

9. Summary

The goal of the Residential Energy Consumption Survey is to provide periodic high-quality national and regional data about household energy consumption and expenditures and related characteristics of housing units and households. The National Interim Energy Consumption Survey (NIECS) in 1978 initiated an ongoing program to solve the many challenging problems that confront attempts to collect accurate survey data on these topics.

A noteworthy feature of RECS from the beginning has been its use of several different sources of data to provide the most accurate information that can be obtained at a reasonable cost. The Household Survey is the central component of the RECS design, but most of the direct information about energy consumption and expenditures of sample households is obtained from their energy suppliers and distributors. For households living in rental units, the accuracy of information about some housing unit characteristics is improved by contacts with their rental agents. Local weather information obtained from the National Oceanic and Atmospheric Administration serves several important purposes. A full understanding of the quality of RECS data requires awareness of how these different sources relate to each other and how the data are integrated to provide a comprehensive picture of the residential energy sector.

Another unusual feature of RECS has been the determination of the total and heated floor space of sample housing units through direct measurement by the survey interviewers. It became evident early on that survey respondents could not provide accurate estimates of floor space (see Chapter 5, section on "Special Data Collection Procedures" and Table 5.1), so the measurement procedure was adopted in the 1980 survey. This is just one example of many procedural improvements that have been introduced over the life of RECS in a constant effort to improve the quality of the survey data and the efficiency of the survey design and procedures.

A few innovations have not been successful. In the 1990 RECS, for example, interviewers attempted to record name-plate information from central air-conditioning equipment so that measures of rated efficiency for the equipment could be obtained by matching against directories of equipment characteristics. The desired information was obtained for fewer than one-fifth of the sample households with central air-conditioning, so the procedure was not repeated in the 1993 RECS (see Chapter 5, section on "Special Data Collection Procedures").

For the analysis of energy consumption, it is important to know not only the total consumption of each fuel, but how consumption is allocated to major end-use categories, such as space and water heating, cooling, and appliances. It has not been feasible to collect such information directly, so an indirect model-based approach has been developed to allocate total consumption of each fuel to these different end uses in each sample household. Much developmental effort has been devoted to improvement of the end-use estimation procedures, both by refinements in the structure of the model and by the introduction of new survey items whose use as independent variables in the model has potential for improving the accuracy of the estimates.

The next section of this chapter summarizes the information about sources and types of errors in RECS estimates that was presented in Chapters 3 through 8. The following section identifies some methodological research currently in progress and some design and procedural changes that

are being considered for the 1996 RECS and beyond. The final section presents some suggestions to data users for making effective use of RECS data, taking into account what is known about their quality and how it has been affected by design and procedural changes during the life of the survey.

Principal Sources of Error

Coverage Error

Vacant units and units occupied only on a seasonal or occasional basis are deliberately excluded from RECS. Such units have accounted for between 9 and 12 percent of all housing units (see Table 3.1) and probably for a smaller proportion of total residential energy consumption. Group quarters and institutions are not considered part of the residential sector and are also excluded. Prior to the 1980 RECS, there was no sample coverage of Alaska and Hawaii, and individual housing units on military bases were not covered.

Consistent with experience in other surveys and censuses, there is evidence that RECS does not achieve full coverage of households in its target population. From 1980 through 1990, RECS sample estimates of the number of U.S. households, following the application of sample weights and adjustments for unit nonresponse, have been adjusted upward by between 6.6 and 9.7 percent in order to agree with benchmark estimates derived from the Census Bureau's Current Population Survey. For 1993, the upward adjustment declined to 4.2 percent (see Table 3.3). The CPS benchmark estimates themselves have been benchmarked to agree with projections based on decennial census counts, but do not reflect known undercounts of households in the census. Except in the 1980 and 1993 RECS, these adjustment factors have been substantially higher for the South than for the other three Census regions.

There is also evidence of differential undercoverage in RECS of housing units classified by year built, with the most recently constructed units being most likely to be missed (Table 3.4). This problem may be due in part to the need to complete the list updating and sample selection operations prior to the survey reference date and in part to errors in carrying out the procedures for updating the sampling frame.

Certain kinds of energy consumption associated with households in the target population are deliberately excluded from the RECS estimates. Some but not all of the energy consumption for outdoor uses, such as lawn mowers and outdoor grills, is excluded. Wood energy consumption is not included in the formal estimates of total energy consumption; however, sufficient information about wood consumption is collected to provide the basis for a rough estimate of what it might contribute if included in the total. No estimates are developed for fuel equivalents of solar energy. Finally, data users should be aware that RECS measures energy consumption at the point where it enters the residence (site consumption); therefore, the survey estimates of consumption of electricity for years prior to 1993 do not reflect the total amount of energy used to generate it (primary consumption). The *Consumption and Expenditures 1993* report includes two tables--5.2 and 5.4--which show both site and primary consumption of electricity, classified by several housing unit and household characteristics.

Nonresponse: Household Survey

Unit response rates for the RECS Household Survey have two components: the proportion of eligible households for which acceptable questionnaires were completed in personal interviews and the proportion for which questionnaires were obtained by mail (in 1993 by mail and also by telephone) following unsuccessful efforts to conduct personal interviews. The latter group can be regarded as partial responses, because the mail and telephone questionnaires contain only a few key items from the interview questionnaire, their main purpose being to get permission from these households to contact their suppliers in order to obtain consumption and expenditure data.

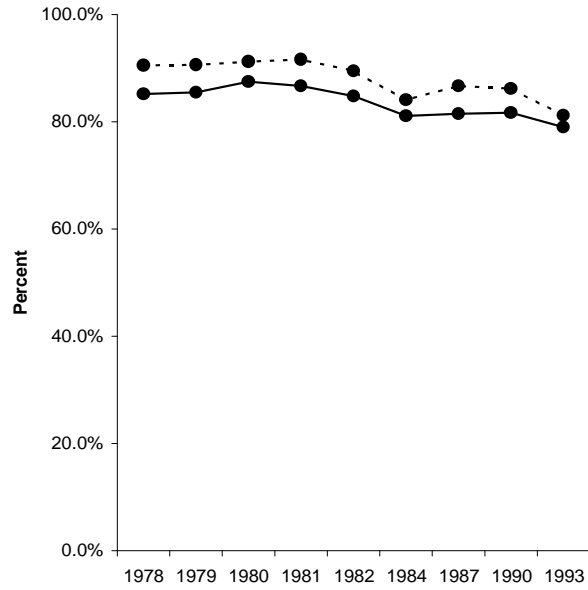
As can be seen in Figure 9.1 and Table 4.1, unweighted interview completion rates remained fairly steady, around 85 percent, from 1978 through 1982. For the next three survey years, they were at a lower level, between 81 and 82 percent, and in 1993 they declined to their lowest level, 79 percent. The percent of questionnaires completed by mail varied in a fairly narrow range, between 3.0 and 5.3 percent, between 1978 and 1990, but in 1993, the percent completed by mail or telephone was only 2.2 percent. As a result, the combined response rate in 1993 was 81.2 percent, nearly three full percentage points below the previous low in 1984. One factor that probably affected response rates for survey years 1982 through 1990 was the presence of a longitudinal component in the sample for those years, with roughly half of the households having been asked to participate in an earlier survey and some also to participate in the Residential Transportation Energy Survey that followed the earlier survey. There is clear evidence that response rates were lower for these "recycled" households (see Table 4.2). However, this was not a factor for the 1993 RECS, whose sample did not include a longitudinal component.

Some fairly consistent patterns have been observed in the relative response rates for different subgroups of the RECS target population (see Tables 4.3a, b, and c). The Northeast region has consistently had the lowest overall response rates. The South has had the highest personal interview response rates, but the lowest mail response rates for most years. Households in urban areas have had the lowest overall response rates and those in rural areas have had the highest rates. Except for 1993, households in structures with five or more housing units have had the lowest personal interview response rates.

As explained in Chapter 7, the estimation procedures for RECS include adjustments for unit nonresponse that are designed to minimize the effects of bias resulting from differential response rates by Census region and urban/rural status. The effectiveness of such adjustments depends on the level of nonresponse and the extent to which the characteristics of responding and nonresponding households within each control group are similar.

Item nonresponse has been relatively low for most items included in the Household Survey. The item nonresponse rate for an item is calculated as the percent of those questionnaires requiring responses to the item for which no response or a refusal was recorded. In the 1990 RECS, only 51 of 416 survey variables had item nonresponse rates of 5.0 percent or more (questionnaires completed by mail were not included in these rates). Not surprisingly, nonresponse rates to the basic question on household income were relatively high for all survey years--for example, they were 14.4 percent in 1990. However, about four-fifths of the households that refused or were unable to respond to the basic income question were willing to say whether their income was

Figure 9.1 Household Survey Completion Rates: 1978-1993



Source: Energy Information Administration, Residential Energy Consumption Survey, 1978-1993.

above or below a single cutoff value (\$35,000 in 1990). Other kinds of items that have had relatively high nonresponse rates include those relating to presence and amount of insulation and those relating to ages of equipment and appliances. Nonresponse for some of these items, such as age of heating equipment, tends to be concentrated among households living in rental apartments. Most of the nonresponse for these topics probably comes from respondents' inability to answer questions about them accurately.

A significant part of each Household Survey interview is devoted to the measurement and recording of information about total and heated floor space. The proportion of housing units for which no usable measurements were obtained has been consistently low, with a maximum of 6.4 percent in the 1984 RECS. However, the proportion of units with only partial information has ranged from 15 to 38 percent (Table 4.5). The most frequent omission is failure to state whether the measurements recorded have been taken inside or outside of the housing unit. (Outside measurements are preferred.)

Imputed values are substituted for most missing items, the main exceptions being questions relating to insulation and temperatures maintained in the home. Hot-deck imputation procedures, in which values for the missing items are randomly selected from groups of households that match on related variables, are used for most items.

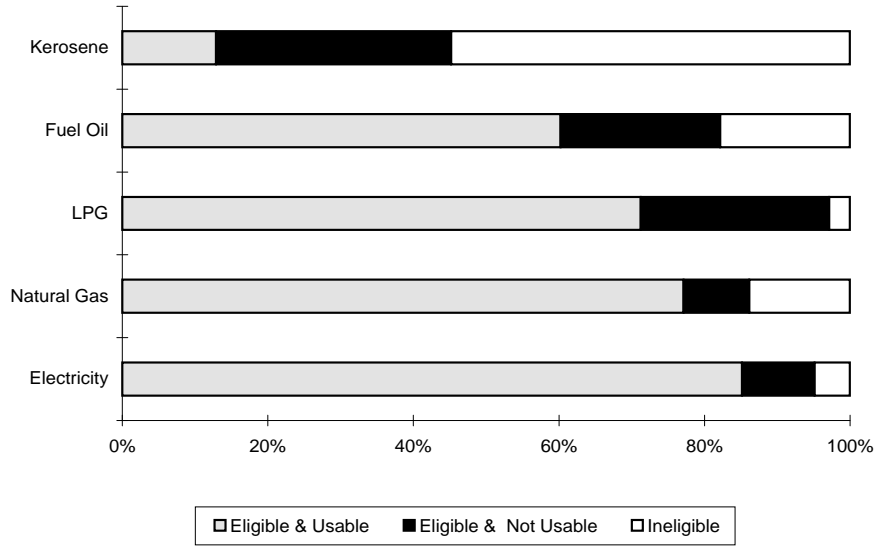
Nonresponse: Supplier Survey

Unlike the unit response rates for the Household Survey, the corresponding rates for the Supplier Survey have remained relatively stable since the beginning of RECS (see Table 4.7). Unit response rates for the Supplier Survey are defined as the percent of eligible households for which usable billing records were obtained. The major variations in eligibility and response rates are by fuel.

Supplier Survey response rates for electricity and natural gas, which together currently account for about 85 percent of the total consumption of the five major fuels, have remained consistently high, in the neighborhood of 90 percent. For fuel oil and LPG, which account for most of the remaining consumption, the rates have varied between 60 and 75 percent, and for kerosene the rate has varied from 19 to 32 percent (separate data for fuel oil and kerosene are available only from 1984 on).

Ineligibility rates for the supplier survey for households using a particular fuel are also a factor affecting the quality of data on consumption and expenditures for that fuel. Households are ineligible either because they do not pay separately for all uses of the fuel or because they purchase it mainly on a cash and carry basis. Thus, their suppliers would not have records containing the desired information. As shown in Figure 9.2, eligibility rates are highest for electricity and LPG, somewhat lower for natural gas and fuel oil, and lowest for kerosene.

Figure 9.2 Supplier Survey Eligibility and Completion Rates, by Fuel: 1993



Source: Energy Information Administration, Residential Energy Consumption Survey, 1993.

For electricity and natural gas, supplier billing data for part of a year can be used to impute a full year's consumption and expenditures. For fuel oil, kerosene, and LPG, billing data are only considered usable if they cover the full year. For households that were not eligible for the Supplier Survey and those for which usable billing data were not obtained, responses to household survey questions on deliveries of fuel oil, kerosene, and LPG are sometimes used to estimate consumption. When this cannot be done, the end-use consumption model is used to estimate total consumption of the fuel, as well as its allocation to different end uses.

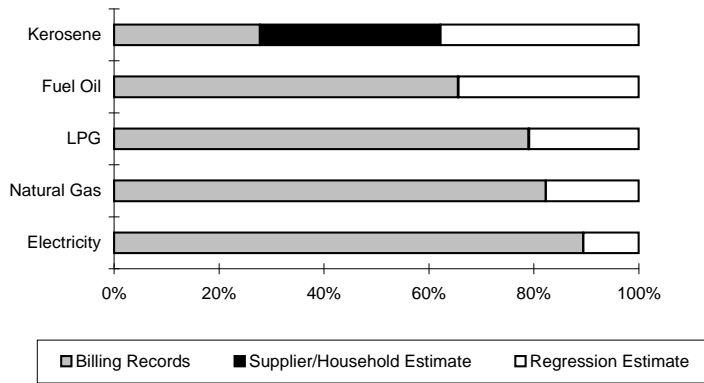
As shown in Figure 9.3 and Table 6.5, the proportion of total consumption that was imputed by the end-use model was lowest for electricity (about 10 percent in 1990 and 1993) and highest for fuel oil and kerosene (each between 25 and 35 percent in those years). Figure 9.4 shows that natural gas and electricity account for about 85 percent of total consumption of the 5 major fuels. Thus, for all major fuels combined, about 75 percent of estimated total consumption was based on Supplier Survey billing data for all or most of the year. The proportion of total consumption imputed by the end-use model was lowest for households living in single family structures and highest for those living in structures with five or more housing units (see Table 4.8).

Measurement Error

In RECS the dividing line between nonresponse error and measurement or response error is not always sharply drawn. The role of the Rental Agent Survey illustrates this point. For households in multiunit structures that have one or more fuels included in their rent payments, information about selected housing characteristics is collected from their rental agents, because experience has demonstrated that the agents can often provide more accurate information about items like year of construction and main heating fuels and equipment. Failure to obtain information from rental agents (nonresponse in the Rental Agent Survey) does not amount to item nonresponse, but it does mean that a response from the Household Survey that is more likely to be in error takes the place of information from the preferred source. Similar considerations apply when no usable information on consumption and expenditures is obtained from suppliers of households that are eligible for the Supplier Survey. As shown in Table 4.6, the proportion of eligible households for which the rental agent survey was completed varied from 57 to 89 percent between 1981 and 1993, and has exceeded 80 percent for each of the last three surveys.

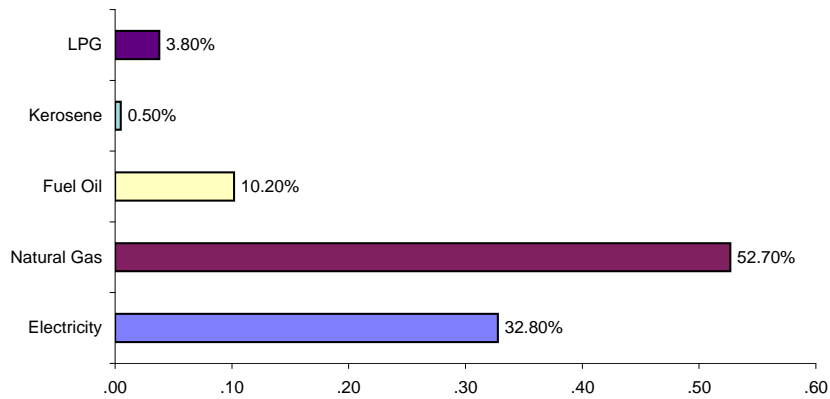
There are no systematic continuing sources of information about measurement error in RECS. As noted in Chapter 5, information about response variance, interviewer variance, and bias has come largely from occasional studies, in some instances restricted to a small set of sample households and often providing indications, rather than direct measures, of these components of total survey error. Such studies were more frequent in the earlier survey years. Two particularly useful sources of information have been longitudinal comparisons for households that were in the longitudinal component of the sample (see Tables 5.3 to 5.5) and reinterviews of a few households that had unusually large differences between consumption reported by suppliers and model-based estimates of consumption (see Figure 5.1).

Figure 9.3 Sources of Data for Estimates of Total Fuel Consumption, by Fuel: 1993



Source: Energy Information Administration, Residential Energy Consumption, 1993.

Figure 9.4 Proportion of Total Energy Consumption, by Fuel: 1993



Source: Energy Information Administration, Residential Energy Consumption Survey, 1993.

A review of the available evidence indicates that certain items are especially difficult for some respondents to answer and are therefore subject to relatively high levels of response variability and possibly bias. Such items include the year of construction of the housing unit (see Tables 5.3 and 5.6) and the number of windows in the unit (Table 5.3). Interviewers have had some difficulty distinguishing single-family attached housing units from single-family units and from units in apartment buildings with two to four units (Table 5.5). An examination of changes based on the Rental Agent Surveys (Table 5.7) suggests that households eligible for that survey often gave incorrect information about the type of main heating equipment and the fuels used for space and water-heating and for air-conditioning.

Comparisons of data on family income from RECS and the Census Bureau's Current Population Survey suggest that RECS, on the average, is understating income, even after the possible effects of conceptual differences are considered (see Table 8.6). The income questions used in the CPS distinguish more different sources of income than those used in RECS and ask for separate information for each member of the household. CPS data, in turn, appear to understate income when compared with data from the Census Bureau's Survey of Income and Program Participation (Jabine 1990, Table 10.1), which uses a still more extensive set of income questions.

There have been no formal evaluation studies of the quality of billing data provided by the energy suppliers for the RECS sample households. One cannot assume that billing records are entirely error-free. For electricity and natural gas, errors may occur when meter readings are estimated or read by the consumer. Even when the meter is read by a company employee, the meter may not be entirely accurate or a value may be incorrectly recorded. Nevertheless, it is unlikely that errors in the billing records have a significant effect on overall quality. As discussed in the next subsection, a more important determinant of the quality of consumption and expenditure data for households with usable supplier data is the set of procedures used to annualize the billing data and to adjust the values for households that use some energy for nonresidential purposes.

Data Processing and Imputation

A complex set of manual and automated processing procedures is used to convert the completed questionnaires from the three component surveys of RECS plus the weather data from NOAA into a set of usable data files. Both manual reviews and computer-assisted edits play a major role in attempts to locate and eliminate errors in the data. Following the initial computer edits, hard copy questionnaires are often consulted when inadmissible or inconsistent values are identified. In a small proportion of cases, respondents are contacted by telephone. These procedures are described in detail in the first section of Chapter 6, "Data Processing Other Than Imputation," and some aspects are summarized in outline form in Figures 6.1 through 6.5.

During the computerized data processing operations, an archival file is created containing records of changes made to individual records at each stage. An evaluation study based on the archival file for the 1984 RECS showed that many changes were needed to correct errors that occurred during data entry (Table 6.1). Based on this finding, sample verification of data entry, which had been adopted as a cost-saving measure for survey years 1981 through 1984, was replaced by 100 percent verification in all subsequent survey years.

In one phase of data processing, weather information obtained from NOAA is used to estimate heating and cooling degree-days for different locations and these estimates are linked to individual sample housing units associated with those locations. Initially, temperature data for each of 345 (as of 1987) NOAA divisions were used for this purpose, but studies (see Chapter 5, section on "Comparisons of Individual Household Data from Alternate Sources" and Table 5.8) indicated that it would be more accurate to use, for each sample housing unit, estimates of degree-days based on the nearest of NOAA's more than 4,000 individual weather stations. This change was made in the 1987 RECS.

Item nonresponse in the Household Survey is low for most items, a major exception being income (see Table 4.4). Missing responses are imputed for most variables and the hot-deck method of imputation is the one most frequently used. A substantial amount of imputation is required for the small proportion of households for which the Household Survey information is collected by mail or telephone, because the content of the mail and telephone questionnaires is deliberately limited to a few key items. The imputation procedure for these questionnaires, which is outlined in Figure 6.6, links them (the "donees") to personal interview questionnaires ("donors") that match on a set of variables that are available on both sets.

The development of consumption and expenditures data requires several types of imputation, all of which are subject to some degree of error:

- For households with usable Supplier Survey data, an elaborate "annualization" procedure is used to convert information for the supplier's billing periods or delivery dates to estimates for a consumption year that contains exactly 365 days and is as close as possible to the RECS reference period for consumption--for example, calendar year 1993 for the 1993 RECS. The annualization procedure is described in Chapter 6, in the section on "Imputation," and the procedure for electricity and natural gas is shown schematically in Figure 6.7.
- For kerosene, usable Supplier Survey data are obtained for fewer than one-third of the sample households. For most households that obtain kerosene on a cash-and-carry basis, estimates of consumption are based on Household Survey questions about number and usual size of purchases during the reference year.
- When neither of these sources of consumption data is available, consumption of the fuel for the household is estimated by a nonlinear regression model. This model, which is also used to allocate total consumption to end uses, is discussed further in the next subsection.
- If a household reports that part of its consumption of a particular fuel was for nonresidential purposes, such as farming or a home business, the reported or imputed consumption is adjusted downward. Adjustments are based on Household Survey questions which ask respondents to choose one of five class intervals containing the estimated proportion of the fuel used for such purposes.

Estimation

The weighting procedures used to produce RECS sample estimates are similar to those used in several other U.S. national household surveys. The final estimation weight for each household (see Chapter 7, section on "Sample Weighting Procedures," and Figure 7.1) is the product of three components: a weight based on the household's overall probability of selection; an adjustment for unit nonresponse; and a ratio adjustment. The ratio-adjustment component serves two purposes: to reduce the effect on the sampling error of the variation between primary sampling units and to reduce the mean square error of estimates by benchmarking them to household counts based on the Census Bureau's Current Population Survey (CPS). Initially, the second part of the ratio-adjustment component benchmarked the sample estimates to 12 control totals consisting of CPS estimates for the four Census regions and three location categories within each region. Subsequently, the adjustments have been refined to provide separate control totals for one-person households, four large States, and the nine Census divisions.

Some special estimation procedures were developed for use in analyses of 1990 RECS data on energy consumption in recently built housing units. These estimation procedures were not applied to all data for the 1990 RECS, but they illustrate the potential for improving the precision of estimates used in specific kinds of analyses (see Chapter 7, subsection on "Special Estimation Procedures for New Homes").

There is no feasible direct method, in a national sample survey, of measuring the allocation of individual households' consumption of each fuel to different end uses like space heating and cooling, water heating, and various appliances. Nevertheless, it is important, for energy policy analysis and other purposes, to have estimates of consumption by major end-use category. Consequently, an indirect, model-based nonlinear regression method of end-use estimation has been developed for RECS (see Chapter 7, section on "End-use Estimation," for a detailed description). The independent variables for the model include many of the housing unit and household variables for which data are collected in the Household Survey, as well as heating and cooling degree-day estimates based on the temperature data obtained from NOAA. The end-use allocation model is also used to estimate total consumption of fuels for which no usable data are available from the Supplier Survey or other sources.

There have been many changes in the details of the end-use estimation methodology since it was first developed. In 1984 the original linear model was replaced by a nonlinear model, with the logarithm of the difference between actual and estimated consumption serving as the error term. For the 1990 RECS, the logarithmic error term was replaced by the difference between the fourth roots of estimated and actual consumption. In addition, there have been many changes in the content of the Household Survey questionnaires and hence in the data items available for use as independent variables in the model. Some new items have been added to the questionnaires primarily in hopes of reducing the mean squared error of the model estimates.

An early evaluation of the model was undertaken for housing units in apartment buildings as part of the 1981 RECS (Chapter 6, Section 6.2, "Evaluation of imputation procedures"). For apartment buildings with one or more RECS sample households, average measured consumption per housing unit was compared with values imputed for the sample households by using the

model. This study indicated that the model-based estimates were low by about 50 percent for electricity in households with air-conditioning and for natural gas in households that used it only for purposes other than space heating. Based on these findings, the end-use model for the 1984 RECS was modified to reflect differences between apartments and other units more explicitly.

Some more recent evaluations of the model have made use of special studies in which utility companies have used recently developed "submetering" procedures to measure the consumption of electricity for different purposes within a household (see Chapter 7, subsection on "Evaluation of End Use Estimation Procedures," and Table 7.2). Because of the limited scope of these studies, they do not permit definitive conclusions, but the results suggest that the model may have been overestimating the consumption of electricity for central and room air-conditioning and space heating, and underestimating its consumption for water heating.

Sampling Error

Sampling errors are estimated by using a balanced half-sample replication method for all published RECS estimates (see Chapter 7, section on "Sampling Errors"). Through the 1982 RECS, 32 half-sample replicates were used; subsequently, the number was increased, and 96 replicates were used for the 1993 RECS estimates of sampling error. The estimates of sampling error for individual items are used as inputs to a generalized variance model which estimates "row and column factors" that are included in the publications.

As shown in Figure 7.3, the row and column factors allow users, with a few exceptions, to determine an approximate value of the sampling error for any cell in a table. Instructions in the introduction and appendices to the published reports explain the use of the row and column factors, including their use to derive estimates of standard errors for ratios and differences of individual table cells (see, for example, EIA 1995a, pp. 18-20 and Appendix B).

The individual records in the RECS public-use microdata files and diskettes do not include the information, such as primary sampling unit identifiers or replication weights, that would be needed to permit users to estimate sampling errors for the variables included in their analyses. Inclusion of such information would lead to an unacceptable risk that the identities of some sample households or housing units could be determined.

Estimates of sampling error have some limitations. They are themselves subject to sampling error. The use in RECS of a sample design which selects a single primary sampling unit from each stratum precludes the possibility of obtaining strictly unbiased estimates of sampling errors. The sampling error determined for a particular estimate by use of the published row and column factors is an approximation to the value that was calculated for that estimate. (Table 7.4 shows some comparisons of direct estimates and approximate values.) The sampling errors for estimates of end-use consumption do not reflect the error of estimation of the model parameters and are therefore underestimates (to a lesser extent, this is also true for total consumption and expenditures).

Targets for the standard errors of estimates of average energy consumption per household in the 1993 RECS were set at 1.25 percent for the U.S. total, 2.75 percent for Census region totals, and 4.50 percent for Census division totals. As shown in Table 7.4, the estimated sampling errors were well below these target values in every instance.

Current Research and Potential Design Changes

Planning for the 1996 RECS is proceeding. Consideration is being given to how best to take advantage of recent developments in computer-assisted techniques for survey data collection and processing, both to reduce costs and to improve quality. Computer-assisted personal interviewing (CAPI) will probably be the principal data collection mode for the 1996 RECS. Another change being explored is conversion to a modern automated survey-processing system, such as Blaise, a system developed by the Netherlands Central Bureau of Statistics. Findings from a recent study of edit changes (Martin 1995), based on comparisons of tabulations of edited and unedited data for selected items from the 1993 RECS, will be helpful in deciding what kinds of edit checks to include in a CAPI version of the questionnaire.

In order to reduce costs and also in anticipation of conducting surveys after 1996 primarily by telephone, the number of topics and individual questions in the 1996 RECS will be considerably smaller than in 1993. Physical measurement of the area of floor space will no longer be part of the survey interview. Other variables, such as number of heated rooms, will take the place of floor space in the end-use consumption model.

Looking beyond 1996, serious consideration is being given to using computer-assisted telephone interviewing (CATI) with random-digit dialing (RDD) as the principal mode of data collection. Use of this mode would require a careful evaluation of its effects on response rates and other aspects of data quality. One question of special significance for RECS is whether it will be possible, with some combination of telephone and mail procedures, to obtain authorization to contact energy suppliers for a sufficiently high proportion of the sample households. A pilot test of the use of CATI/RDD procedures for the collection of RECS data is under way, and preliminary results are expected to be available early in 1996.

Some Suggestions for Data Users

User Options

The primary means of user access to RECS data are through publications and public-use data files. A full list of all EIA consumption survey publications and public-use files, along with instructions for obtaining them, is provided in Appendix A. The main publications for each RECS survey year are the reports on *Housing Characteristics* and on *Household Energy Consumption and Expenditures*. The latter is published in two volumes, the first containing national data and the second containing regional data. There have also been several special publications based on RECS, and some summary data from RECS are published annually in EIA's *Annual Energy Review* and in the Census Bureau's *Statistical Abstract*.

Public-use data files for each survey year are available to users who wish to develop their own tabulations or do other kinds of statistical analyses. Files for the 1987 RECS and subsequent years can be obtained on diskettes or downloaded from the Internet; those for earlier years are available on tapes. The public-use files contain data for individual sample households, including billing data from the Supplier Survey, with all identifiers removed in order to preserve the confidentiality of individual information. For the same reason, selected billing records with unusual values have been deleted, and random errors have been introduced for certain variables, such as degree-days and starting and ending dates of billing periods. Additional information about steps taken to preserve confidentiality is provided with the documentation that accompanies each public-use data tape or diskette.

Learning More About RECS

Each of the regular *Housing Characteristics* and *Household Energy Consumption and Expenditures* reports includes detailed appendices describing how the survey was conducted and discussing various aspects of the quality of the survey data. The reports also contain copies of the data collection forms used and an extensive glossary defining terms and concepts used in the survey. A detailed description of the sample design and selection procedures for all surveys through 1993 is contained in a 1994 report, *Sample Design for the Residential Energy Consumption Survey* (EIA 1994).

All public-use data files include extensive internal documentation. The *User's Guide* for the 1993 public-use files includes: information about the general nature of the survey; technical file specifications; variable listings, including information about variables that changed from the prior survey; unweighted and weighted frequencies for each variable; an explanation of the codes used for imputed variable values; copies of the questionnaires; and a list of "Cautions when Using RECS Data." The *User's Guide* also identifies persons to contact at EIA for additional information.

The above sources, along with this *Quality Profile*, should meet the needs of most users. However, additional information that may lead to a fuller understanding of some aspects of the quality of RECS data is contained in internal operating manuals for the surveys, including interviewer instruction manuals and, starting with the 1984 RECS, separate survey documentation reports covering sample design, data collection, and data processing procedures. The list of references cited in this report includes several articles, contractor reports, and internal memoranda and reports, some of which may be of interest to a few data users.

Using Cross-sectional Data: General Considerations

Effective use of data from any survey requires knowledge of the basic features of the survey design and awareness of how sampling and nonsampling errors may affect conclusions drawn from the data. The following suggestions apply to all users of RECS data, whether they are working with published tabulations or public-use data files:

- To obtain a general overview of the RECS objectives, content and design, review Chapter 2 of this report. It may also be useful to read the section of Chapter 3 on "RECS Target Populations," which provides information about the target populations and the reference periods and dates for each survey year.
- For data elements of particular interest, review the specific questionnaire items relating to these topics and the relevant definitions given in the glossary of each published report.
- Whenever possible, evaluate the statistical significance of any comparisons based on the survey data. For users of published data for 1984 and subsequent survey years, this can usually be accomplished by using the "row and column factors" appearing in each table according to the instructions provided. However, no row and column factors were provided for estimates of end-use consumption in the 1990 *Consumption and Expenditures* report.
- Be aware of the possible effects of coverage, nonresponse, and measurement errors on the estimates. The most accurate data on consumption and expenditures are for electricity and natural gas, because most of the data for these fuels are obtained from billing records obtained in the Supplier Surveys. For the same reason, consumption data for single-family owner-occupied housing units are likely to be more accurate than data for multi-family units and those occupied by renters. Estimated totals for newly-constructed housing units--those completed during the survey year and the years immediately preceding it--are likely to be low because of problems entailed in incorporating new units into the sampling frame. Additional information about nonsampling errors and their effects can be found in Chapters 3 through 8 of this report and in appendices to the *Housing Characteristics* and *Consumption and Expenditures* reports.

Special Considerations for Users of Public-Use Files

Users working with public-use data files should, of course, review the documentation that is provided with them. Additional recommendations are to:

- Use weighted data for all tabulations. Several features of the sample design, such as oversampling of low-income households and newly-constructed housing units in some survey years, require the use of variable weights to produce unbiased estimates. For analytical uses of the data, such as multivariate analyses, users may sometimes find it more convenient to use unweighted data. However, before deciding to do so, it would be advisable to determine the extent of variability of the sample weights for the housing units to be included in the analysis.

- Some users may wish to exclude imputed variable values from their analyses or to reimpute the values using a different procedure. The data files contain information that allows users to determine which values were imputed. Consult the public-use file *User Guide* for specific information on how to do this.
- Be aware of the possible effects of statistical disclosure limitation procedures that have been used to prevent data users from determining the identity of individual sample households. For survey years 1980, 1981, 1982, 1984, and 1993, records for sample households in Alaska and Hawaii were excluded from the public-use files. They were included in the public-use files for 1987 and 1990, but a substantial proportion of the billing records for households in the two States were excluded. More specific information is given in the documentation material that accompanies each file.

For the same reason, that is, to prevent disclosure, it has not been possible for the public-use files to include replication weights or other variables that would allow users to develop their own estimates of sampling error for items of interest. For most survey years, the most that users of these files can do is to obtain a range of possible values for the sampling error of a particular estimate by calculating standard errors for similar items appearing in the survey publications for the same year. For survey years 1981 and 1982 however, each of the publications includes an appendix with a generalized procedure for deriving an approximate sampling error for any item of interest.

Analyzing Changes Over Time

There have been many changes, since the initial NIECS effort in 1978, to the RECS survey design, content, and procedures. These changes, which are described in Chapter 2, in the section on "Evolution of the RECS Design: 1978-1993," have been motivated by efforts to respond to new data needs, improve the quality of the data, and take advantage of new technologies for survey data collection and processing. RECS data users who are interested in analyzing trends in housing unit characteristics and energy consumption need to be aware of these changes and their possible effects on comparisons of data for different survey years. The most important features to keep in mind are:

- An upper bound to the sampling error of the difference between estimates of the same item for different survey years can be obtained by assuming that the two estimates are independent. With this assumption, the appropriate formula is:

$$S_{X_1-X_2} = \sqrt{[S_{X_1}]^2 + [S_{X_2}]^2}$$

where X_1 and X_2 are the estimates for times one and two. Because the estimates for different survey years are positively correlated in most instances, the value derived from this formula will be an overstatement. For information about the extent of the overstatement, see Chapter 7, section on "Sampling Errors," and Table 7.3.

- For the first two survey years, 1978 and 1979, Alaska and Hawaii were not included in the target population for the survey. They have been included in all subsequent survey years, but, as noted above, all or some of the individual records for sample housing units in these two States have been excluded from public-use data files.
- The reference month for household counts was November for all survey years through 1990; in the 1993 RECS, it was changed to July. Through the 1984 RECS, the 12-month reference period for consumption and expenditures ran from April of the survey year through March of the following year. From the 1987 survey year on, the reference period has been the calendar year corresponding to the survey year (see Figure 2.2).
- Consumption data for fuel oil and kerosene were combined through the 1982 survey year; since 1984, data for the two fuels have been collected and presented separately.
- In the 1987 RECS, a significant change was made in the method of associating weather data (heating and cooling degree-days) with sample housing units. Consequently, weather data from 1987 to date are not comparable with data for earlier survey years (see Chapter 5, section on "Comparisons of Individual Household Data from Alternate Sources," and Table 5.8).
- Since its initial development, there have been frequent changes in the end-use consumption model which is used to allocate total consumption to specific end-uses and to impute total consumption when billing or delivery data are unavailable. There have been several significant structural changes in the model and, in virtually every survey year, changes in the data items used as independent variables. It is believed that most of these changes have significantly improved the reliability of cross-sectional estimates of end-use consumption, but, at the same time, they constitute an additional source of error in estimates of change between survey years.
- Estimates of end-use consumption are available for all survey years except 1979 in the following categories:

<u>Category</u>	<u>Fuels</u>
Space heating	All fuels
Water heating	All fuels
Appliances	All fuels
Air-conditioning	Electricity and natural gas

For electricity, in 1990 the appliance category was subdivided into refrigerators, freezers, and all other appliances. In 1993, the all other category was further subdivided to provide separate estimates for lighting, cooking, dishwashers, and clothes dryers.

- Users who are interested in analyzing long-term trends in residential energy consumption should keep in mind the likely effects of short-term fluctuations in average temperatures during the heating and cooling seasons, whether at the national, regional, or divisional level. In the same vein, analyses of variation across regions or divisions may be influenced by departures from long-term averages that differ in direction from one area to another. Trends in expenditures can be affected by fluctuations in both average temperature and energy prices. The 1993 *Consumption and Expenditures* report presented, for the first time, trend data for consumption and expenditures adjusted to control for the effects of price changes and variations in weather (EIA 1995d, pp. 3-7).
- A final suggestion for analysis of changes over time is to review the specific questions used in each survey year for the items included in the analysis in order to determine whether there have been any changes in wording, format, or placement that may have affected comparability between survey years. As noted above, the *User's Guide* for the 1993 public-use files includes complete information on variables that have changed from the previous survey.

User Feedback

EIA and the staff responsible for RECS are anxious to hear from data users. Let us know about your experiences in using the data, any problems you may have encountered, and your suggestions for improving the quality and utility of RECS data. Please contact Robert Latta by telephone (202/586-1385) or E-Mail (rlatta@eia.doe.gov).