3	NS
1981.....	7,550	709	9.4	9.2
1982.....	5,808	536	9.2	NS
1983.....	NS	NS	NS	9.5
1984.....	7,535	783	10.4	NS
1985.....	NS	NS	NS	11.5
1986.....	NS	NS	NS	NS
1987.....	8,007	824	10.3	11.5
1988.....	NS	NS	NS	NS
1989.....	NS	NS	NS	11.3
1990.....	6,607	698	10.6	NS
1991.....	NS	NS	NS	10.9
1992.....	NS	NS	NS	NS
1993.....	9,671	918	9.5	11.1

NA = Not Available.

NS = No Survey conducted.

Sources: Energy Information Administration, *Consumption and Expenditures* (1978); *Housing Characteristics* (1980-1993); and *American Housing Survey* (1981-1993).

Frame Development and Sample Selection Procedures

As described in Chapter 2, RECS uses a four-stage sampling procedure. The first two stages, selection of primary and secondary sampling units (PSU's and SSU's), rely solely on area sampling techniques. The third and fourth stages can involve both list and area sampling. In the third stage, the larger sample SSU's are divided into listing segments and one listing segment is selected for the sample. Smaller SSU's are used as the listing segments. During the spring and summer of the survey year, field workers prepare listings of all housing units in each listing segment.

The fourth-stage procedures also depend on the size (number of housing units listed) of the sample listing segments for which housing unit listings were prepared in the third stage. For smaller listing segments, a systematic sample of housing units for interviews is selected directly from the listings. The larger listing segments are subdivided into groups of housing units called "penultimate clusters." One of these is selected and a systematic sample of housing units for interviews is selected from it.

There is no information to indicate that the area sampling parts of these procedures--selection of PSU's, SSU's and, when necessary, listing segments and penultimate clusters--have been affected

by nonsampling errors in the selection process. In this section, therefore, attention is directed to errors affecting the completeness of coverage that can and may have occurred in the listing part of the sample selection procedures.

Listing and Update Procedures

Because of the longitudinal nature of the RECS samples for the 1982 through 1990 surveys, different listing and sample selection procedures were used for SSU's designated for selection of a new sample and those for which the sample of housing units was to be carried over from the prior survey. In some of the SSU's in the former group, new listing segments, requiring a full initial listing of housing units, were selected. In others, the listing for the old segment was updated and a new sample of housing units was selected. In the carryover group, field workers updated listings for the listing segments from the prior survey.

As noted above, initial listing and listing update procedures have been carried out in the spring and summer of the survey year. Field workers are instructed to list as housing units "Houses and apartments that are under construction--if they are likely to be completed and ready for occupancy by September [of the survey year]" (Response Analysis Corporation 1992a, p. F-11). Primarily because of the possibility of subsequent housing unit additions and deletions, the survey interviewers, who do their interviews in the fall and winter of the survey year, are instructed to look for new or previously missed housing units in the "half-open" interval between each sample housing unit assigned for interview and the following housing unit on the listing, whether or not that unit had been included in the sample. Any such housing units are to be interviewed, in addition to the sample housing units initially assigned. In each of the survey years 1984, 1987, 1990, and 1993 about 1 percent of the Household Survey interviews completed were for housing units added, at the time of interviewing, through the use of the half-open interval technique (EIA 1995f).

Finally it is possible that, at the time of interviewing, some assigned sample housing units may no longer meet the definition of a housing unit because they have been condemned or demolished, are being used solely for nonresidential purposes, or are currently being used as group quarters with living arrangements for institutional residents or inmates or for other groups of 10 or more unrelated persons. It is also possible that a sample housing unit may have been split into two or more separate units; in this event, the survey interviewer is expected to conduct a separate interview for each unit.

Evaluation of the Quality of Listing and Update Procedures

Little direct information is available about the quality of the listing and updating procedures. In the 1979 survey, clusters of housing units were independently relisted by a second interviewer in one-fourth of the survey locations. The survey report states that "In general, the original listings and relistings are in agreement for 90 to 95 percent of the housing units listed." (EIA 1981, p. 77). There have been no checks of this kind in subsequent survey years.

On at least two occasions, observation of field activities by EIA staff members has provided anecdotal information about errors in listing/updating procedures. During the interviewing for the 1990 survey, an observation visit to a listing segment in a small North Carolina community developed evidence which suggested that the housing unit listing for the segment, which had been carried over from the 1987 survey, had not been updated earlier in 1990 (Battles and Thompson 1991). Near the end of the interview period for the 1993 survey, an EIA staff member checked the listings for seven RECS sample locations in the Los Angeles area. In one listing segment, he found that 67 of 85 units that had been listed were outside the boundaries of the listing segment. In the remaining six listing segments, he found a total of 21 addresses that had not been included in the 223 addresses originally listed, for a miss rate of 8.6 percent (Thompson 1994a).

In the 1984, 1987, and 1990 surveys, each of which included a longitudinal panel carried over from the preceding survey, the sample was divided into 4 subsamples, or rotation groups, each consisting of one-fourth of the total sample of SSU's. Because the frame development and sampling procedures varied by rotation group, an analysis of the counts of new housing units in the sample by rotation group provides indirect information relevant to the quality of listing and updating.

Three different frame development and sample selection procedures were used, as follows:

Procedure 1. For every SSU, select a new listing segment, do the housing unit listing and select a sample.

Procedure 2. Check each SSU to identify those with significant amounts of new construction (25 or more units). For those with significant new construction, follow Procedure 1. For other SSU's, update the listings from the preceding survey and select a sample of new units and units not sampled in the preceding survey.

Procedure 3. For all SSU's, update the listings from the preceding survey. The sample consists of units included in the sample for the preceding survey plus a sample of the new units identified in the updating operation.

For all three surveys, Procedures 1 and 2 were each used in one of the four rotation groups and Procedure 3 was used in the other two rotation groups. In the SSU's for which Procedures 1 and 2 were used, significant amounts of new construction were found in 130 of 608 SSU's in 1984, in 205 of 615 SSU's in 1987, and in 197 of 758 SSU's in 1990. (In 1984 and 1987, the preliminary checks for new construction were not undertaken for SSU's in primary sampling units that were entering the sample for the first time.)

Table 3.2 shows the sample counts of new housing units--those built in the survey year and the two preceding years--by rotation group for the 1984, 1987, and 1990 survey years. Under the hypothesis that new listings are likely to provide better coverage of newly constructed housing units than are updates of prior listings, one would expect to find the most new units in rotation groups in which Procedure 1 was used. One would expect the rotation groups for which prior listings were updated (Procedure 3) to have the smallest number of new housing units and the

rotation groups for which new listings were used for segments with significant amounts of new construction (Procedure 2) to be somewhere in between. For the most part, these expectations are borne out by the data in Table 3.2, although the distinction between Procedures 1 and 2 is not clear. On the other hand, the data for *older* houses, shown for the 1990 RECS only (houses constructed between 1985 and 1987), do not follow the pattern observed for new houses. Overall, these data provide support for the hypothesis that new listings provide better coverage of new housing units than do updates of listings from the preceding survey (Jabine 1993, p.8).

Evaluation of Coverage Based on External Data Sources

Comparison with Current Population Survey (CPS) Estimates of Households

Some indication of the completeness of coverage of housing units and households in RECS can be had by comparing CPS estimates of households with the RECS estimates for each survey year prior to the stage at which the latter are "benchmarked" to (adjusted to agree with) the corresponding CPS estimates. The benchmarking adjustment is carried out separately for each of 12 strata defined by Census region and metropolitan statistical area status (MSA central city, MSA other, and non-MSA). The estimates prior to benchmarking are obtained by applying the appropriate weights to sample households based on their overall sample selection probabilities and adjusting for unit nonresponse.

Table 3.3 shows, by Census region and survey year, the percentages by which the benchmarked estimates exceeded the estimates prior to benchmarking. These percentages provide indications of possible net undercoverage of households in RECS. They are not precise measures of undercoverage for several reasons:

- Both the benchmarked estimates and the estimates prior to benchmarking are subject to sampling error.
- Based on comparisons with the decennial censuses and census post-enumeration surveys, it seems likely that the CPS estimates of households are themselves somewhat low.
- The "benchmarking factors" on which the values shown in Table 3.3 are based actually combine the effects of two stages of ratio estimation, only the second of which represents the actual benchmarking to the CPS estimates. The first stage, which applies only to non-self-representing primary sampling units in the RECS sample, is based on Census counts of households by PSU and reduces the component of sampling variance that results from the selection of a sample of PSU's.

At the national level, the undercoverage indicator shown in Table 3.3 stayed within a fairly narrow range--from 6.6 to 9.7 percent--from 1980 through 1990, but in 1993 it declined to 4.2 percent. In each survey year between 1981 and 1990, the indicator was substantially higher for the South than for the other three Census Regions, but in 1993 it declined to a much lower level. The redesign of the sample for the 1993 RECS, using 1990 Census data, may have accounted,

Table 3.2. Sample Counts of New Housing Units by Rotation Group: 1984, 1987, and 1990 RECS

Survey Year	List/Update Procedure ^a	Rotation Group	Sample Count of Housing Units
1984 RECS (Units built in 1982 to 1984)	1	D	56
	2	C ^b	41
	3	E	29
	3	F	29
1987 RECS (Units built in 1985 to 1987)	1	E	58
	2	F ^b	71
	3	C	39
	3	D	39
1990 RECS (Units built in 1988 to 1990)	1	C	49
	2	D	34
	3	E	21
	3	F	34
(Units built in 1985 to 1987)	1	C	65
	2	D	48
	3	E	47
	3	F	65

^aSee description of procedures in text.

^bBecause of the introduction of new Primary Sampling Units, Procedure 1 was used in about 1/5 of the SSU's in 1984 and 1987. Source: Special tabulations of 1984, 1987, and 1990 RECS data.

at least partially, for the changes that occurred in these indicators between 1990 and 1993. In the 1993 redesign, 1,250 out of 1,610 secondary sampling units (SSU's) were newly-selected, whereas in the 1984 redesign, based on the 1980 Census, only 266 of 1,515 SSU's were new. As noted earlier in this chapter, the use of the "half-open interval" technique in the 1984, 1987, and 1990 survey years to add housing units missed by listing and updating procedures accounted for about 1 percent of the household interviews completed in each of those years. Given the undercoverage indicators shown in Table 3.3 for those years, it would appear that interviewers using the half-open interval technique succeeded in identifying only about 10 percent of the previously missed housing units.

Coverage of New Housing Units: Comparisons with Data from Census Bureau Surveys of Construction and New Mobile Homes

The results of the 1990 RECS showed an unexpected reversal in what had been a nearly uninterrupted downward trend in average energy consumption per household by year built. This finding led to a search for ways of using external sources of data to improve the accuracy of the estimates based on the sample housing units that had been built in the period 1980 through 1990.

Table 3.3. Benchmark Adjustment Factors^a for RECS Estimates of Households, by Census Region: 1990-1993

Census Region	Survey Year						
	1980	1981	1982	1984	1987	1990	1993
Northeast.....	4.1	5.5	9.4	7.2	4.7	2.9	1.8
Midwest.....	4.7	3.5	6.9	6.8	7.3	4.3	5.1
South.....	8.1	12.2	12.3	14.7	16.4	15.5	2.6
West.....	9.1	8.6	4.1	0.9	6.5	5.7	8.1
U.S. Total.....	6.6	7.8	8.7	8.4	9.7	8.1	4.2

^aBenchmark adjustment factors are applied to weighted sample estimates, following adjustment for unit nonresponse, to make them agree with Current Population Survey estimates of households. The value shown in each table cell is equal to 100 ($f-1$), where f is the benchmark adjustment factor.

Source: Robert B. Latta, *Analysis of Listing Undercount and Other Undercount Factors for RECS* (October 1994).

Two such sources were the Census Bureau's Survey of Construction and its Survey of New Mobile Home Placements. The special estimation procedures that were developed by using data from those sources are described in Chapter 7. This chapter looks at what these data reveal about RECS coverage of housing units by year built.

Table 3.4 shows Census Bureau and 1990 RECS estimates of housing units, by housing type and main space-heating fuel, built in the periods 1980-1984, 1985-1987, and 1988-1990. The two sets of estimates are not fully comparable. The Census Bureau estimates, based on the Survey of Construction for each year during the periods shown, include housing units used as vacation homes, second homes, and seasonal rentals. They may include some units that have subsequently been demolished or converted to nonresidential use by 1990, when the RECS data were collected. For some of them, the main space-heating fuel may have changed by 1990. The estimates from both sources are subject to sampling error.

Keeping these differences in mind, one still finds that the data in Table 3.4 strongly suggest that undercoverage associated with the 1990 RECS frame development and sample selection procedures was greatest for units built during the most recent period, 1988-1990.

The timing of the 1990 RECS data collection was such that one could not expect coverage of all newly-constructed housing units first occupied during the latter part of 1990, and units built but not yet occupied in 1990 would generally not be included. Nevertheless, the apparent undercoverage of new housing units for the 1988-1990 period is larger than could be fully accounted for by these factors and may be due in part to failure of the listing, updating, and half-open interval procedures to capture such units. The data show that the deficit is greatest for new housing units in multifamily buildings.

Table 3.4. 1990 RECS and Census Bureau Estimates of Housing Units by Year Built: 1980-1990

Year Built and Housing Type	Main Space-heating Fuel	Estimated No. of Housing Units (000)		Ratio of RECS Estimate to Census Estimate
		Census Bureau	1990 RECS	
1980-1984				
Single family	Gas or oil	1,954	2,461	1.26
Single family	Electricity or other	2,429	2,671	1.10
Units in multifamily buildings	All	2,433	2,083	0.86
Mobile homes	All	812 ^a	812	1.00
Total		7,628	8,027	1.05
1985-1987				
Single family	Gas or oil	1,722	1,245	0.72
Single family	Electricity or other	1,619	1,996	1.23
Units in multifamily buildings	All	1,789	1,230	0.69
Mobile homes	All	778	610	0.78
Total		5,908	5,081	0.86
1988-1990				
Single family	Gas or oil	1,916	1,327	0.69
Single family	Electricity or other	1,188	641	0.54
Units in multifamily buildings	All	1,160	324	0.28
Mobile homes	All	622	475	0.76
Total		4,886	2,767	0.57

^aCensus data were not available for this period, so the 1990 RECS estimate has been used.

Source: Energy Information Administration, *Consumption and Expenditures* (1990), Tables B19-B21.

Similar comparisons based on data from the 1993 RECS suggest much improved coverage of new housing units in that survey year, perhaps as a result of the special sampling procedures that were used to ensure a larger sample of newly-constructed housing units. The Census Bureau estimated that 4.070 million housing units were built or, in the case of mobile homes, put in place, from 1991 through 1993. The 1993 RECS estimate for this time period was 4.5 million housing units (plus or minus 0.7 million at the 95 percent confidence level), reversing the pattern shown in Table 3.4 for estimates of the newest homes in the 1990 RECS.