Independent Statistics & Analysis
U.S. Energy Information Administration

Residential Energy Consumption Survey (RECS)

2009 Technical Documentation-Summary

January 2013

Independent Statistics & Analysis
www.eia.gov

U.S. Department of Energy
Washington, DC 20585

This report was prepared by the U.S. Energy Information Administration (EIA), the statistical and analytical agency within the U.S. Department of Energy. By law, EIA's data, analyses, and forecasts are independent of approval by any other officer or employee of the United States Government. The views in this report therefore should not be construed as representing those of the Department of Energy or other Federal agencies.

Table of Contents

Overview and History of RECS 1
Data Products 3
Coverage and Sample Design 4
Sample selection 4
Household Survey 6
Questionnaire content and survey instrument 6
Square Footage Measurements 7
LIHEAP Supplement 7
Data collection and response rates 7
Rental agent survey 8
Editing and data quality 8
Item imputation 8
Weather and geographic data 9
Energy Supplier Surveys (ESS) 10
Overview 10
Data collection and response rates 10
Editing and data quality 10
Consumption and Expenditures Estimates 12
Annualization and imputation 12
End-use estimation 12
Weighting and Sampling Error 14
Base weights and adjustments 14
Relative standard errors (RSE) 14
Confidentiality of information 15

Tables

Table 1. Comparison of 2005 and 2009 RECS sample allocation and results 5

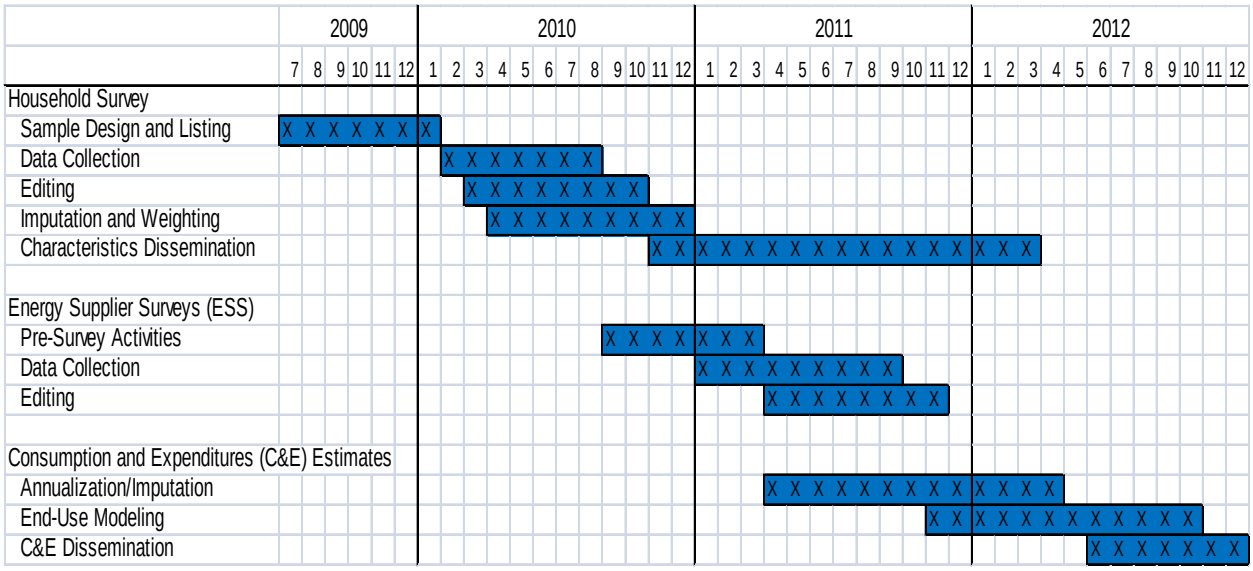
Figures

Figure 1. Approximate timeline for the 2009 RECS 1
Figure 2. Energy end-use equation 13

Overview and History of RECS

RECS is a periodic survey[1] sponsored by the U.S. Energy Information Administration (EIA) that provides detailed information about energy usage in U.S. homes. RECS is a multi-year effort (Figure 1.) consisting of a Household Survey phase, data collection from household energy suppliers, and detailed consumption and expenditures estimation. The Household Survey collects data on energy-related characteristics and usage patterns of a nationally representative sample of housing units. For renters that do not directly pay for their energy usage, a supplementary Rental Agent Survey is conducted. The Energy Supplier Surveys (ESS) collect data on how much electricity, natural gas, propane/LPG, fuel oil, and kerosene were consumed in the sampled housing unit during the reference year. It also collects data on actual dollar amounts spent on these energy sources. EIA uses a non-linear statistical model to produce consumption and expenditures estimates for heating, cooling, refrigeration and other end uses in all housing units occupied as a primary residence in the United States. Originally conducted with paper and pencil, RECS now uses a combination of Computer-Assisted Personal Interview (CAPI), internet, and mail to collect data for the Household and Energy Supplier Surveys.

Figure 1. Approximate timeline for the 2009 RECS



The scope and purpose of RECS differ slightly from similar EIA products that report "residential" energy data. For one, RECS samples homes occupied as a primary residence, which excludes secondary homes, vacant units, military barracks, and common area in apartment buildings. RECS estimates, therefore, do not represent sector level estimates, but are best suited for those wishing to compare across different characteristics of homes within the residential sector. RECS "totals" also exclude consumption and expenditures estimates for which data are difficult to obtain, such as estimates for biomass (wood) or solar consumption.

[1] In the late 1970's, EIA developed RECS from two surveys, the National Interim Energy Consumption Survey and the Household Screener Survey. RECS was administered annually between 1978 and 1982, triennially between 1984 and 1993, and has been a quadrennial survey (conducted every four years) ever since. The 2009 RECS is the 13th survey in the series.

In the spirit of continuous improvement and in collaboration with the data user community, EIA instituted the following survey design revisions, content changes, and variable updates for the 2009 RECS:

- A nearly threefold increase in the sample size. This allowed EIA to release estimates for household characteristics and energy use for 16 States, 12 more than in past rounds of RECS. The total number of responding households increased from 4,382 in the 2005 RECS to 12,083 in the 2009 RECS.
- A new sample frame development methodology. The majority of the sample frame for the 2009 RECS was constructed using the U.S. Postal Service mail address database, the Delivery Sequence File (DSF).
- Expanded data collection on the type and usage of consumer electronics, including televisions and related devices, computers, and personal electronic devices.
- Finer resolution on age of appliances to align with benchmark years for efficiency standards. The "10-19 years old" equipment and appliance age range from previous surveys was split into two responses; it is now split into two age groups, 10-14 and 15-19 years.
- New data items for recent energy efficiency actions taken by the household including caulking, weatherstripping, insulation, and home energy audits.
- Standardized, Internet-based data collection for the Energy Supplier Surveys that allowed companies to respond via online forms, structured spreadsheet, or paper forms via mail.

The RECS is authorized under the Federal Energy Administration Act of 1974 (Public Law 93-275) as amended, and the Energy Policy Act of 1992.

Data Products

EIA has released a variety of RECS products across survey cycles tailored to a wide range of data users. This includes summary level tables of energy-related household characteristics, detailed tabulations of energy consumption intensities across key variables, and public use micro-data for customized analysis of home energy use. Because of the expanded sample size, many of the 2009 RECS products are available at more granular levels.

Articles highlighting key findings from the 2009 RECS are available in addition to the standard table and microdata file. All current and historical RECS products can be found here. (link: http://www.eia.gov/consumption/residential/index.cfm)

RECS data are also used as critical inputs for EIA sector level forecasts and DOE program analysis, including the Annual Energy Outlook and the Annual Energy Review.

Coverage and Sample Design

As in previous survey years, the 2009 RECS used a multistage area probability sample design, where the universe was broken up into successively smaller, statistically selected areas starting from counties and ending with individual housing units. The universe for the 2009 RECS sample design included all housing units that are occupied as primary residences in the 50 States and the District of Columbia. Because EIA benchmarks to occupied housing totals from the U.S. Census Bureau's American Community Survey (ACS), the RECS uses the Census Bureau definition of a housing unit. (link: http://www.eia.gov/consumption/residential/terminology.cfm#h) It includes single-family homes, units in multi-family buildings, and mobiles. The RECS excludes vacant, seasonal or vacation homes, and group quarters such as prisons, military barracks, dormitories, and nursing homes.

The sample design for the 2009 RECS expanded on the design used for the previous round. It retained the structural definitions in the 2005 RECS design but substantially added to the 2005 selections to achieve an almost triple sample size; 12,083 complete interviews compared to 4,382 in 2005.

Sample selection

The starting point of the RECS sample design was the subdivision of the entire United States into geographic domains. This subdivision allowed geographic control of the sample allocated to the Census Divisions and selected states (16 for the 2009 RECS.) Precision constraints for estimates of average energy consumption per household were then applied at the national, Census Region, Census Division, and reportable state levels to develop the sample allocation for the geographic domains. There were 29 geographic domains in the 2009 RECS design. Sixteen of them were reportable states, which were selected based on population and climatic or geographic diversity.

Sample selection began by randomly choosing counties (Primary Sampling Units or PSUs.) The selected counties were then sub-divided into groups of Census blocks called segments and a sample of segments was randomly drawn from the selected counties. Within each selected segment, a list of housing units (sample frame) was created from a combination of the Delivery Sequence File (DSF) and field listing. Addresses in approximately 88% of selected segments were obtained from the DSF. Field listing was used in the remaining segments, which were mostly rural areas where the DSF contained many post office boxes and rural route boxes that not suitable for an in-person survey sample frame.

The final sample of housing units was randomly selected from the housing unit list constructed from the selected area segments. This type of design is called multi-stage area probability sampling. Its proper application ensures that the selected sample statistically represents the entire population of occupied housing units in the United States.

Table 1. Comparison of 2005 and 2009 RECS sample allocation and results

	2005 RECS	2009 RECS
Primary sampling units (Counties)	180	430
Segments	1,449	3,898
Sample size	6,551	18,856
Households responding	4,382	12,083

Household Survey

Questionnaire content and survey instrument

The RECS Household Survey was conducted on a voluntary basis with respondents. The 2009 RECS again used trained professional interviewers to administer a standardized survey instrument programmed in Blaise software on laptops. The interview was conducted in person, allowing for interviewers to also measure the dimensions of the housing unit, data which is used to produce estimates of total, heated and cooled square footage.

Questionnaire content included the type and number of energy-consuming devices, usage patterns, structural characteristics of the home, household demographics, and energy supplier information. The sections of the questionnaire were as follows:

- Section A: Housing Unit Characteristics
- Section B: Kitchen Appliances
- Section C: Home Appliances and Electronics
- Section D: Space Heating
- Section E: Water Heating
- Section F: Air Conditioning
- Section G: Miscellaneous (included lighting, windows, and efficiency-related improvements)
- Section H: Fuels Used
- Section I: Housing Unit Measurements (when permission was provided)
- Section J: Fuel Bills
- Section K: Residential Transportation
- Section L: Household Characteristics
- Section M: Energy Assistance (for a subsample of housing units)
- Section N: Scanning of Fuel Bills

Some changes to the 2009 questionnaire included an increased focus on home electronics, new questions about efficiency-related improvements (including assistance or incentives received by households), and changes to the ordering. A complete list of questions asked during the 2009 RECS Household Survey interviews, as well as previous questionnaires, are available online here. (link: http://www.eia.gov/survey/#eia-457)

Although sample energy bills (Section N) were requested from respondents and scanned as part of the survey, these bills provide only a sample of information needed for the reference period and were not used as the basis for household energy consumption. These scanned bills are used to verify supplier contact information and in some cases to verify data received from the Energy Supplier Surveys.

In previous rounds of RECS an alternate paper questionnaire was mailed to nonresponding households. This paper questionnaire was not used for the 2009 RECS.

Square Footage Measurements

The size of a home is a key component associated with energy consumption. To ensure the highest quality of this characteristic, interviewers measured and recorded dimensions of the housing unit to determine its square footage. Respondents reported the number of stories in the housing unit, characteristics of the basement, attic, or garage, and the shape of each floor. Then the interviewer used a measuring tape to collect the dimensions of each floor of the housing unit.

The RECS square footage definition includes all areas of the home enclosed from the weather. This includes all basements, attics that are finished, heated, or cooled, and garages that are heating or cooled. This definition commonly differs from that used in real estate or from respondent estimates. More detailed information about RECS square footage measurements can be found here. (link: http://www.eia.gov/consumption/residential/reports/2009/methodology-square-footage.cfm)

LIHEAP Supplement

Section M of the 2009 RECS survey instrument included a series of questions to permit analysis of the federal residential energy subsidy programs. The Low Income Home Energy Assistance Program (LIHEAP) is a suite of federally funded programs to help eligible low-income households meet their home heating and/or cooling needs. LIHEAP is administered at the federal level by the Administration for Children and Families (ACF), within the U.S. Department of Health and Human Services.

Data collection and response rates

Household interviews averaged 51 minutes and were conducted with a household member knowledgeable about energy use in the home. In most cases, this was a householder. The data collection field period lasted six months (February to August 2010) and 419 interviewers conducted the data collection. Bilingual interviewers also conducted interviews in Spanish, where necessary.

To ensure the maximum coverage across the RECS population and validity of the survey data, a multi-contact approach was employed. Before the initial personal contacts, a post card alerting the respondent to the upcoming survey was mailed to each selected housing unit. At the conclusion of interviewer training, a letter stressing the purpose and importance of the survey was sent to each selected housing unit. Interviewers made several callbacks at different times of the day and different days of the week in an effort to minimize the number of non-contacted households. Furthermore, targeted letters were sent to gated communities and non-contact households to encourage response.

A total of 18,856 housing units were selected to participate in the 2009 RECS. About 15,300 of these were occupied, primary residences that were eligible for RECS. As data collection took place during 2010, some housing units were ineligible (out of scope) because the household residing there during data collection had not resided there in 2009. After all data collection attempts, 12,083 households completed a personal interview, yielding an overall cooperation rate (AAPOR COOP4) among eligible households of 79%. Very few cases were determined to be incomplete, resulting in an overall response rate (AAPOR RR3) of 79%, as well.

Rental agent survey

For renters that did not pay some or all of their energy bills directly, follow-up surveys were attempted with a rental agent or landlord to get more information about the housing unit. Respondents in the Rental Agent Survey were asked about the equipment and fuels used by tenant households for space heating, water heating, air conditioning, and cooking, as well as the method of bill payment. The interviews with rental agents were conducted concurrently with the household survey. The Household Survey interviews spawned 829 Rental Agent Surveys cases, of which 583 (70%) were successfully completed.

Although the Rental Agent Survey is not a building-level survey, new questions about common areas, such as laundry rooms, receptions areas, and gyms were added in 2009. This is in addition to existing questions on the number of units and floors in apartment building, as well as building square footage.

Editing and data quality

EIA employed many strategies to ensure data quality in the 2009 RECS Household Survey. Checks and edits were built into the Blaise interviewing software to reduce missing or invalid responses. Some asked respondents to verify rare or unusual responses (ex: “It is very unusual for a residence to contain no bathrooms, are you sure that there are no bathrooms in this residence?”). Other checks prevented respondents from providing illogical responses (ex: “Heat pumps do not use wood as a fuel source. Please select the correct type of heating equipment or the correct fuel used.”) Also, help screens within the survey instruments were available to interviewers who offered, as needed, standardized definitions or examples to help respondents provide accurate responses. For questions with numerous or complex response options, show cards were provided so respondents could follow along while the interviewer read them aloud.

Data for completed interviews were transmitted daily using a sample management system. All completed interviews were combined into one Blaise database for further processing. As soon as possible following the interview, the address was verified to ensure it was a valid sample case, interviewer comments were reviewed, and a subset of key questions was checked for completion and consistency. When problems remained, a follow-up telephone contact with a member of the household was made. Additionally, case validation by telephone was used to combat falsification of data by interviewers. At least 20% of each interviewer’s completed cases were verified.

Only the most essential edit checks were done during the actual survey interview in order to minimize burden and awkwardness. Upon receipt, EIA subjected the data to more thorough review including outlier checks, reviews of inconsistent responses, analysis of square footage data, interviewer comment review, and updates to household responses from Rental Agent Survey data.

Item imputation

Item nonresponse occurs when respondents do not know or refuse to answer certain items on the survey. A hot-deck imputation methodology, which fills in a missing value by copying a value from a similar case, was used for all variables with missing values. For each variable with missing values, a statistical model was used to determine a set of related variables. These related variables were used to identify statistically similar cases, which “donated” their values to the case with missing items. After imputation, final editing reviews ensured questionnaire skip patterns were performed correctly.

When an eligible sample case did not participate in the survey because a respondent could not be located or refused to participate during the interview period, all values were missing for the case. To address this unit non-response, sampling weights for responding cases were so adjusted to reduce biases that may be induced by differential non-response within the RECS population. Weighting adjustments are discussed in the section on Weighting and Sampling Error.

Weather and geographic data

EIA gathers weather and certain geographic indicators from other government agencies to complete the characteristics profile of sampled housing units. Average heating degree-days (HDD) and cooling degree-days (CDD) for 2009 are available from the National Climate Data Center (NCDC) for each weather station in the U.S. Each sampled RECS housing unit was matched to a local weather station in the area (most commonly the closest one to the house) then assigned the corresponding HDD and CDD values. Thirty-year HDD and CDD averages were also pulled from the NCDC data.

EIA also matched two official Census Bureau geographic identifiers to each sampled housing unit-Urban/Rural and Metropolitan/Micropolitan Statistical Area identifiers.

Energy Supplier Surveys (ESS)

Overview

Every four years the RECS program describes a snapshot of factors driving residential energy demand. The Household Survey asks respondents about their homes, their energy consuming equipment, and behavior, but most respondents are ill-equipped to provide data about their energy consumption and expenditures for an entire calendar year. To capture the most complete and accurate information EIA conducts the follow-on surveys with energy suppliers.

During the 2009 RECS household interviews, respondents provided names and contact information for their energy suppliers. The list of suppliers accumulated during the household survey served as the frame for each of the four Energy Supplier Surveys (ESS); Electricity (Form EIA-457E), Natural Gas (EIA-457F, LPG/Propane (EIA-457D), and Fuel Oil/Kerosene (EIA-457G). The frame for the 2009 RECS ESS included over 21,000 unique requests for household billing data from more than 1,500 energy suppliers. Fuel accounts for which household respondents did not pay energy bills directly (for example, a renter who pays electricity as part of monthly rent) were excluded from the frame.

Data collection and response rates

The Energy Supplier Surveys are mandatory for all utilities, municipalities, cooperatives, and bulk fuel dealers identified in the Household Survey. For the 2009 RECS, respondents were asked to provide energy billing data between September 2008 and April 2010, including the amount consumed or delivered, and total costs for each billing period. Electricity and Natural Gas suppliers were asked to provide data for each month. Bulk fuel suppliers (LPG/Propane and Fuel Oil/Kerosene) were instructed to provide any deliveries within the 20 month period, intending to cover the survey reference year (2009).

The Energy Supplier Surveys started data collection in February 2011 and concluded in September 2011. For the first time, a web-based data collection strategy was used as the primary data collection mode of response. A secure website and login were provided to energy suppliers, who responded either by using fill-able forms online or uploading an Excel spreadsheet. Some respondents opted to print and mail forms, or submitted printouts of customer billing records. (link: http://www.eia.gov/consumption/residential/methodology/2009/research-polling.cfm)

EIA received complete or partially complete data for 90% of the unique household customer requests, with 90% of the energy suppliers participating.

Editing and data quality

Similar to the Household Survey process, EIA employed a multi-phased approach for editing supplier billing data. First, critical edit checks were built into the Web instrument to ensure required fields were completed and to check for inappropriate formats in certain fields (such as a character in a data field). Upon receipt of data, analysts reviewed data inconsistencies, incomplete responses, outliers, and

respondent comments. In some cases, supplier respondents were contacted to re-submit or verify data. Data were checked across cases within the same supplier, and against cases within similar geographies and housing types. Analysts then reviewed data to ensure consistency with household interview responses.

Although data are collected on a monthly (or individual delivery) basis, RECS produces only annualized consumption and expenditures estimates for the reference year. The following sections describe 1) the processes used to calculate annual amounts using the edited billing data from the Energy Supplier Surveys, and 2) how EIA used the annualized amounts, along with the Household Survey data, to calculate individual energy end-use estimates.

Consumption and Expenditures Estimates

Annualization and imputation

After energy supplier data were collected and edited, a process called annualization needed to occur. This statistical process used the billing data to estimate a housing unit's consumption and expenditures between January 1, 2009 and December 31, 2009.

This step was a complex process for two main reasons. First, the time period covered by billing data did not match up exactly with 365 days in 2009. Second, the content of the billing data may not have matched the energy-consuming activities of the household interviewed. For example, the billing data may have included consumption for non-household uses, such as a farm or business. Like the Household Survey, imputation is also necessary when there is incorrect, incomplete, or missing billing data.

To arrive at final consumption and expenditures estimates for each case, and for each fuel that applied, EIA employed the following adjustments to the edited supplier data:

- Ratio adjusting to align billing data with 365 days in 2009.
- Augmenting partially complete billing data to inflate consumption and expenditures to an annual amount.
- Augmenting billing data that only include some of the end-uses in the household for that fuel. For example, applying an upward adjustment to an apartment's natural gas billing data when the apartment's water heating is supplied by a central, natural gas water heating system.
- Disaggregating billing data for an entire apartment building to the single housing unit sampled
- Discounting billing data that includes non-household consumption (e.g. data that include a farm or business).
- Imputing annualized amounts where billing data are incorrect, incomplete, or missing by using the 2005 RECS end-use regression models.

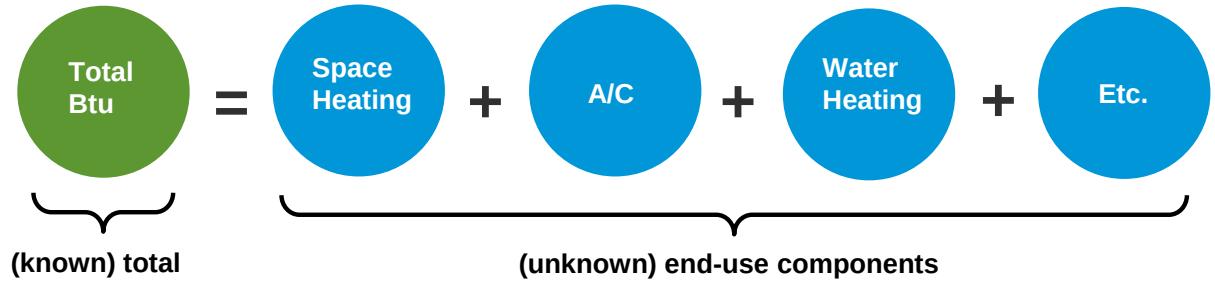
End-use estimation

An energy end-use model was used to break down the total, annualized consumption and expenditures for each sampled case into portions used for space heating, air-conditioning, water heating, refrigerators, appliances, and other uses. There are separate models for electricity, natural gas, fuel oil, LPG/propane, and kerosene.

The equation for each end-use model has total consumption for a fuel (the "known") as the dependent variable, and an appropriate combination of housing unit energy-related characteristics, appliance and

electronics information, household demographic variables, and weather data as independent variables. The model breaks down (or distributes) the total into end-use components (the "unknown.") Each end-use component is a non-linear expression relating the end use to its most relevant explanatory variables.

Figure 2. Energy end-use equation



Weighting and Sampling Error

Base weights and adjustments

The 2009 RECS sample was designed so that survey responses can be used to estimate characteristics, consumption, and expenditures of all occupied, primary housing units in the United States. In order to arrive at national estimates from the RECS sample, base sampling weights for each housing unit, which were the reciprocal of the probability of selection into the sample, were calculated. Therefore, a housing unit with a base weight of 10,000 represents itself and 9,999 similar, but un-sampled, housing units.

The base weights were adjusted for ineligibility (including unknown eligibility) and interview nonresponse. The final RECS analytical weights were poststratified to match control totals for energy consumption by fuel type and other factors. Adjustments were made to ensure that the RECS weights add up to 113,626,229, the total number of occupied, primary housing units in 2009 according to the American Community Survey conducted by the U.S. Census Bureau.

Relative standard errors (RSE)

Because the RECS sample is a subset of the entire U.S. population of occupied housing units, survey estimates will differ from their true population value. It is possible to estimate this difference, because in a statistical sample, the probability of a housing unit being selected is known. Sampling error provides a measure of the accuracy of a particular estimate for a characteristic based on how common and variable it is in the population, given a particular sample size.

Estimates from sample surveys are not exact but are statistical values with some associated sampling error in each direction—the result of generating estimates based on a sample rather than using the entire population.

A standard error is a measure of the reliability or precision of the survey statistic. The value for the standard error can be used to construct confidence intervals and to perform hypothesis tests by standard statistical methods. A relative standard error (RSE) is defined as the standard error (square root of the variance) of a survey estimate, divided by the survey estimate and multiplied by 100. In other words, the RSE is the standard error relative to the survey estimate on a scale from zero to 100. The larger the RSE, the less precise the survey estimate is of the true value in the population.

The 95-percent confidence interval allows us to say that if we conducted the survey 100 times under essentially the same conditions, we would expect that about 95 times out of 100 the true population value would fall within this interval. The confidence interval for a given survey estimate can be determined with the RSE. To calculate the 95-percent confidence interval:

1. Divide the RSE by 100 and multiply by the survey estimate to determine the standard error.
2. Multiply the standard error by 1.96 to determine the margin of error.
3. The survey estimate plus or minus the margin of error is the 95-percent confidence interval.

RSE data are available as a separate tab within Excel versions of each 2009 RECS data table.

In estimating sampling error only, RSEs therefore will underestimate the total survey error. Response error, error introduced by interviewers, and processing error are types of nonsampling error.

Confidentiality of information

The Confidential Information and Protection Statistical Efficiency Act (CIPSEA) is designed to protect the privacy of respondents of Federal surveys, including RECS. Any information collected that might permit the identification of respondents or their households is kept confidential and used only for statistical purposes. The household records that are placed on the public-use data file do not have name or address information, and disclosure measures are taken to mask the data so that a sampled housing unit or its individual occupants cannot be identified to the public.