DOE/EIA-0555(96)/1

Energy Consumption Series

Residential Energy Consumption Survey Quality Profile

March 1996

Energy Information Administration

DOE/EIA-0555(96)/1
Distribution Category UC-950

Energy Consumption Series

Residential Energy Consumption Survey Quality Profile

March 1996

Energy Information Administration
Office of Statistical Standards
Office of Energy Markets and End Use
U.S. Department of Energy
Washington, DC 20585

Contacts

This publication was prepared by the Energy Information Administration (EIA) jointly under the general direction of Dr. Yvonne Bishop, Director of the Office of Statistical Standards (OSS) (202-426-1068) and W. Calvin Kilgore, Director of the Office of Energy Markets and End Use (EMEU) (202-586-1617). The FAX number for all OSS personnel is 202-426-1081.

In EMEU, the project received general direction from Lynda T. Carlson, Director of the Energy End Use and Integrated Statistics Division (EEUISD) (202-586-1112) and Nancy L. Leach, Chief of the Residential and Commercial Branch (202-586-1114). The FAX number for all EEUISD personnel is 202-586-0018.

Detailed technical questions on the topics indicated may be referred to the following people:

Renee Miller OSS	Project Design	202-426-1117 rmiller@eia.doe.gov
Thomas Broene OSS	Project Design	202-426-1108 tbroene@eia.doe.gov
Robert Latta EEUISD	Sampling Design, Sampling Variances End-Use Estimation, RECS Manager	202-586-1385 rlatta@eia.doe.gov
Nanno Smith EEUISD	Public-Use Data	202-586-5481 nsmith@.eia.doe.gov
Hattie Ramseur EEUISD	Publications, Report Production	202-586-1124 hramseur@.eia.doe.gov

This report was written by Thomas B. Jabine, under contract to EIA. He received assistance from William Connors and Barbara Hammond, Allied Technology Group, Incorporated.

Public-Use Data Diskettes containing RECS data are available through the Office of Scientific and Technical Information (615-576-8401) and the National Technical Information Service (703-487-4650). For questions about RECS publications, call 202-586-8800. Information, reports, and household level data files are also available by accessing EIA's Home Page on the Internet at http://www.eia.doe.gov.

Tribute

Dr. Wendel L. Thompson, Residential Energy Consumption Survey (RECS) Manager, is retiring from Federal service on March 29, 1996. He has guided RECS for 17 years and has been instrumental in making it a survey known for its integrity and high quality data.

It is through his untiring efforts that RECS has continued to evolve and improve over the nine survey cycles--since 1979. With unfailing good humor, grace, and patience--under pressure, Wendel has been the friendly and knowledgeable expert answering everyone's questions about residential energy usage. Wendel epitomizes the very finest in customer service, which EIA and the entire Federal government strives to achieve.

"Mr. RECS" will surely be missed by his colleagues and all the RECS users.

We wish Wendel the best in his retirement.

Contents

Page

Figures

Tables

Page

1. Introduction

Purpose and Scope of This Report

The Residential Energy Consumption Survey (RECS) is a periodic national survey that provides timely information about energy consumption and expenditures of U.S. households and about energy-related characteristics of housing units. The survey was first conducted in 1978 as the National Interim Energy Consumption Survey (NIECS), and the 1979 survey was called the Household Screener Survey. From 1980 through 1982 RECS was conducted annually. The next RECS was fielded in 1984, and since then, the survey has been undertaken at 3-year intervals. The most recent RECS was conducted in 1993.

Purpose and Audience

The purpose of this *RECS Quality Profile* is to present, in a convenient form, a report on what has been learned about the quality of RECS data since the survey began. In a broad sense, the term "quality" covers the relevance, timeliness, and accuracy of the survey estimates. The emphasis here will be placed on "accuracy." The report provides information about sampling and nonsampling errors, focusing on the latter. It discusses the types and sources of errors that occur and their possible effects on interpretation of RECS data, especially when used for longitudinal analysis. This information should be helpful to users of RECS data, to those responsible for the design and operation of the survey, and to persons with general interest in survey design and data quality. The final section of Chapter 9 provides specific suggestions for data users on how to gain access to RECS data and use them for cross-sectional and longitudinal analyses.

Scope of the Report

This report includes information about the quality of data from all of the surveys conducted to date, starting with the 1978 NIECS through the 1993 RECS. As background, a summary of the survey design and procedures for the 1993 RECS is provided, as well as a description of quality-related design changes from 1978 through 1993. The report does not cover the Residential Transportation Energy Consumption Survey (RTECS), a separate survey that has been conducted for a subsample of RECS households in the year following each RECS data collection. The final chapter describes some ongoing methodological research and some anticipated design changes for the 1996 RECS.

Sources of Information About Data Quality

This report draws on both published and unpublished sources of information. Following established practices of the Energy Information Administration (EIA), all RECS publications have included a substantial amount of information about survey procedures and about the quality of the data. Appendices to the reports on *Housing Characteristics* and *Household Energy*

Consumption and Expenditures for all survey years have been a primary source of information for this report. It also draws on numerous survey documentation reports, contractor reports, papers presented at conferences, and internal memoranda. These sources are cited in the text of the report and as sources for information presented in tables by showing the author's name and the year of publication. A full list of all references cited follows Chapter 9.

Direct estimates of the sampling error associated with most RECS data are available and form the basis for row and column variance factors that have been included in the data tables found in the basic RECS reports, starting with the 1984 consumption and expenditures report (EIA 1987a). By using these row and column factors, data users can arrive at an estimate of the sampling error associated with the value found in each cell of the table.

Most available information about nonsampling errors is less direct. For example, there is detailed information about household and item nonresponse rates for each survey, but it is seldom possible to determine the resulting level of bias in the survey estimates that are produced after missing values have been imputed and weighting adjustments have been applied in an attempt to minimize the effects of nonresponse. Similarly, estimates for the same variable from two different sources can be compared, but without detailed analysis and reconciliation of differences between individual reporting units, it is generally not possible to be certain which of the estimates is more nearly correct. Nevertheless, the analysis of both direct and indirect indicators of nonsampling error can contribute in significant ways to understanding and interpreting the survey results and to efforts to improve quality.

Information about nonsampling errors comes from several sources:

- Operational or performance data, such as unit and item nonresponse rates, imputation rates, and weighting adjustments
- Methodological experiments and pretests of survey procedures
- Micro-evaluation studies, such as callbacks or reinterviews of respondents who have reported unusual values and energy audits or assessments of sample housing units by specially qualified persons
- Macro-evaluation studies, such as comparisons of RECS data on energy consumption with data from EIA surveys of energy suppliers, and RECS data on household and housing-unit characteristics with data from the Census Bureau's Current Population Survey, American Housing Survey, and Survey of Construction

Relation of This Report to Other RECS Publications

As noted above, the basic publications for each survey have appendices that describe the design and procedures for that survey and provide some information about the quality of the data. More detailed information about survey operations can be found in the operations, procedures, and data user's manuals that have been prepared for each survey, starting with 1982. A recent publication

provides a fuller description of the sample design for each survey through 1993 and changes over time (EIA 1994). A 1987 EIA publication, *Trends in Consumption and Expenditures, 1978-1984* (EIA 1987b) contains longitudinal data and a discussion of how their interpretation may be affected by changes in survey design, content, and procedures.

The particular goal of this *RECS Quality Profile* is to serve both users and survey designers and managers by providing a more systematic and complete presentation of information about data quality for all of the surveys from the 1978 NIECS through the 1993 RECS. To accomplish this, the report follows a structure similar to the one that has been developed in recent years for presenting information about the quality of data from other major household surveys, such as the Survey of Income and Program Participation and the Schools and Staffing Survey (Bureau of the Census 1990, National Center for Education Statistics 1994).

Structure of the Report

Chapter 2 provides an overview of RECS, with a description of the 1993 RECS design and procedures and of the quality-related design changes that have occurred since the first survey in 1978. Chapters 3, 4, and 5 present information about the three major sources of nonsampling error: coverage error, nonresponse, and measurement error, respectively. Chapter 6 discusses the contributions to nonsampling error of data processing and imputation procedures. Chapter 7 looks at the effects of estimation procedures on data quality and reviews the effects of sampling error on interpretation of the data.

Chapter 8 presents the results of studies that have compared RECS data with data from EIA surveys of energy suppliers and with data collected by other organizations, especially the Census Bureau. Chapter 9 summarizes the effects on RECS data of the principal sources of error. It also describes relevant research currently in progress and quality-related design changes planned or being considered for the 1996 RECS. It concludes with suggestions for users about how to take account of data quality in their analyses of RECS data. A list of all references cited follows Chapter 9. Appendix A lists all EIA publications related to RECS.

2. An Overview of RECS

For those who may be unfamiliar with the nature and principal features of the Residential Energy Consumption Survey, this chapter of the *Quality Profile* presents some background information. The first section provides a general overview of RECS, its objectives, and the timing of the periodic surveys. The second section describes the design and methodology of the 1993 RECS. The final section identifies significant changes as the survey design and procedures evolved from the initial survey (known as NIECS, the National Interim Energy Consumption Survey) in 1978 through the 1993 RECS.

A General Overview of RECS

Objectives

RECS, the Residential Energy Consumption Survey, is a periodic sample survey that is designed to provide timely information about energy consumption and expenditures of U.S. households and about energy-related characteristics of these households. RECS data are developed for use by the Congress, Government agencies, researchers, and the general public. The data provide major inputs to EIA's National Energy Modeling System, a forecasting system that has been developed for the Department of Energy. In the 1981, 1984, 1987, and 1993 survey years, RECS included a supplemental sample of low-income households to provide information needed by the Administration for Children and Families, Department of Health and Human Services, to provide data needed to administer its Low-Income Home Energy Assistance Program.

Legal Authority for Survey

RECS is conducted under the authority of the Federal Energy Administration Act of 1974 (Public Law 93-275), as amended, and the Energy Policy Act (EPACT) of 1992. The latter act directs the EIA to conduct a survey of residential energy use at least once every 3 years.

Data Dissemination

Beginning with the 1980 RECS, the data from each survey have been presented in two major publications, a *Housing Characteristics 1993* report and a *Household Energy Consumption and Expenditures 1993* report. Anonymized data files with information for individual households have been released to researchers and other data users on public-use tapes for all RECS survey years and on diskettes for the 1987 and subsequent surveys. Since August 1995, it has been possible for users to download the data files for the 1987, 1990, and 1993 RECS from the Internet by contacting the EIA home page (http://www.eia.doe.gov).

Sources of Data

The information contained in the RECS data files and publications comes from several sources (Figure 2.1). The primary data source is the Household Survey, in which data are collected, mostly via personal interviews, from a sample of several thousand households. For most of these households, a Supplier Survey is undertaken to obtain billing information on energy consumption and costs from the households' suppliers of electricity, natural gas, fuel oil, kerosene, and LPG. For some households occupying rental units in multiunit structures, a Rental Agent Survey is conducted to improve the accuracy of data on the main fuel sources and types of equipment used for space and water heating and air-conditioning.

For every sample household, information on heating and cooling degree-days as measured by a nearby weather station (and humidity, starting in 1990) is obtained from the National Oceanic and Atmospheric Administration (NOAA). Finally, data from the Census Bureau's decennial censuses of population and housing and Current Population Survey are used in the RECS sample design and estimation procedures to improve the reliability and precision of the survey results.

Figure 2.1. Sources of Information for the RECS System

Source	Information Provided (preferred source)	Household Type	Fallback Source
Household survey	Housing-unit and household energy-related characteristics	All housing-units	
Supplier survey	Housing-unit consumption and expenditures by fuel type.	Households that pay supplier directly for one or more delivered fuels	Household survey (kerosene only).
Rental agent survey	Main fuel source for space and water heating, cooking, air-conditioning	Households in multi-unit structures with one or more fuels included in rent	Household survey
NOAA	Weather data for station close to each sample housing-unit.	All housing-units	
Census Bureau	1. Data for formulation of sample strata. 2. Household estimates for benchmarking RECS estimates.		

Source: Energy Information Administration, *Consumption and Expenditures* (February 1993).

Frequency, Reference Periods, and Sample Sizes

The survey has been conducted nine times between 1978 and 1993. As shown in Figure 2.2, it was conducted annually from 1978 to 1982, then 2 years later, in 1984, and subsequently at

3-year intervals. Data collection for the Household Survey has started in the autumn of each survey year and has continued through the winter or early spring of the following year. For the first 8 survey years, through 1990, the "reference month" for households has been November of the survey year. For the 1993 RECS, July 1993 was chosen as the reference month. RECS estimates of the number of households are benchmarked to estimates of households for the reference month from the Census Bureau's Current Population Survey. Through the 1984 survey year, the reference period for consumption and expenditures was the 12-month period starting in April of the survey year and extending through March of the following year. From survey year 1987 on, the reference period has been the calendar year coinciding with the survey year.

Sample sizes are the numbers of sample units initially assigned to interviewers that turned out to be eligible for the survey and for which acceptable questionnaires were obtained by field or telephone interviews or by mail (Figure 2.2). The numbers of units initially assigned and the corresponding eligibility and response rates achieved may be found in Chapter 4, Table 4.1. The larger samples in 1981, 1984, 1987, and 1993 were due in part to the inclusion of supplemental samples of households in low-income areas, as noted above. The sample for 1993 was also supplemented to provide a larger sample of newly-constructed units, defined as those built in 1987 or later.

Figure 2.2. Key Features of RECS, by Survey Year

Survey Year	Reference Month[a] for Households	Reference Period for Consumption and Expenditures	Sample Size[b] (Households)
1978	November	Apr 78 - Mar 79	4,081
1979	November	Apr 79 - Mar 80	4,033
1980	November	Apr 80 - Mar 81	6,051
1981	November	Apr 81 - Mar 82	6,269
1982	November	Apr 82 - Mar 83	4,724
1984	November	Apr 84 - Mar 85	5,682
1987	November	Jan 87 - Dec 87	6,229
1990	November	Jan 90 - Dec 90	5,095
1993	July	Jan 93 - Dec 93	7,111

[a]Month of survey year for which number of households was estimated.
[b]Number of sample households for which acceptable questionnaires were obtained.
Sources: Energy Information Administration, *Housing Characteristics* (1980-1993); *Consumption and Expenditures* (1978-1993).

Overview of the 1993 RECS Design and Procedures

This section summarizes the main features of the 1993 RECS design and procedures. More detailed information is provided in Appendix A, "How the Survey Was Conducted," in the *Housing Characteristics 1993* report (EIA 1995a), the comparable appendix in the *Household Energy Consumption and Expenditures 1993* report (which includes information about the Supplier Survey, EIA 1995d), and the various procedures manuals developed for the 1993 RECS (for example, Response Analysis Corporation 1995a,b).

Content

The data collected for households and housing-units in the 1993 RECS can be classified into seven major categories:

1. Energy consumption and expenditure by the household
2. Housing-unit characteristics, equipment, and appliances most directly related to energy use
3. Socioeconomic characteristics of the household occupying the housing-unit
4. Energy sources, uses and suppliers
5. Ownership and use of vehicles
6. Use of energy assistance programs
7. Participation in demand-side management programs of utility companies.

Providing an estimate of energy consumption and expenditure by source for the household and by end use within the household is the primary goal of RECS. Data in the remaining six categories serve several purposes. In addition to being of interest in their own right, variables such as the year the structure was built, the number of rooms in the housing-unit, the area of heated floorspace, and household income can be used as classifiers in presenting and analyzing estimates of energy consumption and expenditures. Socio-economic data for the household and data on housing characteristics, equipment, and facilities are extensively used in the imputation of missing consumption data and in the model-based allocation of the consumption of each fuel to end-use categories. Data on ownership and use of vehicles are used primarily as inputs to the Residential Transportation Energy Consumption Survey and as the basis for selecting a subsample of RECS households for that survey.

Although there has been substantial variation in specific items within each category, data in the first five categories have been collected in all RECS surveys since 1980. Inclusion of questions about the use of energy assistance programs began in the 1981 RECS, coincident with the Congressional authorization of the Low-Income Home Energy Assistance Program (LIHEAP) in that year. Participation in demand-side management programs (activities sponsored by utilities that are intended to influence the timing and amount of electricity use) was the subject of two questions in the 1990 RECS. Detail on this subject was substantially expanded in the 1993 Household Survey questionnaire, and utilities were asked in the 1993 Supplier Survey to report on the programs they offer to residential customers.

In 1992, the EIA conducted a user-needs study to help determine the content of the 1993 RECS (EIA 1993c). Written suggestions were solicited from data users and EIA staff held 15 meetings with different user groups. The findings from that study, as well as specific requirements

mandated by the Energy Policy Act of 1992, led to several changes in content. Important topics covered by questions added to the 1993 RECS questionnaire were:

- Consumer decision-making behavior, such as purchases of new equipment and the influence of energy efficiency considerations on purchase decisions

- Additional building envelope and thermal characteristics, such as exterior wall materials and number and type of windows

- Lighting: Number of lights, location, bulb type, wattage and control mechanisms; i.e., a more detailed lighting supplement was administered to a subsample of households

- Appliance Usage and Equipment: Levels of usage, especially of equipment using hot water, such as washers and dishwashers; identification of some appliances not previously included on the questionnaire

- Emerging Technologies: Awareness and use of several new energy-conserving technologies, such as low-E window glass and halogen light bulbs

- Business use of homes.

In order to accommodate these new questions without undue increase in response burden, several questions that appeared on the 1990 RECS questionnaire were dropped, taking into account expressions of user interest, past experience in the ability of respondents to answer accurately, and demonstrated utility of specific questions for imputation of missing data and estimation of energy consumption by end use. Several questions on space heating, insulation, and other housing characteristics were among those dropped. A more detailed account of these changes is provided in the supporting statement submitted to OMB for clearance of the forms (EIA 1993b).

Sample Design and Selection

All of the RECS surveys have used a stratified, multistage sample design. Basic principles that guide the sample design are: use of probability sampling at all stages of selection; the ability to produce estimates of acceptable reliability for each of the nine Census divisions and for other subgroups of the target population; and, at the national level, determination of sample sizes at all levels in a way that produces the most reliable estimates possible, given the resources available for the survey. For the 1993 RECS, additional requirements were to oversample newly constructed housing-units and units occupied by households with income below the poverty level, that is, to sample them at rates higher than those used for units not in these categories.

The types of sampling units and the sample selection procedures for each stage of sampling in the 1993 RECS are important (Figure 2.3). The grouping of primary sampling units (PSU's) into strata was carried out within Census divisions, and, within divisions, separately for each of four States with large population (California, Florida, New York, and Texas) and for Alaska and

Hawaii because of their unique weather conditions. Of the 116 strata, 31 consisted of a single metropolitan area that had a large enough population to form a stratum by itself.

The 1993 sample was the first one for which 1990 Census data were available for use in stratification of PSU's and secondary sampling units (SSU's) within sample PSU's, and for assigning selection probabilities at each level. Although a new set of PSU's was selected, for purposes of efficiency this was done in a manner that maximized the overlap with the sample PSU's used in the 1984, 1987, and 1990 surveys. Of the 116 sample PSU's selected for the 1993 RECS, 94 had been included in the three preceding surveys.

Figure 2.3. Sample Design and Selection Procedures for 1993 RECS

DEFINITIONS

Primary Sampling Units (PSU's):

Metropolitan areas or groups of counties in non-metropolitan areas.

Secondary Sampling Units (SSU's):

One or more Census blocks, with a minimum of 50 housing-units in the most recent census.

Listing Segment:

A complete SSU or a selected part of an SSU for which a detailed listing of street addresses is prepared.

Ultimate Cluster:

A small group of (potential) housing-unit addresses selected from the listing for a listing segment.

SAMPLE SELECTION STEPS

1. Divide the United States into PSU's, each consisting of a metropolitan area or one or more non-metropolitan counties.
2. Group PSU's to form strata (some strata consist of a single large metropolitan area).
3. Select one PSU from each stratum.
4. Select several SSU's from each sample PSU.
5. (Larger SSU's only) Divide SSU into listing segments and select one.
6. Prepare a detailed address listing for each SSU or listing segment.
7. Select ultimate cluster of addresses from each SSU or listing segment.

Figure 2.3. Sample Design and Selection Procedures for 1993 RECS (Continued)

SAMPLING FRAME AND SAMPLE SIZE INFORMATION	
PSU's in United States:	1,786
Strata:	116
Sample PSU's:	1 per stratum
Sample SSU's in sample PSU's:	
Core sample:	1,461
Supplement :	149
Total:	1,610
Mean per sample PSU:	13.9
Listed units assigned to field:	
Total:	9,869
Mean per sample SSU:	6.1

Source: Energy Information Administration, *Housing Characteristics* (1993).

The basic criterion for assigning selection probabilities at each stage of sampling was to produce a national self-weighting sample of housing-units, that is, one for which the overall selection probability of each unit, taking all stages of sampling into account, would be the same. However, some departures from this criterion were necessary in order to obtain the desired oversampling of new houses and low-income households, and to ensure that sufficiently reliable estimates could be made for each Census division.

Additional information about the 1993 RECS sample design and selection procedures will be found in Appendix A of the *Housing Characteristics 1993* and the *Household Energy Consumption and Expenditures 1993* reports and in the *Sample Design Procedures Manual* (Response Analysis Corporation 1995a) for the 1993 RECS. The publication *Sample Design for the Residential Energy Consumption Survey* (EIA 1994) provides information about sample design and selection procedures for all survey years and about how they have changed over the life of the survey.

Data Collection Procedures

An overview of the timing and collection modes used in collection of data for the 1993 RECS Household, Rental Agent, and Supplier Surveys is important (Figure 2.4). Data collection began in October 1993 with personal interviewing for the Household Survey. As shown in Figure 2.4,

there were three waves of personal interviewing. In the first wave, interviewers made a minimum of four attempts (initial attempt plus three callbacks), sometimes several more, on various days of the week and at various times of day, to try to establish contact with all assigned households. In the second wave, a similar effort was made, usually by different interviewers, in an attempt to contact households that had not been available during the first wave and to convince some of those that refused interviews in the first wave to reconsider.

Figure 2.4. Data Collection Modes and Timing: 1993 RECS

Survey and Wave	Collection Mode	Collection Period		Remarks
		Start	End	
Household Survey				
Field Wave 1	Personal interview	Oct 93	Apr 94	
Field Wave 2	Personal interview	Jan 94	Apr 94	
Field Wave 3	Personal interview	Feb 94	Apr 94	Selected sample sites with low response in Waves 1 and 2.
Abbreviated Version	Telephone interview	Apr 94	May 94	Authorization forms collected by mail.
Abbreviated Version	Mail	Apr 94	May 94	
Rental Agent Survey	Telephone	Jul 94	Jul 94	
Supplier Survey	Mail, with telephone followup	Mar 94	Oct 94	

Sources: Energy Information Administration, *Housing Characteristics* (1993); *Consumption and Expenditures* (1993).

The third wave was undertaken to try to improve response rates in selected locations that had low completion rates after the second wave.

The telephone and mail phases of data collection for the Household Survey had two goals: (1) to collect, for each household for which a personal interview had not been completed, a limited amount of information about housing-unit and household characteristics and (2) to obtain permission to contact energy suppliers for billing information on their household's consumption and expenditures. Telephone contacts were limited to those households for which telephone numbers were already available or could be obtained from reverse directories. Households responding to telephone interviews were mailed authorization forms for the Supplier Survey to sign and return, and reminder calls were made to those whose forms had not been received.

Results of the telephone phase were:

Eligible households	786
Telephone interviews completed	197
Authorization forms returned	74
Interviewed, no authorization form required	4

Only the last two groups were considered to represent completed households; thus, additional interviews were generated for 78 households, or 10 percent of those eligible. A similarly abbreviated mail questionnaire was mailed to most of the remaining nonrespondents. Completed questionnaires were received from 115 (7.5 percent) of the 1,528 households to whom they were sent.

Eligibility for the Rental Agent and Supplier Surveys was determined on the basis of information collected in the Household Survey, so data collection for those surveys could be undertaken only after data collection for the Household Survey was at or close to completion. The Rental Agent Survey was conducted entirely by telephone, in two stages. In the first stage, the survey contractor's telephone interviewers called to verify each rental agent's address, in order to mail an advance letter explaining the nature and purpose of the survey. In the second stage, data for the survey were collected by telephone. The Supplier Survey was conducted by mail. Information about eligibility and completion rates for these surveys is presented in Chapter 4.

The quality of data collected in personal interviews is affected by the qualifications, training, working conditions, and supervision of the interviewing staff. There were 234 interviewers who completed one or more personal interviews in the 1993 RECS. Of these, 35 percent had completed interviews in a prior RECS. Nearly all of the interviewers were trained in 3-day regional training sessions; a small number were trained in a 2-day session or by telephone. At the end of the training, each interviewer took an open book quiz, which was reviewed immediately thereafter.

Field interviewers were paid on an hourly basis for all time spent working on the survey and were reimbursed for their travel expenses. Contractor staff reviewed the first two interviews completed by each interviewer and provided extensive written feedback. Additional telephone training was provided when warranted. A 20-percent sample of the interviews was verified by telephone or mail to ensure that they had been conducted as intended.

Prior to each stage of telephone calls for the Rental Agent Survey, interviewers and their supervisors were briefed at the survey contractor's telephone center by the project manager. The project manager or a trained telephone supervisor monitored the first several calls made by each interviewer and the interviewers were monitored intermittently for the duration of the calling period (EIA 1995b).

Data Processing and Imputation

The 1993 RECS data collection operations produced information, mostly in hard-copy format, from six sources: (1) completed Household Survey interview questionnaires, (2) completed Household Survey telephone and mail questionnaires, (3) Housing-Unit Record Sheets, (4) Rental Agent Survey questionnaires, (5) Supplier Survey questionnaires, and (6) weather data from NOAA. In a complex set of operations, data from these six sources were processed, first individually and then in merged files, to produce the final RECS data tapes. A brief overview of these operations is provided here; a more extensive treatment, with emphasis on quality-related aspects, will be found in Chapter 6, Data Processing and Imputation and Chapter 7, Estimation and Sampling Error.

Initial processing of inputs from each of the first five sources involved three major steps: (1) check-in and manual review; (2) data entry; and (3) computer-assisted editing. All data entry operations were 100 percent verified. Telephone calls to respondents were made as needed to resolve inconsistencies identified in the manual reviews. In the computer-assisted edits, problems were resolved first by referring back to the questionnaires and then, when necessary, by calling respondents. The inputs from the sixth source, NOAA, were data files that were manipulated to produce the desired weather information to be associated with each sample household.

Some imputation of missing data, especially for items missing from the Household Survey interview questionnaires, was carried out prior to merging the files with data from other sources. Missing values were imputed for about two-thirds of the Household Survey variables, with "hot deck" imputation being the method most frequently used. Other kinds of imputation required the presence of data from more than one source; for example, imputation of items not included on the telephone and mail questionnaires required that the files of those questionnaires be merged with a set of potential "donor records" from the personal interview data file.

Following is a brief summary of some of the special features of processing relating to each of the six data sources:

- Household Survey Interview Questionnaires. In addition to consistency and range checks, the computer-assisted edits included several "special reports" in which computer listings with identifiers and selected data items were produced for housing- units with certain unusual characteristics. The listings and corresponding questionnaires were reviewed, respondents were contacted when necessary, and all changes were entered on the data file. A few examples of the many topics covered by special reports are: households that did not use any heating fuel; households with incompatible combinations of heating fuel and main equipment; and households that reported presence of a heat pump but no central air-conditioning. An elaborate set of computer runs was used to convert recorded linear measurements to estimates of total and heated areas and to translate from inside to outside measurements where the former had been recorded by the interviewer.

- Household Survey Telephone and Mail Questionnaires. The items not included on the abbreviated telephone and mail versions of the questionnaire were imputed, based on random selections from eligible "donor" households from the set of interview

questionnaires. The donor households were those that matched on a set of variables common to the complete and abbreviated questionnaire versions.

- Housing-Unit Record Sheets. For each assigned field interview, the interviewer completed a Housing-Unit Record Sheet (HURS) with his or her name and ID number, the type of living quarters and occupancy, the date, time, and outcome of each visit, and other particulars. These HURS forms were used, along with the original sample control file and information from other sources, to create the HURS file, which contained basic information about the outcome of data collection for each assigned sample address. HURS file records were also created for additional housing-units identified within sample addresses or by the application of the "half-open interval check" (see Chapter 3, Section 3.2). The HURS file had several uses, including the generation of reports on interview completion rates and interviewer characteristics and the development of weighting factors to adjust for unit nonresponse.

- Rental Agent Survey Questionnaires. The Rental Agent Survey covered households in multiunit structures that had one or more fuels included in their rent payments. Records created from the questionnaires were matched to the Household Survey data files for the same housing-units. Responses for items common to both surveys, such as year of construction and main heating fuels and equipment, were compared and changes were made to the Household Survey data whenever it was judged that the rental agent was more knowledgeable than the household respondent.

- Supplier Survey Questionnaires. For most households, the Supplier Survey is the preferred source of RECS information on household consumption and expenditures. After data entry, data from this source underwent a complex series of processing steps. A separate data file was created for each of the five fuels covered: electricity, natural gas, fuel oil, kerosene, and LPG. A series of edits was performed on the data that had been reported by month or another billing period, and the edited data were used to arrive at annualized estimates for all respondents who had adequate monthly data. These estimates were compared with model-based estimates based on household characteristics, and large differences were investigated. Similar model-based estimates were used to develop estimates of total consumption and expenditure by fuel for households not eligible for the Supplier Survey and those for which usable data were not obtained in the Supplier Survey. Finally, total consumption and expenditure of each fuel was allocated to end-use categories on the basis of a nonlinear regression model.

- Weather Data. Temperature data from all official U.S. weather stations were purchased on data tapes from NOAA. Based on analyses of proximity factors and the quality of the weather data, a specific weather station was associated with each RECS SSU. Temperature data for the selected SSU's were used to calculate heating and cooling degree-day estimates. The estimates for each SSU were added to the household and billing data files for all households located in that SSU. As one step in processing the Supplier Survey data, monthly estimates of degree-days were used to develop annual estimates for housing-units for which only part-year data were reported. For humidity,

the data were keyed from a NOAA publication and linked with RECS sample households at the PSU level.

Estimation

As in earlier surveys, the estimation procedures used in the 1993 RECS had four goals:

1. To reflect the overall selection probability of each sample housing-unit by the application of appropriate weights

2. To minimize bias resulting from unit nonresponse by the application of weight adjustments to groups of sample households with similar characteristics

3. To minimize sampling variance by the use of ratio estimates based on data available for all PSU's, whether or not included in the sample

4. To minimize bias resulting from undercoverage by benchmarking the RECS estimates of number of households to more precise estimates derived from the Census Bureau's Current Population Survey.

The overall weight developed for each sample housing-unit reflects the joint effects of sample weighting (Item 1 above), nonresponse adjustment (Item 2), and ratio estimation to reduce sampling variance and bias (Items 3 and 4).

Sampling errors for most estimates included in the published reports from RECS are estimated by the use of a balanced half-sample replication method. The results for individual data cells are used to develop generalized variance factors that are presented in the publications. Estimation procedures and sampling errors are discussed in greater detail in Chapter 7.

Data Dissemination

The *Housing Characteristics 1993* report for the 1993 RECS was released in June 1995 and the *Household Energy Consumption and Expenditures 1993* report was released in October 1995. Preparation of public-use data files with housing-unit data from all sources requires various deletions and other changes to ensure that the identity of individual housing-units and households cannot be determined by users. For example, all geographic identifiers other than Census region and division are removed. These anonymized public-use files are made available on magnetic tape for use with main frame computers and on diskettes for use with personal computers. The public-use files for the 1993 RECS were sent to the National Technical Information Service for distribution in August 1995 and, as noted in Chapter 1, were also made available for downloading via Internet at that time.

Cost of the Survey

The total cost of the 1993 RECS was approximately $3,965,000 over a 3-year period. This includes $475,000 provided by the Administration for Children and Families to support the collection of data used in the administration of the Low-Income Home Energy Assistance Program. The 1993 RECS was the first to use the sampling frame redesigned with the use of 1990 Census data. The cost of the redesign attributed to RECS was $556,000. The sampling frame is updated once every 10 years and the latest update will be used for at least two additional RECS.

The total cost of the 1990 RECS, exclusive of any cost arising from the redesign of the sample after the 1980 Census, was $2,115,000. The percentage breakdown of the total cost by major survey activities was:

Activity	Percent of Total
Planning and administration	5.2
Sample selection, including field listings and updates	18.3
Interviewer training	10.4
Household Survey data collection	28.4
Household Survey data processing	18.9
Rental Agent Survey data collection and editing	1.1
Supplier Survey data collection	5.9
Supplier Survey data processing	8.5
Documentation	3.3
Total	100.0

Evolution of the RECS Design: 1978-1993

Many of the basic design features of the 1993 RECS were present in the 1978 National Interim Energy Consumption Survey (NIECS), which was the first in the series of national sample surveys of residential energy consumption. Features that have remained constant throughout the nine surveys conducted between 1978 and 1993 include:

- Collection of data on household consumption of major fuels as a primary goal
- Collection and integration of data from multiple sources: households, rental agents, energy suppliers and weather stations
- Use of a national multistage probability sample of several thousand households

- Primary reliance on personal interviews to collect data from households

- Use of regression methodology to estimate consumption of each major fuel by end use within each household.

Within this broad framework, however, there have been many changes in specific features of the survey design and methodology. Changes have occurred because experience, experimentation, and research have suggested improvements; because the needs and interests of data users have changed; because there have been changes in the data available for use in sample design and estimation; and because the level of resources available for the surveys has varied from year to year. This section identifies important changes that have occurred in the 15-year evolution of RECS. Their impacts on the quality of RECS data are discussed in the chapters that follow.

Frequency and Timing of the Surveys

- Surveys were conducted annually from 1978 through 1982. The next RECS was conducted in 1984 and, since then, surveys have been undertaken at 3-year intervals, in 1987, 1990, and 1993. (As noted above, the Energy Policy Act of 1992 requires EIA to conduct a survey of residential energy use at least once every 3 years.)

- From 1978 through 1984, consumption data were collected for the 1-year period from April of the survey year through March of the following year. Starting with the 1987 survey, however, the consumption data have been collected for the survey calendar year.

Survey Content

- Although the broad topics of RECS have remained fairly constant, there have been many changes in specific items. Several of these changes reflect efforts to improve the performance of the regression models used to estimate total consumption of each fuel for households with no data from the Supplier Survey and to allocate consumption for each fuel to specific end uses. Some items used in one or more surveys have been dropped because respondents have found them difficult to answer. Examples are questions on the location and thickness of insulation.

- Several items have been added or deleted in response to changes in energy policies and programs and the emergence of new energy technologies. Questions on participation of low-income households in government programs to help cover the costs of home energy and weatherization improvements were added in 1981 and have been a regular feature since then. Questions about use of income tax credits for expenditures on home energy conservation improvements were asked in the 1982 and 1984 RECS but were then dropped when this provision of tax law expired. Questions about participation in demand-side management programs (utility-sponsored activities

designed to affect the amount and timing of customer electricity use) have become more detailed as these programs have expanded in scope and coverage.

- As described in the previous section, a comprehensive 1992 study of RECS user needs led to inclusion of several new items in the 1993 RECS Household Survey questionnaire and the elimination of some items covered in earlier surveys.

Sample Design

- The sample for the 1980 survey was the first one specifically designed for the purpose of a residential energy consumption survey. Coverage of the target population was expanded to include Alaska and Hawaii and residential housing-units on military bases.

- The samples for the 1984 and 1993 RECS were redesigned in order to take advantage of the availability of new data from the 1980 and 1990 Censuses, respectively. In each instance, the new census data were used in the definition, stratification and selection of PSU's and SSU's and to improve the precision of the survey estimates.

- The sample used in 1978 and 1979 was designed to produce data for the four Census regions. The 1980 sample was designed to produce statistically reliable data for nine Census divisions and 10 Federal regions. The requirement for the ten Federal regions was dropped in the 1984 redesign.

- The 1980 and subsequent redesigns have provided for the possibility of including a longitudinal component in successive surveys--that is, a subsample of the housing-units included in each survey could also be included in the next survey. Longitudinal components were included in the samples for the 1982, 1984, 1987, and 1990 surveys but not in the 1993 survey. In each instance, the sample for one-half of the sample SSU's consisted of housing-units from the previous survey, plus a sample of newly constructed units. Inclusion of longitudinal components has two advantages: it increases the precision of estimates of change between successive surveys and it provides a basis for longitudinal analyses, at the housing-unit level, of changes in consumption and household characteristics.

- The core sample for each RECS has been close to a self-weighting sample (the same overall selection probability of selection for each sample housing-unit), with adjustments to obtain the required minimum precision for each Census division. Supplemental samples of housing-units occupied by low-income households were provided for in the 1981, 1984, 1987, and 1993 designs. Because of an interest in changing trends in the energy efficiency of new houses, the 1993 sample design also provided for oversampling of new housing-units, defined as those completed in 1987 or later.

Data Collection Procedures

- Each Household Survey from 1978 through 1990 included a multiwave effort to obtain personal interviews for the greatest possible number of sample households, followed by the use of abbreviated mail questionnaires to seek limited information from the remaining households. In the 1993 RECS, a telephone followup procedure, also using an abbreviated version of the questionnaire, was inserted between the personal interview and mail phases.

- In the 1978 survey, Household Survey respondents were asked to estimate their total floor space, and there were also questions about the number of rooms and the size of the largest room. In 1979, the question on total floor space was dropped. From 1980 on, interviewers used tape measures to determine the physical dimensions of sample housing-units.

- Because of increasing interest in knowing the rated efficiency of major appliances, in the 1990 RECS interviewers attempted to obtain relevant information from the nameplates of central air-conditioning units. This procedure had only limited success and was not repeated in the 1993 RECS.

- Small cash incentive payments were used in connection with the Household Survey interviews in 1978 and 1979 and dropped thereafter. Cash incentives were used with the mail questionnaires through the 1982 survey. No incentives were used in the 1984 RECS; from 1987 on, a small token incentive, such as a key chain, has been sent out with the mail questionnaires.

- Interviewers for the Household Survey were trained by mail for the 1978 and 1979 surveys. In 1980, interviewers received in-person training in small group sessions at several locations throughout the country. For the 1981, 1982, and 1984 RECS, the 1980 training procedure was used for new interviewers, but training for those with prior RECS experience consisted of self-study materials and practice interviews. For the 1987 RECS, most of the interviewers received in-person training at five different locations. For the 1990 RECS there was no in-person training; all interviewers received their training via self-study materials, instructional videotapes, and completion of practice interviews, with evaluation by contractor staff. Interviewer training for the 1993 RECS reverted to the 1987 in-person model, with training at four different locations.

Data Processing

- Data entry operations were 100-percent verified in the 1978, 1979, and 1980 surveys. From 1981 through 1984, key questionnaire items were verified 100 percent and the remaining items were checked for a 25-percent sample of households. The cost of correcting data entry errors in subsequent stages of processing proved to be substantial, so 100-percent verification of data entry for all items has been used since the 1987 RECS.

- Procedures for associating weather data acquired from NOAA with individual sample housing-units have evolved over the life of the survey. A major change occurred in the 1987 RECS, when degree-day information for each housing-unit in an SSU was taken from the individual weather station nearest (in terms of distance and other factors) to the SSU, provided usable data were available for that station. Prior to 1987, degree-day data had been based on clusters of weather stations within a NOAA weather division.

Estimation Procedures

- For each RECS, part of the estimation procedure has been to benchmark the survey estimates of the number of households in various subgroups to agree with independent estimates derived from the Census Bureau's Current Population Survey (CPS). Initially, this ratio estimation procedure was carried out separately for 12 cells based on Census region and location type (central city, suburban, rural). Early analyses showed that RECS estimates of one-person households were low, so in the 1982 RECS a preliminary step was added in which the estimates were benchmarked to CPS estimates for three categories: single-person households with male householders, single-person households with female householders, and all other households. In the 1993 RECS, a further benchmark adjustment was introduced to ensure agreement of RECS and CPS household estimates for the nine Census divisions and for four large States: California, Florida, New York, and Texas.

- For all surveys through 1990, the CPS-based estimates used as benchmarks for the RECS estimates of households were developed for November of the survey year, representing the approximate midpoint of the data collection period for the Household Survey. In the 1993 RECS, the benchmark month was changed to July of the survey year, to coincide with the midpoint of the reference period for energy consumption.

- Measured total consumption of each fuel is allocated to end-use categories by use of a model-based estimation procedure that relates end-use consumption to a variety of housing and household characteristics for which information is obtained in the survey. The allocation model has been gradually refined in successive surveys. A major shift occurred in the 1984 RECS, when a nonlinear regression model replaced the linear model used in the earlier surveys (Carroll 1987). The number of separate end-use categories estimated for electricity was increased in 1990 and again in 1993.

Data Dissemination

- Starting with the 1980 RECS, there have been two major publications of the results from each survey, the first covering *Housing Characteristics 1993* and the second covering *Household Energy Consumption and Expenditures 1993*. In 1984, for the first time, estimates of energy consumption by end use were included in the regional supplement to the *Household Energy Consumption and Expenditures 1993* report.

Prior to then, the end-use data had been released in various special publications (EIA 1983c, EIA 1984c, Thompson 1987).

- Public-use data tapes (i.e., tapes with anonymized data for individual sample households) have been released for all surveys. Starting with the 1987 RECS, public-use data sets have also been released on diskettes for use on personal computers. In August 1995, public-use data sets for the 1987, 1990, and 1993 surveys were made available for downloading via the Internet.

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

[a]See description of procedures in text.
[b]Because 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

[a]Benchmark 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	1160	324	0.28
Mobile homes	All	622	475	0.76
Total		4,886	2,767	0.57

[a]Census 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.

4. Nonresponse

Nonresponse occurs in each of the three RECS surveys: Household, Rental Agent, and Supplier Surveys. *Unit nonresponse* occurs when no information at all is obtained from an assigned sample unit or when so little information is obtained that the questionnaire is classified as unusable. *Item nonresponse* occurs when a usable questionnaire is obtained but the desired information is missing for one or more items. Item nonresponse can occur because a respondent is unable or unwilling to answer a specific question, because the interviewer fails to ask it, or because a data entry clerk fails to key the response.

Both kinds of nonresponse affect the quality of the survey results. The magnitude of their effects can seldom be determined precisely. They depend in part on the efficacy of the imputation and estimation techniques that are used to try to limit the extent of bias due to nonresponse. They also depend on how large the nonresponse rates are and the degree to which the characteristics of nonrespondents differ from those of respondents.

This chapter presents information about the amount of, and trends in, nonresponse in the three component surveys, the characteristics of nonrespondents, the reasons for nonresponse, and the techniques that are used to minimize nonresponse rates. The imputation and estimation techniques used to deal with nonresponse are covered in detail in Chapters 6 and 7, respectively; however, a brief summary is given at this point. Figure 4.1 summarizes these techniques by survey component and type of nonresponse.

For the *Household Survey*, weighting adjustments are used to minimize the biases caused by unit nonresponse. For most variables based on the Household Survey questionnaire, missing values are imputed by using one of several different deterministic or probabilistic imputation techniques. For a few variables, some or all of the missing values are not imputed. Housing characteristics not fully imputed in the 1990 RECS included several variables related to conservation practices, such as thermostat settings, participation in demand-side management programs, and the use of insulation, caulking, and weatherstripping. Also not fully imputed in 1990 were missing responses to questions on the age and other characteristics of appliances and equipment and on the number of years survey respondents expected to remain in their present homes.

For households in multi-unit structures with one or more fuels included in the rent, the *Rental Agent Survey* is designed to provide the most accurate information possible about the main space heating fuel and equipment and the main fuels for water heating and air-conditioning. Information provided by rental agents is the preferred source for this kind of information--that is, it is usually considered to be more accurate than the corresponding answers obtained from household respondents. When information is not available from the Rental Agent Survey due to unit or item nonresponse, the fallback procedure is to rely on information reported in the Household Survey.

For households that pay suppliers directly for one or more delivered fuels, the *Supplier Survey* provides information from billing records about housing unit consumption and expenditures by fuel type. When no Supplier Survey information for a specific fuel is obtained for a household, the household's consumption and expenditures for that fuel are imputed on the basis of household

Figure 4.1. Treatment of Nonresponse by RECS Component

Survey Component	Type of Nonresponse	
	Unit	Item
Housing unit	Weighting adjustments	Varies by item: • Impute • Publish as NA
Rental agent	Use corresponding data from housing unit survey	
Supplier	Impute from housing characteristics	Varies by fuel: • Impute from part-year data (electricity, NG) • Impute from housing characteristics (other fuels)

Source: Energy Information Administration, *Housing Characteristics* (1993); *Consumption and Expenditures* (1993).

characteristics by using a nonlinear regression method. When partial information is obtained in the Supplier Survey, the imputation technique depends on the fuel type. For electricity and natural gas, annual consumption and expenditures are sometimes imputed on the basis of part-year data from the Supplier Survey. For other fuels, part-year data on billings are not used; consumption and expenditures are imputed in the same way as if no data had been obtained in the Supplier Survey.

The next section of this chapter covers nonresponse in the Household Survey, with subsections on unit and item nonresponse. The following two sections present comparable information for the Rental Agent and Supplier Surveys, respectively. The final section summarizes the nature and consequences of nonresponse in these three components of RECS.

Nonresponse in the Household Survey

Unit Nonresponse

The information on overall unit response rates for RECS survey years 1978 through 1993 has been summarized (Table 4.1). *All rates shown in Table 4.1 and subsequent tables in this chapter are unweighted*--that is, they are based on counts of eligible sample units and do not reflect variations in the overall selection probabilities of individual units. The implications of using unweighted rates are discussed later in this section.

Table 4.1. Household Survey Eligibility and Response Rate by Survey Year: 1978-1993

Category	Survey Year								
	1978	1979	1980	1981	1982	1984	1987	1990	1993
Number of units assigned to interviewers	4,849[a]	4,935	7,338	7,668	5,903	7,658	8,232	6,757	9,869
Percent ineligible:									
Not housing units	NA	2.8	1.4	1.5	1.6	1.6	2.7	2.2	2.0
Vacant or seasonally vacant	7.1	7.0	8.1	9.2	9.1	10.2	10.0	10.3	9.3
Number of units eligible	4,507	4,453	6,634	6,841	5,272	6,752	7,183	5,909	8,753
Percent completed by interview	85.2	85.5	87.5	86.7	84.8	81.1	81.5	81.7	79.0
Percent not completed by interview:									
Refusal	NA	NA	8.4	8.2	11.5	14.9	14.0	12.1	15.3
No one home	NA	NA	2.2	3.2	1.9	2.5	3.1	3.8	2.7
Other	NA	NA	2.0	1.8	1.7	1.5	1.4	2.4	3.0
Percent completed by mail[b]	5.3	5.1	3.7	4.9	4.7	3.0	5.2	4.5	2.2
Percent completed by interview or mail	90.5	90.6	91.2	91.6	89.5	84.1	86.7	86.2	81.2

[a]Data unavailable for assigned units that were not housing units.
[b]Data for 1993 include households completed by mail and telephone.
NA = Not Available.
Sources: Energy Information Administration, *Housing Characteristics* (1980-1993); *Consumption and Expenditures* (1978-1979).

As explained in Chapter 2, the multi-wave, multi-mode data collection procedure used in the Household Survey is designed to maximize the level of survey response for the eligible units in the initial sample. Following multiple attempts to complete a personal interview for each unit, mail questionnaires are sent to most of the remaining addresses. In the 1993 RECS, a telephone followup was inserted between the personal interview and mail phases of data collection.

Substantially abbreviated versions of the personal interview questionnaires are used in the telephone and mail phases of data collection, their main purpose being to identify suppliers and obtain respondent waivers so that consumption and expenditure data for these units can be obtained in the Supplier Survey. The mail and telephone versions also include some of the basic items used to estimate end-use consumption for these housing units. One consequence of using abbreviated questionnaires for the mail and telephone followups is that item nonresponse is automatic for all data items not included on those questionnaires.

Personal interview and mail/telephone response rates are shown separately in Table 4.1, each being calculated as a percent of all housing units eligible for interviews. Personal interview completion rates remained fairly stable, in the neighborhood of 85 percent, from 1978 through 1982, but declined to a lower level, in the neighborhood of 80 percent, for all subsequent surveys. For most survey years, mail response rates were close to 5 percent, the exceptions being 1980, 1984, and 1993. In 1993, the proportion of all eligible units completed by mail and telephone combined was only 2.2 percent, and the overall response rate, 81.2 percent, was the lowest yet experienced in the RECS Household Survey.

Two kinds of explanations can be sought for the decline in unit response rates for RECS in the four most recent survey years. It might be argued that there has been a general decline in the willingness of persons to respond to national household surveys for which participation is voluntary. Such a trend can be countered to some extent by increased efforts to motivate response and to contact persons or households that are difficult to reach, but there are limits to what can be done to counteract outright refusals. The question of whether such a trend has actually occurred is controversial. Based on a review of eight periodic demographic surveys, a subcommittee of the Office of Management and Budget's Federal Committee on Statistical Methodology recently concluded that:

> ... there was little evidence of declining response rates over time ... There was some evidence that refusal rates were increasing in demographic surveys; however, the analysis revealed that there are no changes in *overall* response rates. This could be due to a greater effort in data collection. (Gonzalez, Kasprzyk, and Scheuren 1994)

Response trends in a specific survey can also be affected by changes in the survey's design and procedures. There have been several changes in Household Survey procedures that might be expected to have some impact on unit response (Figure 4.2).

Of the five design features shown in Figure 4.2, only the first two, involving the use of incentives, might be expected to *increase* unit response rates. The first RECS, in 1978, used a $2 cash incentive that was given to each household at the time it was first contacted, plus another $2 included with the questionnaires mailed to households for which personal interviews had not been obtained. To evaluate the effectiveness of the initial $2 incentive, an experiment was undertaken in the 1979 RECS. Primary sampling units in large cities, where response rates are usually lower, were excluded from the experiment. In the remaining 80 primary sampling units, the incentive was used in 60 randomly selected secondary sampling units and was not used in the remaining 20. The $2 incentive had no significant effect on unit and item nonresponse rates,

but more of the respondents who received the incentive agreed to sign waivers allowing EIA to collect data from their energy suppliers and a slightly higher proportion of them permitted interviewers to measure floor space inside the house, rather than outside (Thompson 1985b).

Figure 4.2. Design Changes That May Have Affected Unit Response Rates

Design Feature	Survey Year								
	1978	1979	1980	1981	1982	1984	1987	1990	1993
Initial incentive payment[a]	Y	Y	Y	N	N	N	N	N	N
Mail questionnaire incentive[b]	Y	Y	Y	Y	Y	N	Y	Y	Y
Longitudinal component	N	N	N	N	Y	Y	Y	Y	N
Returning households identified for interviewers[c]	---	---	---	---	N	Y	N	N	---
Low-income supplement	N	N	N	Y	N	Y	Y	N	Y

[a]In 1979, the incentive payment was provided to a subset of the sample households as part of an experiment.
[b]From 1978 to 1982, a cash incentive was used. From 1987 on, a token gift has been used.
[c]Applicable only in years with a longitudinal component.
--- = No longitudinal component.
Source: Energy Information Administration, *Housing Characteristics* (for years shown).

Nevertheless, the initial $2 incentive was discontinued after the 1980 RECS. The $2 incentive that accompanied the mail questionnaires was continued through 1982. In 1984, no incentive was used; in the 1987, 1990, and 1993 surveys a non-cash token gift was included with the mail questionnaires. The absence of any mail incentive, cash or non-cash, in 1984 may explain the low mail response rates observed in that year.

Among the factors that might be expected to lead to *lower* response rates was one that clearly had this effect--the introduction, starting in 1982, of a longitudinal survey design component, involving the collection of data from some households in two successive surveys. A subset of the households affected were also asked, in the interval between the two surveys, to participate in the Residential Transportation Energy Consumption Survey (RTECS). Many of the RTECS households were requested to keep track of their vehicle mileage and fuel purchases for a period of several months. Table 4.2 shows personal interview and mail response rates, separately for new households and those included in the sample previously, for the four years in which RECS included a longitudinal component. Clearly the new units had lower refusal rates and higher overall response rates, especially in 1984, 1987, and 1990.

The data do not distinguish, among households previously included in the sample, between those that participated in RTECS and those that did not (Table 4.2). A special study (Hersey and McCarthy 1986) provided such a comparison for the 1984 RECS. The sample for the 1983 RTECS included all households in the 1982 RECS that reported having driven their vehicles 12,500 miles or more and a subsample of the remaining households. Therefore, the most direct and relevant comparison is of the response rates for the RTECS participants and non-participants among the latter group, those households that reported fewer than 12,500 vehicle miles driven. In that group, the 1984 RECS participation rates were 75.4 percent for 644 households that had been included in the sample for the 1983 RTECS and 80.6 percent for 375 households that had not been selected.

Table 4.2. Household Survey Response Rates by Survey Year and Prior Sample Status: 1982-1990

Survey Year and Prior Status[a]	Response Rates (percent of eligible units)			Personal Interview Nonresponse Rates (percent of eligible units)	
	Personal Interview	Mail	Total	Refusals	Other
1982					
Previously in sample	83.9	4.8	88.7	12.7	3.4
New units	85.9	4.6	90.5	10.3	3.9
1984					
Previously in sample	78.2	2.8	81.0	17.6	4.2
New units	84.1	3.2	87.3	12.2	3.7
1987					
Previously in sample	79.5	5.2	84.7	16.3	4.2
New units	83.5	5.1	88.6	11.7	4.8
1990					
Previously in sample	79.4	4.5	83.9	13.9	6.7
New units	84.0	4.5	88.5	10.4	5.6

[a]The category "Previously in Sample" includes a few units missed in the previous survey or constructed subsequently.
Source: Energy Information Administration, *Housing Characteristics* (for years shown).

Another factor, which was operative only in the 1984 RECS, was a sample assignment procedure which made it possible for the survey interviewers to know which of the units assigned to them had already been included in the 1982 RECS. There is no direct evidence about how this knowledge may have affected interviewers' efforts, but it has been suggested that interviewers might not have tried as hard to obtain interviews for these "recycled" housing units as they did for new ones (Response Analysis Corporation 1987).

A change in the relative sample selection probabilities for different subsets of the RECS target population can affect *unweighted* unit response rates. An example of this is provided by the low-income supplements that were included in RECS in survey years 1981, 1984, 1987, and 1993. In these years, housing units in secondary sampling units that were classified by interviewers as low-income areas were oversampled, with the result that a larger proportion of the total sample was located in such areas. Data for the 1987 RECS (Slider 1995) show that response rates were somewhat higher in the areas that were oversampled, so that the overall unweighted response rates were higher in 1987 than they would have been if there had been no low-income supplements in these years. *Weighted* response rates, of course, would not be affected by differences in sampling rates for different segments of the population.

There is some evidence that personal interview response rates may be higher for interviewers who have had experience in prior RECS survey years. A special tabulation for the first round of interviewing in the 1990 RECS showed that interviews were completed for 73.7 percent of the sample households assigned to interviewers who had also worked in the 1987 RECS, as compared with only 64.7 percent for interviewers who had not worked in 1987 (Response Analysis Corporation 1991, Table 11). However, these findings could also be explained by higher turnover rates in areas where respondents tend to be less cooperative.

The proportion of experienced interviewers (those who had worked on earlier RECS surveys) was in the neighborhood of 60 percent for survey years 1981, 1982, and 1984. It declined to about 45 percent in 1987 and to about 35 percent in 1990 and 1993. To some extent, this trend tracks the decline in personal interview response rates in the most recent survey years, but 1984, with a high proportion of experienced interviewers and a low response rate, is an exception.

The 5 percent drop in response rates between 1990 and 1993 resulted primarily from a conscious decision to reduce the target response rate for RECS from 85 to 80 percent. In order to maintain the desired sample size, the level of followup effort for households not completed in the initial wave of interviewing was reduced in comparison with prior survey years. Another factor that *may* have influenced response rates for the 1993 RECS was the substantial increase in the length of the Household Survey questionnaire compared with preceding surveys. The basic household questionnaire contained 117 pages in 1993, compared to 63 in 1990 and 50 in 1987. Potential respondents might not have had a very clear idea, when approached to participate in the survey, of the length of the questionnaire or the time it might take to complete it. However, some of the experienced interviewers, knowing of the increased length of the 1993 questionnaire, might not have tried quite as hard to enlist the cooperation of initially reluctant respondents, knowing that there was an increased risk of their breaking off the interview prior to its completion.

In summary, the drop of 5.4 percentage points in overall response rates between the 1982 and 1984 survey appears to have resulted from the confluence of several factors: the elimination of incentives, cash or noncash, for mail questionnaires; the inclusion of a longitudinal component and a low income supplement; and the use of a procedure which allowed interviewers to identify housing units previously interviewed. Several steps were taken to try to increase response rates in the 1987 RECS. These measures appear to have had some success in 1987 and 1990, but did not bring the rates back to the levels achieved prior to 1984. In the 1993 RECS, the absence of a longitudinal component was not accompanied by the reduction in refusal rates that might have

been expected; on the contrary, the proportion of eligible housing units refusing personal interviews increased to its highest level. The increased length of the 1993 Household Survey questionnaire may have played a role. This increase, in combination with a substantial decline in the success of mail and telephone followups, led to the lowest overall completion rate so far experienced in the RECS Household Survey.

The effects of nonresponse on the quality of survey estimates depend not only on the overall level of nonresponse but also on its distribution among subgroups of the RECS target population. Trends in overall response rates (personal interview, plus mail/telephone) for selected housing unit characteristics--Census region, urban-rural location, and type of structure--for survey years 1979 through 1993 are important (Table 4.3).

Even with the declining trend in overall response rates, some fairly consistent patterns in relative rates are evident. The Northeast Region has consistently had the lowest overall response rates. The South has consistently had the highest response rates for personal interviews but has had the lowest mail response rates in nearly all survey years. Households located in urban areas (the central cities of metropolitan statistical areas) have consistently 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. Households in single-family structures (including mobile homes) have had the highest overall response rates except in 1990, when the rates for units in structures with two to four housing units were the same as those for units in single family structures. As explained further in Chapter 7, the weighting procedures used to produce RECS estimates include weighting factors that are specifically designed to minimize the effects of nonresponse bias arising from differential response rates by Census region and urban/rural status.

Item Nonresponse

Responses to individual items on a completed questionnaire can be assigned to one of the following categories:

1. No entry required for item (skip based on prior item)

2. Entry required for item
   a. Item left blank
   b. Allowable nonresponse
      (1) Don’t know or not sure
      (2) Refused
   c. Non-standard response
   d. Standard response

Table 4.3a. Household Survey Response Rates by Region, Urban Status, and Structure Type: 1979-1993

Census Region and Survey Year	Response Rates (percent of eligible units)			Personal Interview Nonresponse Rates (percent of eligible units)	
	Personal Interview	Mail Questionnaire[a]	Total Response	Refusals	Other
Northeast					
1979	82.0	6.0	88.0	NA	NA
1980	83.8	4.9	88.7	10.5	5.7
1981	83.2	6.3	89.5	10.5	6.3
1982	81.7	5.2	86.9	13.1	5.2
1984	81.2	2.0	83.2	15.1	3.7
1987	79.0	5.7	84.7	16.3	4.7
1990	77.5	5.9	83.4	13.8	8.7
1993	75.8	2.6	78.3	16.6	7.7
Midwest					
1979	86.7	4.4	91.1	NA	NA
1980	87.4	3.7	91.1	9.0	3.6
1981	86.7	5.2	91.9	8.8	4.4
1982	84.4	5.4	89.9	12.5	3.0
1984	79.7	4.1	83.8	16.5	3.8
1987	80.7	5.9	86.6	15.1	4.2
1990	83.1	4.3	87.4	11.8	5.1
1993	80.4	2.9	83.3	15.5	4.1
South					
1979	87.6	3.8	91.4	NA	NA
1980	89.8	3.1	92.9	6.8	3.4
1981	88.9	3.4	92.3	6.2	4.9
1982	86.5	3.2	89.7	9.7	3.8
1984	83.5	2.1	85.6	12.8	3.7
1987	84.0	4.2	88.2	11.7	4.3
1990	84.9	3.1	88.0	10.3	4.8
1993	81.0	1.8	82.8	13.8	5.2
West					
1979	84.2	7.5	91.7	NA	NA
1980	87.9	3.5	91.4	8.8	3.3
1981	86.9	5.3	92.2	8.4	4.7
1982	85.9	5.4	91.3	11.2	2.9
1984	79.4	4.0	83.4	15.7	4.8
1987	81.8	5.1	86.9	13.3	4.9
1990	80.4	5.1	85.6	13.0	6.6
1993	77.9	1.9	79.8	16.0	6.1

[a]Data for 1993 include mail and telephone questionnaires.
NA = Not Available.
Source: Energy Information Administration, *Housing Characteristics* (1979-1993).

Table 4.3b. Household Survey Response Rates by Region, Urban Status, and Structure Type: 1981-1993

Location Type and Survey Year[a]	Response Rates (percent of eligible units)			Personal Interview Nonresponse Rates (percent of eligible units)	
	Personal Interview	Mail Questionnaire[b]	Total Response	Refusals	Other
Urban					
1981	82.4	6.5	88.9	10.0	7.6
1982	80.8	6.1	86.8	13.6	5.6
1984	79.4	3.5	82.9	15.8	4.8
1987	79.8	5.2	85.0	14.4	5.8
1990	79.0	4.9	83.8	12.3	8.7
1993	77.6	1.7	79.4	15.5	6.8
Suburban					
1981	85.8	5.9	91.7	9.7	4.6
1982	85.0	4.6	89.6	12.5	2.5
1984	79.3	3.7	83.0	16.7	4.0
1987	80.4	6.0	86.4	15.6	4.0
1990	80.6	5.2	85.9	13.7	5.7
1993	77.6	2.9	80.5	17.1	5,4
Rural					
1981	91.1	2.6	93.7	5.4	3.5
1982	89.7	3.2	93.0	7.5	2.8
1984	86.2	1.4	87.6	10.9	2.9
1987	85.6	4.1	89.7	10.9	3.5
1990	87.2	2.8	89.9	9.3	3.5
1993	84.2	1.5	85.7	11.1	4.7

[a]Data available for 1979 and 1980 are based on a different location type classification.
[b]Data for 1993 include mail and telephone questionnaires.
Source: Energy Information Administration, *Housing Characteristics* (1981-1993).

Table 4.3c. Household Survey Response Rates by Region, Urban Status, and Structure Type: 1979-1993

Structure Type and Survey Year	Response Rates (percent of eligible units)			Personal Interview Nonresponse Rates (percent of eligible units)	
	Personal Interview	Mail Questionnaire[a]	Total Response	Refusals	Other
Single Family					
1979	87.6	4.3	91.9	9.7	2.6
1980	88.9	3.3	92.2	6.4	2.7
1981	88.3	4.1	92.4	8.1	3.6
1982	86.2	4.4	90.6	11.6	2.3
Mobile Home					
1979	88.1	2.8	90.9	7.5	4.4
1980	90.0	2.1	93.0	8.4	2.7
1981	89.2	2.6	91.8	6.1	4.7
1982	87.4	2.0	89.5	8.9	3.6
Single-Family/Mobile Home					
1984	83.8	1.2	85.0	12.2	4.0
1987	82.3	5.4	87.7	14.5	3.2
1990	82.3	4.4	86.7	12.8	4.9
1993	79.6	2.3	81.9	15.8	4.6
Buildings with 2-4 units					
1980	80.1	7.1	87.2	NA	NA
1981	86.4	4.8	91.2	7.2	6.4
1982	85.0	4.2	89.2	10.2	4.8
1984	81.4	2.9	84.3	12.7	5.9
1987	80.1	3.9	84.0	12.4	7.5
1990	83.2	3.5	86.7	8.0	8.8
1993	74.7	2.5	77.1	17.0	8.4
Buildings with 5 or more units					
1980	76.3	9.5	85.8	NA	NA
1981	78.6	9.9	88.5	10.5	10.9
1982	76.7	7.9	84.5	13.0	10.3
1984	79.4	3.8	83.2	14.4	6.2
1987	79.4	5.4	84.8	13.0	7.6
1990	77.4	6.0	83.4	11.7	10.9
1993	78.6	1.7	80.2	12.0	9.4

[a]Data for 1993 include mail and telephone questionnaires.
NA = Not Available.
Sources: Energy Information Administration, *Housing Characteristics* (1980-1993); *Consumption and Expenditures* (1979).

Some of these categories can be illustrated by the 1990 RECS item on age of water heating equipment (Item C-5). The item was skipped (Category 1) if the preceding question "Does the main equipment for heating water for your home also heat water for other buildings or housing units?" was answered "Yes" or "Don't know." If item C-5 was asked, acceptable entries included "Don't know" (Category 2b(1)) and "As old as the house/original equipment" (Category 2c). Standard responses (Category d) were ranges for number of years. For this item, blanks and "Don't know" responses would be counted as item nonresponse.

A few items specifically permit the interviewer to check "Refused"; on the 1990 questionnaire these were the items on family income, account numbers of fuel suppliers, and relationship to respondent of person to whom fuel bills are addressed. Except for family income, refusals to these items had no direct effect on the survey estimates.

For the purpose of this chapter, item nonresponse will consist of categories 2a and 2b, blanks, and allowable nonresponse. Nonstandard responses generally provide some kind of usable information. In the example given above, if the response to the item on age of water heating equipment were "As old as the house/original equipment," the response to the item on age of house or building would be used to assign an age range to the water heating equipment. Nonresponse rates are calculated by dividing the number of households with nonresponse (2a plus 2b) by the number of households for which a response was required (total number of questionnaires minus those in category 1).

Item nonresponse in the Household Survey has been relatively low for most items on questionnaires completed by personal interview. As noted above, many questions are not included on the mail and telephone questionnaires. To the extent that these excluded items apply to all housing units, the additional item nonresponse from this source has varied from 2.7 to 5.9 percent--mail response as a percent of total response--depending on the survey year (see Table 4.1).

A tabulation of the 1990 RECS Household Survey personal interview data file prior to computer edits shows that, of 416 survey variables based on questionnaire entries, 51 (12 percent) had item nonresponse rates of 5.0 percent or more. Variables related to household measurements were excluded from this count; nonresponse for this topic is discussed separately below.

Table 4.4 lists the 10 questionnaire items with the highest nonresponse rates in the 1990 RECS. Of the 10 items listed, only 3--age of water heater and two items related to household income--required entries on more than 10 percent of the questionnaires. The basic household income item, which asked each sample household to report its total income in one of 25 class intervals, had a nonresponse rate of 14.4 percent, with half of the nonresponse being accounted for by refusals. Nonrespondents to the basic item were asked whether their income was above or below $35,000, the cutoff used for deciding whether to ask about participation in income-tested government assistance programs. Nonresponse to this question was 21.4 percent, so there was no income information of any kind available for 3.1 percent of the sample households.

Table 4.4. Ten Variables with Highest Item Nonresponse Rates: 1990 RECS

Question No.	Description	Number of Housing Units			Nonresponse Rate
		Entry Required for Item	Type of Response		
			Don't Know	No Entry	
L-5	Amount received from government for home heating costs	255	72[a]	10	32.2
H-18	Amount per gallon paid for kerosene	209	51	10	29.2
B-5	Month of change in main heating fuel	137	--	38	27.7
I-8	Proportion of kerosene bill for nonhousehold uses	4	--	1	25.0
K-10	Household income over or under $35,000	695	38	111[b]	21.4
P-7	Sales of agricultural products	100	15	4	19.0
K-10	Household income	4,840	250	445[c]	14.4
L-4e	Other form of government payment for home heating costs	239	1	33	14.2
C-5	Age of water heater	4,089	400	171	14.0
B-5	Year of change in main heating fuel	137	7	12	13.9

[a]Response category "Not sure."
[b]Includes 100 refusals.
[c]Includes 347 refusals.
Source: Preliminary tabulation of 1990 RECS data prior to edit changes.

Items on age of equipment had relatively high nonresponse rates, as follows:

Water heater (shown in Table 4.4)	14.0
Main heating equipment	12.9
Central air-conditioning equipment	12.9
Most used window or wall unit	9.3
Most used refrigerator	6.5
Second most used refrigerator	5.1

The higher nonresponse rates in this group were for the ages of equipment that was less readily accessible at the time of the interview. Perhaps for the same reason, water heater capacity had a nonresponse rate of 10.1 percent. Although no items on insulation were among the 10 with the highest nonresponse rates, 7 of 18 items relating to presence and extent of insulation had nonresponse rates in the range of 5 to 11 percent.

Households living in rental units in multiunit buildings are more likely to be unable to respond to questions on age and other characteristics of equipment. In the 1990 RECS more than half (57 percent) of the "don't know" responses to the item on age of main heating equipment came from such households. Recent occupancy of the housing unit is another factor associated with some types of nonresponse. The 1990 data show that 91 percent of the "don't know" responses to the question on main heating fuel used at the time of the 1987 RECS were given by members of households that had moved into their current residences subsequent to 1987.

The highest item nonresponse rate, 32.2 percent, was recorded for a question on the total dollar amount received from the government for assistance with home heating costs. Other dollar amount items with high nonresponse rates were amount per gallon paid for kerosene, 29.2 percent, and sales of agricultural products, 19.0 percent.

The extent of missing information on household size merits special attention. RECS interviewers were instructed to measure the dimensions of all "area enclosed from the weather," using a retractable 50-foot metal tape measure. Outside measurements were preferred, but interviewers were allowed to measure from the inside when necessary. They were instructed to record which of these options was employed for each housing unit. They were asked to prepare a rough diagram of the floor plan, indicating which areas were heated, which were unheated, and the dimensions of each.

Table 4.5 shows, for survey years 1980 through 1993, the extent to which the interviewers succeeded in providing some or all of the household size information. (Comparable data on total and heated floor space were not obtained in the 1978 and 1979 surveys.) The data in Table 4.5 cover only housing units for which personal interviews were conducted. The data for survey years 1982 and 1984, the first two years in which RECS included a longitudinal component, exclude a substantial number of units for which the measurements were taken from the data file for the previous survey.

Table 4.5. Housing Unit Survey, Completeness of Household Size Information: 1980-1993

Category	Survey Year						
	1980	1981	1982	1984	1987	1990	1993
Number of personal interviews completed[a]	5,804	5,937	3,648[b]	4,895[c]	5,856	4,828	6,918
Percent with:							
Complete measurement information	81.5	82.2	62.4	56.0	73.0	75.4	66.4
Partial Information:	15.4	15.4	32.3	37.5	25.8	21.4	28.4
Unknown if inside or outside otherwise complete	(9.9)	(9.2)	(27.3)	(31.7)	(20.7)	(15.9)	(22.8)
Some measurements missing	(5.5)	(6.2)	(5.0)	(5.9)	(5.1)	(5.5)	(5.7)
No usable information	3.2	2.4	5.3	6.4	1.2	3.2	5.1
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

[a]Excludes housing units for which mail or telephone questionnaires were completed.
[b]Excludes 827 units for which measurements from 1981 survey were used.
[c]Excludes 584 units for which measurements from 1982 survey were used.
Source: Energy Information Administration, *Housing Characteristics* (1980-1993).

The substantial year-to-year variation in the proportion of cases for which full information was obtained has resulted primarily from wide variations in the proportion of cases for which all required measurements were obtained, but interviewers did not specify whether they were inside or outside measurements. That proportion peaked in the 1982 and 1984 surveys, at 27.3 and 31.7 percent, respectively. In those same years, the proportions of cases with no usable information, 5.3 and 6.4 percent, were also higher than in other years, although 1993 was not far behind, with 5.1 percent in this category. The proportion of cases with some measurements missing (some of which may also have failed to distinguish heated and unheated areas) has been remarkably stable, in the vicinity of 5 or 6 percent.

These variations in the proportion of cases for which the method of measurement was not clearly specified may have resulted in part from changes in how interviewers were asked to record this information. Variations used from 1980 through 1993 have been as follows:

Survey Year	Location of Question on Measurement Procedure	Response Categories
1980	Start of measurements module	Inside, outside
1981	Same as 1980	Inside, outside, other
1982	End of measurements module	Inside, outside, and 4 other options
1984	Same as 1982	Same as 1982
1987	After *each* floor, attic and basement	Inside, outside, other
1990, 1993	Same as 1987	Same as 1987

The information on where the measurements were made is needed in order to standardize area measurements for all housing units to outside dimensions. For housing units known to have been measured on the inside, scale factors are used to convert their values to outside dimensions. For this purpose, all housing units for which the measurement method was not specified were assumed to have been measured on the outside, this being the much more commonly used method, at least for single-unit structures.

Nonresponse in the Rental Agent Survey

As explained in the introduction to this chapter, the Rental Agent Survey covers households in multiunit structures that have one or more fuels included in their rent. Its main purpose is to improve the quality of information about types of fuels used for space heating, water heating, and air-conditioning and the main type of heating equipment used. In 2 survey years, 1980 and 1984, permission was also requested from some of the rental agents to collect consumption and expenditure data for buildings from their fuel suppliers. In those two years, both personal and telephone interviews were conducted; in other years, the rental agents have been contacted entirely by telephone. Each rental agent is requested to provide information for all of the sample housing units for which the agent is responsible, whether the units are located in a single building or in more than one building.

Table 4.6 shows response rate information for the Rental Agent Survey for survey years 1981 through 1993 and partial information for 1979 and 1980. The proportion of interviewed households eligible for the Rental Agent Survey has varied from 10.7 to 15.4 percent. The higher eligibility rates in 1984 and 1987 were probably associated with the inclusion of low-income supplements in those years. The average numbers of housing units per rental agent interviewed were also higher in those years, as well as in 1981 and 1993, when there were also low-income supplements.

Table 4.6. Rental Agent Survey Eligibility and Response Rates by Survey Year: 1979-1993

Category	Year[a]							
	1979	1980	1981	1982	1984	1987	1990	1993
No. of household interviews completed			6,269	4,724	5,682	6,229	5,095	7,111
Eligible for rental agent survey								
Number	NA	NA	746	540	826	961	646	764
Percent			11.9	11.4	14.5	15.4	12.7	10.7
Rental agent survey completed								
Number	NA	551	466	308	549	856	550	625
Percent of eligible	NA	NA	62.5	57.0	66.5	89.1	85.1	81.8
No. of rental agents interviewed	109[b]	283	203	168	210	303	281	285
Mean no. of housing units per rental agent	NA	1.95	2.30	1.75	2.61	2.83	1.96	2.19

[a]Rental Agent Survey was conducted in 1978, but no data on response are available.
[b]Of 141 identified.
NA = Not Available.
Sources: Energy Information Administration, *Housing Characteristics* (1980-1993); *Consumption and Expenditures* (1979).

The response rates in Table 4.6 represent the proportion of eligible housing units for which Rental Agent Survey information was obtained. Because of more focussed and systematic efforts to contact rental agents, beginning with the 1987 survey, the response rates achieved for 1987, 1990, and 1993--all in the 80 to 90 percent range--were substantially higher than those that had been achieved in the three prior survey years.

The probable consequence of nonresponse in the Rental Agent Survey is that the information for the items covered in that survey will be less accurate for the housing units for which interviews were not completed. Data from the Rental Agent Survey are compared with corresponding data from the Household Survey, and the former source is generally considered more accurate with respect to main fuels used and main heating equipment. Data for survey years 1981 through 1987 show that one or more changes were made on the basis of this comparison for 30 percent of the housing units in 1981, 26 percent in 1982, 32 percent in 1984, and 42 percent in 1987. Most of the changes were for main heating fuel and equipment and main water-heating fuel.

Nonresponse in the Supplier Survey

The Supplier Survey covers households that pay the supplier directly for one or more of five fuels: electricity, natural gas, fuel oil, kerosene, and LPG. It does not cover fuels included in the payment of rent. Fuels purchased on a cash and carry basis, primarily kerosene, are also excluded. For each fuel, eligible housing units are asked to sign waiver forms allowing their suppliers to provide billing information to the survey contractor.

Table 4.7 shows eligibility and response rates for the Supplier Survey for survey years 1980 through 1993. For survey years 1980, 1981, and 1982, a single Supplier Survey form was used for fuel oil and kerosene; hence the data for fuel oil and kerosene are combined for these survey years. From 1984 on, the two fuels are shown separately. Except for kerosene, most of the households not eligible for the Supplier Survey occupied rental units in buildings with two or more units. A large proportion of the households that used kerosene were ineligible for the supplier survey because they purchased it on a cash and carry basis. Respondents for these households were asked to estimate the amount of kerosene used or purchased during the past 12 months. For the 1982 survey year, following a substantial increase in the proportion of households using kerosene as a supplemental heating fuel, followup telephone calls were made to such households to request estimates of the amounts used during the 1982-1983 heating season. Estimates were obtained from 65 of 96 eligible households (EIA 1984b, p.101).

Table 4.8 shows, for the 1980 to 1984 survey years, eligibility rates for the Supplier Survey for electricity and natural gas by type of structure. For housing units in single-family structures, eligibility rates were high for both fuels. They were lower for housing units in structures with 2 to 4 units and lowest in structures with 5 or more units. The corresponding data for fuel oil for 1984 (the only year for which separate data were available for fuel oil and kerosene) show an even more pronounced pattern. The housing-unit eligibility rates were 98.0 percent for single-family structures, 37.6 percent for structures with 2 to 4 units, and only 2.2 percent for structures with 5 or more units. Most large apartment buildings that use fuel oil for heating have central systems, with no metering of consumption by individual units.

The primary measure of completeness of response for the Supplier Survey is the proportion of eligible housing units for which usable records were obtained. As shown in Table 4.7, these rates have been consistently high, in the neighborhood of 90 percent, for electricity and natural gas. They have been somewhat lower for fuel oil and LPG and very low for kerosene. Except for natural gas, the lowest rates for all fuels occurred in the 1984 survey.

The primary reason for nonresponse in the Supplier Survey for electricity, natural gas, and LPG has been failure to obtain permission from respondents to contact their suppliers. For these fuels, the proportion of eligible respondents not providing waivers has consistently been between 5 and 10 percent. Other reasons for failure to obtain usable records include: supplier unknown or not contacted; supplier refused to participate (a very small proportion for all fuels); supplier could not find household in records; and supplier did not provide records covering a sufficient part of the desired reference period. Records for electricity and natural gas were considered unusable if they covered less than 5 months and included seasonal use (heating or cooling) or covered less than 2 months. Records for fuel oil, kerosene, and LPG were considered unusable if they

Table 4.7. Supplier Survey Eligibility and Response Rates by Fuel and Survey Year: 1980-1993

Fuel Type and Survey Year	Number of Households Using Fuel[a]	Percent of Total Households		Percent of Eligible Households for which:		
		Eligible for Supplier Survey[b]	Eligible and Usable Records Obtained	Usable Records Obtained	Unusable Records	
					No Waiver Obtained	Other Reasons
Electricity						
1980	6,048	93.0	82.5	88.7	6.9	4.4
1981	6,263	92.2	80.8	87.6	8.9	3.5
1982	4,721	92.4	83.4	90.3	7.0	2.7
1984	5,677	91.0	79.5	87.4	8.1	4.5
1987	6,228	92.5	83.0	89.7	7.2	3.1
1990	5,094	94.0	85.1	90.5	5.7	3.8
1993	7,108	95.1	85.6	90.0	5.6	4.4
Natural gas						
1980	3,725	83.9	75.2	89.6	6.8	3.6
1981	3,850	81.7	71.7	87.8	9.2	3.0
1982	2,951	82.3	74.3	90.3	6.8	2.9
1984	3,599	79.0	70.0	88.6	8.1	3.3
1987	3,990	83.3	74.0	88.8	7.6	3.6
1990	3,255	85.8	76.6	89.3	5.7	5.0
1993	4,069	86.2	77.0	89.4	6.2	4.4
Fuel oil/kerosene						
1980	1,132	91.7	54.6	59.5	7.9	32.6
1981	1,122	93.2	46.7	50.1	7.3	42.6
1982	863	94.7	48.3	51.0	5.6	43.4
Fuel oil						
1984	918	68.1	43.2	63.4	12.2	24.4
1987	951	75.3	55.6	73.8	8.4	17.8
1990	700	80.9	58.1	71.8	7.5	20.7
1993	865	82.1	60.5	73.8	5.7	20.5
Kerosene						
1984	421	50.5	9.7	19.2	8.5	72.3
1987	414	36.5	11.6	31.8	1.4	66.8
1990	278	37.4	10.1	27.0	2.9	70.1
1993	272	45.2	12.9	28.4	5.8	65.8
LPG						
1980	574	95.7	65.5	68.4	8.0	23.6
1981	627	97.3	61.3	63.0	8.2	28.8
1982	413	94.0	67.3	71.6	6.9	21.5
1984	525	93.7	58.5	62.4	7.9	29.7
1987	543	94.5	64.5	68.3	9.3	22.4
1990	461	95.2	65.3	68.6	6.4	25.0
1993	684	97.1	71.6	73.8	5.2	21.0

[a]Includes households for which mail questionnaires were obtained.
[b]For kerosene, excludes households for which estimated purchases were reported in the Household Survey.
Source: Energy Information Administration, *Consumption and Expenditures* (1980-1993).

Table 4.8. Supplier Survey Eligibility Rates for Electricity and Natural Gas by Structure Type: 1980-1984

Fuel Type and Year	Type of Structure			
	Single Family	2-4 Units	>5 Units	Mobile Home
Electricity				
1980	98.6	85.2	68.4	99.5
1981	98.8	82.7	65.7	93.1
1982	98.4	80.1	69.5	92.8
1984	98.2	81.9	68.5	89.2
Natural Gas				
1980	98.3	66.4	25.6	78.4
1981	98.4	73.2	17.7	83.2
1982	98.3	55.5	27.5	85.1
1984	98.1	61.2	21.9	88.2

Source: Energy Information Administration, *Consumption and Expenditures* (1980-1984).

did not cover a full year. The tighter requirement for the latter three fuels has led to somewhat higher rejection rates for records submitted by their suppliers. For kerosene and LPG, a significant part of the nonresponse has occurred because companies were unknown or were not contacted for other reasons.

For each fuel, the proportion of eligible households for which consumption was imputed or estimated was lowest in the 1993 RECS. The high estimation/imputation rates for kerosene occur partly because only one-half to one-third of the households using kerosene were eligible for the Supplier Survey and partly because Supplier Survey data were successfully obtained for fewer than one-third of the eligible households each year.

When supplier survey information is not obtained for an eligible sample household, the consumption of the affected fuel for that household is estimated or imputed by a nonlinear regression model that makes use of reported data for relevant characteristics of the housing unit. The same methods are used to estimate or impute consumption for households not eligible for the supplier survey. Consumption and cost information derived from billing records is not necessarily 100 percent accurate for every housing unit (see further discussion in Chapter 6); nevertheless, the precision and reliability of such information is likely to be substantially greater than that of values imputed for housing units that were not eligible for the Supplier Survey or for which the supplier did not respond. Based on the data shown in Table 4.7, ranges of the proportion of user households for which consumption and cost have been imputed (or, for kerosene, based largely on user estimates) are:

Fuel Type	Years	Proportion Imputed
Electricity	1980-1993	14.4 to 20.5 percent
Natural Gas	1980-1993	23.0 to 30.0 percent
Fuel Oil	1984-1993	39.5 to 56.8 percent
Kerosene	1984-1993	87.1 to 90.3 percent[a]
LPG	1980-1993	28.4 to 41.5 percent

[a]Includes units for which consumption was estimated by the household respondent.

Summary

This chapter has presented information about unit and item nonresponse in the three data collection components of RECS: the Household Survey, the Rental Agent Survey, and the Supplier Survey. A summary of the highlights follows.

Household Survey

- Overall (personal interview, plus mail and telephone) unit response rates have varied from a high of 91.6 percent in the 1981 RECS to a low of 81.2 percent in 1993 (Table 4.1).

- Between the 1978 and 1990 survey years, the percentage of eligible housing units for which mail questionnaires were completed varied between a high of 5.3 percent in 1978 and a low of 3.0 percent in 1984. In the 1993 RECS, mail or telephone questionnaires were completed for only 2.2 percent of the eligible units (Table 4.1). The mail and telephone questionnaires are abbreviated versions of the personal interview questionnaires; hence item nonresponse is automatic for all items not included on these questionnaires.

- The inclusion of a longitudinal component to RECS in survey years 1982 to 1990 led to a increase in unit nonresponse each year for households that had been included in the prior survey (Table 4.2). The increase was greatest for households that participated in the Residential Transportation Energy Consumption Survey in the interval between the RECS survey years.

- The highest unit nonresponse rates have consistently occurred in the Northeast Region, in central cities, and in buildings containing more than one housing unit (Table 4.3).

- Item nonresponse rates for questionnaires completed by personal interview have been relatively low, with only 12 percent of the items having nonresponse rates of 5.0 percent or more in 1990. The highest nonresponse rate for an item required of all sample households was 14.4 percent for family income (Table 4.4).

- The proportion of interviewed housing units with complete information on size measurements has varied from a high of 82.2 percent in 1981 to a low of 56.0 percent in 1984. The most common omission has been failure of the interviewer to indicate whether measurements were made inside or outside the housing unit (Table 4.5).

Rental Agent Survey

- The proportion of households eligible for the Rental Agent Survey has varied from 10.7 to 15.4 percent of all households for which questionnaires were obtained by interview or mail. Response rates for eligible households have ranged from a low of 57.0 percent in 1982 to a high of 89.1 percent in 1987 and have exceeded 80 percent in each of the last three survey years (Table 4.6).

Supplier Survey

- The proportion of households eligible for the Supplier Survey varies by fuel and type of structure. Eligibility rates have consistently been over 90 percent for electricity and LPG, and close to or above 80 percent for natural gas. Rates for fuel oil have been somewhat lower. Most of the households not eligible for the Supplier Survey for these fuels are living in rental units located in buildings with two or more units (Table 4.8).

- Low eligibility rates for the Supplier Survey for kerosene occur largely because many of the households using that fuel purchase it only on a cash and carry basis.

- Among households eligible for the Supplier Survey, generally between 5 and 10 percent have been unwilling to sign a waiver to allow the survey contractor to contact their suppliers (Table 4.7).

- Taking into account the joint effects of eligibility and the extent of success in obtaining usable records for eligible households, supplier data have been obtained consistently for close to or better than 80 percent of all electricity users, over 70 percent of natural gas users, and close to or better than 60 percent of LPG users. For fuel oil, the corresponding rates have varied from 43 percent in 1984 to 60 percent in 1993, and for kerosene they have been in the vicinity of 10 percent (Table 4.7). By this measure, the best results for each of the 5 major fuels was obtained in the 1993 Supplier Survey.

5. Measurement Error

This chapter presents information about measurement errors associated with the data collection phase of RECS. Measurement error can contribute to the total error of survey estimates in the form of bias or nonsampling variance. Direct or indirect quantitative information about measurement errors in RECS can be obtained in several different ways.

Special data collection procedures, which generally cost more than standard interviews but are believed to provide more precise information, can be used to collect information for selected households. These procedures often involve various kinds of direct observation and physical measurement, as opposed to merely asking for information from survey respondents. Such procedures include energy assessments in which data on household characteristics are collected by trained technicians. Other examples are the collection of nameplate data in order to obtain more precise information on the characteristics of central air-conditioning units and the collection of information on thermostat settings and temperatures by direct observation rather than by asking respondents to report them. Another useful procedure has been to conduct personal or telephone interviews with "outliers"--i.e., households identified in the data processing phase of the survey as having reported unusual or apparently inconsistent values for selected items. These kinds of procedures are reviewed in the first section.

Comparisons of data for the same household from different sources provide another kind of information about measurement error. Cross-sectional comparisons involve data for a household for the same time period from the household, rental agent, and supplier surveys; longitudinal comparisons involve data for the same household from successive survey years. Weather data assigned to households in a specified geographic area can also come from more than one source. Results from these types of comparisons are presented in the second section.

The level of measurement error can also be affected by the design and format of the survey questionnaire and by the type of training administered to interviewers. Information on these topics is presented in the final section.

RECS estimates of end-use energy consumption within households are obtained indirectly from survey data by allocating total consumption to various uses on the basis of a nonlinear regression model. These estimates and their evaluation through submetering studies are covered in Chapter 7, "Estimation and Sampling Error." Macro-comparisons, that is, comparisons of RECS estimates with comparable data from other surveys conducted by EIA and from surveys conducted by other agencies and organizations, are discussed in Chapter 8, "Comparisons of RECS Estimates with Other Data."

Special Data Collection Procedures

Energy Assessments

In 1979, following the National Interim Energy Consumption Survey (NIECS), an Energy Assessment was undertaken by Technology and Economics, Inc., of Cambridge, Massachusetts, in a subsample of 44 of the NIECS sample households. Trained technicians visited the households, all but two of which were single-family households. They measured floor areas, counted windows, examined insulation and noted the characteristics of space-conditioning equipment and selected appliances. Their observations were compared with the responses to the NIECS interviews (Blumstein, York, and Kemp 1981).

This Energy Assessment was undertaken as a pilot test for a continuing program of assessments that was being considered as a regular part of RECS (Response Analysis Corporation 1980, Part 6). The plan was to perform such assessments for a subsample of the households included in each Household Survey. However, resources available for RECS proved to be insufficient to implement this plan and there have been no further assessments of this type. The generality of the Assessment findings was limited by the use of a small convenience sample, lack of fully standardized procedures, and limited training for the technicians. In addition, the data collection instrument was not designed for direct comparisons with corresponding NIECS data items and there was no followup to reconcile differences between the two sources of information. Nevertheless, the Assessment provided useful information about possible sources of measurement errors in NIECS and subsequent surveys.

There were large differences between the NIECS and Assessment data on square feet of floorspace (Table 5.1). For 14 of the 27 households that had usable measurements from both sources, differences were 25 percent or more of the Assessment values. NIECS respondents had been asked to give their best estimates of floorspace; in the Assessment the technicians made measurements. Some of the discrepancies may have been due to a conceptual difference: NIECS respondents were asked to report square feet of living space, while the Assessment technicians were asked to measure "conditioned space," including only rooms and other enclosed areas with some direct means of heating. On this basis, one might expect the NIECS values to be somewhat larger; nevertheless, for 9 of the 27 households, the NIECS values were 25 percent or more below the Assessment measures.

Despite their limitations, the Assessment findings on floorspace demonstrated that respondent estimates of floorspace were likely to be subject to unacceptably large errors. Consequently, from RECS survey year 1980 on, measurement by survey interviewers has been the preferred method of obtaining information on floorspace. The measurement procedure used in the Assessment was itself not fully satisfactory and has been replaced in RECS by procedures that are believed to be easier to use and more reliable.

Table 5.1. Comparison of NIECS and Energy Assessment Data on Floorspace

Percent Difference $\left(\frac{\text{NIECS - EA}}{\text{EA}} \times 100\right)$	Number of Households		
	NIECS > EA	NIECS < EA	Total
0 to 9.9	8	1	9
10.0 to 24.9	2	2	4
25.0 to 49.9	2	4	6
50.0 and over	3	5	8
Total	15	12	27[a]

[a]There was no space information from one or both sources for 17 of the EA households.
Source: Blumstein, York, and Kemp, *An Assessment of the National Interim Energy Consumption Survey* (1981), Table 8.

Some other findings from the comparison of NIECS and Energy Assessment data were:

- There were many differences between counts of windows, both by type and overall for the household.
- The presence or absence of attic insulation was reported accurately, but there were substantial differences in reports of the thickness of insulation used.
- Reports of fuel used for heating and other purposes were generally in agreement. An exception was fuel used for dryers; of 33 households for which the dryer fuel was reported in both NIECS and the Assessment, there were differences for 7, all of which reported electricity in the Assessment and gas in NIECS.
- Several differences were observed in the numbers of refrigerators and separate food freezers reported and in the characteristics of refrigerators, such as temperature controls and automatic defrost/frost free features.

These findings from comparisons of NIECS and Energy Assessment Data were taken into account in the determination of content and formulation of questions for subsequent surveys.

Collection of Nameplate Data

In the 1990 Household Survey, interviewers were asked, for single family houses with central air-conditioning, to record manufacturer's name, model number, year manufactured, and other information from the nameplate of the outside unit (Hall 1992). The main purpose of collecting this information was to obtain a measure of rated efficiency for each housing unit's central air-conditioning equipment. This was to be done by matching the make and the model year and number against semi-annual directories of equipment characteristics issued by the

Air-conditioning and Refrigeration Institute (ARI). For all successful matches, the seasonal energy efficiency ratio (SEER) for the equipment was entered into the RECS data file.

Table 5.2 shows the results of the attempts to acquire SEER's for central air-conditioning equipment. No nameplate data were obtained for 26.5 percent of the 1,820 households with central air-conditioning, either because they were located in multiunit buildings or because they had responded by mail. Directory matches were attempted for the remaining 1,337 households; SEER's were obtained for only 24.8 percent of those households, or 18.2 percent of all sample households with central air conditioning. The most frequent reasons for failure to find a SEER in the directories were failure to match on manufacturer's name and failure to match on model number.

A subsequent effort was made, for the 331 households for which SEER's had been obtained, to obtain capacity values from the ARI directories. Values were located for 279 (84.3 percent) of these households. In view of the high cost and limited success of the nameplate data collection and matching operations in the 1990 RECS, they were not undertaken in 1993.

Table 5.2. Results of Matching Nameplate Data Against ARI Directories to Obtain Seasonal Energy Efficiency Ratios: 1990 RECS

Outcome of Match	Number of Units	Percent	
		Of Total	Of Attempted Matches
Households with central air conditioning	1,820	100.0	
No match attempted	483	26.5	
Mail questionnaire			
Multi-unit building	99	5.4	
	384	21.1	
Match attempted	1,337	73.5	100.0
Successful, SEER obtained	331	18.2	24.8
No SEER obtained	1,006	55.3	75.2
No match on make			
No model year	574	31.5	42.9
No match on model	23	1.3	1.7
No SEER available[a]	283	15.6	21.2
	126	6.9	9.4

[a]Unit manufactured prior to 1980 or no SEER in directory.
Source: Hall, *Nameplate Data Collection in the 1990 RECS* (1992).

Checking Thermostat Readings

Since 1981, the RECS Household Survey questionnaires have included questions on average temperatures maintained in the home in the wintertime under three conditions: during daytime when someone is home, during daytime with no one home, and during sleeping hours. If respondents say they cannot report temperatures but can give thermostat settings, the latter are accepted. These self-reported temperatures are characterized in the survey reports as follows:

> The self-reported temperatures, especially for some respondents, are impressions of typical temperatures and may not represent actual temperatures, or the averages of actual temperatures in the home. (EIA 1993a, p. 148)

There have been no attempts in RECS to collect information about indoor temperatures or thermostat settings by direct observation. However, a study in a small city in New York State provided some information on the accuracy of self-reported thermostat settings (Luyben 1982). Data were collected for one sample of households by personal interviews and for another sample by telephone. In the telephone survey, respondents were first asked to report their thermostat settings and then to go to their thermostats and check the reported values. The mean of the checked values was significantly higher than the reported values by 0.6 degrees Fahrenheit.

In the personal interviews, the interviewers recorded observed values of thermostat settings and temperature readings. Temperature readings exceeded thermostat settings by a mean of 0.8 degrees. The mean of the observed settings was significantly higher than the mean of the checked settings from the telephone survey households: 68.3 versus 67.0 degrees.

Detection and Evaluation of Outliers

The detection and analysis of outliers can be a useful technique for understanding survey responses, identifying and controlling survey errors, and improving survey processes. Outliers are reported values that lie at the extremes of a univariate or multivariate distribution of variables included in the survey. The analysis of outliers can include recontacts with survey respondents to determine whether there were errors in the values initially reported and whether there were special circumstances to explain the unusual observations.

In March of 1984, in-depth reinterviews were conducted with eight households that had participated in the 1981 RECS and for which data on consumption were available from suppliers (EIA 1984b, Appendix G, Erickson 1984). The method used to identify these eight households as outliers was to impute their consumption of specific fuels, using the regression models normally used to impute missing data on consumption and to compare the imputed values of consumption with the values actually provided by the suppliers. A purposive sample of eight households showing large differences in either direction for consumption of electricity, natural gas, or fuel oil was selected for the interviews. These were households whose consumption appeared to be far out of line with what might have been expected on the basis of housing unit characteristics and household behaviors that had been reported in the initial interviews.

There were two sets of four interviews each. Set A was conducted by a pair of interviewers, and Set B by a single interviewer. All used conventional ethnographic interviewing techniques. Their primary goal was to arrive at an explanation for the unusually high or low consumption, but they also investigated several broad topics, including family interactions, recreation, home improvements and attitudes toward utility companies, nuclear power, conservation, rising costs of energy, and family finances.

The main findings of the eight in-depth interviews are presented in Figure 5.1. The interviewers were successful, for the most part, in finding reasonable explanations for the extreme values in consumption of specific fuels. The explanations proved to be more or less equally divided between reporting errors in the initial RECS interviews (six of the eight households) and unusual circumstances affecting consumption, some of which may not have been fully reflected in the imputation model (also present in six of the eight households).

An important motivation for undertaking these interviews was to determine whether the questionnaire for the 1984 RECS could be expanded to include information that would help to explain patterns of unusually high or low consumption. After the findings were reviewed, no questions were added but consideration was given to other changes in survey procedures, such as improved interviewer training, additional processing steps, and followup interviews. One outcome has been the inclusion of "model-based outlier checks" as a standard part of data processing. In processing the 1990 RECS data, for example, there was a manual review, sometimes involving telephone calls to respondents, of data for all households for which the model-based estimate of fuel consumption was more than three times or less than one-third of the value based on Supplier Survey data (Response Analysis Corporation 1992b, p.7-14).

Another outlier investigation associated with the 1981 RECS had to do with data on temperature settings, a topic that had been included in the questionnaire for the first time in that survey year (Thompson 1982, Day 1982). The survey contractor made telephone calls to 9 respondents who reported maintaining (with a thermostat, radiator valve, or other control) nighttime temperatures higher than their daytime temperatures (presumably a reversal from normal behavior) and to 9 respondents who reported nighttime temperatures substantially lower than their daytime temperatures.

In the first group, nighttime higher than daytime, eight of the nine respondents called changed their responses in ways that reduced the differences; however, all but two of the group confirmed that they purposely maintained higher temperatures at night and provided explanations for that behavior. In the second group, nighttime much lower than daytime, all but one confirmed their original responses, although some said they were uncertain about the precise temperature levels. Most of the explanations involved use of electric blankets or a warm combination of non-electric blankets. The findings suggested that responses to questions about temperatures maintained in the housing unit are subject to sizable response errors.

An analysis of outliers in the 1984 RECS by Latta (1988) suggests the potential power of this method. The goal of Latta's analysis was to improve the nonlinear regression model used to impute missing entries for heated floorspace. He used data for sample housing units with complete information to estimate the parameters of the proposed model and observed that there

were several extreme outliers, units with large differences between reported and imputed values. An examination of the data listings for these units showed substantial clustering by PSU and interviewer, leading to a hypothesis that a few interviewers may have been making systematic errors in identifying which portion of total floorspace was heated. One set of outliers consisted of four townhouse units, each with three floors whose dimensions had been recorded as 7 by 30 feet, a rather unlikely set of measurements. Because this analysis was undertaken well after completion of the 1984 RECS, there was no followup on these particular cases. However, the findings indicate that comparison of reported and imputed values followed by review and followup of outliers is a promising technique for quality improvement.

Comparisons of Individual Household Data from Alternate Sources

The design of RECS provides built-in opportunities to investigate the nature and size of measurement errors through the analysis of multiple observations of the characteristics of individual housing units or households. The longitudinal component of the sample design provides observations for the same housing units at different times. The collection of overlapping data for selected items from three sources--households, rental agents and suppliers--provides duplicate observations for the same housing units at the same time or covering the same time period. The interpretation of data from multiple observations is not necessarily straightforward; the sources of the observations must be carefully considered to decide what they tell us about the effects of response bias or response variability on the survey estimates.

Longitudinal Comparisons

As described in Chapter 2, the RECS samples for 1982, 1984, 1987, and 1990 each contained a subsample of housing units which had been included in the sample in the preceding survey year. Because a large proportion of the questionnaire content is repeated in successive survey years, responses to comparable items for the same unit in the 2 years can be compared. The interpretation of observed differences is not obvious. For some housing unit characteristics, such as year built and type of housing unit (mobile home, single-family detached, etc.), there should be no differences from one survey year to another, so that differences are almost certainly due to errors in data collection or processing. For other housing characteristics, such as appliances, types of fuels used, and even number of rooms, real changes can occur. Real changes in household characteristics, like number of persons and family income, can occur whether or not a different household occupies the housing unit in successive survey years.

Table 5.3 provides information about differences in selected items for housing units that were occupied by the same households in the 1980 and 1982 RECS (Thompson 1985a). In an effort to determine the reasons for individual differences, telephone calls were made to households for which the responses for 1980 and 1982 differed for one or more of the selected items. Only 71 percent of the differences were checked in this way: 12 percent were eliminated to reduce burden on households with more than three items showing differences and 19 percent were associated with households that could not be reached by telephone. Thus, the final column of the table, showing the percent of differences "unexplained," includes some differences for which interviewers and respondents could not provide any explanation and some which were not covered by telephone calls to respondents.

Table 5.3. Differences Between Responses Reported by the Same Household in the 1980 and 1982 RECS

Item	Base[a]	Differences		Percent of differences:		
		Number	%	Explained by Real Change	Explained by Errors[e]	Unexplained
1. Number of windows[b]	1,398	337	24.1	3	65	32
2. Year the house was built[b]	1,296	296	22.8	0	12	88
3. Main home heating equipment[b]	1,394	206	14.8	19	19	62
4. Number of stories[c]	999	145	14.5	1	74	25
5. Year moved in	1,397	194	13.9	0	21	79
6. A/C equipment present	1,398	179	12.8	38	36	26
7. Number of rooms[b]	1,398	165	11.8	9	44	47
8. Use a home freezer	1,395	149	10.7	59	23	18
9. Basement heated or unheated[c]	1,111	110	9.9	8	71	21
10. Number of refrigerators	1,381	128	9.3	45	37	18
11. Full basement/part basement[c]	1,111	96	8.6	0	64	36
12. Availability of natural gas[d]	370	32	8.6	21	33	46
13. Use a clothes dryer	1,397	115	8.2	40	13	47
14. Type of living quarters	1,400	107	7.6	0	63	37
15. Number of bathrooms[b]	1,387	91	6.6	15	41	44
16. Main home heating fuel	1,398	87	6.2	42	24	34
17. Main water heating fuel	1,393	84	6.0	17	19	64
18. Presence of a basement[c]	1,111	45	4.1	3	71	26
19. Main cooking fuel	1,400	40	2.9	0	12	88

[a]Base excludes households for which 1980 RECS response is imputed or unknown and those for which 1982 RECS response is unknown.

[b]Some responses are grouped for these items. For a difference to be counted, it must be >3 windows; the years 1975-79 were combined into one category; hot water pipes and radiators were combined as one heating system; full and 1/2 baths were combined and each counted as one bathroom; the difference between number of rooms must be >1.

[c]Single-family homes.

[d]Single-family or mobile homes that do not use natural gas.

[e]Errors by respondents, interviewers, coders, and data entry operators.

Source: Thompson, *Utility of Paying Respondents: Evidence from the RECS* (1985).

For 9 of the 19 items shown in Table 5.3, real change accounted for at least 10 percent of the observed differences. These are minimum values, because some of the unexplained differences may have resulted from real changes. Such real changes appear to be largely related to the acquisition of new appliances and heating or cooling equipment and to changes in the availability of natural gas, making possible changes in the main heating fuel used.

For 6 of the 19 items, the attempt at reconciliation of differences confirmed that at least half of them resulted from errors in data collection or processing. These were items for which one would expect few, if any, real changes to occur: number of windows, number of stories, type of living quarters, presence of a basement, basement heated or unheated, and full basement or part basement. The general conclusion to be drawn from these findings is that, at the level of the individual housing unit, real changes over time are difficult to distinguish from differences due to measurement errors. Because essentially the same data collection procedures were used in both years, we can also conclude that estimates for some housing unit characteristics, notably number of windows, are subject to high response variability.

Measures of total and heated floorspace were also compared for a subsample of 355 housing units included in the 1980 and 1982 RECS (EIA 1984b, p. 114-115). The results for 300 housing units that had usable square footage data for both years are shown in Table 5.4. Averages for the total and single-family detached units were fairly close for the two survey years. However, the median absolute percent differences between values for individual units for the two years were relatively large, 11.7 percent overall for total square footage. They were larger, at 15.6 percent, for heated square footage, probably because of uncertainties about the interpretation of the concept of a "heated area," possibly also in part because of some real changes in this item.

Longitudinal comparisons of 1982 and 1984 data, also with telephone calls to explain differences, were undertaken following the 1984 RECS. In this instance, only seven topics were selected for analysis: main home heating fuel, main water heating fuel, air- conditioning equipment and fuel, clothes dryers, home freezers, dishwashers, and availability of natural gas. Telephone contacts were successfully completed for 505 (76 percent) of the 668 differences that were found for these seven topics. Real changes explained 42 percent of these differences; virtually all of the rest resulted from errors in the 1982 or 1984 values or, in a few instances, errors in both years (EIA 1987d).

Records for the longitudinal differences for which respondents were successfully contacted are included in the public use files for the 1982 and 1984 RECS. There is a separate record for each difference showing the topic number, a code for the interpretation of the difference (year 1 correct, year 2 correct, neither year correct, real change, or cannot determine) and a code identifying the reason for the error, if one occurred. As noted below, some longitudinal comparisons have been made for 1984-1987, and comparisons are possible for 1987-1990, but no followup contacts were made, following the 1987 or 1990 RECS, to determine reasons for differences.

Table 5.4. Differences in Square Footage Reported for the Same Household in the 1980 and 1982 RECS

Item	Total	Housing Type			
		Single-Family Detached	Mobile Home	Multi-unit Building	Housing Type Responses Differ in 1980 & 1982
Number of cases[a]	300	208	14	70	8
Average Square Feet per Housing Unit					
1980	1,797	2,116	803	1,082	1,503
1982	1,821	2,142	721	1,147	1,282
Median Percent Difference in Square Footage	11.7	11.8	7.2	12.2	11.3
Average Heated Square Footage per Housing Unit					
1980	1,536	1,780	798	966	1,469
1982	1,521	1,751	711	1,039	1,194
Median Percent Difference in Heated Square Footage	15.6	16.9	7.2	14.4	13.4

[a]Units that had good square footage data for both years.
Source: Energy Information Administration, *Consumption and Expenditures* (1982).

Table 5.5 shows results of a comparison of 1984 and 1987 RECS data on type of housing unit for units that were included in both surveys. The final column shows the index of inconsistency for each category. The index of inconsistency is a measure of the percent of total variance for an item that is accounted for by response variance. As a rough rule of thumb, response variance is considered to be low when the index is less than 20, moderate for values between 20 and 50, and high when it is greater than 50. (For further discussion and a formula for calculating the index, see Groves (1989).) The value of the index for the "single family attached" category is at the upper end of the moderate range, indicating that there were frequent difficulties in distinguishing such units from single family detached units and from those in apartment buildings with two to four units.

Table 5.6 shows a comparison, also based on housing units included in both the 1984 and 1987 RECS, for reports on year of construction of the housing unit (Battles 1991a). This tabulation is limited to housing units that were occupied by the same household in 1984 and 1987 and for which the householder was the respondent in both years. The values of the index of inconsistency are in the moderate to high ranges, higher for the most part than the values

observed for type of housing unit in Table 5.5. As one might expect, the values are smaller for the most recent periods. They are also smaller for units built prior to 1940, presumably because the time period covered by that category is much longer.

Table 5.5. Longitudinal Households[a]: Housing Type Reported in 1984 and 1987 RECS

Housing Type Reported in 1984	Housing Type Reported in 1987					Index of Inconsistency
	Mobile Home	Single-Family Detached	Single-Family Attached	Apartment Bldg.		
				2-4 Units	5+ Units	
Mobile Home	**115**	9	0	0	0	7.7
Single-Family Detached	9	**1,265**	16	20	1	9.5
Single-Family Attached	0	26	**53**	14	2	46.7
Apt. Bldg. 2-4 Units	0	10	21	**209**	10	19.2
Apt. Bldg. 5+ Units	0	0	6	10	**269**	5.9

[a]Tabulation excludes 15 cases where it was determined that different housing units had been interviewed and one case where the basement had been converted to an apartment.

Source: Energy Information Administration, *Housing Characteristics* (1987).

Table 5.6. Longitudinal Households[a]: Year of Construction Reported in 1984 and 1987 RECS

Year of Construction Reported in 1984	Year of Construction Reported in the 1987 RECS							Total Units	Index of Inconsis-tency
	Before 1940	1940-1949	1950-1959	1960-1969	1970-1974	1975-1979	1980-1984		
Before 1940	**333**	42	26	22	12	6	1	442	30.4
1940 to 1949	27	**59**	19	7	2	0	2	116	57.7
1950 to 1959	23	15	**134**	34	7	4	2	221	49.3
1960 to 1969	9	11	40	**145**	25	13	4	247	47.8
1970 to 1974	8	6	6	22	**95**	17	7	161	48.2
1975 to 1979	3	0	5	3	26	**114**	9	160	32.7
1980 to 1984	1	0	4	1	3	7	**52**	68	29.8
Total Units	404	133	236	234	170	161	77	1,415	

[a]Housing units occupied by the same household in 1984 and 1987.

Source: Battles, *Effects of the Adjustment of 1990 Census Data on the 1990 RECS Control Totals Obtained from the Current Population Survey* (December 1991).

Cross-Sectional Comparisons

As a result of the multi-stage clustered sample design used in RECS, the sample of housing units sometime contains two or more units from the same multiunit structure. When this occurs, there are some housing unit characteristics, such as the year in which the structure was built and the main space and hot water heating fuels, that one would expect to be the same for every unit in the structure. When differences are found for these characteristics in "inter-case comparisons" of different units in the same structure, they can be taken as indications of response error.

Blumstein, York, and Kemp (1981) used data from NIECS (the 1978 RECS) to make such inter-case comparisons. There were some difficulties in determining which sample housing units came from the same structure, but by matching on structure characteristics and identifiers, the investigators succeeded in identifying 78 structures with more than one sample housing unit. These structures included 305 sample housing units for which interviews were completed. When different responses were found on items, such as year built for housing units in the same structure, the most frequent response was assumed to be the correct one. On this basis, and leaving out responses of "don't know" and those that were missing, the apparent error rates were:

Item	Total	Number of Responses Other Than Most Frequent	Percent Apparently Incorrect
Year Built	274	44	16
Main Heating Fuel	300	15	5
Main Water Heating Fuel	300	24	8

The 16-percent gross error rate for year built is much lower than the rate of 42 percent that can be derived for the data on year built shown in Table 5.6. The data in Table 5.6 were based on a longitudinal rather than a cross-sectional comparison and included all types of housing units, not just those in multiunit structures.

As described in earlier chapters, for housing units in multiunit structures for which one or more fuels are included in the rent, the Rental Agent Survey provides data for selected items that can be compared with data for the same items from the Household Survey. As part of regular data processing operations, the two sets of data are compared. When there are differences, the response considered more likely to be correct is accepted. Except for supplemental heating fuels, this is normally the response given by the rental agent.

Table 5.7 summarizes changes made on the basis of responses from the Rental Agent Surveys for survey years 1981 through 1987. These data suggest what error rates for these items *might have been* if the Rental Agent Surveys had not been conducted and the Household Survey responses had been accepted. They also provide an indication of the level of error for these items for housing units that were eligible for the Rental Agent Survey but for which no information was obtained in that survey. The levels of nonresponse to the Rental Agent Surveys were shown in Chapter 4, Table 4.6.

Table 5.7. Changes Resulting from Comparison of Rental Agent and Household Survey Responses: 1980-1987

Item	Survey Year				
	1980	1981	1982	1984	1987
Main Heating Fuel					
Number of Changes	31	58	31	75	62
Percent of Base[a]	NA	15.8	12.2	14.7	9.2
Main Heating Equipment					
Number of Changes	NA	52	40	68	206
Percent of Base[a]	NA	14.1	15.7	13.3	30.7
Supplemental Heating Fuel					
Number of Changes	27	18	5	41	29
Percent of Base[a]	NA	4.9	2.0	8.0	4.3
Water-Heating Fuel					
Number of Changes	40	82	36	103	120
Percent of Base[a]	NA	21.1	13.2	19.4	14.8
Air-Conditioning Fuel					
Number of Changes	6	1	2	14	61
Percent of Base[a]	NA	16.7	4.5	11.8	39.6
All Items					
Number of Units in Rental Agent Survey	551	466	308	549	856
Percent with >1 Changes	NA	30.0	26.0	32.4	41.8

[a]Base for the first 3 items in the number of units whose rental agents paid for the main heating fuel. For the fourth and fifth items, it is the number whose agents paid for the fuel in question.

NA = Not Applicable.

Source: Energy Information Administration, *Housing Characteristics* (for the years shown).

With two striking exceptions, the proportions of eligible units (those included in the base for each item) whose Household Survey responses were changed based on Rental Agent Survey responses were relatively stable over the years shown. The proportions were lowest for supplemental heating fuel because the household respondent is usually considered to be more knowledgeable for that item. The exceptional cases were main heating equipment and air-conditioning fuel, for which the proportion of changes in the 1987 RECS was substantially greater than in any prior survey year. These changes may have resulted, at least in part, from changes in the questions relating to these two items on the Rental Agent Survey questionnaires. On the 1984 Rental Agent questionnaire there was a single item for main heating equipment, listing 13 possible alternatives. On the 1987 questionnaire, two separate lists of heating equipment were provided, one for units using electricity as the main heating fuel and one for units using any other fuel. For main central air-conditioning fuel, there was a minor change: the response categories on the 1984 questionnaire were electricity, gas from underground pipes, and LPG, in that order, whereas

on the 1987 questionnaire the order was changed to gas from underground pipes, LPG, and electricity. There were no changes for these items on the Household Survey questionnaires.

For the most part, there has been little overlap between the data items collected in the Household and Supplier Surveys. For delivered fuels that are paid for directly by the household, consumption and cost data are collected from the suppliers. For fuels whose costs are included in rent payments, consumption and cost are imputed on the basis of housing unit and household characteristics; this is also done when a sample household is eligible for the Supplier Survey for one or more fuels but a response cannot be obtained from the supplier(s). To evaluate these imputation procedures, some data on consumption for whole buildings containing sample rental units were collected in the 1981 RECS. The findings from that study are described in the section on "Imputation" in Chapter 6.

Following the 1993 RECS, responses to new Household Survey questions about availability of and participation in demand-side management (DSM) programs were evaluated by comparing them with responses to similar questions that had been included in the Supplier Survey. Several kinds of DSM programs are offered by utilities to encourage customers to modify their patterns of energy use, the goals being to reduce overall demand or shift some uses away from peak load periods. The comparisons showed substantial Household Survey underreporting of the availability of DSM programs. Of the households interviewed, 36.1 percent reported that at least one type of DSM program was offered to them by their electric utility, natural gas utility or some other group. By contrast, 80.6 percent of the suppliers providing electricity to the same households reported that they offered some type of DSM program. The proportion of households actually participating in electric or natural gas DSM programs was much smaller, but again there was considerable disagreement between the response to the Household and Supplier Surveys. There were many differences in both directions, but the net result was that participation appears to have been overreported in the Household Survey (EIA 1995d, pp. 152-153).

Alternative Sources of Weather Data

The data record developed for each RECS sample housing unit includes data on weather conditions in the vicinity of the housing unit. Of particular importance are data on heating and cooling degree-days, both for the survey reference year and for a recent 30-year period (current and normal degree-days). Such data have several important uses:

- When supplier data are not obtained for a housing unit for one or more fuels, the data on heating and cooling degree-days are important inputs in the models used to impute consumption of those fuels.

- For all housing units, the data are used as inputs to the models used to estimate end-use consumption.

- In longitudinal analyses, variations in degree-days and departures from normal are important determinants of variations in consumption.

Ideally, degree-day data for each housing unit would be obtained by measurement of temperatures at the site of the housing unit. Because this is not practical, a reasonable alternative is to use data from the more than 4,000 individual weather stations maintained by the National Oceanic and Atmospheric Administration (NOAA). Two basic methods of using these data are possible: (1) For each housing unit, use data from the individual weather station that is closest, in some sense, to that unit, or (2) Use average data based on all stations in the NOAA weather division in which the housing unit is located. NOAA has divided each of the 48 contiguous States into divisions, usually consisting of groups of counties, that have similar weather conditions. As of 1987, there were 345 NOAA divisions, an average of about seven per State.

A priori reasoning suggests that method (1), using the data from the closest individual weather station, would provide more accurate measures of degree-days. Temperatures can vary substantially within a multi-county area, especially when influenced by changes in elevation or proximity to large bodies of water. However, higher costs and some operating problems are associated with method (1). Selection of the "closest" station, taking into account distance and other relevant factors, requires manual processing operations which must be repeated for each survey as new PSU's or SSU's are introduced into the sample. Data are incomplete for some of the individual stations, so that imputation of missing data or substitution of another nearby station may be necessary.

Based on these considerations, EIA elected to use NOAA division data on degree-days for all survey years through 1984. Two evaluations of the effects of using alternative methods were undertaken prior to the 1987 RECS. In a 1982 study by the Energy Resources Group (Blumstein *et al*.), one site was chosen in each of the 103 PSU's included in NIECS and its station data were compared with averages for the NOAA division in which the site was located. The sites were chosen to meet two requirements: high population density and presence of an individual weather station. The data used in this evaluation were 30-year averages.

This comparison showed a median absolute difference in degree-days of five percent between the data for the site averages and the NOAA division averages. Most of the large differences (in excess of 13 percent) were in California, where they averaged 30 percent. Reasons for these large differences included large divisions, with boundaries drawn to coincide with drainage basins rather than areas of homogeneous climate, and climatic patterns that vary substantially over short distances. The study investigators recommended that an alternative method be used to derive degree-day values for housing units in California and in other locations that showed large differences between individual station data and division averages.

The 1986 evaluation (Mooney and Carroll) was undertaken by the main survey contractor, Response Analysis Corporation. It was initially limited to the five States that had shown the largest differences between station and division data in the 1982 evaluation. Instead of selecting one individual station to represent a PSU, a separate selection was made for each SSU. The comparisons were based on data for the 1984 survey reference year, April 1984 through March 1985. The investigators concluded that "... using individual station data on the SSU level, rather than NOAA divisional data, more accurately represents local temperature conditions." (Mooney and Carroll, p.27)

The evaluation also examined the effect of temperature data from alternative sources on the model-based imputation and end-use allocation procedures used in RECS. Based on the data for the five States included in the initial evaluation, the following conclusions were reached:

> ... end use models run with division data are biased in several ways. First, because of the fact that we overestimate degree days to a larger degree for low users of the fuels, the models overestimate the amount of fuel used for space heating. Second, because the degree to which we mis-estimate degree days varies by household, consumption amounts at the household level are mis-estimated. Finally, we underestimate consumption for "imputation" households because the division model allocates too little consumption to non-space heating uses. (Mooney and Carroll, p.40)

Subsequent to the five-State evaluation, all 1984 RECS SSU's were assigned to individual weather stations rather than divisions and the new degree-day values that resulted were assigned to individual households. The models used to impute consumption and to allocate it to end uses were rerun with the new degree-day values, and the results were compared with those derived by using the division averages for degree-days (Response Analysis Corporation 1988). These comparisons showed that:

- As shown in Table 5.8, there was a reduction of 3.6-percent in heating degree-days at the national level, with particularly large reductions in the Mountain and Pacific divisions. Conversely, there was an increase of 12.1 percent in cooling degree-days at the U.S. level, with increases of 10 percent or more in 7 of the 9 Census Divisions.

- The changes in overall consumption were relatively small, because only households lacking supplier data are affected.

- The only fuel with a substantial change in consumption was fuel oil, for which use of the station data led to a reduction of 1.6 percent at the U.S. level.

- End use allocations shifted somewhat. At the national level, the use of station data led to a 1.6 percent decline in space heating consumption which was offset by a 3.6 percent increase in water heating consumption.

As a result of the above findings, RECS degree-day data for the 1987 and subsequent surveys have been based on records provided by NOAA for individual weather stations. The "closest" weather station is identified for each SSU, mainly on the basis of distance, but also taking into account differences in elevation, proximity to large bodies of water, and the extent of missing data for the preferred station. Users should be aware that, because of this change, estimates of degree-days from the 1987 and subsequent survey years are not directly comparable with estimates from earlier surveys.

Table 5.8. Comparison of Heating Degree-Days Using NOAA Division Method Versus Station Method, April 1984 Through March 1985

Census Division	Million Households	Heating Degree-Days			Cooling Degree-Days		
		Division Method	Station Method	Percent Difference	Division Method	Station Method	Percent Difference
United States	86.328	4,686	4,518	-3.6	1,153	1,293	12.1
New England	4.269	6,398	6,331	-1.0	524	621	18.4
Middle Atlantic	14.029	5,663	5,460	-3.6	683	822	20.3
East North Central	15.203	6,524	6,427	-1.5	685	777	13.4
West North Central	6.414	6,619	6,499	-1.8	976	1,076	10.2
South Atlantic	14.777	2,951	2,979	0.9	1,768	1,819	2.8
East South Central	5.784	3,651	3,512	-3.8	1,433	1,583	10.5
West South Central	8.764	2,443	2,444	0.1	2,361	2,431	2.9
Mountain	4.512	5,728	5,158	-10.0	1,102	1,550	40.6
Pacific	12.577	3,508	3,019	-13.9	873	1,148	31.5

Source: Energy Information Administration, *Consumption and Expenditures* (May 1987).

Questionnaire and Interviewer Effects on Measurement Error

Since the inception of RECS, there have been continuing efforts to reduce measurement error by making improvements in questionnaires and in the training and supervision of interviewers. For each survey year, the Household Survey questionnaire and other survey instruments have been pretested and subjected to reviews by EIA and contractor staff and other persons with expertise in questionnaire design. As described in Section 5.1, there has been some use of in-depth interviews in attempts to explain unusual consumption patterns. This section cites some additional examples of relevant activities.

Pretests and Questionnaire Reviews

In preparation for the 1990 RECS, the draft Household Survey questionnaire was pretested by three interviewers, one of them an experienced RECS interviewer, in nine households. Each of the interviewers completed a detailed evaluation form, with comments on each section of the questionnaire, and participated in a debriefing session. In addition to suggestions for clarifying specific questions, some of the points raised in the overall report on the pretest (Miksovic 1989) were (some, but not all, of these suggestions were adopted in the final version of the questionnaire):

- The questionnaire includes some rather abrupt switches from one topic to another. Transition sentences should be provided at these points.

- The interviewers felt that many of the questions were very wordy, especially some that included the phrase "... other apartments, condos, households, businesses, or farm buildings." It was recommended that the phrase "... and farm buildings" be put in parentheses, to be used by interviewers only when it seems appropriate.

- The format of the questionnaire made it difficult for interviewers to refer to the relevant instructions while they were outside making and recording measurements of building dimensions. Several possible improvements were suggested.

- Improvements in format were suggested, such as use of larger print, highlighting skip instructions in various ways, and using different print types.

A user-needs study prior to the 1993 RECS (EIA 1993c) identified a widespread interest in the collection of additional information on lighting, to track use of new lighting equipment technology and to allow estimation of a measure of energy end-use intensity for lighting. Collection of accurate information on lighting facilities and use in the home poses some challenging cognitive problems and could require substantial additional time in the interview. In 1992, contractor staff conducted a series of in-depth interviews, which were recorded on both audio and video tape, to explore various means of asking for the desired data. For the first set of interviews, seven members of the contractor staff served as respondents. Each was interviewed twice, with a week intervening. For the second set of interviews, 5 non-staff respondents were recruited. Two of them were a husband and wife who were interviewed separately and then together, to attempt to reconcile differences in their reports. All respondents were asked to report on use for the day preceding the interview. They were asked to report on all lights (fixtures controlled by one switch, as opposed to individual bulbs) that had been used for at least 15 minutes and to report how long each one had been turned on (Daniels 1992).

There were fairly substantial week-to-week percentage differences in reported use for the respondents who were interviewed twice. Some of these differences could have been real; some could have been caused by response variability. Nevertheless, the relative stability of rankings of the seven respondents in terms of total hours of use suggested that the data could have been used with fair reliability to classify households as high, medium, and low lighting users. The couple who were interviewed separately and then together had significant differences, which could not be completely reconciled in the joint interview. In general, most respondents found it difficult to estimate the time of use for each light, and it was not obvious that allowing them to report time of use in broad categories made it any easier than asking for an answer in hours or fractions thereof. The room-by-room inventory approach used in the pretest was estimated to require an average of at least eight minutes per respondent, even without additional probing that might be necessary to obtain reasonably accurate responses.

On the basis of this test and other considerations, two sets of questions on lighting were included in the 1993 Household Survey questionnaire. A short module, asked for all households, requested them to report the total number of lights and the number of fluorescent lights used on a typical November weekday for: more than 12 hours per day, between 4 and 12 hours per day, and between 1 and 4 hours per day. A supplementary set of questions was asked only for a subsample of 474 households. It called for a more detailed accounting covering each indoor light used in the home, on a typical November weekday, for at least 15 minutes. Respondents were given options on whether to report lights by room, activity, or time-of-day usage and were allowed to report the time used for each light in actual number of hours or in class intervals based on number of hours.

A comparative evaluation of the quality of lighting data obtained from these two modules will be undertaken. For the subsample of households that responded to the supplemental module, it will be possible to analyze any differences in their responses to the two modules.

In 1992, at the request of a member of the American Statistical Association's Energy Statistics Committee, a survey researcher reviewed the cognitive features of the 1990 RECS Household Survey questionnaire, with special emphasis on questions requiring respondent recall (Biemer 1992). A major finding of this review was that many of the questions in this category did not include a reference period as a basis for the response or used a vague, unbounded or ill-defined reference period. In response to this analysis, all recall questions planned for use in the 1993 RECS were carefully reviewed and several changes were made. For example, the 1990 question:

H-10 About how many deliveries [of LPG] does your household usually get in a year?

was changed for 1993 to:

J-14 About how many deliveries did your household get in the past 12 months?

Other Questionnaire Effects: Conservation Improvements

In the 1984 RECS, the Household Survey questionnaire included several questions about conservation improvements that had been made to the housing unit since September 1, 1982, such as storm doors and windows, additional insulation, caulking, weather stripping, and heating system improvements. For all such improvements, respondents were asked to report the month and year in which they were installed, which could have been any month between September 1982 and the month of the Household Survey interview, generally in the Fall of 1984.

Following these questions there was a general question asking whether any improvements of this kind had been added or installed and paid for during calendar year 1983. This question was designed to identify households eligible for the energy tax credit that was permitted on their Federal income tax returns for that year. Households answering "yes" to this question were asked whether or not they had actually taken the energy tax credit on their returns.

A comparison of the general question about improvements in 1983 with responses to the earlier questions about specific improvements that were eligible for the tax credit showed that, of the 1,328 households that answered "yes to the general question, 567 (42.7 percent) had not reported any specific improvements as being added or installed in 1983. This could have been legitimate in some cases; the specific questions allowed for reporting of only a single month and year. If caulking or weather stripping, for example, had been added at two different times during the approximately 2-year reference period, only one of these would have been reported. Also, some improvements might have been installed in 1983 but not paid for until 1984. Nevertheless, the high incidence of apparent inconsistency suggests that responses to the questions about specific improvements or the general question, or perhaps both, were subject to substantial bias or response variability. The general question on improvements in 1983 was complex, with several

subquestions imbedded in it, and interviewers reported difficulties in administering it (EIA 1987a, pp.112-113).

Interviewer Training

Typically, about 300 interviewers have conducted personal interviews for each of the RECS Household Surveys. Partly because of the longer interval between surveys in recent years, interviewer turnover between surveys has increased. In the 1987 RECS, 57 percent of the interviews were conducted by interviewers who had not participated in the 1984 survey. In the 1990 RECS, 60 percent of the interviews were done by interviewers with no experience in the 1987 RECS. This turnover, along with numerous changes in questionnaire content, means that the effectiveness of training and supervision of interviewers can be an important element in determining the quality of the survey results.

For the 1980 RECS, interviewers were trained for two days in small group sessions for 10 to 20 persons each. For the 1981, 1982, and 1984 RECS, new interviewers were trained in the same way as in 1980, but the training of experienced interviewers consisted of completion of self-study materials and practice interviews. For the 1987 RECS, all interviewers were trained in four large group sessions, each lasting two and one-half days.

The cost of training was becoming a substantial element in the total budget for RECS, and means were sought to reduce training costs for the 1990 RECS. The solution adopted was to use home-study materials for all interviewers, including a videotaped presentation in several sections, an interviewer's manual, a quiz and practice interviews, the last two of these to be sent in to the survey contractor for evaluation. Use of these training methods resulted in a significant reduction in training costs, estimated at about 30 percent in constant dollars. Part of the savings were applied to more extensive office reviews of practice and initial interviews, followed by contacts with all interviewers to provide feedback from these reviews (Leach 1991).

Given the rather substantial change in training procedures that was introduced in 1990, it was considered important to try to evaluate the relative effectiveness of the old and new procedures. Two methods were used: administration of an evaluation questionnaire to the interviewers and comparative analyses of the extent of edit changes and imputation in the data processing.

Completed evaluation questionnaires were obtained from 257 of about 290 interviewers who completed the training for the 1990 RECS. Most of those responding had had prior exposure to both large and small-group training sessions for RECS or other surveys. When asked to compare the effectiveness of and their preferences for four kinds of training--small group in-person, large group in-person, self-study only, and self-study plus video--a large proportion, 78 percent, thought small group in-person training was the most effective and 60 percent identified it as their first preference. Self-study plus video was a distant second for both effectiveness (15 percent) and preference (22 percent). When first and second ratings were pooled, self-study plus video received favorable ratings for both effectiveness and preference from 55 percent of the interviewers.

The interviewers' overwhelming preference for in-person training may be influenced in part by factors other than its effectiveness. For most interviewers, these training sessions provide an opportunity to travel away from their home base and to meet fellow interviewers with whom they would otherwise not have much contact.

Interviewers were asked how well they felt they understood each of 11 topics covered in the video presentation. Most were rated "well understood" by at least 80 percent of the interviewers. Exceptions included:

- Fuels and equipment (61 percent understood very well). One written comment suggested that the treatment of combination equipment was inadequate.

- Measurement of total and heated floorspace. This has always been one of the most difficult aspects of the survey for interviewers. Some of the written comments praised this section of the videotape, indicating it was more realistic than what could have been done in a classroom training session.

- Recording of air-conditioner nameplate data. This was a new requirement for the 1990 RECS, so it was difficult to anticipate what kinds of problems might arise and discuss them in the training.

In response to a question about the degree of difficulty of the training exercises and final quiz, nearly all of the interviewers considered them "about right." However, many interviewers did quite badly on the exercises and quiz, especially on topics such as what to measure and what not to measure, who are eligible respondents, and households versus group quarters. These same topics had caused many problems in training and in actual field work in prior survey years.

In addition to finding out how interviewers reacted to the new training procedures and what improvements they had to suggest, it was felt important to seek an objective measure of the effects of the new procedures on actual interviewer performance. The method of analysis adopted was to compare the levels of edit and imputation changes for 14 RECS variables for the 1987 and 1990 RECS in total and for experienced and inexperienced interviewers in each survey year. An experienced interviewer was defined as one who had participated in RECS in the immediately preceding survey year. Inclusion of interviewer identifiers on the RECS file of individual household data for each survey year made it possible to distinguish the work of experienced and inexperienced interviewers. The 14 variables were chosen from among those that were included in the same form in both survey years, had extensive edit checks, and had required the most imputation (Response Analysis Corporation 1991).

The indicator of interviewer performance used in the analysis, admittedly an indirect measure, was the proportion of sample households for which changes had been made in each of the 14 variables following data entry, as the result of editing and imputation procedures. Changes were detected by comparing initial and final values for each variable; intermediate changes were not taken into account. Interviewer errors were only one source of such changes; they could also result from respondent and data entry errors. Moreover, some interviewer errors were detected and corrected in manual reviews prior to data entry.

Table 5.9 shows, for each of the 14 variables selected, the percent of sample households with changes for each survey year for experienced and inexperienced interviewers. Overall, the data do not reveal a substantial difference between the change indicators for the 1987 and 1990 surveys. For a few variables there were fairly large differences in the proportion of changes between the two years: same heating fuel as prior survey (down 5 percent in 1990), year house built (down 6 percent), and presence of a basement (up 4 percent). Except for the possibility of insufficient attention in the 1990 training to the item on basements, more detailed analyses did not reveal any obvious reasons for these changes.

As shown in Table 5.9, the mean proportion of changes for the 14 variables was slightly higher for inexperienced interviewers in both survey years. This finding does not necessarily prove that their performance was not as good as that of experienced interviewers. An alternative explanation would be that interviewer turnover is larger in areas like central cities, where there is likely to be a higher incidence of respondent error and item nonresponse requiring imputation. Table 5.10 provides data relevant to this hypothesis. As shown in the last column of that table, there is a clear association between housing characteristics and the extent of changes subsequent to data entry. Households in center cities, those living in apartment buildings, and those who were not owners had the largest number of changes. These explanatory variables are correlated, and some of the differences for renters of apartments may be explained by changes made to variables related to heating fuel and equipment based on responses to the Rental Agent Survey.

However, the data in Table 5.10 on percent of interviews completed by experienced interviewers, by housing type, provide only moderate support to the supposition that there are higher proportions of inexperienced interviewers conducting interviews with the types of households for which changes subsequent to data entry are most frequent. The proportions of experienced interviewers are nearly the same for owner and non-owner occupied housing units. They were slightly lower for apartments than for single-family units and lower for households living in metropolitan areas.

Taken overall, the results of the interviewer questionnaire and the analysis of processing changes did not demonstrate any clear or substantial differences in effectiveness between the 1987 and 1990 RECS training. However, two features of the 1993 RECS seemed to favor the use of in-person over self-study training for that survey: first, the inclusion of several new items in the Household Survey questionnaire and, second, as a result of the 1993 sample revision, 22 out of 116 primary sampling units had not previously been included in RECS, so the proportion of interviewers lacking previous RECS experience was likely to be higher than usual. Consequently, two and one-half days in-person training sessions were held for all interviewers at four sites, followed by a small make-up session and some telephone training for replacement interviewers.

Table 5.9. Percent of Household Survey Interviews with Imputation or Edit Changes Subsequent to Data Entry, by Interviewer Experience[a]: 1987 vs. 1990

RECS 1990 Question	Total		1987 RECS		1990 RECS	
	1987	1990	Exp.	Not Exp.	Exp.	Not Exp.
B-2 Main Heating Fuel	3	2	3	3	2	2
B-9 Main Heating Equipment	6	5	5	6	4	5
C-3 Water Heating Fuel	5	5	5	4	4	5
K-10 Income	11	14	10	13	16	13
B-3 Same Heating Fuel as 1987[b]	10	5	9	10	6	5
I-1 Budget Plan for Main Fuel	2	4	2	2	3	4
A-6 Year House Built	15	9	14	16	8	9
P-11 Has Basement	1	5	1	1	3	6
L-12 Public Housing	1	3	1	1	4	3
K-7 Race	*	2	*	*	2	2
K-1 Non-Householder Age	1	1	1	1	1	1
P-12 Amount of Basement Heated	3	5	3	3	5	5
K-1 Householder Age	1	1	1	1	1	2
K-6 Marital Status	1	2	1	1	1	2
Mean: 14 Items	4.3	4.5	4.0	4.4	4.3	4.6
Number of Interviews	5,856	4,828	2,530	3,326	1,965	2,873

* = Rounds to less than 0.5 percent.

[a]Interviewers were counted as experienced if they had worked on the immediately preceding RECS.

[b]The 1987 question was "Same Heating Fuel as 1984."

Note: Exp. = Experienced Not Exp. = Not Experienced

Source: Response Analysis Corporation, *Quality Assessment of Videotape Training: Conclusions and Recommendations* (September 1991).

Table 5.10. Interviewer Experience and Extent of Edit and Imputation Changes, by Type of Household: 1990 RECS

Household Category	Interviews		Mean Proportion of Edit and Imputation Changes: 14 Variables[b]
	Number	Percent by Experienced Interviewers[a]	
All households	4,828	40.5	4.5
Metropolitan Status			
Center City	1,543	38.1	5.8
Other Metropolitan	1,994	34.7	4.4
Nonmetropolitan	1,291	52.4	3.8
Housing Type			
Single-Family			
• Detached or Mobile Home	3,346	40.9	3.2
• Attached	289	42.6	5.4
Apartment	1,193	38.7	7.9
Home Ownership			
Owner	3,201	40.7	3.1
Other	1,627	40.1	7.6

[a]Interviewers were counted as experienced if they had worked on the immediately preceding RECS.

[b]Mean, for 14 selected variables, of the proportion of households for which the value was changed by edit or imputation subsequent to data entry.

Source: Response Analysis Corporation, *Quality Assessment of Videotape Training: Conclusions and Recommendations* (September 1991).

6. Data Processing and Imputation

This and the next chapter of the Quality Profile cover the processing operations that follow the collection of RECS data from households, rental agents, and energy suppliers. This chapter is about the quality aspects of initial manual reviews of questionnaires, coding, data entry, computer-assisted edits, and imputation. Chapter 7 covers the weighting procedures used to develop sample estimates, the model-based procedures for allocating consumption of each fuel to specific end uses, and the estimation of sampling errors.

The primary outputs of the processing operations discussed in this chapter are the data files delivered by the survey contractor to EIA. These data files are of three kinds:

1. *Household Files*, which contain all information collected for every *interviewed* household

2. The *HURS* (Household Unit Record Sheet) *File*, which contains information on type of living quarters and occupancy, sampling information, and data collection outcomes for *all* sample housing units, including those that did not respond to the survey

3. *Utility Bill Data Files*, which contain individual household billing data, annualized consumption and expenditures data, and associated weather data for each of the five major fuels.

Unlike the Household and Utility Bill Files, the HURS File does not include the weights that would be needed to produce estimated totals and weighted unit response rates.

Prior to delivering these files to EIA, the survey contractor removes specific identifiers and related information that would make it possible for EIA to identify individual respondents. For the same reason, individual bill data are inoculated with random errors.

An overview of the processing operations covered in this chapter was provided in the first section of Chapter 2, under "Data Processing and Imputation." As noted there, the procedures are intricate and detailed, consisting of a large number of distinct processing steps. Initially, data from each of the six major sources--Household Survey interview questionnaires, Household Survey telephone and mail questionnaires, Housing Unit Record Sheets, Rental Agent Survey questionnaires, Supplier Survey questionnaires and billing records, and weather data--are processed independently to make them computer-readable, eliminate as many errors as possible, and impute values for some of the items that are missing or incorrect (other kinds of imputation require matching with records from other sources to provide donor information). Then data from the six sources are compared and combined in various ways to produce the three major output files.

This chapter describes the general structure of the data processing operations, with emphasis on features that affect the quality of the final output. Quantitative data are presented when available. Most of the material focuses on the procedures used in the 1990 and 1993 RECS; however, some relevant data from earlier surveys are presented and significant procedural changes over the

history of RECS are discussed. Readers who want a step-by-step detailed description of all processing activities should consult the *Data Editing and Manipulation Procedures Manuals* for the 1990 and 1993 RECS (Response Analysis Corporation 1992b, 1995c).

Several steps in data processing, especially the manual and computer edits, are designed to detect and, insofar as possible, to reduce or eliminate errors. The purpose of imputation is to replace missing entries or those believed to be in error with values that are closer, at least on the average, to the correct ones. However, errors can also be introduced at each stage of data processing. The overall processing system is designed to optimize the quality of the final product. The underlying philosophy guiding this effort has been that because the RECS sample of households is relatively small, the use of substantial manual and computer resources to make the final files as "clean" as possible is justified.

Data Processing Other than Imputation

In order to establish a context for discussing the quality aspects of data processing, this section begins with a general overview of the nature and flow of the main processing steps. The overview is followed by a review of the quality-related features of manual operations and computer-assisted operations. Finally, some results of a special analysis of changes resulting from computer-assisted data processing operations in the 1984 RECS are presented, along with some summary data for these kinds of changes in the 1987 and 1990 RECS.

Structure of the Processing Operations

Figures 6.1 and 6.2 show the main processing steps, including imputation, for the RECS data sources other than the weather data obtained from the National Oceanic and Atmospheric Administration (NOAA). As noted above, questionnaires and other inputs from each of these five sources are first processed independently of each other. These processing steps are shown for each data source in Figure 6.1. For the sake of completeness, imputation, which will be discussed in the following subsection of this chapter, is included in the figure. The first three steps--receipt and check-in, manual coding and editing, and data entry--are primarily clerical operations. The edits listed under Step 4 are computer-assisted. Computer-generated listings of suspect variable values and associated information are reviewed manually by editors and, when necessary, by supervisors or specialists, to determine whether and how to change the suspect values. In some instances, respondents are recontacted by telephone.

Steps 5 and 6 are computer procedures that apply only to the Household Survey data. Computer-generated updates are used for Household Survey interview questionnaires to correct certain common interviewer or respondent errors without manual examination of the questionnaire. These updates are used only in a few instances where a preliminary review has shown that a particular kind of correction is virtually certain to be appropriate for all households that have a specified combination of entries.

Figure 6.1. Steps in Internal Processing of Data from Each Major Source[a]

Processing Step	Data Source				
	Household Survey			Rental Agent Survey	Supplier Survey
	Interview	Phone, Mail	HURS[b]		
1. Receipt, check-in	X	X	X	X	X
2. Manual editing and coding	X	X	X	X	X
3. Data entry	X	X	X	X	X
4. Computer-assisted edits					
a. Range checks	X	X	X	X	X
b. Internal consistency checks	X	X	X	X	X
c. Special reports	X				
5. Computer updates	X				
6. Same-source imputations	X	X			

[a]Processing of weather data, which follows a different pattern, is excluded from this exhibit.
[b]Housing Unit Record Sheet.
Source: Response Analysis Corporation, *1990 RECS Data Editing & Manipulation Procedures Manual* (September 1992).

Same-source imputations for the Household Survey questionnaires (step 6) are those which do not require inputs from other sources. They may be based on the values of other variables for the same household or on data for other sample households.

The initial processing of weather data, not shown in Figure 6.1, consists of the extraction and manipulation of temperature data from data tapes obtained from NOAA. The first step is to associate a weather station with each secondary sampling unit (SSU, see Chapter 2, Figure 2.3), taking into account physical proximity and the completeness of data available for the stations. For each weather station associated with one or more SSU's, temperature data are extracted from the NOAA tapes, missing data are imputed, and long-term and reference year values of heating and cooling degree-days are developed.

Figure 6.2 lists the principal processing operations that require comparing or merging data from more than one source:

A. The annualization of Supplier Survey bill data for individual housing units starts with data reported by suppliers for billing periods, most commonly months, and uses these data to develop estimates of total consumption and expenditures for each fuel for the 12-month survey reference period. As part of this process, which is described later in this chapter, the degree-day information developed from the NOAA tapes is used to adjust data for billing periods that overlap the start or end of the calendar-year reference period and to impute data for parts of the year for which no billing data are available.

Figure 6.2. Processing Operations Involving More than One Data Source

A.	Annualizing billing data
B.	Comparison of Household and Rental Agent Survey responses
C.	Imputation of missing data for telephone and mail households
D.	Model-based outlier checks of Supplier Survey data
E.	Creation of output files

Source: Response Analysis Corporation, *1990 RECS Data Editing & Manipulation Procedures Manual* (September 1992).

B. As discussed in previous chapters, Rental Agent Survey responses for such variables as main heating and cooling equipment and fuels are usually considered to be more reliable than those of the occupants of rental units in multiunit structures and are substituted for the latter when disagreements exist.

C. The imputation of data items not included on the telephone and mail versions of the Household Survey questionnaire requires matching the two sets of sample households--those that responded by interview and those that responded by telephone or mail--on variables that are common to all questionnaire versions to find interview households that are suitable to serve as "donors" for imputation.

D. The model-based outlier checks of estimates of annual consumption, which were discussed in Chapter 5, require merging of Household and Supplier Survey data.

E. The ultimate goal of all processing steps up to this point is the creation of the output files, especially the Household Files, which, in their final form, include data from all six sources.

Manual Operations: Quality Considerations

Questionnaires from each of the three surveys, following receipt and control operations, are subjected to an extensive manual combined editing and coding operation. For each type of questionnaire, the first step is to verify the accuracy of the basic identification information. Then the editors check each questionnaire item for completeness and logical consistency with responses to closely related items. In preparation for data entry, they enter codes next to card column numbers on the questionnaire.

The extensive and detailed nature of the editing/coding operations may be seen by examining the detailed instructions for processing interview questionnaires from the 1993 Household Survey (EIA 1995b). In addition to the basic tasks of consistency checking and coding, editors are instructed to:

- Convert numerical entries for such items as amount of wood burned to the desired dimensions, rounding when necessary.

- Keep lists of write-in responses to "other" categories for several questions which have this response option. These lists are used later in processing to recode some of the "other" responses to other existing or newly created categories.

- Refer unusual types of entries and other problems to a special coding and editing section or to designated technicians.

Editing and coding of the housing unit measurement section of the Household Survey questionnaire require certain particularly complex tasks, such as dealing with measurements for floors having shapes that are not simple rectangles.

There is no formal verification system for the manual editing and coding operations. The subsequent computer-assisted edits provide an opportunity to detect some of the errors that may have been overlooked or introduced by the editors. One example of findings from such checks is provided in the next subsection.

Data keying is performed by a separate EIA contractor. Batch tapes of keyed data are transmitted to EIA and loaded to its main computer, where they are used by Response Analysis Corporation, the main contractor, to create unedited data files for each source and perform the computer edits. For the 1981, 1982, and 1984 surveys, key Household Survey questionnaire items were 100-percent verified and the remaining items verified for a 25-percent sample of households. However, a review of the changes that had been made during processing operations for the 1984 Household and Supplier Surveys showed that keying errors were leading to substantial numbers of computer edit rejects (Jabine 1987). The costs of processing these rejects were deemed to exceed the savings from sample verification of data entry and there was also no guarantee that the computer edits and special reports would detect all keying errors. Consequently, beginning with the 1987 RECS, all keying has been 100-percent verified.

Computer-Assisted Edits

Figure 6.3 shows the different kinds of computer-assisted edit checks used in RECS. *Range checks* are applied to values for individual variables from all of the survey components of RECS. Simple range checks ensure that no illegal or impossible variable values are included in the final records. Outlier checks identify, for clerical review, values for continuous variables that may be correct but are unusually high or low for that variable.

Internal consistency checks are also applied to data from all of the RECS survey components. Most commonly, these checks examine relationships of responses for different variables for the same household. Some checks involve comparisons of data for the same household from the current and prior rounds of RECS. For example, housing unit area measurements may be compared for the current and immediately preceding round. Such checks can be used only in those survey years for which the sample includes a longitudinal component and only for the housing units that were in the sample both times.

Figure 6.3. Computer-Assisted Edits Used in RECS

A. Range checks
   1. Simple range checks
   2. Outlier checks

B. Consistency checks
   1. Internal to source
      a. Same unit, same survey
      b. Same unit, prior survey
      c. Different units, same survey
   2. Across sources

C. Special reports

Source: Response Analysis Corporation, *1990 RECS Data Editing & Manipulation Procedures Manual* (September 1992).

Another kind of internal consistency check compares data for different households from the same survey component. For example, during the processing of the Supplier Survey records, values for each household are compared with those for some of the other sample households. Some of these comparisons are with all households using the same fuel and others are with all households obtaining that fuel from the same company. In these between household checks, large differences are treated as outliers and are reviewed to determine whether changes are necessary.

The primary consistency checks across survey components are items B and D shown in Figure 6.2. Item B, Comparison of Household and Rental Agent Survey responses, is the final step in processing data collected in the Rental Agent Survey. For the 1990 RECS, Memo #951 provides instructions for resolving the differences appearing on computer-generated listings. In addition to reviewing the Household and Rental Agent questionnaires for units with differences, editors are instructed to review other sources of information, including Supplier Survey data and questionnaires for neighboring households in the same building.

Item D, Model-Based outlier checks of Supplier Survey data, requires the use of Household and Supplier Survey data for the same units. A nonlinear model, with parameter values based on the previous survey, uses data on housing and household characteristics from the Household Survey to predict fuel consumption for each unit. The predicted values are compared with annualized estimates of consumption based on bill data obtained in the Supplier Survey. In the 1990 RECS, whenever the model-based estimate was more than three times or less than one-third of the annualized consumption estimate from the bill records, the data from the two sources were checked.

For the Household Survey records, in addition to the range and consistency checks, a series of computer-generated *special reports* lists information for households with unusual responses or combinations of responses on several different topics. Figure 6.4 lists the topics for which special reports were prepared in the 1990 RECS. Each of these reports identifies the households that qualify for inclusion and gives their values for variables that are relevant to the topic of the report.

Figure 6.4. 1990 RECS: List of Topics for Special Reports

1.	Households with Wood-burning Inconsistencies
2.	Households Not Using a Heating Fuel
3.	Households Reporting That Their Main Heating Fuel Changed in November 1987 or Later
4.	Missing Supplemental Fuels and Equipment
5.	Incompatible Fuel/Equipment Combinations
6.	Family Members' Ages and Relationships
7.	Central Fuel Inconsistencies
8.	Inconsistencies in the Number of Rooms Cooled
9.	Inconsistencies with the Family Grid
10.	Inconsistencies with the Foldout Page
11.	Households That Use Heat Pumps
12.	Lighting
13.	Households with No Windows, Doors, or Electricity
14.	Swamp Coolers
15.	Households Using Gas Air-Conditioners
16.	Air-Conditioner Section Skip Patterns
17.	Unusable Measurements
18.	Households with Unknown Heating Equipment and Heating Fuel
19.	Households That Selected Underground Gas as an Alternative Heating Fuel
20.	Households with Marital Inconsistencies
21.	Basement Insulation vs. Basement Heating
22.	Recoded Variables
23.	Changes Made to the Mail Questionnaire Donor Selection Variables
24.	Various Miscellaneous Checks
25.	Households That Used the Answer "Other"

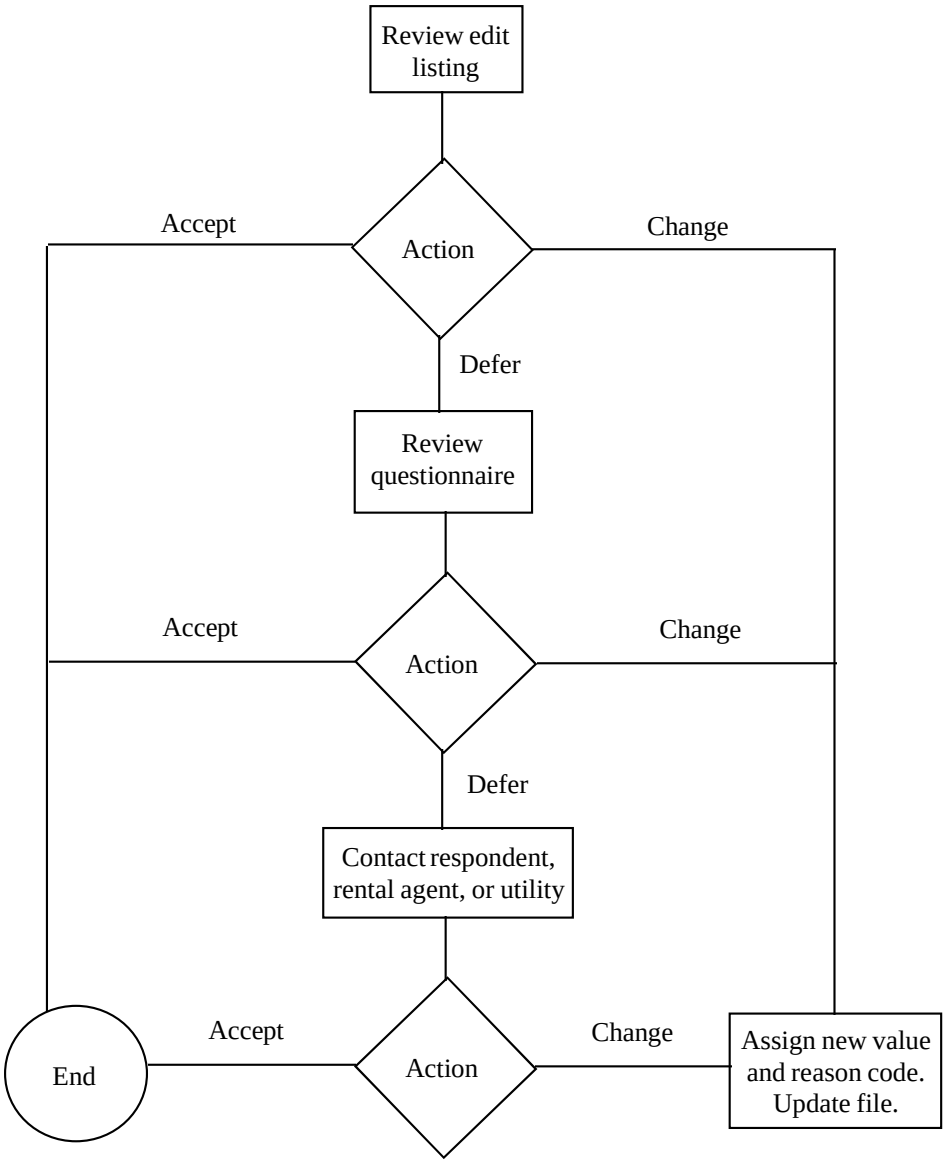
Source: Response Analysis Corporation, *1990 RECS Data Editing & Manipulation Procedures Manual* (September 1992).

Processing Computer Edit Outputs

A more or less standard approach, shown in Figure 6.5, was used to resolve rejects and questionable values included in the computer-generated outputs of range checks, consistency checks, and special reports. In some instances, editors can decide whether to accept or change a response on the basis of a review of the information in the computer listing and the original questionnaire. Failing this, decisions require examination of other relevant questionnaires. At this stage, an editor may be able to determine, for example, that a coding or keying error led to the inclusion of an incorrect value in the computer record. As a final resort, in a small proportion of cases, editors may attempt to contact a household, rental agent or supplier. In order to maintain an audit trail, whenever an editor decides to change a value in the record, information about the nature of the change and the basis for making it is recorded in an archive file.

For the 1990 RECS, instructions for processing each of the special reports are contained in a series of RECS-90 Memos. Most of the memoranda consist of general rules for making changes or for referring certain types of problems to a supervisor. In a few instances, however, the memoranda include quantitative information about the number of households included in a report and the manner in which apparent inconsistencies and other possible errors were dealt with.

Figure 6.5. Processing Computer Edit Outputs



Source: Response Analysis Corporation, 1992b.

A special report was run to examine consistency and skip checks for the "family information grid" (questions K-1 and K-3). This report generated lists of households with 14 different kinds of potential errors--for example, households reporting more than one spouse or households with inconsistencies between the number of persons listed in the chart and the variable for the number of persons in the household. RECS-90 Memo #602 provides information about the disposition of the 292 households that appeared on the 14 listings. Changes were made for 227 (78 percent) of these households, in most instances because the editor determined that there had been coding or keying errors. The largest number of changes, 127, were made in cases where the family grid showed the householder to be married but a separate question on marital status (K-6) indicated otherwise. Most of the discrepancies were the result of coding errors: a person identified as partner or fiancee in the family grid was incorrectly coded as spouse. The relevant questionnaire items were modified in the 1993 RECS in an effort to reduce the frequency of this kind of error.

Another set of special reports dealt with households that reported using gas air conditioners. Because the use of gas as a fuel for air conditioning is rare, all of the 117 sample households that reported it were listed and their data were reviewed. Some cases were resolved by reviewing data for the same household from the 1987 RECS, where similar checks had been undertaken, or from the Rental Agent or Supplier Surveys. Respondent contacts were attempted with 53 of the 117 households. As a result of all of these efforts, the weighted estimate of the number of households using gas air-conditioners was reduced from 1.3 million to 0.4 million (Response Analysis Corporation 1992b, RECS-90 Memo #306 and EIA 1993a, p. 150).

The percentages of sample households for which telephone contacts have been successfully completed at *any* stage of the manual and computer edit procedures have been as follows (*Consumption and Expenditure* reports for years shown):

Survey Year	Percent Contacted
1981	14
1982	10
1984	6
1987	1
1990	4

The gradual reduction in telephone contacts with Household Survey respondents during data processing, especially from 1987 on, reflected concerns that such contacts may have been hurting response rates for the Residential Transportation Energy Consumption Survey and for households in the longitudinal panel for RECS.

Analysis of Processing Changes in the 1984 RECS

As noted above, information about changes made to individual records in the initial unedited Household and Supplier data files is systematically maintained in archival files. The archival files for the 1984 RECS provided the basis for a detailed analysis of processing changes (Jabine 1987). The initial portion of the study was based on tables that had been produced as a matter

of routine, showing the number of changes and their distribution by "reason" for each of 562 Household Survey and 59 Supplier Survey variables. Summary findings for all variables combined for each survey are shown in Table 6.1 (the Billing Files contained Supplier Survey data for the five fuels).

Table 6.1. 1984 RECS: Changes to the Household and Billing Files, by Reason

Reason	Changes to Household File		Changes to Billing Files	
	Number	Percent	Number	Percent
Keying error[a]	1,868	9.1	2,066[a]	50.0[a]
Coding error[a]	3,699	18.0		
Clerical error (prior to coding)	NA	NA	374	9.0
Interviewer error	1,118	5.5	NA	NA
Respondent error	236	1.2	122	3.0
Interviewer or respondent error	422	2.1	NA	NA
Data processing error (after keying)	202	1.0	1	*
Phone call to respondent household	514	2.5	20	0.5
Phone call to utility/supplier	256	1.3	496	12.0
Other phone call or information	143	0.7	14	0.3
Rental agent (master meter) information	1,251	6.1	--	--
Kerosene survey information	NA	NA	--	--
Editor's judgement	9,807	47.8	1,016	24.6
Additional information from questionnaire	545	2.7	25	0.6
None of the above	411	2.0	--	--
Total	20,472	100.0	4,134	100.0

[a]Changes due to keying errors that could not be distinguished from changes due to coding errors.
-- = None in this category.
NA = Not Applicable.
* = Less than 0.05 percent
Source: Jabine, *Review of Computer Edit & Update Performance Statistics for the RECS, Final Report* (December 1987).

Knowledge of certain limitations and other features of the data shown in Table 6.1 is necessary for an informed interpretation:

- When review of a computer edit reject led to changes in a string of two or more consecutive variables in the same portion of the record, an archival file record was created only for the first variable. Consequently, the numbers in the table are undercounts of the total changes made.

- Not all of the changes made during processing were made because of errors; some of them were built into the processing system. About one-tenth of the changes to the Household File were recodes for questions that included an open-end "other" category. Recodes of the "other" responses were necessary because the final code structures could not be established until after the coding and keying operations had been completed and an analysis of the nature of the open-end responses was possible. Another one-tenth resulted from a special computer program that was used to insert leading zeros in square footage measurement variables for those households in which data from the prior (1982) survey were being used.

- The reason codes that were entered in the archival files are a mixture of two different dimensions: the source of the (presumably) incorrect value that was changed and the source of the information that allowed the editor to determine the (presumably) correct value. The reason "editor's judgment," which was assigned to nearly half of the Household File changes and one-quarter of the Billing File changes, does not provide much useful information in isolation, but its meaning becomes clearer when it is associated with a code showing the specific processing step in which the change was made (see Table 6.2).

Notwithstanding these limitations, it was clear from the data that keying and coding errors were a major source of computer edit rejects, accounting for more than one-fourth of the changes to the Household File and one-half of the changes to the Billing Files. The 193 Household Survey "key" variables for which data entry was verified 100 percent had an average of only 0.44 changes, but the remaining 369 variables that were subject to sample verification averaged 4.83 changes per variable. These findings prompted the decision to revert to 100 percent verification of data entry for the 1987 and subsequent surveys. Use of 100 percent verification has led to a substantial reduction in keying errors, although not necessarily to their complete elimination.

Overall, changes to the Household File averaged 3.6 per household and changes to the Billing Files averaged 0.5 per household (a household was counted once for each fuel for which Supplier Survey data for that household were available). Analysis of variables with large numbers of changes showed that 25 of the 562 Household Survey variables accounted for 42 percent of all changes. Most of these 25 variables were located in two areas of the questionnaire: the portion dealing with main and secondary heating equipment (11 variables) and the portion in which area measurements of the housing unit were recorded (6 variables). For the Billing Files, 10 of 59 variables accounted for 81 percent of the changes. Many of these changes were made to beginning dates for billing periods.

In the second part of the 1984 study of changes, codes were added to the archival records for the Household File to show the stage of data processing at which each change was made and the identification number of the interviewer for the household. Table 6.2 shows a distribution of the Household File changes, by reason, for each step of the processing operations. The nature of most of the processes shown in the table is self-evident. The recode changes were made primarily for two reasons: to assign initial "other" responses to the final set of categories adopted for an item and to make the area measurement data carried over from the 1982 RECS consistent with the 1984 format for these variables. Initial imputations of missing data were not included as changes in the archive files. The "imputation-related" changes shown in Table 6.2 are changes that were necessary to make imputed variables consistent with related variables. The table updates reflect changes that were made to eliminate anomalies detected in preliminary tabulations. Utility and final updates shown in the last two columns of the table were made at the final stages of computer processing for a variety of reasons.

Many features of the processing operations are clear from Table 6.2, for example:

- Nearly all of the changes made to items rejected by range checks were to correct coding and keying errors;

- Consistency checks frequently required contacts with respondents and other sources of information or exercise of an editor’s judgment to determine the correct values;

- Special reports followed a pattern similar to that of consistency checks, but fewer changes were made to correct coding and keying errors, presumably because most of them had already been detected by range and consistency checks.

The study report also includes analyses of changes by process for individual variables with large numbers of changes. Analyses of changes attributed to interviewer errors were used to guide the training of 1984 RECS interviewers who were scheduled to serve as interviewers or supervisors in the 1987 RECS.

The report included some recommendations for refinements to the archive files: (1) include a separate record for each variable changed; (2) include a code to show at what stage of processing the change was made; (3) include both old and new values for each change; and (4) replace the 1984 RECS reason code with two codes, one showing the source of information on which the decision to change the value was based and one showing the probable source of the (presumed) incorrect value.

Archive files have been created for all subsequent survey years, but the "reason" codes were replaced by a set of "level-of-effort" codes, putting less emphasis on the source of the error and more on the level of effort required to correct it. Each change was assigned a single code reflecting the greatest level of effort needed to reach a decision. Thus, for example, a change based both on examination of questionnaires and recontact with respondents would be assigned a code reflecting the type(s) of respondent contacts. All decisions about changes required reviewing both edit outputs and the original questionnaire on which the edited record was based. Hence, codes for examination of questionnaires were assigned only when a decision required

Table 6.2. 1984 RECS: Percent Distribution of Changes for Each Process, by Reason

Reason	Process									
	Range Checks	Consistency Checks		Recodes	Special Reports	Imputation Related	Rental Agent	Table Updates	Utility Updates	Final Updates
		Round 1	Round 2							
Interviewer Error	1.0	10.1	16.3	0.1	6.7	0.2	--	--	--	--
Respondent Error	--	0.1	2.3	--	6.5	--	--	--	--	--
Interviewer or Respondent Error	0.1	4.4	4.1	--	3.1	0.9	--	--	--	--
Coding Error	38.6	50.8	23.5	0.1	6.3	0.2	--	--	--	0.7
Keying Error	58.5	14.5	7.5	*	0.4	0.2	--	--	--	--
Data Processing Error	0.4	0.1	5.1	--	0.9	0.1	--	--	--	9.2
Respondent Call	--	3.4	5.2	0.1	4.5	0.5	--	--	61.3	0.7
Utility Call	--	0.2	1.2	--	6.6	1.0	--	--	1.6	0.7
Other Call	--	0.5	0.9	--	2.9	--	--	--	19.4	--
Rental Agent Information	0.2	0.7	9.9	--	2.4	5.9	100.0	--	3.2	4.2
Editor's Judgement	1.1	14.4	22.4	87.2	46.8	90.8	--	100.0	14.5	84.5
Other Questionnaire Information	--	*	--	12.5	--	0.2	--	--	--	--
Other	0.1	0.8	1.6	--	12.9	--	--	--	--	--
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Number of Changes	1,727	4,234	2,998	4,370	2,539	3,664	634	40	124	142

* = Less than 0.05 percent.
-- = None in this category.
Source: Jabine, *Review of Computer Edit & Update Performance Statistics for the RECS, Final Report* (December 1987).

Table 6.3. Changes to the Household File by Source of Information or by Reason: 1987 and 1990 RECS

Source of Information or Reason	Number of Changes		Changes per 100 Households[a]		
	1987	1990	1987	1990	Change
Contacts with:					
Household	129	37	2.2	0.8	-1.4
Fuel Supplier	81	6	1.4	0.1	-1.3
Rental Agent or Interviewer	67	196	1.1	4.1	3.0
Multiple Sources	22	3	0.4	0.1	-0.3
Examination of:					
Supplier Data	86	202	1.5	4.2	2.7
Rental Agent Data	1,399	568	23.9	11.8	-12.1
Other Information	275	29	4.7	0.6	-4.1
Multiple Sources	668	133	11.4	2.8	-8.6
Application of Inference Editing Rules:					
Less Than 5 Minutes	5,338	6,828	91.2	141.4	50.2
5 Minutes or More	342	164	5.8	3.4	-2.4
Other					
Correction of Prior Processing Error	243	210	4.1	4.3	0.2
Post-Imputation Change	64	b	1.1	b	b
Recode of Open-End Response or Special Rule	3,625	433	61.9	9.0	-52.9
Dependent Change[c]	13,130	351	224.2	7.3	-216.9
Total					
With Dependent Changes	25,469	9,160[d]	434.9	189.7	-245.2
Excluding Dependent Changes	12,339	8,809[d]	210.7	182.5	-28.2

[a]Number of completed personal interviews was 5,856 in 1987 and 4,828 in 1990.
[b]Category not used in 1990.
[c]A different definition may have been used in 1990, see text.
[d]Excludes one change in category "Dummy Editor II Update."
Source: Reason for Change Tabulations, RECS Personal Interview Editing: 1987 and 1990.

looking at a questionnaire other than the one that the record to be changed was derived from. Codes assigned to changes based solely on reviews of the outputs of range and consistency checks or special reports identified separately those instances where more than 5 minutes time was needed to reach a decision. The level-of-effort codes used in the 1987 and 1990 RECS are listed and explained in RECS Memo #007.

Summary tabulations of changes to the Household File records based on personal interviews for 1987 and 1990 are shown in Table 6.3. In order to account for differences in sample size for the two years, the data are presented in terms of changes per 100 households.

The differences between 1987 and 1990 in the figures for changes of all kinds are dominated by a precipitous drop in the number of "dependent changes." This probably resulted from a change in the procedures for generating archival records and assigning level-of-effort codes, but it has not been possible to determine the exact nature of the change.

Aside from the effect of the different treatment of dependent changes in 1987 and 1990, the overall frequency of changes on a per-household basis was lower by about 13 percent in the latter year. Overall, the frequency of recontacts with respondents was about the same in both years, with a decline in the number of contacts with household respondents and suppliers being balanced by an increase in contacts with rental agents and interviewers. There was a substantial increase in changes based on the application of inference and editing rules, but a decline in the number requiring 5 minutes or more to reach a decision. There was a substantial decline in 1990 in the number of changes representing recodes of open-end and "other" responses or the application of special rules. Much of the decline appears to have been associated with changes made because of inconsistencies between entries in the part of the questionnaire that summarized uses and methods of payment for each fuel and earlier questions on these same topics.

Imputation

At several stages during data processing, interviewers' or respondents' initial entries on questionnaires are changed or deleted, or values (including 0) are supplied for items initially left blank. In RECS, most of this imputation follows completion of the initial round of computer-assisted edits and special reports. Following imputation, consistency checks are repeated to ensure that imputed values are consistent with other related variables.

This section discusses the imputation procedures used for the Household Survey, the Supplier Survey, and the weather information obtained from NOAA. Items missing on Rental Agent Survey questionnaires are not imputed; in general, the Household Survey responses for these missing items are accepted. The model-based allocation of energy consumption and expenditures to end uses, such as space heating, water heating, and appliances, is considered to be estimation, not imputation, because respondents and suppliers are not asked directly for this information. These allocation procedures are discussed in Chapter 7 in the section on "End-use Estimation."

Imputation for the Household Survey: Interview Questionnaires

The frequency of item nonresponse, based on unedited data files from the 1990 RECS, was discussed in Chapter 4. Excluding variables related to household measurements, 51 of 416 variables based on questionnaire entries had item nonresponse rates of 5.0 percent or more. Of the 10 variables with the highest nonresponse rates (see Table 4.4), only 3--age of hot water heater and two items related to household income--required entries for more than 10 percent of the households.

The treatment of each missing item requires two decisions. The first is whether or not to impute a value for it. Since 1982 missing values have been imputed for roughly two-thirds of the Household Survey variables in each survey year (see Table 6.4). Items not imputed are those for which it is judged that there is not enough information for related variables to provide the basis for an imputation procedure that is likely to reduce the effects of nonresponse bias. For the 1990 RECS, items not imputed included questions on the presence, type and amount of attic and floor insulation, indoor temperatures, and the presence of wall insulation (EIA 1992, p.200).

For each item that is to be imputed, a choice of the most appropriate method is required. Except in 1987, hot-deck imputation, in which the missing value is obtained from a household that matches on variables related to the missing item, has been the most commonly used method. Based on an intensive review of imputation procedures prior to data processing for the 1987 survey, some variables were shifted from hot-deck to other methods of imputation. For example, the entire household grid (demographic characteristics of household members) was shifted from hot-deck to an allocation method and deductive imputation methods were adopted for some variables related to main heating fuel and equipment. Some items were shifted back from deductive to hot-deck for the 1990 RECS and hot-deck imputation was used for most of the new variables in 1993.

Table 6.4. Imputation Methods Used for Household Survey Variables: 1981-1993

Imputation Method	Percent of Items Subject to Imputation					
	1981	1982	1984	1987	1990	1993
Not imputed	23	35	32	36	32	32
Imputed	77	65	68	64	68	68
Hot-Deck	58	52	56	27	42	51
Random	13	9	9	15	13	7
Other[a]	6	4	3	22	13	9
Total	100	100	100	100	100	100
Number of Items[b]	356	443	447	422	429	559

[a]Includes regression, deductive, allocation, and modal imputation methods.
[b]Excludes items for which missing values, if any, were determined by explicit editing rules during the early stages of processing.
Source: Energy Information Administration, *Housing Characteristics* (1981-1993).

In 1993, hot-deck imputation was the only procedure used for the 26 most frequently imputed variables (EIA 1995a, Table A8). The use of other procedures is limited to variables for which there is relatively little item nonresponse.

Other imputation methods that are or have been used include regression, random, deductive, allocation, and modal techniques:

- A *regression* equation, developed from questionnaires with usable data, is used to estimate the total square footage of each sample housing unit for which actual measurements are not obtained or are unusable. Variables used to predict the area of the unit include such housing unit characteristics, as type of housing unit, year built, number of rooms, number of bathrooms, and type of heating equipment, and such household characteristics as income and number of persons. A full statement of the equation and the variable definitions used in the 1993 RECS is given in the *Housing Characteristics 1993* report (EIA 1995a, pp. 230-231).

- *Random-selection* imputation procedures are used for two purposes: to supply missing dates, such as the year and month a housing unit was occupied, and to supply missing values that are conditional on other known values, such as the number of storm windows in a house with a known total number of windows. A value is assigned at random from the appropriate distribution of values for households that report fully.

- *Deductive* procedures are used when the amount of missing data is small and other available information provides reasonably conclusive evidence of what the missing value should be. These procedures are used primarily when information is missing on fuels used for specific purposes and methods of payment for fuels used.

- *Allocation* procedures are used for imputation of missing information on household members, such as age, sex, and relation to householder. Rules for assigning missing values are based on the configuration of known information on these variables for other household members.

- *Modal* imputation procedures, which were used in RECS only in the earlier survey years, assign the most commonly reported value to the missing variable. Typically, a negative response would be assigned to a question on the presence of a relatively rare item, such as the use of secondary water-heating equipment. In the more recent survey years, modal imputation procedures have been replaced by hot-deck or random methods in order to obtain better variance estimates.

Imputation for the Household Survey: Telephone and Mail Questionnaires

Nearly all household survey variables are imputed for the relatively small proportion of households for which questionnaires are obtained by telephone or mail rather than by personal interview (see Chapter 4, Table 4.1). These imputed values of housing unit and household characteristics are combined with the actual Supplier Survey and weather data obtained for these

households and their localities. Thus, the main purposes of obtaining mail and telephone questionnaires for households not responding to interviews are to identify the energy suppliers for the unit, to obtain vehicle information needed to apply sample selection procedures for the Residential Transportation Energy Consumption Survey, and to provide the data needed to select a donor interview questionnaire for imputation of housing unit and household characteristics.

Figure 6.6 shows the steps in the modified hot-deck procedure that was used to impute data for telephone and mail questionnaires in the 1993 RECS. Most of the procedures were computerized. The procedure ensured that no interview questionnaire was used more than once as a donor for a telephone or mail questionnaire. Donor questionnaires were selected manually only for the small proportion of questionnaires for which a suitable donor was not identified by the computerized scoring rules. In the 1990 RECS, only 3 of the mail questionnaires required donors that did not match on all of the sorting variables used in Operation 1; donors were selected from other Census regions for these donees (Response Analysis Corporation 1992b, p.4-22).

Figure 6.6. Imputation Procedure for Household Survey Mail and Telephone Questionnaires: 1993 RECS

Operation		Sorting/Matching Variables
1.	Sort both donor (interview) and donee (mail and telephone) questionnaires into groups based on basis of specified variables.	Census region Type of housing unit structure Space-heating fuel Hot-water fuel Presence of air-conditioning Type of air-conditioning
2.	For each donee, pick the best donor from the corresponding sort group, using a scoring procedure based on specified additional variables.	Income Number of persons in household Number of vehicles Age of householder Tenure (owned/rented) Number of rooms Model year of newest vehicle Household type (married couple/other)
3.	Assign donor values for all Household Survey variables, except number of vehicles, to the donee household.	

Source: Energy Information Administration, *Housing Characteristics* (1993).

Imputation of Supplier Survey Data: Procedures

The level of imputation for Household Survey questions on housing and household characteristics is relatively low, but it has been somewhat higher for consumption and expenditure data for the 5 major fuels, for which the preferred source of information is the actual bill data obtained from energy suppliers. Two kinds of imputation are required. In a process called "annualization," bill data obtained from suppliers are used to arrive at estimates of consumption for a 1-year period.

For those households for which usable bill data are not obtained, annual consumption is imputed on the basis of housing unit and household characteristics. The extent of need for the latter type of imputation depends on eligibility and completion rates for the Supplier Survey. As can be seen in Table 4.7, Chapter 4, Supplier Survey completion rates have been relatively high for electricity and natural gas, somewhat lower for fuel oil and LPG, and quite low for kerosene. Table 4.8 shows that imputation of consumption of electricity and natural gas is seldom needed for single family houses but is needed more often for housing units in multiunit structures.

The procedures for estimating annual consumption differ by fuel type. For electricity and natural gas, the reporting unit for each household in the Supplier Survey is the *billing period*, and an "annualization" process is used to convert the data by billing period to an estimate for a 365-day period. For the 1993 RECS, utilities were asked to provide data for the sample households for all billing periods starting on or after December 1, 1992, and ending prior to the date at which they were asked to complete the form, generally in the late winter or early spring of 1994. They were asked to report the beginning and ending date for each billing period, the amount consumed, the cost, and whether the amount was based on a reading by the customer or on a reading or estimate by the company. Suppliers were instructed to provide bill data only for the specific account for which a waiver was obtained in the Household Survey. Thus, if the sample household did not occupy the housing unit for all of 1993, bill data would, in most instances, be obtained only for the portion of the year during which they occupied it.

For electricity and natural gas, estimation of annual consumption from billing period data was attempted only in the following circumstances:

1. The household paid for some or all uses of the fuel and had 146 or more days of bill data; or

2. The household paid for appliance and/or water heating use, did not pay for space heating or space cooling use, and had 60 or more days of bill data.

When these criteria were met, the procedure for annualization of a household's bill data followed the steps shown in Figure 6.7. First, an annualization period was defined, consisting of consecutive billing periods with a start date as close as possible to January 1, 1993 (the first day of the reference year for consumption), and with the total number of days as close as possible to 365. Second, a consumption year was defined, containing exactly 365 days and matching the annualization period as closely as possible. Third, *predicted* values of consumption for the annualization period and the consumption year were derived from a nonlinear model by using available information about the household's uses of the fuel, the number of days in the period, and the number of heating and cooling degree-days in the period. Finally, the actual consumption reported for the annualization period was adjusted by the ratio of the values of predicted consumption for the consumption year and the annualization period. Expenditures for the consumption year were estimated by applying the unit cost for the annualization period to the estimated value of consumption for the consumption year. A special adjustment procedure was applied to consumption and cost estimates in those instances where the household paid for some but not all uses of electricity or natural gas. In 1993, such households accounted for 0.7 percent of total annual electricity consumption and 1.9 percent of natural gas consumption (EIA 1995d, Table B7).

Figure 6.7. Imputation of Annual Household Consumption from Billing Period Data for Electricity and Natural Gas

Step	Criteria/Procedure
1. Select billing periods for use in estimation. These billing periods cover the "annualization period" (AP).	(1) Continuous data. (2) Start date close as possible to January 1, 1993. (3) Total days close as possible to 365.
2. Define the consumption year (CY).	(1) Must contain 365 days. (2) Match AP as closely as possible.
3. Calculate predicted consumption (C_p) for AP and CY.	Prediction model based on prior survey. Inputs include household uses of fuel, number of days in period, and number of heating and cooling degree days in period.
4. Calculate inputed consumption C_I for CY.	C_I (CY) = Actual consumption for AP x $\frac{\text{Predicted consumption for CY}}{\text{Predicted consumption for AP}}$

Source: Response Analysis Corporation, *1990 RECS Data Editing & Manipulation Procedures Manual* (September 1992).

The inclusion of information on heating and cooling degree-days as part of the prediction models for the annualization procedure started in the 1990 RECS. Prior to 1990, the prediction models relied only on the total number of days in the annualization period and the consumption year (Response Analysis Corporation 1992b, p. A-195).

For fuel oil, LPG, and kerosene, the reporting unit in the Supplier Survey is the *delivery*. Suppliers were asked to report all deliveries from October 1, 1992, through the date at which they completed the form. For each delivery they were asked to report the type of fuel, the amount, the price per unit of volume, and the total price. They were also asked to report the beginning and ending dates of the period covered by the recorded deliveries. If the beginning and ending dates covered all of calendar year 1993, only those deliveries occurring during 1993 were included as part of consumption. If the data on deliveries did not cover a full year, the Supplier Survey data for that household for fuel oil, LPG, or kerosene were not used. It would be possible to develop an imputation procedure that made use of part-year data on deliveries, but the number of households with part-year data is so small that the addition of such a procedure would have a low payoff.

The Household Survey questionnaire included some questions on deliveries and use of fuel oil, LPG, and kerosene. When no usable Supplier Survey data were available, these responses could sometimes be used to estimate annual consumption. This occurred frequently for kerosene, but rarely for fuel oil and LPG.

As explained more fully in Chapter 7 in the section on End-Use Estimation, a separate nonlinear regression model was developed for each fuel, based on data for sample households that had a

full or nearly full year of acceptable bill data, and was used to allocate total consumption of each fuel to 5 use categories: space heating, water heating, air-conditioning, refrigerators, and other appliances. This same model was used to estimate total consumption of a fuel when the sample household used it and did not have usable bill data from the Supplier Survey or, for kerosene, from the Supplier or Household Survey. The regression imputation procedure included the addition of a random error component, making it possible to calculate estimates of sampling error without separating imputed and unimputed data.

Finally, an imputation adjustment was made for each fuel for any household reporting in the Household Survey that some of its bills covered non-household uses of that fuel, for example, for a farm or home business or another household. In such instances, total consumption was reduced by a scale factor developed on the basis of responses to a Household Survey question about the proportion of the bills for that fuel which covered the non-household uses. The percentages of estimated total annual consumption for such households for each fuel in 1990 and 1993 were:

Fuel	1990	1993
Electricity	2.8	4.9
Natural gas	0.6	0.6
Fuel oil	1.3	2.0
Kerosene	0.1	1.0
LPG	3.2	4.4

Imputation of Supplier Survey Data: Quality Implications

As noted in Chapter 3, the goal of RECS is to collect data, for a sample of households, on energy consumption of each major fuel used for residential purposes during a specified time period (for the 1993 RECS, calendar year 1993). Chapters 4 and 5 have revealed several factors that pose problems for the acquisition of precise data for each household, most of them related to nonresponse or incomplete response to the Supplier Survey. The primary factors, the procedures used to deal with them, and their effects on the accuracy of consumption estimates are summarized in Figure 6.8.

For electricity and natural gas, the ideal situation would be to have, for each household, metered values of total consumption, for household uses only, for the calendar year covered by the survey. Because the metering and billing practices of utilities seldom meet these precise requirements, various kinds of compromises and approximations are required. For fuel oil, LPG and kerosene, direct records of consumption do not exist, so information about delivered amounts during the consumption reference period is used as a proxy.

Figure 6.8. Sources of Error in RECS Consumption Data

Fuels Affected and Source of Error	Estimation Procedures
Electricity and natural gas	
No separate metering for household	Model-based imputation based on household survey data.
Billing periods do not coincide with reference year.	Annualization procedure.
Billing periods cover only part of reference year.	Annualization procedure.
Estimated bills.	Influences choice of annualization period.
Household pays for some but not all uses.	If bill amounts are annualized, the results are adjusted upwards.
Fuel oil, LPG, and kerosene	
Data available for deliveries, not actual consumption.	Estimation based on delivered amounts.
All fuels	
Changes in occupancy during reference year.	Household generally treated as if occupied for full year.
No supplier survey data obtained.	Model-based imputation based on household survey data.
Nonresidential uses included in bills.	Amounts scaled back based on household survey estimate of proportion nonresidential.

For all fuels, bill data from the Supplier Survey normally only cover the period during which the sample housing unit was occupied by the household that was present at the time of the Household Survey interview. In cases of part-year occupancy, the imputation procedures treat such housing units as though they were occupied and consumed fuels at the same rate for the entire reference year. Because some of these units were actually not occupied or even not ready for occupancy for part of the year, this approach to imputation overstates their consumption. However, this overstatement may be at least partly offset by the failure to obtain consumption data for units that were occupied for part of the consumption year but were vacant during the interview period for the Household Survey.

Table 6.5 shows, for 1990 and 1993, the proportion of total annual consumption of each fuel that was estimated or imputed by the various methods just described. For fuels other than kerosene, bill data for all or most of a year were the basis for roughly two-thirds to four-fifths of the estimated amounts. For electricity and natural gas, less complete bill data accounted for about 8 or 9 percent of the total amounts. The proportion of consumption based on regression estimates varied from one-tenth for electricity to slightly more than one-third for fuel oil.

Table 6.5. Basis of Estimates of Annual Consumption: 1990 and 1993 RECS (Percent of Total Consumption of Each Energy Source)

Source of Consumption Estimates	Fuel and Year									
	Electricity		Natural Gas		Fuel Oil		Kerosene		LPG	
	1990	1993	1990	1993	1990	1993	1990	1993	1990	1993
Actual Billing Records										
Covering All Uses										
330 or more days[a]	80.5	80.2	74.3	74.1	64.7	67.4	28.5	27.8	71.6	79.1
146 to 329 days	7.5	8.5	6.4	6.3	NA	NA	NA	NA	NA	NA
60 to 145 days	0.2	0.1	0.1	*	NA	NA	NA	NA	NA	NA
Covering Some Uses	0.7	0.7	2.0	1.9	NA	NA	NA	NA	NA	NA
Estimate from Supplier or Household[b]	NA	NA	NA	NA	0.1	0.9	43.0	34.4	0.3	0.1
Regression Estimate	11.0	10.5	17.3	17.7	35.3	31.7	28.5	37.8	28.1	20.9
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Percent of Total Consumption Accounted for by Fuel	32.5	32.8	53.1	52.7	10.6	10.2	0.7	0.5	3.0	3.8

[a]For fuel oil, kerosene, and LPG, billing records were used only if they covered 365 days.
[b]For kerosene, the estimate was supplied by the household.
* = less than 0.05 percent.
NA = Not Applicable.
Note: Because of rounding, percents may not sum to 100.0.
Sources: Energy Information Administration, special tabulations of RECS data files for 1990 and 1993.

For kerosene only, a substantial proportion of total consumption (more than two-fifths in 1990 and about one-third in 1993) was estimated from information on deliveries and purchases supplied by household respondents. Slightly more than one-fourth was based on bill data from the Supplier Survey and the remainder was based on regression estimates.

The basis for estimates varied substantially by type of living quarters. For electricity in 1993, for example, the percent of consumption based on regression estimates by type of structure was:

Type of Structure	Percent Based on Regression Estimate
Mobile home	10.8
One-family detached	7.5
One-family attached	12.6
2 to 4 housing units	25.7
5 or more housing units	22.9

This kind of variation occurs because many of the households in multiunit structures are living in rental units for which some of the utilities are included in the rent and are therefore not eligible for the Supplier Survey. Similar patterns with more pronounced differences by type of structure can be observed for natural gas, fuel oil, and LPG.

As shown on the last line of Table 6.5, the percent of total consumption accounted for by each fuel varies substantially, from 52.7 percent for natural gas in 1993 to 0.5 percent for kerosene in the same year. Using these percents and the data in the body of Table 6.5, it is possible to calculate the basis for estimates of annual consumption for all fuels combined:

Source of consumption estimate	1990	1993
Bill data for all or most of year	74.8	75.4
Partial bill data	7.3	7.4
Estimate from supplier or household	0.3	0.3
Regression estimate	17.5	17.0

Imputation of Missing Weather Data

Weather data obtained from the National Oceanic and Atmospheric Administration (NOAA) for use in RECS include daily temperature data for each of NOAA's weather stations. These data, which are used primarily to estimate heating and cooling degree-days for sample housing units, are sometimes missing for one or more days. Through the 1984 RECS, average temperatures for all weather stations in a NOAA division were used for this purpose. Starting with the 1987 RECS and subsequently, an individual weather station has been selected to provide temperature data for each cluster of sample housing units. The extent of missing temperature data is one of the factors considered in selection of a weather station to be associated with each cluster; data quality is considered acceptable if data are missing for fewer than 15 days of the consumption reference year. Once the stations are selected, missing temperature data are imputed by making use of the relationship between division temperatures and station temperatures for the previous year. For each survey year, for those clusters that remain in the sample from earlier years, the

selection of the associated weather stations is reviewed to take into account changes in the extent of missing temperature data (Response Analysis Corporation 1992b, pp. 8-31 to 8-35).

Evaluation of Imputation Procedures

Because of concern about the high proportion of imputed consumption data for housing units in apartment buildings, a special study was undertaken as part of the 1981 RECS. Permission was obtained from selected apartment building managers to obtain actual fuel records for their buildings, each including one or more RECS sample housing units. Total consumption for each building was allocated equally to the apartment units in that building, and the estimates for the sample units were compared with imputed values assigned by the regression modeling procedures used for units lacking Supplier Survey data. These comparisons indicated biases in some imputed values (EIA 1983b, p.102):

Households Using	Imputed Values Are:	Adjust. Factor
Electricity with air-conditioning	Too low by 50 percent	1.84
Electricity, no air-conditioning	Too high by 10 percent	None
Natural gas for space heating	About right	None
Natural gas, but not for space heating	Too low by 50 percent	2.04

Records of use of fuel oil and LPG in apartments were insufficient in number to make reliable estimates of bias in their imputed values. As a result of this study, the adjustment factors shown above were applied to imputed values of electricity and natural gas consumption in apartments. The same adjustment factors were used in the 1982 RECS. For the 1984 RECS, the regression imputation model was revised to reflect differences between apartments and other units more explicitly, so that these final adjustments were no longer necessary.

RECS questions about temperature setting behavior in the household have been among those for which missing values have not been imputed. Battles and Harrison (1992), using 1990 RECS data, experimented with several regression models in an attempt to identify some of the household and housing unit characteristics that relate to temperature setback behavior when natural gas, electricity, or fuel oil is used for space heating. They also hoped to develop discrete temperature models that could be used to impute missing temperatures. Some of their findings were:

- Low income was a significant factor in a household's decision to reduce temperatures in homes heated with fuel oil, but not in those heated with natural gas or electricity.

- For all three fuels, homes with higher daytime temperatures were likely to have higher setback temperatures as well. Homes that had recently installed a clock thermostat were likely to have lower setback temperatures.

- The colder the climate, the higher the proportion of households that set back temperatures.

In spite of these and other significant findings, the fit of the models developed was not thought to be good enough to use them to impute missing values, so they were not imputed for these items in the 1993 RECS.

7. Estimation and Sampling Error

This chapter includes sections on the weighting procedures used to develop sample estimates, the model-based procedures for allocating total consumption of each fuel to specific end uses, and the estimation and presentation of information about sampling errors of the estimates. Each section begins with a description of the procedures used in the 1993 RECS. Information about current procedures is followed by an account of procedural and design changes aimed at improving the quality of the estimates that have been introduced since the first survey (NIECS) in 1978.

Sample Weighting Procedures

Weighting Procedures for the 1993 RECS

The sample weighting procedures used in RECS closely parallel those used in other U.S. national household surveys, such as the Census Bureau's Current Population Survey. Figure 7.1 provides an overview of the sample weighting procedures that were used to produce estimated totals for the 1993 RECS target population from the sample data, following completion of all editing operations and imputation for item nonresponse.

The overall weight assigned to each household is equal to the product of the weighting factors assigned to it in the four steps shown in Figure 7.1.

Figure 7.1. RECS Sample Weighting Procedures: 1993

Step	Description	Purpose	Auxiliary Data
1	Apply sampling weights	Reflect sample selection probabilities	None
2	Adjust for unit nonresponse	Reduce effects of nonresponse bias	None
3	Ratio estimate, Stage 1	Reduce between-PSU sampling variance	Uses data from 1990 Census
4	Ratio estimate, Stage 2	Reduce mean square error	Uses Current Population Survey estimates as control totals

Source: Energy Information Administration, *Housing Characteristics* (1993).

Step 1: Application of Sampling Weights. Each household record is assigned a weight equal to the reciprocal of its overall probability of selection. The overall probability of selection is the product of the selection probabilities at all stages of sampling: selection of primary sampling units (PSU's), selection of secondary sampling units (SSU's), selection of listing segments within SSU's, and selection of addresses from the listings. In some instances, the selection probability

at one or more stages can be one, as in the case of large metropolitan area PSU's that are selected with certainty.

Sampling weights vary across SSU's for two reasons. The first is that the allocation of sample SSU's to Census divisions is not directly proportional to the number of households by division. Proportionately more sample SSU's are assigned to divisions with fewer households in order to ensure that estimates of acceptable reliability can be made for each division.

The second source of variation in sampling weights is oversampling of targeted population groups. In the 1993 RECS, two groups of special interest were oversampled: low income households and new houses. For the first group, oversampling was accomplished by using higher sample selection rates in SSU's determined by interviewers to be in low-income areas, especially in areas where the main heating fuel was something other than natural gas. For the second group, a supplemental sample of SSU's was selected from Census tracts or block groups with a high percentage of units constructed in the 6-year period prior to the 1990 Census.

In the 1993 RECS, for the first time, sampling weights could also vary within SSU's. Housing units judged to be new by field workers during the listing operation were sampled at a higher rate than other units.

Step 2: Adjustment for Unit Nonresponse. The basic procedure for 1993 was to form a set of weight-adjustment cells consisting of sample households with similar attributes. For each cell, an adjustment factor was calculated by dividing the total number of assigned sample households, including households not interviewed, by the number of interviewed households. If the factor was greater than 2.0, similar cells were collapsed, according to predetermined rules, to form a cell for which the weighting factor was 2.0 or smaller.

The variables used to form the weight-adjustment cells for the 1993 RECS were as follows:

Geographic Domains. These included the nine Census divisions, with Alaska and Hawaii treated as separate domains. Within the four largest Census divisions, the large metropolitan areas that had been selected with a probability of one were also treated separately.

Weighting Classes. Within the geographic domains, subdomains were formed consisting of SSU's and individual housing units that had the same basic sampling weights. As noted previously, basic weights varied primarily because of procedures for oversampling newly constructed housing units and those in low-income areas.

Weather Zones. These domains were based on long-term heating and cooling degree-day averages for counties.

Housing Unit Type. Individual housing units were grouped by type of structure: single family detached, single family attached, multifamily with two to four housing units, and multifamily with five or more units.

A large number of weight-adjustment cells were formed on the basis of these four characteristics. When collapsing was necessary, it was done in the reverse order of the characteristics listed above, that is, starting with the combination of cells representing different housing unit types.

Step 3: Stage 1 Ratio Adjustments. The stage 1 ratio adjustment factors for the 1993 RECS were based on 1990 Census data for the PSU's in strata that were not self-representing; that is, they did not consist of a single PSU that was selected with certainty. A separate adjustment factor was created for each of 36 groups of non-self-representing strata, defined by the 4 Census regions and 9 space-heating fuel categories. The adjustment factor for each group was the ratio of the 1990 Census count of households for all PSU's in the group to an estimate of that count based only on the sample PSU's in the group.

A restriction was placed on the calculation of the adjustment factor that if the denominator, that is, the estimate of Census households in a region and fuel category, was less than 1 million, that fuel category had to be combined with one or more other categories so that the denominator of the calculated adjustment factor was at least 1 million.

Each of the adjustment factors was applied to the weights for all RECS sample households in the corresponding region and space-heating fuel category or categories. Adjustment factors for the 1993 RECS varied from 1.1688 for natural gas in the Northeast Region to 0.7897 for LPG in the Midwest Region.

In sampling terminology, the goal of the first-stage ratio adjustment is to reduce the between PSU component of the sampling variance by using Census information on heating fuels that is available for all PSU's. It can be looked at as a method of compensating for chance factors that may lead to the selection of samples of PSU's whose proportion of households using specified heating fuels at the time of the 1990 Census was higher or lower than the corresponding proportion for all PSU's. Since the distribution of households by heating fuel does not change rapidly, one can expect that these samples of PSU's would deviate in the same direction with respect to the distribution at the time of the survey.

Step 4: Stage 2 Ratio Adjustments. For the 1993 RECS, the second-stage ratio adjustments consisted of four separate steps. In each of these steps, the sampling weights were ratio adjusted so that the sum of the weights for specific categories agreed with the control totals obtained from the Current Population Survey (CPS). Because estimated household counts are available only from the March CPS each year, the control totals for the survey reference month (July 1993 for the 1993 RECS) were derived by linear extrapolation from the CPS estimates for March 1992 and March 1993. The first of the stage 2 ratio adjustment steps started with the weights resulting from Step 3 (stage 1 ratio adjustments). The next three steps in stage 2 started with the weights resulting from the previous step.

The rationale for these adjustments is the expectation that the mean square error of the RECS estimates can be reduced by benchmarking them to the more precise estimates available from the CPS. The CPS estimates are believed to be more precise than the RECS estimates prior to benchmarking for two reasons: the CPS uses a sample of households that is roughly 10 times the size of the RECS sample, and the CPS sample estimates have themselves been benchmarked

to post-censal projections of Census household counts. There is considerable evidence from the CPS and other surveys that survey coverage, especially for some population subgroups, is significantly below the coverage of the population that is obtained in the decennial censuses (for further discussion of this point, see Cox, 1995).

The four steps in the stage 2 ratio adjustments were as follows:

Step 4.1 Weights derived from Steps 1 to 3 were adjusted so that their sum equalled the extrapolated CPS household counts for each of 4 large States--California, New York, Texas and Florida--and for each of the 9 Census divisions.

Step 4.2 Weights derived from Steps 1 through 4.1 were adjusted so that their sum equalled the extrapolated CPS counts for 12 categories defined by the 4 Census regions and 3 "MSA (metropolitan statistical area) status" classifications: central city of MSA, remainder of MSA, and non-MSA.

Step 4.3 Weights derived from all preceding steps were adjusted so that their sum equalled the extrapolated CPS counts in three categories: one-person households occupied by males, one-person households occupied by females, and all other households.

Step 4.4 Step 4.1 was repeated, so that the final weights resulted in exact agreement with the CPS-based household counts for the four large States and nine Census divisions.

Step 4, with its series of successive adjustments to different sets of marginal totals, can be regarded as a raking procedure designed to minimize differences between RECS and CPS estimates of the distribution of households by geographical and other classifiers.

Changes in Sample Weighting Procedures

The basic structure of the sample weighting procedures, as shown in Figure 7.1, has been the same for all survey years. However, there have been several changes in the details, aimed mostly at improving the precision of the survey estimates.

The basic sampling weights (Step 1 in Figure 7.1) have varied as necessary to reflect the specific sample selection procedures used in each survey year. Oversampling of households in low-income areas occurred in the 1981, 1984, 1987, and 1993 survey years. Oversampling of new housing units occurred for the first time in 1993.

Because of the procedures used to oversample new housing units, 1993 was the first survey year for which it was possible for an SSU to have households with different basic sampling weights. This change led to a significant revision of the procedure for calculating the factors used to adjust for unit nonresponse. In all prior survey years, a separate adjustment factor was calculated for each SSU by dividing the total number of assigned sample households in the SSU by the number for which interviews had been completed. If the factor was greater than 2.0, the effect of the adjustment was spread across other SSU's in the same PSU. As noted above, for the 1993

RECS, the initial weight-adjustment cells were formed across SSU's, combining housing units with the same basic sampling weights from a group of SSU's with similar characteristics. This procedure also made it possible to take account of another characteristic of individual housing units, namely type of structure, in deriving the nonresponse adjustment factors.

The stage 1 ratio adjustment (Step 3 in Figure 7.1) has been essentially the same for all survey years, with minor changes made to conform with changes in the space heating fuel categories used in the most recent Census of Population and Housing. The stage 2 ratio adjustment procedure (Step 4 in Figure 7.1) has been modified to make use of a successively larger number of control totals from the CPS.

For survey years 1978 through 1982, a single set of adjustments was made to 12 geographic control totals consisting of CPS estimates of housing units for the 4 Census regions and 3 location categories--central city, remainder of metropolitan statistical area, and nonmetropolitan. Examination of the resulting RECS estimates of the number of one-person households for these years showed that they were consistently about 3 percent below comparable CPS estimates (Response Analysis Corporation 1983). Consequently, for the 1984 RECS, an intermediate ratio adjustment was introduced using national CPS estimates as controls for the number of households in three categories: one-person, male; one-person, female; all others. One more stage was introduced in the 1993 RECS. As explained above, the first step was based on CPS estimates for four large States and the nine Census Divisions, and this step was repeated following the intermediate steps, ensuring exact agreement between RECS and CPS estimates for these geographic domains.

Application of the second-stage ratio estimate procedures is dependent on the availability of the control totals from CPS. Because of uncertainty as to whether March 1991 CPS estimates would be adjusted for undercoverage in the 1990 Census, the necessary CPS data were not available when the 1990 RECS weights were first developed in September 1991. Consequently, the control totals for the survey reference month, November 1990, were developed by forward extrapolation from the March 1989 and March 1990 CPS estimates. In November 1991, the March 1991 CPS estimates were released and revised second-stage ratio adjustments were developed by interpolation between March 1990 and March 1991. For most of the 12 region/location cells, the change in the RECS estimates was 1 percent or less, but the RECS estimates in the Northeast Region increased by 3 percent for the central city metropolitan domain and by 2 percent for the nonmetropolitan domain (Battles 1991b).

Special Estimation Procedures for New Homes

Initial estimates of average energy consumption per household by year built from the 1990 RECS showed a striking reversal of a previously consistent trend for newer homes to consume less energy. The estimated average consumption for homes built in the 1988-1990 period was 103.1 million Btu, 53 percent above the estimate of 67.6 million Btu for homes built in the 1985-1987 period. Both estimates were based on relatively small samples of households, 225 for 1985-1987 and only 138 for 1988-1990 (EIA 1993a, Table 11, p.28).

In an attempt to better understand the factors associated with the apparent trend reversal, data from the Census Bureau's Survey of Construction and Survey of New Mobile Home Placements were used as ancillary data to produce new estimates of average consumption per household for new homes. Two different estimation procedures were developed: a *post-stratification procedure* and a *ratio-adjustment procedure*. They are described in detail in the *Consumption and Expenditures Report* for 1990 (EIA 1993a, pp. 173-181).

The post-stratification procedure used nine strata defined by a combination of Census region, type of home, and main space-heating fuel. The post-stratified estimate was a weighted average of the RECS estimates of average consumption per household for the nine strata, with the weights being the Census Bureau estimates of the proportion of housing units in each stratum. The ratio-adjustment procedure was based on Census Bureau estimates of the distribution of new homes by Census region and the increase in average heated floor space by region for homes built in 1988-1990 compared with those built in 1985-1987. Ratio-adjusted RECS estimates of average consumption per household were based on adjustments that eliminated or reduced RECS-Census differences for these two characteristics.

Both of the revised estimates showed a substantially smaller increase in average energy consumption for homes built in 1988-1990 compared with those built in 1985-1987:

Estimate	Average Consumption (millions of Btu)		Percent Increase
	1985-87	1988-90	
Original RECS estimate	67.6	103.1	53
Post-stratified estimate	74.5	89.7	20
Ratio-adjusted estimate	70.6	90.3	30

Standard errors of the post-stratified estimates were appreciably smaller than those of the original RECS estimates for both periods (Latta 1993, p.14). Standard errors were not computed for the ratio-adjusted estimates.

The estimation procedures used in this instance were designed to improve the precision of a specific class of RECS estimates, and it is doubtful whether their application across the board for all estimates would be feasible or desirable. However, they illustrate the potential for using post-stratification and allied techniques for improving estimates used in specific kinds of analyses.

Special Weighting Procedures for Buildings

Appendix B to the *Consumption and Expenditures 1990* report presented, for the first time, some tabulations of RECS data that used the building, rather than the household or housing unit, as the unit of analysis (EIA 1993a, Table B6, p.152). Additional data on residential buildings were

presented in a 1995 EIA report, *Buildings and Energy in the 1980's* (EIA 1995b). According to the building definition used in EIA's Commercial Buildings Energy Consumption Survey, most housing units correspond to separate buildings; however, this was not the case for units in multiunit apartment buildings.

Estimates of the number of buildings were obtained by dividing the sampling or base weight for each RECS sample housing unit in a multiunit building by the total number of housing units in that building. This information had been collected in the Household Survey for sample housing units in buildings with five or more housing units, but had not been collected for sample households in buildings with two to four housing units. Therefore, for the latter group, a constant divisor was used in each Census region, based on data from the 1990 Census of Housing and Population.

RECS estimates of the number and total floorspace of residential buildings are subject to two kinds of biases:

- The number of buildings is underestimated because RECS interviews are not conducted in vacant housing units. The amount of underestimation is likely to be similar to the housing unit vacancy rate, which was estimated by the Census Bureau's American Housing Survey to be about 9 percent in 1989.

- The size of multiunit buildings is understated because the floorspace of common areas, such as hallways, stairwells, elevators, and lobbies is not accounted for.

RECS estimates of energy consumption for multiunit residential buildings are probably also understated because they are made by applying appropriate weights to metered or estimated consumption for individual units in those buildings. Consumption for heating, cooling, and other uses in common areas of these buildings is not accounted for.

No attempt has been made to adjust the published estimates for these sources of bias. Their net effect on estimates of energy intensity (thousand Btu per square foot) is unknown.

End-Use Estimation

Introduction

In addition to knowing the total residential energy consumption for each of the five major fuels--natural gas, electricity, fuel oil, LPG and kerosene--energy analysts and policymakers need information about the allocation of these amounts to different end uses, such as space heating, water heating, air-conditioning, and appliances. However, utility bills, the primary source of data on total consumption, are not broken down by end use, nor is there any practical means by which such information could be obtained directly from each sample household. Consequently, an indirect, model-based nonlinear regression technique is used in RECS to provide estimates of the consumption of each fuel by end use for each sample household. The same technique is used

to estimate total consumption for those households and fuels for which no usable utility bill data have been obtained.

There are three main steps in the modeling and estimation process:

1. For each fuel, parameter values in a preliminary model for end-use allocation are estimated by using data only for sample households that used the fuel, have usable billing data, do not have imputed values for key independent variables in the model, and meet other quality requirements. The process is iterated, eliminating (with some exceptions) independent variables whose estimated coefficients do not differ from zero by at least four standard deviations and, if necessary, removing outliers, i.e., households with large differences between estimated and reported consumption, from the data base. Model parameters for natural gas are estimated first, because certain relationships in the natural gas model are carried over to other fuels.

2. The final model from step 1, with estimated parameter values, is used to impute missing values for total consumption for each fuel.

3. For all sample households, the final model is used to estimate total consumption of each fuel by end-use category. For each household, the end-use consumption estimates are "normalized"; that is, they are adjusted proportionately to sum to the reported or imputed value of total consumption.

The RECS end-use estimation techniques have been gradually developed and refined since the first survey in 1978. Through the 1982 RECS, the techniques were considered experimental and the results were published in special reports and articles (EIA 1983c, Thompson 1987). Starting with the 1984 RECS, the estimates of end-use consumption by fuel have been published routinely in the *Consumption and Expenditures* reports (the data for 1984 were published in Volume 2, Regional Data, of the report). However, refinements have continued, as described in more detail below. The nature of the estimation procedure is such that the regression equations used in each survey year are unlikely to be identical to those used in the preceding survey year.

The specific estimation equations for each end use depend on the kinds of information collected in the Household Survey or from other sources that are relevant to that end use. For space-heating, for example, such variables as heating degree-days, type, size, and age of the housing unit, amount of heated space, thermostat settings, type of heating equipment, and amount and type of secondary space heating are all likely to be associated with variations in consumption. While many of these variables are of interest in their own right, the inclusion of questions needed to provide inputs to the end-use estimation equations has always been a major consideration in the choice of content for the RECS Household Survey questionnaire.

Although the full equations used for the 1984 survey were presented in the regional supplement to the *Consumption and Expenditures* report and the full equations for the 1987 and 1990 surveys were presented in the corresponding national *Consumption and Expenditures* reports, data users are cautioned with respect to interpretation of the coefficients that are associated with the independent variables.

As with any large regression, care should be taken in interpreting the individual coefficients in the equations. The variables that are used in the equations may be highly correlated with variables, which are not used in the equations. Thus the value of the coefficients will reflect both the impact of the included variables and the impact of any correlated excluded variables. For instance, the natural gas equations did not contain variables that used the type and R-value of insulation directly, but the impact of the type and amount of insulation is included through variables which indicate the presence of insulation (EIA 1993a, p. 198).

End-Use Estimation for the 1993 RECS

The 1993 end-use estimation model consisted of five nonlinear regression equations: one for each of the major energy sources. In each equation, the dependent variable was the total consumption for that fuel for the survey reference year. The equation expressed total consumption for the fuel as the sum of three or more components, corresponding to the end uses for which separate estimates were to be obtained, plus an error term. Each of these components was expressed, in turn, as a complex nonlinear function of household variables available from RECS.

For each fuel, the estimation equation included a space-heating and water-heating component. For fuel oil, LPG, and kerosene, there was one additional component, called "appliances," covering all other uses of that fuel. For natural gas, there was an additional component for air-conditioners and a residual category for appliances. For electricity, there were additional components for air-conditioners, refrigerators, freezers, lighting, cooking, dishwashers, clothes dryers, and all other appliances.

To illustrate the structure of the nonlinear regression model and its components, Figure 7.2 shows the basic equation used in the 1993 RECS for electricity and the formulation for one of its components, the one covering electricity consumption for freezers. The units of measure in Figure 7.2 are thousands of Btu’s. Although there are many variations in the variables by component and fuel, the basic structure of all components is similar. Typically, a component, such as the electricity freezer component shown in Figure 7.2, consists of a base term and one or more multiplicative adjustment terms. The base term models the energy consumption for a "standard" situation.

For the freezer component, the base term (FZBASE × CDDBASE) is a function of the number of freezers and cooling degree-days. There are four adjustment terms. The first two adjustment terms (MANUADJ and UPRTADJ) are functions of the type of freezer (manual defrost versus frost free and chest versus upright). The third adjustment term (AGEADJ) is a function of the age of the freezer. For the few households with two or more freezers, the adjustment terms are determined by the type and age of the largest freezer. (In effect, the model assumes any additional freezers are of the same type and age as the largest freezer. To save interviewing time, only the type and age of the largest freezer was recorded.)

Figure 7.2. Selected Components of the Nonlinear End-Use Consumption Model for Electricity Used in the 1993 RECS

Regression equation for electricity:

$$Y_{EL} = X_{SH} + X_{WH} + X_{AC} + X_{RFG} + X_{FZ} + X_{LGT} + X_{CK} + X_{DW} + X_{CD} + X_{OTAPL} + e$$

where Y_{EL} = *actual* annual consumption of electricity

X_{SH}, X_{WH}, X_{AC}, X_{RFG}, X_{FZ}, X_{LGT}, X_{CK}, X_{DW}, X_{CD}, and X_{OTAPL} are end-use components for space heating, water heating, air conditioning, refrigerator, freezer, lighting, cooking, dishwashing, clothes dryer, and all other appliances, respectively

and $e = Y^{1/4}_{EL} - \hat{Y}^{1/4}_{EL}$

with $\hat{Y}_{EL}$ being the *estimated* annual consumption of electricity.

Details for the Electricity Freezer Component: 1993 RECS

$$X_{FZ} = FZBASE \times CDDBASE \times MANUADJ \times UPRTADJ \times AGEADJ \times TOTADJ$$

where TOTADJ is an adjustment factor applied to all electricity components, based upon the price of electricity, demographic characteristics of the household, geographic location, and type of housing unit

and	FZBASE	= 2345 × (Number of freezers)
	CDDBASE	= 1 + (0.0170 × (CDD65)$^{1/2}$)
	MANUADJ	= 1 - (0.2019 × MANUFZ)
	UPRTADJ	= 1 + (0.1123 × UPRTFZ)
	AGEADJ	= 1 + (0.2718 × FZ20PLUS) - (0.3203 × FZ4MNUS)
and	CDD65	= cooling degree-days to the base 65 degrees Fahrenheit
	MANUFZ	= 1 if largest freezer is a manual defrost freezer and 0 otherwise
	UPRTFZ	= 1 if largest freezer is an upright freezer and 0 otherwise
	FZ20PLUS	= 1 if largest freezer is 20 years old or more and 0 otherwise
	FZ4MNUS	= 1 if largest freezer is 4 years old or less and 0 otherwise.

The last adjustment term (TOTADJ) is used for all components in the electricity model. It includes variables that should have an effect on all electricity components. Examples of this are the price of electricity, the family income level, and other demographic characteristics of the household. The model assumes that the effect of variables used in TOTADJ is the same for all components. For instance, high income may imply bigger homes with bigger freezers, bigger appliances of other kinds, and more appliances. Thus, high income may be associated with higher electricity consumption for all end-uses.

Using a nonlinear formulation of the model, the freezer component requires the estimation of only six coefficients to model the effect of climate (number of cooling degree-days), type of freezer, and age of freezer on the electricity consumption of freezers. The model assumes that factors interact proportionally. For example, the effect of the age of a freezer on its electricity consumption is proportionally the same for all freezer types and for all climates. The resulting

equation projects that, for freezers of the same type located in the same climate, the newest freezers use 32 percent less electricity than freezers in the next age group, while the oldest freezers use 27 percent more.

The model-fitting procedure was designed to minimize the sum of the squared error term over all households included in the analysis for each fuel. For the 1990 and 1993 RECS, as shown in Figure 7.2, the error term was equal to the difference between the fourth root of the actual consumption and the fourth root of the estimated consumption. The error term defined in this way was found to be more nearly normally distributed with a constant variance than alternative formulations of the error term, such as the simple difference or the difference between logarithms or square roots of the actual and estimated consumption.

Because the regression equations are nonlinear, the parameter values cannot be estimated with standard multivariate linear regression techniques. They were estimated by using a nonlinear regression procedure in the statistical computer package, SAS.[1]

As noted in the introduction to this section, the regression analysis for each fuel was based on a subset of the sample households using that fuel. Households were excluded from the analysis for many reasons, the principal ones being: they did not pay the supplier directly for all uses of the fuel (so that usable billing data were not available); other problems with the consumption data, such as data covering only part of a year or inclusion of nonresidential uses in the billing data; or imputed values of key independent variables, such as occurred for many variables when the Household Survey data were obtained by mail.

Table 7.1 shows the number and percents of sample households used in the regression analysis for each fuel type in the 1990 RECS. Of the households using each fuel, the proportion included in the analysis varied from slightly less than one-half for fuel oil to about two-thirds for electricity. A draft Technical Note on the 1990 regression analysis (Harrison 1993) provides additional detail on the data sets included in the analyses by housing unit type and major end-use category. That report identifies several fuel/end-use categories for which the number of sample households used was small, for example, natural gas air-conditioning (only eight observations were available), use of all fuels as secondary fuels for water heating, and use of fuel oil or kerosene for any purpose other than heating. Although separate models were not developed for each housing type (adjustment terms were developed to model differences by housing type), the report also notes that "Because there were more observations for households living in single-family detached homes, the regression analysis should give the best estimates for [these households]."

This relatively brief description has covered the highlights of the complex nonlinear regression models used in the 1993 RECS to allocate consumption of each fuel to end-use categories and to impute total consumption of the fuel when necessary. Substantial additional detail, including the equations used for each fuel and end use, is provided in: Appendix D, "End-Use Estimation Methodology," of the *Consumption and Expenditures 1990* report (EIA 1993a); Appendix C, "End-Use Estimation and Methodology," of the *Consumption and Expenditures 1993* report (EIA 1995d); and in the Technical Note cited above (Harrison 1993).

[1] Statistical Analysis System (SAS) (Cary, NC).

Table 7.1. Number and Percent of Sample Households Used in Regression Analyses, by Fuel Type: 1990 RECS

Category	Fuel Type									
	Nat. Gas		Electricity		Fuel Oil[b]		LPG		Kerosene	
	No.	Pct.	No.	Pct.	No.	Pct.	No.	Pct.	No.	Pct.
Households Using Fuel	3,255	100.0	5,094	100.0	700	100.0	461	100.0	278	100.0
Used in Analysis	1,917	58.9	3,392	66.6	336	48.1	257	55.7	163	58.6
Not Used, by Reason[a]										
Didn't Pay for Some Used Directly	612	18.8	356	7.0	145	20.7	26	5.6	1	0.4
Quality of Consumption Data Not Acceptable	506	15.5	1,028	20.2	161	23.0	152	33.0	81	29.1
Key Independent Variables Imputed	220	6.8	318	6.2	50	7.1	22	4.8	15	5.4
Other[c]	NA	NA	NA	NA	8	1.1	4	0.9	18	6.5

[a]Each household not used is counted under the first applicable reason.
[b]Data for Model A only, see Source, p. 45.
[c]Did not purchase or, for kerosene only, did not use for space heating.
NA = Not Applicable.
Source: Latta, *Poststratification Estimation* (May 1993).

Changes in End-use Estimation Methodology: 1978-1993

The RECS energy consumption end-use estimation procedures and models have never been precisely the same from one survey to the next. Changes occur for several reasons. First, 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 regression analyses. Many of the questionnaire changes have been motivated by the desire to collect new information that could be used either to reduce the size of the error term in the basic equations or to introduce new end-use categories for which separate estimates could be made. In the 1990 RECS, the appliance category for electricity was subdivided into refrigerators, freezers, and other appliances.

In the 1993 RECS, the addition of new questions on lighting and electric appliances made it possible to further subdivide the appliance category for electricity to provide separate end-use estimates for lighting, cooking, clothes dryers, and dishwashers.

Second, even if the questionnaire content and the model specification were to remain unchanged from one survey to the next, the estimates of the model parameters would change, due in part to sampling variability of the estimates and in part to real changes in the underlying relationships of the independent variables. It is also possible that some parameter estimates that met the basic test for significance in one survey year might not qualify in a subsequent survey year. However,

the initial significance test criterion has been loosened somewhat, in order to improve comparability of end-use estimates over time and to avoid eliminating variables that appear to have intrinsic validity as part of the model.

Third, there have been some basic changes in the structure of the model used for end-use allocation of energy consumption. A major change occurred in the 1984 RECS when the linear model used in prior surveys, for which parameter values could be estimated by standard multivariate least squares regression, was replaced by a nonlinear model requiring a different estimation method. The decision to adopt a nonlinear model was reached after extensive experimentation with and evaluation of both linear and nonlinear models and associated procedures to estimate end-use consumption by fuel, using data from the initial survey years (see following subsection on "Evaluation of End-use Estimation Procedures").

There were three major reasons for the change to a nonlinear model: (1) the ability to formulate a more realistic model; (2) the ability to formulate the model in a way that avoids negative estimates for households that have a combination of factors all pointing to lower energy consumption; and (3) the ability to formulate the error term in a way that results in its distribution being approximately normal, with a constant variance.

To understand these advantages, it may be useful to consider a possible linear formulation of one component of the nonlinear model now in use. Figure 7.3 shows a linear formulation of the model for electricity and its freezer component. Like the nonlinear formulation, the linear one uses actual annual consumption of electricity as the dependent variable. The error term for the linear model, following the usual practice, is the difference between actual and estimated annual electricity consumption. The coefficients for the terms of the freezer component--a_1, a_2, a_3, a_4, a_5, and a_6--would be estimated using linear regression. Knowledge of the characteristics of freezers gives the expectation that a_1, a_2, a_4, and a_5 would be positive and that a_3 and a_6 would be negative.

The model in Figure 7.3 does not include interaction terms for the climate, type, and age of the freezers. Interaction terms could be added. The model does not include terms for the effects of the price of electricity or the family income on the freezer component. (Lower electricity prices and/or higher income could be associated with larger freezers.) Again, interaction terms could be added.

The use of many interaction terms in the linear model may result in a formulation that looks more realistic, but the actual estimated coefficients may result in unrealistic estimates for some combinations of type of freezer, age of freezer, climate, income level, and electricity price. Some estimates could even be negative. The use of a nonlinear model allows the formulation of a more realistic model using far fewer terms than would be needed with a linear model. The use of fewer terms in the model reduces the possibility of unrealistic estimates for some combinations.

Analysis of the residual terms from the linear model previously used shows that the error terms were not normally distributed with constant variance. In fact, the error terms were skewed in the positive direction and the variance of the error terms increased as the projected energy consumption increased. The use of weights can alleviate the effect of trends in the variance of error term with either linear or nonlinear regression. However, weights alone do not alleviate the effect of the skewness of the error terms.

Figure 7.3. An Alternative Linear Formulation of the Model Components Shown in Figure 7.2

Regression equation for electricity:

$$Y_{EL} = X_{SH} + X_{WH} + X_{AC} + X_{RFG} + X_{FZ} + X_{LGT} + X_{CK} + X_{DW} + X_{CD} + X_{OTAPL} + e$$

where Y_{EL} = *actual* annual consumption of electricity

X_{SH}, X_{WH}, X_{AC}, X_{RFG}, X_{FZ}, X_{LGT}, X_{CK}, X_{DW}, X_{CD}, and X_{OTAPL} are end-use components for space heating, water heating, air-conditioning, refrigerator, freezer, lighting, cooking, dishwashing, clothes dryer, and all other appliances, respectively

and $e = Y_{EL} - \hat{Y}_{EL}$

with $\hat{Y}_{EL}$ being the *estimated* annual consumption of electricity.

Details for the Electricity Freezer Component: 1993 RECS

$$\begin{aligned} X_{FZ} = {} & a_1 \times (\text{Number of freezers}) \\ & + a_2 \times (\text{Number of freezers}) \times (\text{CDD65})^{1/2} \\ & + a_3 \times (\text{Number of freezers}) \times \text{MANUFZ} \\ & + a_4 \times (\text{Number of freezers}) \times \text{UPRTFZ} \\ & + a_5 \times (\text{Number of freezers}) \times \text{FZ20PLUS} \\ & + a_6 \times (\text{Number of freezers}) \times \text{FZ4MNUS} \end{aligned}$$

where CDD65 = cooling degree-days to the base 65 degrees Fahrenheit
MANUFZ = 1 if largest freezer is a manual defrost freezer and 0 otherwise
UPRTFZ = 1 if largest freezer is an upright freezer and 0 otherwise
FZ20PLUS = 1 if largest freezer is 20 years old or more and 0 otherwise
FZ4MNUS = 1 if largest freezer is 4 years old or less and 0 otherwise.

The error term adopted for the nonlinear model introduced in 1984 was the difference between the logarithms of actual and estimated consumption. This error term was closer to being normally distributed, but its variance was still not constant for all energy and household types. This problem was dealt with by using a weighted regression method in which households in categories with high error variances were given lower weights. For example, for natural gas a weight of 1.0 was given to most households, but a weight of 0.2 was assigned to households using natural gas which:

1. Did not use it as a main space-heating or water-heating fuel.

2. Did not use it as a main water-heating fuel, did use it as a main space-heating fuel, and the main equipment was a natural gas floor furnace, wall furnace, pipeless furnace, or room heater.

In the 1990 RECS, the logarithmic error term used in 1984 and 1987 was replaced by the one shown in Figure 7.2--that is, the difference between the fourth roots of actual and estimated consumption of each fuel. Investigation of four alternative error terms--linear, logarithmic, square root, and fourth root--had shown that the last of these came closest to meeting the basic

requirements for normality and constant variance. With the introduction of the new error term, the weighted regression procedures used in 1984 and 1987 were no longer necessary.

There have also been some conceptual changes involving the definition of certain end-use components of the model. In the 1984 RECS, electricity used to run fans for central forced-air heating systems was assigned to the space-heating component for electricity. In subsequent survey years, electricity used for this purpose was assigned to the appliance component rather than the space-heating component. This change was made so that households which did not use electricity for space heating would not have any consumption of electricity assigned to the space-heating component. A similar change was made for electricity used to operate whole-house fans, ceiling fans, window fans, and evaporative (swamp) coolers. In 1984, electricity used for these purposes was included in the air-conditioning component; since 1987, it has been included in the appliance component.

Evaluation of End-Use Estimation Procedures

In recent years, new technologies have made it possible to measure the consumption of electricity for individual appliances and other uses within the home, a process often referred to as submetering (Windell 1986). Conceivably, similar technologies could also make it possible to measure amounts of natural gas used for space heating, hot water heating, cooking, and other uses. An ideal means of evaluating the RECS end-use estimation models for electricity and natural gas would be to measure end-use consumption of these fuels in a subset of RECS sample households and compare these direct measurements with estimates generated by the nonlinear regression models for the same households. However, there are two obstacles to such a project: the high cost per household of installing the monitoring equipment and the difficulty in enlisting an acceptably high proportion of households in a national probability sample to agree to participate in such a study. Consequently, efforts to evaluate the RECS end-use estimation procedures have so far relied on less direct methods. Three studies that have been undertaken are described in this subsection.

As noted earlier in this section, end-use allocation for the first five survey years, 1978 through 1982, was based on a linear model, with the nonlinear model being introduced in the 1984 survey. The independent variables included in the linear model varied during the five survey years for which it was used, as new items were added to the Household Survey questionnaire.

Following the 1984 RECS, an exploratory study was undertaken to examine the effects on the end-use estimates of using different models (Carroll 1987). The study also looked at the effects of the models on estimates of total consumption for each fuel, because the same models were being used to impute total consumption for households for fuels for which direct data were not available. For the survey years 1978, 1980, 1981, 1982, and 1984 (the 1979 survey was not included in the study), estimates of total consumption and consumption by end use were developed by using two different nonlinear models: one (called the NIECS-based model) using only those variables, which were available from all five surveys, and the other (called the 1984 RECS-based model) using all of the variables available from the 1984 RECS, with substitution of proxy variables for those for which data were not collected in earlier surveys. For example,

the study report states that "Income dummy variables proved to be effective proxies for the [unavailable air-conditioning] use data in the 1978 and 1980 surveys."

The only results available from this study are from a preliminary report which does not include all of the basic tabulations. Some of the author's conclusions were as follows:

- Compared to the linear models, the nonlinear models consistently allocated more consumption to space cooling and less to appliances.

- The overall predictive power of the NIECS-based nonlinear model (which included fewer independent variables) was slightly lower than that of the 1984 RECS-based nonlinear model, accounting for 10 percent less of the total variance. However, there was relatively little difference in mean consumption amounts estimated by the two models.

- Differences between the end-use allocations estimated by the NIECS-based and RECS-based nonlinear models were small, except for space cooling.

The other two studies were about end-use allocation of residential consumption of electricity, and both of them made use of residential submetering data collected by electric utilities. Battles (1990) reports on a comparison of nonlinear model-based estimates of electricity consumption by end use from the 1987 RECS with estimates based on submetering data collected for various studies by eight electric utilities. The comparisons covered four end uses of electricity: space heating, room air-conditioning, central air-conditioning, and water heating. The method of comparison for each of the eight utilities was to select a subset of RECS sample households from the same Census division that matched as closely as possible on known characteristics of the households for which the utility had obtained submetering data. All households in the study were in single family housing units. Other characteristics taken into account for all or some of the utilities were heating and cooling degree-days, tenure, floor area, and the use of certain appliances. The RECS end-use estimates for this subset of sample households were then compared with the corresponding estimates for the households that had been submetered by the utility, taking into account the sampling errors associated with the RECS estimates.

Given the large sampling errors associated with the RECS estimates and the fact that the households studied by the utilities did not constitute a probability sample of the same population, the results of the comparisons can only be roughly indicative of possible biases in the RECS model-based estimates. Battles concluded that the RECS model-based estimates were "reasonable estimates" when compared to the utilities' submetered estimates. However, she stated that:

> This study does, though, reveal some areas where further investigation may be warranted. All of the submetered estimates for both air-conditioning and space heating are lower than the RECS CDA [conditional demand analysis] comparative estimates and all water-heating submetered estimates are higher than the RECS CDA comparative water-heating estimates. The consistency in differences is important. (Battles 1990, p.12)

The other study that made use of submetered data on consumption of electricity (Response Analysis Corporation 1992c,d) compared the utility data with RECS estimates of end-use consumption based on the model used in the 1990 RECS. As explained earlier in this section, the 1990 RECS was the first to use an error term based on the fourth roots of estimated and actual consumption, as opposed to the logarithmic error term used in 1984 and 1987. The study used submetering data and information on household and demographic characteristics that had been obtained for samples of households by five utilities. Only those households for which both kinds of information were complete were included in the study: sample sizes by utility varied from 13 for the City of Austin to 182 for Pacific Gas and Electric. All of the utilities provided end-use load data on water heating and all provided data on one or more of the following: central air-conditioning, room air-conditioning, space heating, total HVAC (heating, ventilation and air-conditioning), refrigerators, and total appliances.

Because the utilities did not collect all of the household and demographic information that is available for RECS sample households, a separate modified end-use estimation model was developed for each of the five utilities, making use only of the variables that were available for that utility. The submetered end-use data for that utility were then compared with three sets of estimates:

1. Estimates for the utility's sample households based on the modified RECS model

2. Estimates for a selected set of RECS sample households, similar in their characteristics to the utility's sample households, based on the modified RECS model

3. Estimates for the same set of RECS sample households, based on the full RECS model.

Assuming that estimates based on the modified RECS model do not differ significantly from those based on the full model, the comparison of the submetered data with set 1 provides the best indicator of how well the statistical end-use model allocates consumption. In some instances it appeared that the assumption was not valid, so adjustments were made to the estimates in set 1 on the basis of the relationship between sets 2 and 3.

Table 7.2 shows the percentage differences between the submetered end-use data and the *adjusted* model-based estimates for the same households (set 1). The findings for four of the five utilities were consistent with the indications from the previous study that the RECS end-use estimation model might be overestimating consumption for central and room air-conditioning and space heating and underestimating consumption for water heating. Findings for the Bonneville Power Administration were in the opposite direction. The authors of the report suggest that the RECS end-use models might be improved by developing a separate model for each region, on the grounds that the factors that determine space-heating and air-conditioning consumption in different climates may be quite different.

Table 7.2. Percentage Difference[a] Between Modeled and Submetered End-Use Estimates

End-Use Estimate	Utility				
	Austin	BPA[b]	PGE[c]	Santee Cooper	SCE[d]
Central Air-Conditioning	40%	NA	370%	13%	33%
Room Air-Conditioning	NA	NA	17%	NA	69%
Space Heating	54%	-10%	NA	33%	NA
HVAC	84%	-22%	NA	27%	NA
Water Heating	-25%	4%	-23%	-8%	NA
Refrigerators	NA	-23%	0%	NA	18%
Appliances	-27%	5%	NA	-1%	NA

[a]Percent difference = $\frac{\text{(Modeled estimate - Submetered estimate)}}{\text{Submetered estimate.}} \times 100$

[b]Bonneville Power Administration.
[c]Pacific Gas and Electric.
[d]Southern California Edison.
NA = Not Available.
Source: Response Analysis Corporation (1993).

Sampling Errors

The sampling error of each published statistic is estimated by using the balanced half-sample replication method. The estimated sampling errors are used to check the validity of statements made in the text of survey reports and as the basis for suppressing estimates whose relative standard errors are 50 percent or more. Due to space limitations, the estimated sampling errors for the individual table cells are not published, but the estimates provide the basis for the derivation of generalized variance functions, which are published and permit users to compute an approximate relative standard error for each published estimate.

This section describes the procedures used for the estimation and publication of sampling errors in the 1993 RECS, followed by information about changes in the methodology used in earlier surveys. The section concludes with a discussion of the accuracy of sampling error estimates and the extent to which RECS has achieved its goals for the precision of key estimates of energy consumption. Readers who would like additional detail about the derivation and use of sampling errors in RECS may refer to the introductions and appendices of the *Housing Characteristics* and the *Consumption and Expenditures* reports for each survey year--for example, pp. 18-20 and 231-236 in the *Housing Characteristics* report for 1993 (EIA 1995a).

Estimation of Sampling Errors for the 1993 RECS

The half-sample replications were formed from 78 "super strata," each containing pairs of sample households:

- Thirty-eight of the super strata consisted of pairs of non-self-representing strata from the same Census Division. Strata from the four most populous States (California, New York, Texas, and Florida), which had been formed so as not to cross State lines, were always paired with other strata from the same State. Within Census Divisions and also within the four most populous States, strata for Metropolitan Statistical Areas (MSA's) were paired with other MSA strata and non-MSA strata were paired with other non-MSA strata. The pairs in each of the 31 super strata consisted of the sample households in the primary sampling units selected from each of the two strata.

- Thirty-one of the super strata consisted of large metropolitan areas that had been selected with certainty. The pairs consisted of sample households in two sets of the secondary sampling units that had been selected from the metropolitan area.

- The nine remaining super strata each consisted of a single non-self-representing stratum. The pairs consisted of sample households in two sets of the secondary sampling units that had been selected from the sample PSU in that stratum. These non-self-representing strata were not combined with other non-self-representing strata because of restrictions on combining strata with differing attributes, for example, strata in different Census divisions.

Ninety-six half samples were formed from the 78 super strata by selecting, in each instance, one of the pairs of sample households from each super stratum. The selection was balanced--that is, it was carried out in such a way that each pair member from a super stratum was included in 48 of the 96 half samples. To produce sample estimates from each of the 96 half samples, the sampling weights were ratio-adjusted upwards so that the sum of the weights was equal to the control totals (housing-unit counts derived from the Current Population Survey) for each of the nine Census divisions and four States--California, New York, Texas, and Florida.

The estimated variance for each sample estimate was the mean squared deviation of the 96 half-sample estimates from the full sample estimate. Because the ratio adjustments to control counts were applied to each half sample, the estimated housing-unit counts for the nine Census divisions have zero variance. For estimates of housing-unit counts that are close to the control counts, such as the number of housing units using electricity or the number with refrigerators, the sampling errors are very small.

Generalized Variances

Showing the estimated sampling error for each published statistic would roughly double the space required for publication of tabulations. As an alternative, generalized variance functions, which permit users to determine an approximate value of the relative standard error (RSE) for each table cell, are included in the publications. Figure 7.4 shows an example of a 1990 RECS publication table with "row and column factors" which can be used as shown in the example to determine the relative and absolute standard errors and a confidence interval for any cell in the table.

Figure 7.4. Example of the Use of RSE Row and Column Factors to Derive Approximate Standard Errors

Characteristics	Major Energy Sources	Electricity	Natural Gas	Fuel Oil	Kerosene	Liquified Petroleum Gas	RSE Row Factors
RSE Column Factors	0.9	0.8	0.7	0.7	1.5	1.9	
Total U.S. Households	12.0	23.6	5.6	7.8	9.4	11.2	1.3
Urban Status							
Urban	11.9	24.6	5.7	7.8	9.5	11.4	1.7
Central City	11.2	24.3	5.8	7.3	9.8	14.3	3.1
Suburban	12.4	24.8	5.6	8.0	9.2	11.2	1.8
Rural	12.2	20.8	5.2	7.8	9.4	11.0	2.6
Climate Zone							
Under 2,000 CDD and Over 7,000 HDD	10.2	21.6	5.0	7.7	9.5	10.1	3.5
5,500 to 7,000 HDD	10.1	24.6	5.3	7.9	9.2	11.2	2.8
4,000 to 5,499 HDD	12.0	23.5	6.3	7.8	9.9	11.2	3.6
Under 4,000 HDD	13.8	24.3	5.8	8.0	9.0	11.5	2.5
2,000 CDD or More and Under 4,000 HDD	15.9	23.0	5.9	Q	10.3	12.9	5.2
Type of Housing Unit							
Single-Family	11.9	23.3	5.5	8.0	9.5	11.0	1.5
Detached	11.8	23.2	5.4	8.0	9.5	11.0	1.6
Attached	12.9	24.5	6.1	8.1	9.8	Q	3.6
Mobile Home	13.0	21.6	5.3	8.2	9.4	11.7	2.9
Multifamily	11.9	25.6	6.1	7.2	8.5	13.5	3.8
2 to 4 Units	10.7	26.5	6.1	8.1	Q	13.1	4.3
5 or More Units	13.4	24.9	6.0	5.9	Q	Q	3.3

Row Factor (Urban) = 1.7
Column Factor (Electricity) = 0.8

Approximate RSE (Average Electricity Expenditure in the Urban Area) = (1.7) * (0.8) = 1.36 percent.

Approximate Standard Error (Average Electricity Expenditure in the Urban Area) = (.0136) * (24.6) = 0.33 Dollars per Million Btu.

Approximate 2 Standard Errors (95 percent confidence interval) = (1.96) * (0.33) = 0.6 Dollars per Million Btu.

Therefore, with 95 percent confidence, the average electricity expenditure in the Urban area is between 24.0 and 25.2 Dollars per Million Btu (24.6 $\pm$ 0.6).

Source: Energy Information Administration, Office of Energy Markets and End Use, 1990 Residential Energy Consumption Survey.

The publication also explains how to use the RSE's for the appropriate table cells to determine the RSE's for percentages based on household counts and for the ratios and differences of two statistics (under the assumption that they are independent). The row and column factors for each publication table are derived by using the estimated RSE's for the table cells to estimate the parameters of a log-linear model,

$$\log(RSE_{ij}) = m + a_i + b_j.$$

The row factor for the ith row is the geometric mean of the RSE's in that row and the column factor for the jth column is an adjustment factor with geometric mean equal to one. Special procedures are used for cells with very large or very small RSE's or missing values.

The row and column factors are derived separately for each publication table. Consequently, an estimate that appears in more than one table may have different RSE's arrived at by using different sets of row and column factors. Any of these values should provide a useful approximation to the relative standard error for that item as estimated by the replication method (EIA 1986b).

Estimates of Change Between Surveys

When comparing statistics between survey years, assuming independence of the estimates from different surveys will, in most instances, lead to an *overestimate* of the sampling error of a difference or ratio of estimates of the same variable for different years. This occurs because the samples for different years are not in fact fully independent. For most survey years, the sample PSU's have been the same as those used in the previous survey year, and even in those years when new samples of PSU's were selected, the selection procedure was designed to maximize the overlap between the old and new samples. In addition, the samples for survey years 1982 through 1990 included longitudinal components, so that in each of these years approximately one-half of the sample housing units had also been included in the sample for the preceding survey year (except for 1982, when the overlapping units had been included in the 1980 survey). For most survey variables, one would expect estimates for different years from these overlapping samples to be positively correlated, leading to reductions in the sampling errors of their differences or ratios.

Better estimates of the sampling error of change between survey years under these conditions are possible with the balanced half-sample replication method, provided that the same sets of half samples are used for the two years in question and the differences are estimated from each half sample. Sampling errors were estimated by this method and compared with sampling errors estimated under the assumption of independence for selected variables and pairs of survey years from 1978 through 1984 (EIA 1987c, pp. 217-225). Some of the results are shown in Table 7.3. For virtually all of the variables and pairs of years shown, the more precise estimates of sampling errors of differences (Method 2) are substantially smaller than sampling errors calculated under the assumption of independence (Method 1). The reductions for the 4-year interval, 1980 to 1984, are less than those for either of the 2-year intervals, 1980 to 1982 and 1982 to 1984. The reasons for these smaller reductions are not entirely clear, because the pattern of sample overlap

for these periods was relatively complicated. However, a reduction in the correlation over time for longer intervals may have been a contributing factor.

Table 7.3. Comparison of Standard Error of Difference Estimated by Two Methods for Changes in Average Consumption per Household Between Survey Years

Years and Fuel	Average Consumption[a]			Standard Error		
	Year 1	Year 2	Difference[b]	Method 1[c]	Method 2[d]	Percent Difference[e]
1980 and 1982						
All fuels	114.2	102.9	-11.2	2.3	1.2	49
Electricity	30.1	28.9	-1.2	1.0	0.4	59
Natural gas	95.7	88.1	-7.6	2.2	1.4	36
Fuel oil/kerosene	100.8	73.4	-27.0	3.5	2.0	42
LPG	47.6	39.4	-8.2	3.4	2.1	38
1982 and 1984						
All fuels	102.9	104.7	1.7	2.2	1.1	48
Electricity	28.9	28.8	-0.2	0.9	0.5	49
Natural gas	88.1	89.9	1.8	2.1	1.4	32
Fuel oil/kerosene	73.4	71.9	-1.5	3.4	2.7	22
LPG	39.4	40.1	0.7	3.3	2.6	21
1980 and 1984						
All fuels	114.2	104.7	-9.5	2.3	1.7	25
Electricity	30.1	28.8	-1.4	0.8	0.4	45
Natural gas	95.7	89.9	-5.8	2.2	1.8	20
Fuel oil/kerosene	100.8	71.9	-28.9	3.4	3.4	1
LPG	47.6	40.1	-7.5	3.4	3.2	6

[a]Average consumption per household using fuel, in millions of Btu.
[b]Due to rounding, may not be consistent with values shown in table for Years 1 and 2.
[c]Assumes estimates for Years 1 and 2 are independent.
[d]Reflects correlation between estimates for Years 1 and 2.
[e]Percent reduction from using Method 2. Due to rounding, may not be consistent with values shown for Methods 1 & 2.
Source: Energy Information Administration (1987).

Changes in Methodology

The same basic method of estimating sampling errors, balanced half-sample replication, has been used in all survey years. There have been several changes in estimation procedures and the methods of presenting information about sampling errors in RECS publications:

- The number of half samples used has varied. From 1978 through 1982, 32 half samples were used, except in 1979, when there were 72. In the 1984, 1987, and 1990 surveys, 128 samples were used and, as noted earlier, 96 were used in the 1993 RECS.

- In the 1978 NIECS, the same overall weights were used for each half sample, so that the effects of nonresponse adjustments and ratio estimates to control totals were not reflected in the estimated sampling errors. In the following year and subsequently, separate weights were developed for each half sample.

- The composition of the "super strata" used to form half samples has varied, mainly because of changes in the design of the RECS sample. In the first three surveys, all non-self-representing strata were paired to form super strata. Subsequently, it was decided that certain restrictions should be placed on pairing, such as not pairing strata from different Census divisions. Non-self-representing strata that could not be paired under these restrictions were treated as separate super strata, with the half samples being formed from two sets of secondary sampling units in the sample primary sampling unit for each super stratum.

- In the 1978, 1979, and 1980 survey reports there were two separate sets of tables, one containing the sample estimates and the second containing the estimated sampling errors corresponding to many of the estimates shown in the first set. Various methods were suggested for estimating sampling errors not shown in the second set of tables. Tables of individual sampling errors were dropped from the reports after 1980 and a series of procedures, based on various methods of estimating generalized sampling errors, was provided in one of the appendices to each report. The *Consumption and Expenditures* report for 1984 introduced the method of showing row and column factors in the data tables, and this approach has been followed since then.

Limitations of Sampling Error Estimates

Estimates of sampling error are themselves subject to sampling error. Their sampling error would be minimized if all possible half samples were used in the balanced half-sample replication estimates, but the cost of doing so would be prohibitive, so a subset is used. The larger the subset, the closer the estimated sampling errors will be to the value obtained by use of all possible half samples.

For the super strata formed by collapsing non-self-representing strata, sampling errors are overestimated because the reduction in sampling error resulting from stratification is not fully reflected. On the other hand, sampling errors for the super strata that consist of a single non-self-representing stratum are underestimated, because the estimates do not reflect the between primary sampling unit component of the variance for these strata. No data are available on the net effect of these biases, which are inherent in the estimation of sampling errors for a sample design that selects a single primary sampling unit from each stratum.

Sampling errors for estimates of energy consumption and expenditures by end use are understated, because the parameters of the nonlinear end-use allocation model are not estimated separately for each half sample. Thus, the estimated sampling errors do not reflect the error of estimation of the model parameters. Sampling errors for the 1990 end-use estimates were calculated, but not published, for this reason. Sampling errors or row and column factors have been published for 1984, 1987, and 1993. In the *Consumption and Expenditures* report for 1993 (EIA 1995d), the row and column factors for end-use estimates were footnoted to indicate that they were underestimates.

The generalized estimates of sampling errors that have appeared in the published RECS reports since 1984 are approximations to the values estimated directly for published data cells. As noted above, the direct estimates were presented only for selected items in the 1978, 1979, and 1980 RECS reports and have not been published in subsequent reports. A detailed analysis of the differences between the direct estimates of sampling error and the approximations based on row and column factors was undertaken for the 1983 Nonresidential Buildings Energy Consumption Survey (EIA 1986b, pp. 200-203). The measure of accuracy chosen for that analysis was the root mean square, along a table column, of differences of the base-10 logarithms of the approximate and direct estimates of the relative standard errors. For most table columns, these values were found to correspond to percentage differences between 20 and 60. The differences from the direct estimates were fairly evenly distributed between positive and negative values.

Sampling Error Targets for Key Estimates

The 1993 RECS sample was designed to produce estimates of average energy expenditures with sampling errors no greater than specified target levels: 1.25 percent for the national estimate, 2.75 percent for estimates by Census region and 4.50 percent for estimates by Census division. As shown in Table 7.4, actual sampling errors, whether estimated directly or by using the appropriate row and column factors, were all below these target values. Achievement of values that were well below the targets in some instances resulted in part from the supplementation of the 1993 core sample with special samples designed to strengthen sample coverage of newly constructed housing units and low income households.

Design Effects

Notwithstanding best efforts to develop a sampling frame consisting of heterogeneous clusters, the sampling errors of estimates based on a complex multistage cluster sample like the one used in RECS are usually greater than the sampling errors that would have been obtained if a simple random sample of the same size had been used. The cluster design is, of course, preferred, because it can produce the desired level of reliability at a lower cost than a simple random sample can.

A recent analysis of the RECS sample design produced estimates of the optimum number of primary sampling units (PSU's), secondary sampling units per primary sampling unit, and households per secondary sampling unit for several categories of RECS variables (EIA 1994). Table 7.5 shows the design effects, expressed as the ratio of the variance or standard deviation to the variance or standard deviation for a simple random sample of the same size, for the optimum size clusters and for those actually used in the 1993 RECS. Number of PSU's and cluster sizes for the optimum designs were based on core samples of 5,095 housing units. As the table shows, the design effects on the standard deviation ranged from 1.21 to 1.26 for the designs using optimum cluster sizes and from 1.39 to 1.60 for the actual 1993 sample design. Design effects for estimates of consumption and expenditures were somewhat higher than those for other types of variables.

Table 7.4. Sampling Errors for Estimates of Average Consumption per Household: 1993 RECS

Area	Relative Standard Error (Percent)		
	Target	Approximation[a]	Direct Estimate
United States	1.25	1.04	1.08
Census Region			
Northeast	2.75	1.92	2.61
Midwest	2.75	1.84	2.24
South	2.75	2.00	2.10
West	2.75	2.08	2.19
Census Division			
New England	4.50	3.28	4.23
Middle Atlantic	4.50	2.32	3.14
East North Central	4.50	2.00	3.04
West North Central	4.50	3.68	2.88
South Atlantic	4.50	2.80	3.16
East South Central	4.50	3.44	3.83
West South Central	4.50	3.68	4.17
Mountain	4.50	3.44	3.89
Pacific	4.50	2.64	2.72

[a]Approximate values based on row and column factors.
Sources: Energy Information Administration, *Consumption and Expenditures* (1993a); Energy Information Administration (1994).

Table 7.5. Design Effects for RECS, Using Optimum and Actual Cluster Sizes

Sample Design and Type of Variable	Number of PSU's m	SSU's per PSU $\bar{n}$	Housing Units per SSU[a] q	Design Effect on	
				Variance[b]	Standard Error
Optimum Design					
Consumption and Expenditures	201.2	16.83	1.50	1.586	1.259
Housing Unit Characteristics	152.8	23.10	1.44	1.472	1.213
Appliances	92.9	30.60	1.79	1.492	1.221
Demographic	94.6	31.14	1.73	1.480	1.217
Actual 1993 Design					
Consumption and Expenditures	116	13.88	3.06	2.558	1.599
Housing Unit Characteristics	116	13.88	3.06	2.337	1.529
Appliances	116	13.88	3.06	1.937	1.392
Demographic	116	13.88	3.06	1.974	1.405

[a]For actual 1993 design, includes only the base sample.
[b]DEF = $h_1 \bar{n} q + 1 + h_2 (q - 1)$ where h_1 = within PSU measure of homogeneity.
h_2 = within SSU measure of homogeneity.
Source: Energy Information Administration (1994).

8. Comparisons of RECS Estimates with Other Data

The previous chapters have presented information about sampling errors and various sources of nonsampling errors associated with RECS estimates. Whenever possible, quantitative information about nonsampling errors has been included, based on operating statistics, pretests, methodological experiments, and special evaluation studies in which the accuracy of *individual* responses has been evaluated by reinterviews or other means. The subject of this chapter is a less direct but nevertheless useful source of information about the quality of RECS estimates: comparisons of *aggregate* estimates from RECS with data from other sources believed to be at least roughly comparable with regard to population coverage and definition of variables.

Typically, comparisons of aggregate data from different sources proceed as follows:

- The analyst looks for differences in design that may cause the estimates to differ. These might include different definitions of the target population, different reference dates or periods, and different definitions of the variables to be estimated. If there are reasonable grounds for doing so, the analyst may adjust one or both of the estimates to make them more nearly comparable with each other.

- If one or both estimates are based on probability samples, the analyst develops confidence intervals for differences between the (adjusted) estimates from the two sources.

- If (adjusted) estimates are significantly different, the analyst will look for additional factors that may explain the differences.

When significant differences are observed, it is sometimes not readily apparent which of the estimates is more accurate. Nevertheless, such comparisons are often valuable. In some instances, such comparisons have suggested ways of strengthening the RECS survey design and procedures. Results of the comparisons are presented in RECS publications in the belief that they will help users to understand the strengths and limitations of the survey data and thus to use them more effectively.

Two kinds of comparisons will be discussed. The next section is about comparisons between RECS estimates of end-use consumption and estimates from surveys of fuel suppliers, mostly conducted by EIA, of amounts of energy supplied to the residential sector. The following section covers comparisons of RECS data on housing unit and household characteristics with data from the decennial census and from other household surveys, such as the Current Population Survey, the American Housing Survey and the Consumer Expenditure Survey. Comparisons between RECS estimates and administrative counts of program participants are also presented.

Comparisons of RECS and Supplier Survey Estimates of Consumption

The collection of data from energy suppliers is an important component of RECS. However, the RECS Supplier Survey collects billing data only for households that are in the RECS sample. In addition to conducting surveys of end-use consumption in the residential, commercial, and manufacturing sectors, EIA conducts several surveys of energy suppliers who provide various types of fuels for consumption by these and other sectors of the economy. In EIA's supplier surveys, respondents are asked to provide data on total amounts of fuel supplied to all customers during specified time periods and, to the extent possible, to disaggregate these amounts by class of customer.

There have been several studies comparing estimates of consumption by fuel type from EIA's end-use consumption surveys with supplier survey estimates of amounts supplied to the residential and commercial sectors. Our focus here will be on the comparisons for the residential sector. Results of these comparison studies have been published in several special reports (EIA 1986a, EIA 1990, Miller 1995, Allied Technology Group 1995). For the 1990 and 1993 RECS, comparisons of RECS and EIA supplier survey data have been published in an appendix to the *Consumption and Expenditures* report (EIA 1993a, Appendix C, EIA 1995d, Appendix D).

Three EIA supplier surveys have been the primary basis for the comparisons:

- The *Annual Electric Utility Report*, Form EIA-861 (prior to 1984, Form EIA-826 was used)
- The *Annual Report of Natural and Supplemental Gas Supply and Disposition*, Form EIA-176
- The *Annual Fuel Oil and Kerosene Sales Report*, Form EIA-821

The first two of these annual surveys cover all known suppliers; the third is based on a sample.

Differences in Defining the Residential Sector

Each of the three supplier surveys asks respondents to report separate estimates for several sectors or classes of customers, one of which is the residential sector. However, the supplier survey definitions of the residential sector differ, both conceptually and operationally, from the one that is used in RECS. The electric utilities reporting on Form EIA-861 are allowed to use discretion to determine which of their end-use customers are classified as residential. In practice, the determination is likely to be based on the utilities' rate structures, which, in turn, are based on customers' relative rates of consumption. As noted in a 1990 report:

> The utility specifies how much fuel it supplied to residential, commercial, industrial, and other customers by totaling the quantity supplied under these rate classes. Utilities are not required to maintain records on the economic activities

> of their customers, so their rate structures may not correspond to economic definitions of the end-use sectors. To the extent there is not a one-to-one correspondence between the economic activity of the customers and the rate schedule at which they are billed, there will be a misclassification of end-use sector supply data. (EIA 1990, p.13)

The same report points out that an individual customer's classification--or rate--schedule can vary during the year as its consumption varies. Similar considerations apply to natural gas distribution companies reporting on Form EIA-176. Fuel oil distributors reporting on Form EIA-821 are specifically instructed to exclude farms and large apartment buildings from the residential sector.

The difference in their definitions of the residential sector is only one of several ways in which RECS and the three supplier surveys identified above differ with respect to coverage, timing, and definition of data items collected. Consequently, one should not necessarily expect RECS estimates of residential consumption to agree closely with estimates of amounts of fuel supplied from any of the supplier surveys. Differences also occur because of sampling and nonsampling errors in the estimates. Figure 8.1 summarizes the main features of RECS and the supplier surveys that affect the comparisons.

Other Differences in Coverage

RECS coverage, as described in Chapter 3, is limited to U.S. housing units occupied as primary residences. Vacant units and units used seasonally or occasionally as second homes are excluded. Suppliers, on the other hand, are asked to report total amounts supplied to customers, without any exclusions. As shown in Table 3.1, Chapter 3, the vacant and seasonal housing units excluded from RECS have accounted for between 9.2 and 11.5 percent of total U.S. housing units between 1981 and 1993, according to biennial estimates from the Census Bureau's American Housing Survey. Their proportionate share of total residential energy consumption is probably somewhat smaller.

The classification of some master-metered apartments as commercial in the supplier surveys works in the opposite direction--that is, it leads to supplier survey estimates that are lower than the RECS estimates for the residential sector. For electricity and natural gas, the effects of this factor are hard to quantify, because suppliers are not consistent in their classification of apartments. In a study undertaken in the mid-1980's the issue was explored for natural gas by contacting public utility commissions and large utilities in 5 midwestern States--Illinois, Indiana, Michigan, Ohio, and Wisconsin--where there is substantial use of natural gas for heating. Based on limited data that these sources were able to provide, it was estimated that 3.4 percent of natural gas supplied to the residential sector in this 5-State area was being reported in other sectors in the supplier survey (EIA 1986a, Table 59, p.72). For fuel oil, the situation is somewhat clearer, because the supplier survey instructions specifically request that respondents exclude apartments from the residential sector. The 1993 RECS estimated that multifamily housing units accounted for 17 percent of all fuel oil consumed by the residential sector (EIA 1995d, Table 5.2).

Figure 8.1. Sources of Differences Between RECS Estimates of End-Use Consumption and EIA Supply Survey Estimates of Energy Supplied to the Residential Sector

Source	RECS	Supply Surveys[a]
Differences in Coverage		
Occupancy	Vacant and seasonal units excluded.	No exclusions.
Apartments	Included.	May be excluded if commercial rate applies.
Farm and Other Residences with Business Uses	Included in survey, business uses excluded from consumption.	Household may be excluded if commercial rate applies. If included, no basis for eliminating consumption for business uses.
Differences in Timing		
Reference Period	Different from calendar year through 1984.	Calendar year for all surveys.
Storable Fuels	Measures amounts used for metered fuels; amounts supplied for others.	Measure amounts supplied during reference period.
Sampling Error	Estimates of sampling error available.	None for electricity and natural gas. Fuel oil based on sample survey but sampling errors of estimates used in comparisons are not available.

[a]This figure covers the following EIA annual supply surveys: Electricity: Form EIA-861 (Form EIA-826 prior to 1984); Natural Gas: Form EIA-176; Fuel Oil and Kerosene: Form EIA-821.

Some customers of energy suppliers combine residential and nonfarm or farm business uses of fuel in the same account. In RECS, business uses are excluded from estimates of residential consumption on the basis of respondents' answers to questions about the proportion of their total consumption of each fuel that is used for business. For the electricity and natural gas supplier surveys (Forms EIA-861 and EIA-176), respondents are asked to classify consumers who use fuels for both residential and commercial purposes according to their predominant use, so the net effect of such mixed uses is difficult to determine. For the fuel oil supplier survey (Form EIA-821), farms are excluded from the residential sector.

Differences in Timing

Through survey year 1984, the reference period for RECS consumption and expenditures data ran 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. For subsequent survey years, RECS consumption data have been collected for a calendar year. All of the EIA supply surveys collect data on a calendar year basis.

Consequently, for RECS survey years through 1984, one might expect to see consumption/supply survey differences in amounts of heating fuels used/supplied in parts of the country for which heating degree-days for January through March varied appreciably from one year to the next. Comparisons of expenditure data would be most affected in periods when there were rapid fluctuations in energy prices. In a special study of consumption and supply estimates for survey

years 1978 through 1982, procedures were developed to adjust the data by Census division for natural gas and fuel oil from both sources for these differences in timing, as well as for the different treatment of apartments in RECS and the supply surveys. These procedures were successful in reconciling differences for fuel oil, but only partially successful in reconciling differences for natural gas (EIA 1986a, Part 7).

A 1990 study that compared measures obtained from consumption and supply surveys noted that "Since fuels (except electricity) can be stored, the amount of product supplied to a sector in a given period is not necessarily equal to the amount consumed" (EIA 1990, p.2). However, this "storability" factor would not be likely to have significant effects on consumption/supply comparisons for the residential sector. For electricity and natural gas, metered amounts are reported both in RECS and in the relevant supplier surveys. For the other fuels, since it would be impractical for households to report their actual consumption in RECS, deliveries are used as a proxy for consumption. Thus, the RECS data for all fuels are comparable in this regard to those obtained in the supplier surveys.

Sampling and Nonsampling Errors

The RECS estimates of total consumption of each fuel are subject to sampling error, and estimates of their sampling errors have been calculated. The supplier surveys for natural gas and electricity include all known suppliers, so the results of these surveys are not subject to sampling error. The supplier survey for fuel oil and kerosene is based on a sample of distributors, and sampling errors of direct sample estimates have been calculated. However, for the comparisons presented below, the sample survey data have been benchmarked to supply data from a different source, and sampling errors for these benchmarked estimates have not been calculated.

All of the estimates of end-use consumption and amounts supplied are subject to various kinds of coverage, nonresponse, measurement, and data-processing errors. Nonsampling errors of RECS estimates have been discussed at length in Chapters 3 through 7 of this report.

Comparisons of Consumption and Supply Data at the National Level

Table 8.1 shows comparisons of RECS and supplier survey data at the U.S. level for electricity, natural gas, and fuel oil for all RECS survey years except 1979 and 1981. The key item in the table for each year and fuel is the ratio of the supply estimate to the consumption estimate. The ratios differ from 1.000 by more than twice their standard errors for 7 of the 21 yearly comparisons.

Table 8.1. Residential Consumption and Supply of Electricity, Natural Gas, and Fuel Oil: 1978-1993

Consumption & Supply Survey Estimates	1978[a]	1980	1982	1984	1987	1990	1993
				ELECTRICITY			
RECS Consumption (billion kWh)	724	721	710	728	808	888	962
EIA Supply Data (billion kWh)	671	717	730	778	850	924	995
Ratio of Supply to Consumption	0.927	0.994	1.028	1.069*	1.052*	1.041	1.034
Standard Error of Ratio	0.043	0.019	0.029	0.026	0.017	0.027	0.019
				NATURAL GAS			
RECS Consumption (billion ft^3)	5,461	4,840	4,680	4,830	4,687	4,737	5,131
EIA Supply Data (billion ft^3)	4,891	4,752	4,633	4,555	4,315	4,391	4,957
Ratio of Supply to Consumption	0.896*	0.982	0.990	0.943	0.921*	0.927*	0.966
Standard Error of Ratio	0.055	0.038	0.039	0.033	0.034	0.032	0.034
				FUEL OIL[b]			
RECS Consumption (million gallon)[b]	15,802	11,220	8,230	9,080	8,850	7,100	7,380
EIA Supply Data (million gallon)[b]	15,091	10,290	8,274	7,602	8,106	6,050	6,590
Ratio of Supply to Consumption	0.955	0.917	1.005	0.837*	0.916	0.852*	0.893
Standard Error of Ratio[c]	0.075	0.057	0.058	0.054	0.054	0.058	0.062

* = Ratio differs from 1.000 by more than twice its standard error.
[a]Totals for 1978 do not include data for Alaska and Hawaii.
[b]For 1978, 1980, and 1982 includes kerosene.
[c]Underestimate; does not reflect sampling error of supply survey estimate.
kWh = Kilowatthours.
Sources: Consumption: Energy Information Administration, *Consumption and Expenditures* (for years shown); Supply: *State Energy Data* (for years shown).

- For *electricity*, the supplier survey estimates were below the RECS consumption estimates in 1978 and 1980; subsequently they have been moderately higher than the RECS estimates. They were significantly higher than the RECS consumption estimates in 1984 and 1987.

- For *natural gas*, the supply estimates were below the RECS consumption estimates in all years and were significantly lower in three of the seven years.

- For *fuel oil*, the supply estimates were below the RECS consumption estimates in all years except 1982, when the ratio was 1.005. The supply estimates were significantly lower in two of the seven years.

The largest *change* in the supply/consumption ratio between RECS survey years was for fuel oil, where the ratio declined from 1.005 in 1982 to 0.837 in 1984. Form EIA-821 was used for the first time in 1984; it succeeded Form EIA-172, which had been used from 1979 through 1982. The statistical procedures and methodologies associated with the new form differed from those used earlier; consequently, the supply estimates for 1984 and subsequent years are not considered directly comparable with those for prior years (EIA 1995g, p. 348).

Table 8.2 compares RECS consumption and supplier survey estimates at the U.S. level for kerosene and LPG for 1990 and 1993. As the table shows, the 1993 RECS consumption estimate for kerosene was significantly below the supplier survey estimate. None of the other three differences was statistically significant.

Table 8.2. Residential Consumption and Supply of Kerosene and LPG: 1990 and 1993

Consumption and Supply Survey Estimates	Kerosene		LPG	
	1990	1993	1990	1993
RECS Consumption (Quadrillion Btu)	.07	.05	.28	.38
Supply Data (Quadrillion Btu)	.06	.08	.36	.40
Difference (RECS - Supply)	.01	-.03*	-.08	-.02
Two Standard Errors (RECS Standard Error)	.02	.01	.06	.07

* = Absolute value of difference exceeds twice its standard error.
Source: Energy Information Administration, *Consumption and Expenditures* (1990 and 1993).

Of the five major fuels, kerosene and LPG are the least frequently used and together accounted for only about 4 percent of total residential consumption in 1993. RECS estimates of their consumption are subject to large relative sampling errors, so that comparisons of consumption and supply estimates cannot determine whether small observed differences are statistically significant. In addition, as was shown in Table 6.5, Chapter 6, for RECS sample households, the proportion of kerosene use derived from supplier billing records was less than 30 percent in 1990 and 1993, compared to much higher proportions for the other four fuels. The primary source of the kerosene supply data is the same as for fuel oil, the *Annual Fuel Oil and Kerosene Sales Report*, Form EIA-821. EIA does not survey suppliers of LPG; the supplier data for LPG appear annually in the *State Energy Data Report* and are based on data provided by the American Petroleum Institute.

Comparisons at the Census Division Level

A 1995 report (Allied Technology Group 1995) compares RECS consumption estimates by Census division with supplier survey estimates for all five major fuels (data for fuel oil and kerosene were combined) for the years 1984, 1987, and 1990. Because of the relatively large sampling errors of RECS estimates at the Census division level, only large estimated differences--generally more than 10 percent of the supply estimate, and often more than 20 percent for smaller divisions and less frequently used fuels--are statistically significant.

The most consistent differences found in this study occurred in the Middle Atlantic Division in the comparisons for fuel oil plus kerosene and for LPG. The data for these comparisons are shown in Table 8.3. For fuel oil plus kerosene (kerosene is only a small part of the total for the two fuels), the Middle Atlantic Division accounts for roughly one-half of total U.S. consumption. The RECS consumption estimates were well above the supplier survey estimates in all three years. As noted above, supplier survey respondents for Form EIA-821 were specifically

instructed to exclude apartments and farms from the residential sector. In 1990, an estimated 24 percent of the consumption of fuel oil in the Middle Atlantic Division was by households that were in buildings with two or more housing units.

Table 8.3. Residential Consumption and Supply of Selected Fuels, Middle Atlantic Division: 1984, 1987, and 1990

Consumption & Supply Survey Estimates	1984	1987	1990
		FUEL OIL AND KEROSENE	
RECS Consumption (trillion Btu)	650	610	513.1
EIA Supply Data (trillion Btu)	410.1	457.0	340.1
Ratio of Consumption to Supply	1.585*	1.335*	1.590*
Standard Error of Ratio[a]	0.178	0.092	0.109
		LPG	
RECS Consumption (trillion Btu)	10	10.0	12.0
Supply Data (trillion Btu)	22.2	26.3	27.3
Ratio of Consumption to Supply	0.450*	0.380*	0.440*
Standard Error of Ratio	0.171	0.122	0.198

* = Ratio differs from 1.000 by more than twice its standard error.
[a]Underestimate; does not reflect sampling error of supplier survey estimate.
Source: Allied Technology Group, *Revised Analysis Report: Comparison of Data from Energy Consumption and Supply Surveys* (March 1995).

Table 8.4. Estimates of Electricity Consumption per Residential Unit from RECS and the Edison Electric Institute (EEI): 1970-1984

Year	RECS (kWh per household)	EEI (kWh per customer)	Ratio RECS/EEI
1970		7.066	
1971		7.380	
1972		7.691	
1973	8.530[a]	8.079	1.06
1974		7.907	
1975	8.630[a]	8.176	1.06
1976		8.360	
1977		8.693	
1978	9.450	8.849	1.07
1979	9.150	8.843	1.03
1980	8.840	9.025	0.98
1981	8.750	8.825	0.99
1982	8.480	8.743	0.97
1983		8.814	
1984	8.440	8.978	0.94

[a]Data from predecessor surveys to RECS that were conducted by the Washington Center for Metropolitan Studies.
Sources: Energy Information Administration, *Consumption and Expenditures* (1984); EEI data are from the *Statistical Yearbook of the Electric Utility Industry*.

For LPG, total consumption was much smaller, and the Middle Atlantic Division accounts for less than 5 percent of total U.S. consumption. The RECS consumption estimates were consistently below the supply estimates, which are based on data provided by the American Petroleum Institute.

Supplier Data from Non-EIA Sources

The *Consumption and Expenditures* report for 1984 includes a comparison of RECS estimates of average electricity consumption per household with a data series on average residential electricity consumption per customer compiled by the Edison Electric Institute (EEI) (EIA 1987a, pp. 288-89). The data from the two sources are shown in Table 8.4. The EEI data were based on quarterly surveys of investor-owned utilities, Tennessee Valley Authority distributors, some State and Federal projects, and large municipal utilities, supplemented by data from secondary sources to complete the coverage (EIA 1989c, pp. 24-25). It is likely that many of the factors that were relevant to comparisons of data on total residential consumption of electricity from RECS and EIA's supplier surveys (including sampling error of the RECS estimates) would also contribute to differences between the RECS and EEI data series. One additional factor might be differences in the denominators. For RECS, the denominator is always a single household; for EEI, some of the customer accounts may have included more than one household.

Given these differences in the sources of data, the differences between the two sets of estimates are relatively small. However, there is a clearly evident trend for the ratio of the two series to decline between 1978 and 1984. The EEI estimates were relatively stable during this period, at the same time that the RECS estimates of consumption per household declined by about 10 percent.

Comparisons of RECS Data on Housing Unit and Household Characteristics with Data from Other Sources

Data items identical or roughly comparable to those included in RECS have been collected in several surveys conducted by other agencies, especially the Census Bureau. The existence of such comparable items does not mean there is unnecessary overlap among the statistical programs. The surveys in question and the decennial census have purposes that are quite different from those of RECS. RECS provides in-depth information about residential energy consumption and expenditures, whereas the Census Bureau's American Housing Survey covers a broad array of characteristics of the nation's housing stock and provides more detailed data for subnational areas. The Decennial Housing Census provides small-area data for a few basic housing items. Some data that are potentially comparable to RECS estimates are also provided by administrative data systems, such as those established for the Food Stamp and Low-Income Home Energy Assistance (LIHEAP) Programs. The comparisons discussed in this section are organized by data source, starting with the American Housing Survey, continuing with other surveys and the Housing Census, and concluding with administrative record sources.

The American Housing Survey: Comparisons with NIECS

Prior to 1980, the American Housing Survey (AHS) was conducted annually and was called the Annual Housing Survey. The most systematic comparison of RECS and AHS data, undertaken by the University of California's Energy Research Group, used data from the 1978 RECS (NIECS) and the 1978 AHS (Blumstein *et al*., 1982). There were 18 variables that were essentially the same in both surveys:

Year structure built	Have thermostat
Main heating equipment type	Have air-conditioning
Main heating fuel	Have hot running water
Cooking fuel	Have roof insulation
Household income	Have storm windows
Property value	Have storm doors
Tenancy type	Have complete plumbing
Water heating fuel	Number of AC units
Number of household members	Number of rooms

Some additional variables were similar but provided data for different time periods in the two surveys.

When the comparisons were made, estimates of sampling errors were only available for a few of the NIECS variables, so it was often not possible to determine which of the NIECS/AHS differences were statistically significant. Unlike the AHS, the NIECS did not cover Alaska and Hawaii, but the study report does not mention whether the AHS data were adjusted to take account of this difference in coverage. For a few of the variables compared, the AHS estimates included vacant units, which were excluded from NIECS. Some highlights of the comparisons were:

- At both the national and regional levels, there was a clear tendency for the NIECS family income distribution to show a higher proportion of families in the upper income categories. This tendency was especially pronounced in the South region, which showed the following distribution:

1977 family income class	Percent of families	
	NIECS	AHS
Below $5,000	14.7	20.9
$5,000 - 9,999	22.0	22.5
$10,000 and over	62.3	56.5

  However, in the comparisons of RECS and CPS income data, presented later in this section, the difference was in the opposite direction. The CPS uses more detailed income questions than either RECS or the American Housing Survey.

- A similar tendency was noted for the distribution of property values for owner-occupied housing units.

- The proportion of households with one member was smaller for NIECS (18.8 percent) than for the AHS (22.2 percent).

The general conclusion of the study was that for most variables there was reasonably good agreement between the NIECS and AHS estimates.

The American Housing Survey: Other Comparisons

The *Consumption and Expenditures* report for 1993 (EIA 1995d) includes a comparison of the distributions of occupied housing units by year built, as estimated from the 1993 RECS and the 1993 AHS. The results are shown in Table 8.5. The two distributions are in reasonably good agreement, but the proportion of units built between 1970 and 1979 as estimated by RECS was significantly below the corresponding AHS estimate.

The Current Population Survey (CPS)

As explained in Chapter 7, Section 7.1, estimated household counts from the annual March supplement to the CPS are used to derive the benchmark values for the stage two ratio adjustments that are part of the RECS estimation procedure. Hence, for the categories used as benchmarks, RECS and CPS estimates are in close agreement. For the first 5 survey years, 12 control totals were used, defined by the four Census regions and three location categories -- central city, remainder of metropolitan statistical area, and nonmetropolitan. However, comparisons of RECS and CPS estimates of the number of households by number of persons for 1980, 1981, and 1982 showed that the proportion of single-person households in RECS was consistently low for both owners and renters (Response Analysis Corporation 1983).

Consequently, for the 1984 RECS stage 2 ratio-estimation procedure, additional benchmark categories were introduced for one-person households occupied by males, one-person households occupied by females, and all other households.

Data on household income are also collected annually in the March supplement to CPS. The CPS procedures for collecting data on income are more elaborate than those used in RECS. The RECS questionnaire asks respondents whether or not any *family members* had income in each of several categories (earnings, self-employment, Social Security, etc.) and then asks them to assign their total *family* income to one of a large number of income class intervals. Income of persons living in the household who are not members of the family is supposed to be excluded. The CPS questionnaire calls for actual dollar amounts in each of several income categories separately for each *household* member age 15 and over. The time references also differ: RECS asks for income in the 12 months preceding the interview date (generally in the fall of the year), whereas the March CPS asks for income in the prior calendar year.

Table 8.5. RECS/AHS Comparisons of Occupied Housing Units by Year Built: 1993

Year of Construction	Percent of Housing Units[a]		
	AHS	RECS	RECS - AHS
1939 or earlier	21.0	21.0	0.1
1940 to 1949	8.0	7.1	-0.8
1950 to 1959	13.0	13.5	0.5
1960 to 1969	15.2	15.5	0.3
1970 to 1979	22.0	18.8	-3.2[b]
1980 to 1984	7.6	8.8	1.2
1985 to 1989	8.4	9.1	0.7
1990 to 1993	4.8	6.1	1.2

[a]Percents may not add to 100.0 due to rounding.
[b]Difference is statistically significant at the 95 percent confidence level.
Source: Energy Information Administration, *Consumption and Expenditures* (1993), Appendix B.

Detailed comparisons of RECS and CPS income data for 1980, 1984, and 1990 were undertaken by Response Analysis Corporation (1994) as part of an analysis of alternative measures of energy burden--that is, the share of income used to pay energy bills. Estimates of median income for the 3 years were as follows:

Year	RECS	CPS	Percent difference (RECS - CPS)/RECS
1980	$16,172	$17,434	-7.8
1984	$19,488	$22,200	-13.9
1990	$26,364	$29,306	-11.2

The values shown for CPS represent total income of all household members. For 1990, it was possible to calculate median *family* income for CPS; that value, $27,915, was closer to the RECS estimate, the difference being -5.9 percent of the RECS value.

Table 8.6 shows comparisons of RECS and CPS income distributions for 1989 and 1990. The RECS distributions are based on family income and the CPS distributions are based on household income. Compared to RECS, the CPS distributions place a significantly higher proportion of households in the two top income classes. The differences might have been smaller if the CPS distributions had been based on family income, excluding nonfamily members in the sample households.

Table 8.6. RECS/CPS Family Income Comparisons: 1987 and 1990

Income Category[a]	Percent of Households[b]			
	1987		1990	
	RECS	CPS	RECS	CPS
Less than $5,000	6.8	6.9	5.6	5.2
$5,000 to 9,999	12.7	11.5	11.4	9.7
$10,000 to 14,999	13.9	10.6	12.1	9.5
$15,000 to 19,999	10.0	10.0	9.0	8.8
$20,000 to 24,999	9.7	9.2	9.6	8.9
$25,000 to 34,999	17.9	16.1	16.2	15.8
$35,000 to 49,999	14.8	17.2	17.8	17.5
$50,000 and over	14.3	18.5	18.4	24.7

[a]Income of family members for RECS, household members for CPS.
[b]Percents may not add to 100.0 due to rounding.
Sources: Energy Information Administration, *Housing Characteristics* (1987), Appendix C; *Consumption and Expenditures* (1990), Appendix C.

The Decennial Housing Census

Most housing characteristics that appeared in both the 1980 RECS and the 1980 Census of Housing were in reasonably good agreement. One exception was the number of households using wood as their main heating fuel (Carlson 1985). Estimates from the two sources were as follows:

Data Source	Households Using Wood as Main Heating Fuel	
	Estimate	Two Standard Errors
RECS (Nov. 1980)	4,700,000	800,000
Census (April 1980)	2,575,560	7,060

The 1980 Annual Housing Survey estimated that 1,377,000 housing units (± 101,000) used wood as their main heating fuel in 1980. However, unlike the RECS and Housing Census inquiries on main heating fuel, which were quite similar, the AHS inquiry did not provide a separate response category for wood.

Possible reasons for the difference between the RECS and Housing Census estimates include:

- Timing. According to RECS estimates, the proportion of households using wood as their main heating fuel rose steadily from 2.5 percent in 1978 to 6.4 percent in 1981. As noted above, the reference date for the 1980 RECS was 7 months later than the Census date.

- The RECS questionnaire gave greater emphasis to the use of wood as a fuel. It had several specific questions about wood, covering all types of uses and amounts used. Questions about secondary heating fuels and equipment were included. Wood is often used in conjunction with other heating fuels.

The Consumer Expenditure Survey (CES)

Since 1980, the CES, which is conducted by the Census Bureau for the Bureau of Labor Statistics, has provided annual estimates of household expenditures in a large number of categories, including natural gas, electricity, and fuel oil. Branch (1994) has compared CES estimates of expenditures for these fuels with RECS estimates for 1984, 1987, and 1990. The results are shown in Table 8.7. There was an apparent error that affected the published ratios for electricity and the total for all major fuels for 1987; the values shown in the table differ from those published by Branch.

The CES estimates of expenditures on electricity were above the RECS estimates for all 3 years. Because electricity accounts for more than half of the total for the three fuels combined, the CES estimates of totals for all major fuels also exceed the RECS estimates in each year. The publication that was the source of the CES estimates does not provide sampling errors but, based on the RECS sampling errors, the 1984 and 1987 ratios for electricity are clearly significantly different from 1.00 at the 95-percent confidence level. The ratios shown for the other fuels and for electricity in 1990 are probably not significantly different from 1.00.

The RECS estimates used for these comparisons were estimates of consumption of each fuel by households that paid for all of their uses of that fuel. Branch states that this population "... more closely matches the population covered in CE estimates for energy expenditures," but does not explain what differences, if any, there are. The CES estimates used for the comparisons were adjusted to eliminate energy expenditures associated with vacation properties.

Other factors that might be associated with differences in the two sets of estimates include:

- CES estimates are for the calendar year in each of the 3 years. The RECS estimates for 1984 covered the period from April 1984 through March 1985.

- CES estimates may include some expenditures by households that do not pay for all of their uses of a particular fuel. These households were excluded from the RECS estimates that were used for the comparisons.

- Military households on post are included in RECS but not in the CES.

- For 1984 and 1987, the CES fuel oil expenditures were compared with RECS expenditures for fuel oil and kerosene combined.

- About 15 percent of households use budget plans to pay their suppliers; these plans allow them to spread their costs more evenly over the year. RECS consumption estimates are based on amounts actually supplied, whereas CES estimates are based on amounts paid.

Table 8.7. Comparison of Aggregate Expenditures for Selected Fuels, Consumer Expenditure Survey (CES) and RECS: 1984, 1987, and 1990

Expenditure Category	CES[a] (in billions)			RECS (in billions)			Ratio: CES/RECS		
	1984	1987	1990	1984	1987	1990	1984	1987	1990
Natural Gas	$26.5	$21.8	$23.8	$25.0	$21.7	$23.3	1.06	1.00	1.02
Electricity	58.0	64.7	73.0	51.8	58.5	68.6	1.12	1.11	1.06
Fuel Oil[b]	7.4	5.4	6.2	7.4	5.8	6.5	1.00	0.93	0.95
Major Household Fuels, Total	91.9	91.9	103.0	84.3	86.0	98.4	1.09	1.07	1.05

[a]CES estimates were adjusted to exclude expenditures for owned or rented vacation property.
[b]For 1984 and 1987, RECS estimates for fuel oil include estimates for kerosene.
Sources: Branch, *The Consumer Expenditure Survey: A Comparative Analysis* (1994); Energy Information Administration, *Consumption and Expenditures* (for years shown).

Comparisons of RECS and Administrative Data

As part of its income inquiry, RECS asks respondents about receipt of food stamps. In the 1981 and 1982 surveys, they were asked about receipt during the calendar year prior to the survey; subsequently they have been asked about receipt during the 12 months prior to the survey interview. Since most interviews take place in the fall of the survey year, the latter approach is roughly equivalent to asking about receipt during the fiscal year that runs from October of the year preceding the survey year to September of the survey year.

Table 8.8 shows the results of a comparison, for selected survey years, of RECS estimates of the number of households receiving food stamps with counts based on records maintained by the Agriculture Department's Food and Nutrition Service, which administers the Food Stamp Program (Thompson 1994b). Estimates from RECS were below the program counts for all years shown and, except for the 1982 and 1984 RECS, the survey estimates were significantly lower.

Table 8.8. Comparison of Number of Households Receiving Food Stamps, RECS Estimates and Program Counts: Selected Years

Year Food Stamps Received[a]	Number of Households (000)		Ratio: RECS/USDA
	RECS	USDA	
1980	6,777	7,718	0.88*
1981	6,724	7,249	0.93
FY 1984	7,348	7,580	0.97
FY 1987	5,568	7,122	0.78*
FY 1990	6,010	7,787	0.77*

[a]The 1981 and 1982 RECS asked about receipt of food stamps during the prior calendar year. Subsequent surveys asked about receipt during the past 12 months, which is roughly equivalent to the fiscal year because interviewing is done in the fall.

* = Ratio differs from 1.00 by more than twice its standard error.

Sources: Energy Information Administration, RECS: Survey data for 1981, 1982, 1984, 1987, and 1990; USDA: Food and Nutrition Service, Public Information Data Bank and National Data Bank, January 1993.

A similar comparison with program data has been made for RECS estimates of the number of households receiving assistance under the Low-Income Home Energy Assistance Program, which was authorized by 1981 legislation and is currently administered by the Administration for Children and Families, Department of Health and Human Services (Thompson 1994b). The results of the comparison, which is based on assistance for home heating costs only, are shown in Table 8.9. In this instance, the RECS estimates are significantly below the program counts for all years. For the 3 years shown in both tables, the observed ratio of RECS estimates to program counts was lower for energy assistance than it was for food stamps.

These findings for RECS are consistent with experience from other household surveys which have attempted to collect data on income recipiency from public income transfer programs. Comparisons with administrative data for 1983 and 1984 showed that the Census Bureau's Survey of Income and Program Participation, which uses a considerably more detailed set of income questions, was identifying about 90 percent of the households receiving food stamps and that a somewhat smaller proportion of the total amounts disbursed was being reported. For calendar 1983, the Current Population Survey estimate of the total value of food stamps received was about 71 percent of the figure provided by the Food and Nutrition Service (Jabine 1990, Table 10.1).

Possible reasons for such underreporting in surveys include respondent reluctance to report receipt of welfare payments, respondent misclassification of the source of income, and survey undercoverage of low-income households. Differences in the frequency and method of receipt may affect the level of reporting. Households receive food stamps every month and take them or a debit card to the store where they buy their food. Energy assistance, on the other hand, is received once a year and the payment may be sent directly to a utility with only a notice to the household recipient (Thompson 1994b, p.6).

Table 8.9. Comparison of Number of Households Receiving Low-Income Home Energy Assistance, RECS Estimates, and Program Counts: Selected Years

Year Heating Assistance Recevied[a]	Number of Households (000)		Ratio: RECS/HHS
	RECS	HHS	
FY 1982	3,908	5,990	0.65*
FY 1984	5,293	6,444	0.82*
FY 1987	4,770	6,495	0.73*
FY 1990	4,156	5,460	0.76*

[a]The RECS questionnaire asks about receipt of LIHEAP assistance during the fiscal year preceding the survey interview.

* = Ratio differs from 1.00 more than twice its standard error.

Sources: Energy Information Administration, RECS: Survey data for 1982, 1984, 1987, and 1990; HHS: Low Income Home Energy Assistance Program reports to Congress for the fiscal years shown.

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.

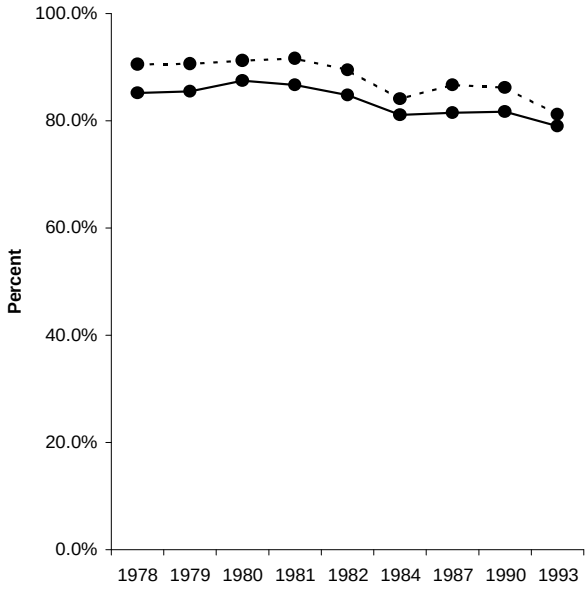
As can be seen in Figure 9.1 and Table 4.1, unweighted interview completion rates remained fairly steady, around 85 percent, from 1978 through 1982. For the next three survey years, they were at a lower level, between 81 and 82 percent, and in 1993 they declined to their lowest level, 79 percent. The 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.

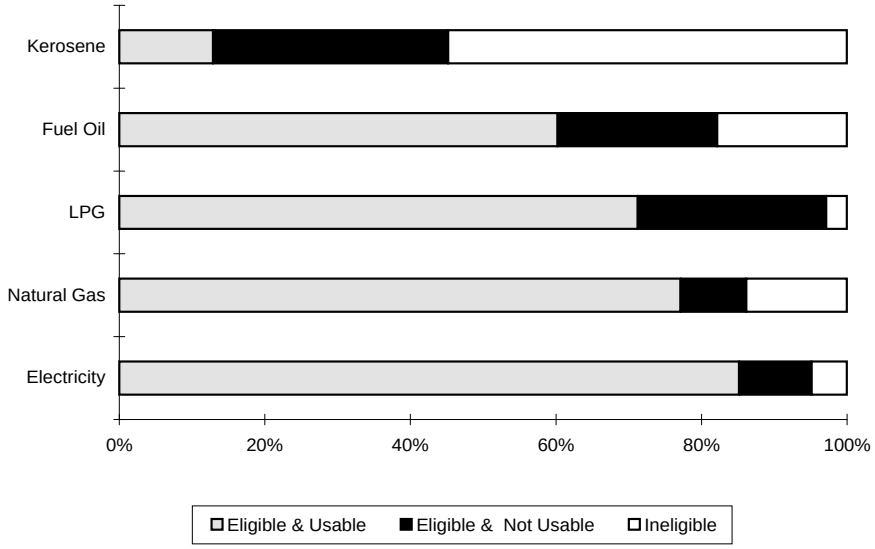
Nonresponse: Supplier Survey

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.

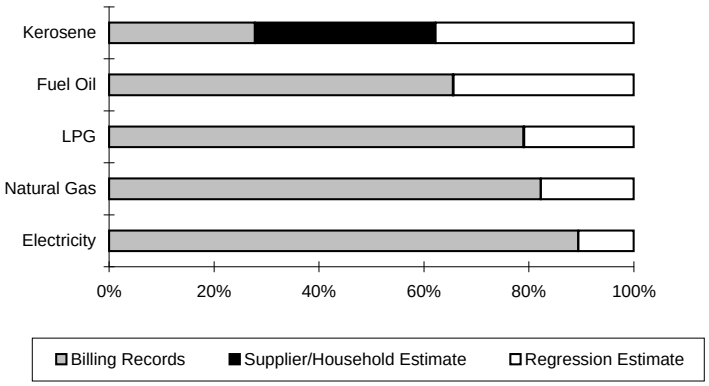
As shown in Figure 9.3 and Table 6.5, the proportion of total consumption that was imputed by the end-use model was lowest for electricity (about 10 percent in 1990 and 1993) and highest for fuel oil and kerosene (each between 25 and 35 percent in those years). Figure 9.4 shows that natural gas and electricity 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.

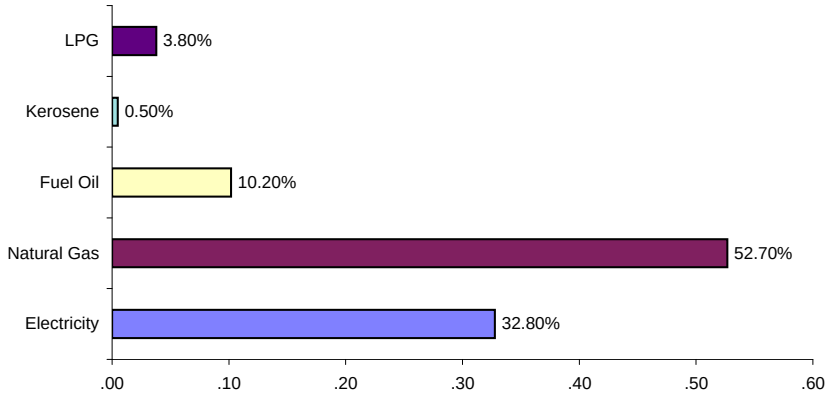
There are no systematic continuing sources of information about measurement error in RECS. As noted in Chapter 5, information about response variance, interviewer variance, and bias has come largely from occasional studies, in some instances restricted to a small set of sample households and often providing 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 inadmissable 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:

Category	Fuels
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).

Appendix A

References

Allied Technology Group, Inc. (1995). *Revised Analysis Report: Comparison of Data from Energy Consumption and Supply Surveys*, Task Assignment No. 94-004. Rockville, MD, March 21.

Battles, Stephanie J. (1990). *Comparison Between Residential End-Use Submetering and Conditional Demand Estimates for a National Survey*, Presented at the ACEEE 1990 Summer Study on Energy Efficiency in Buildings.

Battles, Stephanie J. (1991a). *Comparison of Short Run Changes in Residential Energy Consumption and Expenditures in Black and Nonblack Households*, Presented at the Socioeconomic Energy Research and Analysis Conference, June 28.

Battles, Stephanie J. (1991b). Technical Note No. 32. *Effects of the Adjustment of 1990 Census Data on the 1990 RECS Control Totals Obtained from the Current Population Survey*, December 30.

Battles, Stephanie J. and Harrison, Ivy M. (1992). *Setback and Discrete Temperature Behavior in the United States as Related to Housing Unit and Household Characteristics*, Presented at the International Association of Energy Economics Conference, France.

Battles, Stephanie J. and Thompson, Wendel (1991). Technical Note No. 31. *Observation of a 1990 RECS Interviewer and the Implications for the 1993 RECS*, December 30.

Biemer, Paul P. (1992). Brief evaluation of an early draft of the questionnaire for the 1993 RECS. Letter to Lynda Carlson, October 9.

Blumstein, Carl J., Kemp, William, McGuire, C. Bart, Valentine, Susan D., and York, Carl M. (1982). *Supplemental Studies of the National Interim Energy Consumption Survey (NIECS)*, ERG-82-4, Energy Resources Group. University of California, Berkeley, July.

Blumstein, Carl J., York, Carl M., and Kemp, William (1981). *An Assessment of the National Interim Energy Consumption Survey*, ERG-81-1, Energy and Resources Group. University of California, Berkeley, July.

Branch, E. Raphael (1994). "The Consumer Expenditure Survey: A Comparative Analysis." *Monthly Labor Review*, U.S. Department of Labor, December.

Carlson, Lynda T. (1985). Letter to Arthur Young, Chief, Housing Division. Bureau of the Census, July 18.

Carroll, David (1987). "RECS Tabulations Report." Memorandum to Martha Johnson, March 5.

Cox, Brenda G. (1995). Transcript of remarks at meeting of the American Statistical Association Committee on Energy Statistics. Washington, DC, April 21.

Daniels, Sally (1992). "Draft of Report of Lighting Questions Pretest." Memorandum to Wendel Thompson, November 16.

Day, Dawn (1982). "Discrepancies Between Day and Night Temperatures in RECS-2." Memorandum to Wendel Thompson, November 3.

Energy Information Administration (1980). *Residential Energy Consumption Survey: Consumption and Expenditures, April 1978 Through March 1979, Part 1: National Data*. DOE/EIA-0207/5. Washington, DC, July 1980.

Energy Information Administration (1981). *Residential Energy Consumption Survey: 1979-1980 Consumption and Expenditures, Part 1: National Data*. DOE/EIA-0262/1. Washington, DC, April 1981.

Energy Information Administration (1982a). *Residential Energy Consumption Survey: Housing Characteristics, 1980*. DOE/EIA-0314(80). Washington, DC, June 1982.

Energy Information Administration (1982b). *Residential Energy Consumption Survey: Housing Characteristics, 1980-1981*. DOE/EIA-0321(80). Washington, DC, September 1982.

Energy Information Administration (1983a). *Residential Energy Consumption Survey: Housing Characteristics, 1981*. DOE/EIA-0314(81). Washington, DC, August 1983.

Energy Information Administration (1983b). *Residential Energy Consumption Survey: Housing Characteristics, 1981-1982*. DOE/EIA-0321(81). Washington, DC, September 1983.

Energy Information Administration (1983c). *Residential Energy Consumption Survey: Regression Analysis of Energy Consumption by End Use*, DOE/EIA-0431. Washington, DC, October 1983.

Energy Information Administration (1984a). *Residential Energy Consumption Survey: Housing Characteristics, 1992*. DOE/EIA-0314(82). Washington, DC, August 1984.

Energy Information Administration (1984b). *Residential Energy Consumption Survey: Housing Characteristics, 1982-1983*. DOE/EIA-0321(82). Washington, DC, November 1984.

Energy Information Administration (1984c). *Residential Energy Consumption and Expenditures by End Use for 1978, 1980, and 1981*, DOE/EIA-0458. Washington, DC, December 1984.

Energy Information Administration (1985). *Technical Documentation, Public Use Tape, 1982 Residential Energy Consumption Survey and 1983 Residential Transportation Energy Consumption Survey*, Washington, DC.

Energy Information Administration (1986a). *An Assessment of the Quality of Selected EIA Data Series: Energy Consumption Data*, DOE/EIA-0292(85). Washington, DC, April 2.

Energy Information Administration (1986b). *Nonresidential Buildings Energy Consumption Survey: Consumption and Expenditures 1985*, DOE/EIA-0318(83), Washington, DC, September 1986.

Energy Information Administration (1986c). *Residential Energy Consumption Survey: Housing Characteristics 1984*, DOE/EIA-0314(84). Washington, DC, October 1986.

Energy Information Administration (1987a). *Residential Energy Consumption Survey: Consumption and Expenditures April 1984 Through March 1985, Part 1: National Data*, DOE/EIA-0321(84). Washington, DC, March 1987.

Energy Information Administration (1987b). *Residential Energy Consumption Survey: Consumption and Expenditures April 1984 Through March 1985, Part 2: Regional Data*, DOE/EIA-0321(84). Washington, DC, May 1987.

Energy Information Administration (1987c). *Residential Energy Consumption Survey: Trends in Consumption and Expenditures 1978-1984*, DOE/EIA-0482. Washington, DC, June 1987.

Energy Information Administration (1987d). *Technical Documentation, Public Use Tape: 1984 Residential Energy Consumption Survey and 1985 Residential Transportation Energy Consumption Survey*. Washington, DC.

Energy Information Administration (1989a) *Housing Characteristics 1987*, DOE/EIA-0314(87). Washington, DC, May 1989.

Energy Information Administration (1989b). *Household Energy Consumption and Expenditures 1987, National Data*, DOE/EIA-0321(87). Washington, DC, October 1989.

Energy Information Administration (1989c). *An Assessment of the Quality of Selected EIA Data Series: Electric Power Data*, DOE/EIA-0292(87). Washington, DC.

Energy Information Administration (1990). *Energy Consumption by End-Use Sector, A Comparison of Measures by Consumption and Supply Surveys*, DOE/EIA-0533. Washington, DC, April 6, 1990.

Energy Information Administration (1992). *Housing Characteristics 1990*, DOE/EIA-0314(90). Washington, DC, May 1992.

Energy Information Administration (1993a). *Household Energy Consumption and Expenditures 1990*, DOE/EIA-0321(90). Washington, DC, February 1993.

Energy Information Administration (1993b). *Supporting Statement for Forms EIA-457 A-H: Residential Energy Consumption Survey*, Submitted to the Office of Management and Budget for clearance of the 1993 RECS, April 2, 1993.

Energy Information Administration (1993c). *User-Needs Study for the 1993 Residential Energy Consumption Survey*, DOE/EIA-0555(93)/2, Washington, DC, September 1993.

Energy Information Administration (1994). *Sample Design for the Residential Energy Consumption Survey*, DOE/EIA-0555(94)/1, Washington, DC, August 1994.

Energy Information Administration (1995a). *Housing Characteristics 1993*, DOE/EIA-0314(93). Washington, DC, June 1995.

Energy Information Administration (1995b). *Buildings and Energy in the 1980's*, Energy Consumption Series, DOE/EIA-0555(95)/1, Washington, DC, June 1995.

Energy Information Administration (1995c). *1993 Residential Energy Consumption Survey: Public Use Diskettes, User's Guide*, Washington, DC, July 1995.

Energy Information Administration (1995d). *Household Energy Consumption and Expenditures 1993*, DOE/EIA-0321(93). Washington, DC, October 1995.

Energy Information Administration (1995e). Special tabulations showing energy consumption by fuel, building type and basis for annual estimate, 1990 and 1993 Residential Energy Consumption Surveys.

Energy Information Administration (1995f). Special tabulation showing the numbers of sample housing units added by the half-open interval for the 1984, 1987 and 1990 Residential Energy Consumption Surveys.

Energy Information Administration (1995g). *State Energy Data Report, 1993*, DOE/EIA-0214(93). Washington, DC, July 1995.

Erickson, Rita J. (1984). Letter to Dr. Dawn Day, Response Analysis Corporation, March 6.

Gonzalez, Maria E., Kasprzyk, Daniel, and Scheuren, Fritz (1994). "Nonresponse in Federal Surveys: An Exploratory Study," *Amstat News*, No. 208, April.

Groves, Robert M. (1989). *Survey Errors and Survey Costs*, New York: John Wiley & Sons, Inc.

Hall, Christy (1992). Technical Note No. 33, "Nameplate Data Collection in the 1990 Residential Energy Consumption Survey," March 18.

Harrison, Ivy (1993). "Nonlinear Regression Analysis 1990 RECS Energy Consumption Data," draft, 1993.

Hersey, James C. and McCarthy, John (1986). *Analysis of Response Rates in Residential Energy Consumption Surveys, Final Report*, SRA Technologies, Inc., June 1986.

Jabine, Thomas B. (1987). *Review of Computer Edit and Update Performance Statistics for the Residential Energy Consumption Survey, Final Report*, Washington, DC, December 1987.

Jabine, Thomas B. (1993). "The EIA RECS Household Energy Consumption Data for 1990," Independent Review submitted to the Energy Information Administration, September 30, 1993.

Jabine, Thomas B. (1994). *Quality Profile for SASS*, NCES 94-340, National Center for Education Statistics, U.S. Department of Education, 1994.

Jabine, Thomas B. (assisted by King, Karen E. and Petroni, Rita J.) (1990). *SIPP Quality Profile*, Bureau of the Census, U.S. Department of Commerce, 1990.

Latta, Robert B. (1988). Technical Note No. 7, *Square Footage Imputations for RECS*, July 13, 1988.

Latta, Robert B. (1993). Draft Technical Note, *Poststratification Estimation*, May 7, 1993.

Latta, Robert B. (1994). *Analysis of Listing Undercount and Other Undercount Factors for RECS*, Technical Note, October 28, 1994.

Leach, Nancy L. (1991). "An Assessment of the Use of Videotape Interviewer Training for the Residential Energy Consumption Survey (RECS): The First Step." Presented at meeting of the American Statistical Association Committee on Energy Statistics, Washington, DC, October 24-25, 1991.

Luyben, Paul D. (1982). "Prompting Thermostat Setting Behavior," *Environment and Behavior*, Vol. 14, No. 1:113-128, January 1982.

Martin, Antoinette W. (1995). "An Examination of the Effects of the Automated Edits on Housing Characteristics for 1993," unpublished report, Energy Information Administration, November.

Miller, Renee (1995). "Why do Consumption Estimates Differ?", presented at the meeting of the American Statistical Association Committee on Energy Statistics Committee, Washington, DC, June 23, 1995.

Miksovic, Ann (1989). "Results of the Field Pretest of the RECS-90 Questionnaire." Memorandum to Robert Latta, June 20, 1989.

Mooney, Geraldine M. and Carroll, David (1986). "The 1986 Weather Methodology Evaluation," Princeton, NJ: Response Analysis Corporation, December 1986.

Response Analysis Corporation (1980). "Residential Energy Consumption Survey: Final Report on Study Design and National Interim Energy Consumption Survey," April 30, 1980.

Response Analysis Corporation (1983). "RECS/CPS Differences in Estimated Number of One person Households, Notes for Discussion," Memorandum, October 26, 1983.

Response Analysis Corporation (1987). "RECS/RTECS Methodological Study." Draft report, Princeton, NJ, November 24, 1987.

Response Analysis Corporation (1988). "Comparison of RECS4 End Use Estimates Using Individual Weather Station Data and Division Average Weather Station Data," Princeton, NJ, August 23, 1988.

Response Analysis Corporation (1991). "Quality Assessment of Videotape Training: Conclusions and Recommendations," September 30, 1991.

Response Analysis Corporation (1992a). "1990 Residential Energy Consumption Sample Design Procedures Manual," Princeton, NJ, March 1992.

Response Analysis Corporation (1992b). "1990 Residential Energy Consumption Survey Data Editing and Manipulation Procedures Manual," Princeton, NJ, September 1992.

Response Analysis Corporation (1992c). "End-Use Validation Study: Draft Report on Modified Models," Princeton, NJ, August 1992.

Response Analysis Corporation (1992d). "End-Use Validation Study: Final Report on Findings," Princeton, NJ, December 1992.

Response Analysis Corporation (1994). *Characterizing the Impact of Energy Expenditures on Low Income Households: An Analysis of Alternative Energy Burden Statistics: Technical Report*, Submitted to Energy Information Administration, Second draft, May 1994.

Response Analysis Corporation (1995a). *1993 Residential Energy Consumption Sample Design Procedures Manual*, Princeton, NJ, 1995.

Response Analysis Corporation (1995b). *1993 Residential Energy Consumption Survey Data Collection Procedures Manual*, Princeton, NJ, 1995.

Response Analysis Corporation (1995c). *1993 Residential Energy Consumption Survey Data Editing and Manipulation Procedures Manual*, Princeton, NJ, 1995.

Slider, Chuckie (1995). Memorandum to Wendel Thompson, Energy Information Administration, "RECS Quality Profile," Response Analysis Corporation, November 20.

Thompson, Wendel (1982). Letter to Reuben Cohen, Response Analysis Corporation, regarding apparently inconsistent reporting of indoor temperatures, September 15, 1982.

Thompson, Wendel (1985a). "Residential Energy Consumption Survey: From a 2-Year to a 3-Year Cycle." Presented at meeting of the American Statistical Association Committee on Energy Statistics, Washington, DC, October 24-25, 1985.

Thompson, Wendel (1985b). "Utility of Paying Respondents: Evidence from the Residential Energy Consumption Surveys." Presented at American Association for Public Opinion Research Annual Conference, May 17, 1985.

Thompson, Wendel (1987). "End-Use Consumption of Residential Energy," *Monthly Energy Review*, pp. vii-xiv, July 1987.

Thompson, Wendel (1994a). "Results of Trip to Los Angeles." Internal memorandum to Nancy Leach, February 25, 1994.

Thompson, Wendel (1994b). "Notes for 1993 RECS Reports," Selected pages from a memorandum to the files, August 2, 1994.

Windell, Phillip A. (1986). "The ELCAP Residential Base Study Sample: A Basic Characterization," *Proceedings from the ACEEE 1986 Summer Study on Energy Efficiency in Buildings: Program Evaluation*, American Council for an Energy-Efficient Economy, August 1986.

Appendix B

Related EIA Publications on Energy Consumption

For information about how to obtain these publications, see the inside cover of this report. Please note that the prices quoted here are subject to change.

In addition to the reports listed below, public-use data tapes and data diskettes for the residential, residential transportation, and commercial sectors are available from the National Technical Information Service (NTIS). To obtain information on how to order the tapes/diskettes, you may call NTIS at 703-487-4807, FAX number 703-321-8547. Data diskettes can also be obtained from the Office of Scientific and Technical Information (OSTI). For OSTI ordering information, call 615-576-8401.

Residential Sector

Housing Characteristics

Note: The survey name was dropped from the beginning of the report title starting with the 1987 data reports.

Housing Characteristics, 1993; June 1995, DOE/EIA-0314(93), GPO Stock No. 061-003-00912-3, $23.00.

Housing Characteristics 1990; May 1992, DOE/EIA-0314(90), GPO Stock No. 061-003-00754-6, $23.00.

Housing Characteristics 1987; May 1989, DOE/EIA-0314(87), GPO Stock No. 061-003-00619-1, $13.00.

Residential Energy Consumption Survey: Housing Characteristics 1984; October 1986, DOE/EIA-0314(84), GPO Stock No. 061-003-00499-7, $12.00.

Residential Energy Consumption Survey: Housing Characteristics, 1982; August 1984, DOE/EIA-0314(82), GPO Stock No. 061-003-00393-1, $7.00.

Residential Energy Consumption Survey Housing Characteristics, 1981; August 1983, DOE/EIA-0314(81), GPO Stock No. 061-003-00330-3, $6.50.

Residential Energy Consumption Survey: Housing Characteristics, 1980; June 1982, DOE/EIA-0314, GPO Stock No. 061-003-00256-1, $11.00.

Residential Energy Consumption Survey: Characteristics of the Housing Stock and Households, 1978; February 1980, DOE/EIA-0207/2, GPO Stock No. 061-003-00093-2, $4.25.

Residential Energy Consumption Survey: Conservation; February 1980, DOE/EIA-0207/3, GPO Stock No. 061--003-00087-8, $6.00.

Preliminary Conservation Tables from the National Interim Energy Consumption Survey; August 1979, DOE/EIA-0193/P (no GPO Stock No.).

Characteristics of the Housing Stock and Households: Preliminary Findings from the National Interim Energy Consumption Survey; October 1979, DOE/EIA-0199/P (no GPO Stock No. available).

Consumption and Expenditures

Note: The survey name was dropped from the beginning of the report title starting with the 1987 data reports. The titles were changed to *Household Energy Consumption and Expenditures 1987, Part 1: National* and *Part 2: Regional*.

Household Energy Consumption and Expenditures 1993; October 1995, DOE/EIA-0321(93), GPO Stock No. 061-005-00932-8, $21.00.

"Household Energy Consumption and Expenditures 1990," *Monthly Energy Review,* August 1993, DOE/EIA-0035(93/08).

Household Energy Consumption and Expenditures 1990; February 1993, DOE/EIA-0321/1(90), GPO Stock No. 061-003-00795-3, $22.00.

Household Energy Consumption and Expenditures 1990\S; DOE/EIA-0321/2(90), GPO Stock No. 061-003-00796-1, $21.00.

Household Energy Consumption and Expenditures 1987, Part 1: National Data; October 1989, DOE/EIA-0321/1(87), GPO Stock No. 061-003-00635-3, $15.00. Note: Energy end-use data are included in this report.

Household Energy Consumption and Expenditures 1987, Part 2: Regional Data; DOE/EIA-0321/2(87) (no GPO Stock No. available), $16.00.

Residential Energy Consumption Survey: Consumption and Expenditures, April 1984 Through March 1985, Part 1: National Data; March 1987, DOE/EIA-0321/1(84), GPO Stock No. 061-003-00519-5, $9.50.

Residential Energy Consumption Survey: Consumption and Expenditures, April 1984 Through March 1985, Part 2: Regional Data; May 1987, DOE/EIA-0321/2 (84), GPO Stock No. 061-003-00528-4, $17.00. Note: Energy end-use data are included in this report.

Residential Energy Consumption Survey: Consumption and Expenditures, April 1982 Through March 1983, Part 1: National Data; November 1984, DOE/EIA-0-321/1(82), GPO Stock No. 061-003-00411-3, $7.00.

Residential Energy Consumption Survey: Consumption and Expenditures, April 1982 Through March 1983, Part 2: Regional Data; December 1984, DOE/EIA-0-321/2(82), GPO Stock No. 061-003-00414-8, $9.50.

Residential Energy Consumption Survey: Consumption and Expenditures, April 1981 Through March 1982, Part 1: National Data; September 1983, DOE/EIA-0-321/1(81), GPO Stock No. 061-003-00340-1, $6.00.

Residential Energy Consumption Survey: Consumption and Expenditures, April 1981 Through March 1982, Part 2: Regional Data; October 1983, DOE/EIA-0321/2(81), GPO Stock No. 061-003-00357-5, $8.00.

Residential Energy Consumption Survey: Consumption and Expenditures, April 1980 Through March 1981, Part 1: National Data; September 1982, DOE/EIA-0321/ 1(80), GPO Stock No. 061-003-00278-1, $7.50.

Residential Energy Consumption Survey: Consumption and Expenditures, April 1980 Through March 1981, Part 2: Regional Data; June 1983, DOE/EIA-0321/2(80), GPO Stock No. 061-003-00319-2, $7.00.

Residential Energy Consumption Survey: 1979-1980 Consumption and Expenditures, Part 1: National Data (Including Conservation); April 1981, DOE/EIA-0262/1, GPO Stock No. 061-003-00191-2, $6.50.

Residential Energy Consumption Survey: 1979-1980 Consumption and Expenditures, Part II: Regional Data; May 1981, DOE/EIA-0262/2, GPO Stock No. 061-003-00189-1, $8.50.

Residential Energy Consumption Survey: Consumption and Expenditures, April 1978 Through March 1979; July 1980, DOE/EIA-0207/5, GPO Stock No. 061-003-00131-9, $7.50.

Single-Family Households: Fuel Oil Inventories and Expenditures: National Interim Energy Consumption Survey; December 1979, DOE/EIA-0207/1, GPO Stock No. 061-003-00075-4, $3.50.

Other Publications on the Residential Sector

Energy Consumption Series—*Sample Design for the Residential Energy Consumption Survey*, August 1994, DOE/EIA-0555(94)/1, GPO Stock No. 061-003-00865-8, $6.50.

Energy Consumption Series—*User-Needs Study of the 1993 Residential Energy Consumption Survey*, September 1993, DOE/EIA-0555(93)/2, GPO Stock No. 061-003-00819-4, $13.00.

"End-Use Consumption of Residential Energy" *Monthly Energy Review* (Article), pp. vii-xiv, July 1987, DOE/EIA-0035 (87/07).

Residential Energy Consumption Survey: Trends in Consumption and Expenditures 1978-1984 June 1987, DOE/EIA-0482, GPO Stock No. 061-003-00535-7, $12.00.

Residential Conservation Measures; July 1986, SR/EEUD/86/01 (no GPO Stock No.).
An Economic Evaluation of Energy Conservation and Renewable Energy Tax Credits; October 1985, Service Report (no GPO Stock No.).

Residential Energy Consumption and Expenditures by End Use for 1978, 1980, and 1981; December 1984, DOE/EIA-0458, GPO Stock No. 061-003-00415-6, $4.50.

Weatherization Program Evaluation, SR-EEUD- 84-1; August 1984 (available from the Office of the Assistant Secretary for Conservation and Renewable Energy, Department of Energy).

Residential Energy Consumption Survey: Regression Analysis of Energy Consumption by End Use; October 1983, DOE/EIA-0431, GPO Stock No. 061-00300-347-8, $5.00.

National Interim Energy Consumption Survey: Ex- ploring the Variability In Energy Consumption; July 1981, DOE/EIA-0272, GPO Stock No. 061-003-00-205-6, $5.00.

National Interim Energy Consumption Survey: Ex-ploring the Variability in Energy Consumption--A Supplement; October 1981, DOE/EIA-0272/S, GPO Stock No. 061-003-00217-0, $4.50.

Energy Use by U.S. Households; November 1980, DOE/EIA-0248 (brochure, no GPO Stock No.).

Cross-Sector

Energy Consumption Series-Measuring Energy Efficiency in the United States Economy: A Beginning, October 1995, DOE/EIA-0555(95)/2, GPO Stock No. 061-003-00935-2, $6.50.

Energy Consumption Series-Buildings and Energy in the 1980's, June 1995, DOE/EIA-0555(95)/1, GPO Stock No. 061-003-00914-0, $6.00.

Energy Consumption by End-Use Sector: A Comparison of Measures by Consumption and Supply Surveys; April 6, 1990, DOE/EIA-0533 (no GPO Stock No. available), $2.50.

Natural Gas: Use and Expenditures; April 1983, DOE/EIA-0382, GPO Stock No. 061-003-00307-9, $5.50.

Public-Use Tapes

Note: All tapes are available through the NTIS.

Residential and Residential Transportation Sectors

Residential Energy Consumption Survey: 1987 and Residential Transportation Energy Consumption Survey, 1988, Order No. PB90-501461, $220.

Residential Energy Consumption Survey: 1984 and Residential Transportation Energy Consumption Survey, 1985; Order No. PB87-186540, $220.

Residential Energy Consumption Survey: 1982 and Residential Transportation Energy Consumption Survey, 1983; Order No. PB85-221760, $220.

Residential Energy Consumption Survey: Consumption and Expenditures, 1980-1981; Monthly Billing Data; Order No. PB84-166230, $220.

Residential Energy Consumption Survey: Housing Characteristics, 1981; Consumption and Expenditures, 1981-1982; Monthly Billing Data; Order No. PB84-1-20476, $220.

Residential Energy Consumption Survey: Housing Characteristics, Annualized Consumption and Expenditures, 1980-1981; Order No. PB83-199554, $220.

Residential Energy Consumption Survey: Household Transportation Panel Monthly Gas Purchases and Vehicle and Household Characteristics, 6/79-9/81; Order No. PB84-162452, $220.

Residential Energy Consumption Survey: Household Screener Survey, 1979-1980; Order No. PB82-114877, $220.

Residential Energy Consumption Survey: Household Monthly Energy Consumption and Expenditures, 1978-1979; Order No. PB82-114901, $220.

National Interim Energy Consumption Survey (Residential), 1978; Order No. PB81-108714, $220.

Public-Use Diskettes

Note: Diskettes are available through the Office of Scientific and Technical Information (OSTI) and NTIS.

Residential Energy Consumption Survey 1990 Data, OSTI-ASCII (3 diskettes) or dBase (2 diskettes) format, order by title, $10.00 per diskette, NTIS-ASCII format, Order No. PB93-506103 or dBase format, Order No. PB93-506095.

Residential Energy Consumption Survey 1987 Data, **OSTI** - ASCII or dBase format, order by title, $10 per diskette, $40 set of four. **NTIS** - ASCII format: Order No. PB-91-505115, $130, and dBase format: Order No. PB-91-505107, $130.