Independent Statistics & Analysis
U.S. Energy Information Administration

Residential Energy Consumption Survey (RECS):

Using the 2015 microdata file to compute estimates and standard errors (RSEs)

May 2017
(revised February 2019)

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 U.S. Department of Energy or other federal agencies.

Table of Contents

Overview

EIA makes available a public-use microdata file for each RECS survey cycle. The 2015 file is a valuable tool for users conducting detailed analysis of home energy use. This document provides some background on the RECS design, as well as useful tips and examples that will guide users through the use of the RECS microdata.

RECS sample design

The RECS sample was designed to estimate energy characteristics, consumption, and expenditures for the national stock of occupied housing units and the households that live in them. The 2015 RECS allows for separate estimation for Census regions and divisions. (The return to the traditional sample size for the 2015 RECS does not allow for state-level estimation, as was available for the expanded 2009 RECS.) To produce estimates for these geographies and the total U.S., the sample cases were properly weighted to represent the population, including the residences not in the sample. In a sense, a case's weight indicates the number of households that the particular case represents.

Base sampling weights, which are the reciprocal of the probability of being selected for the RECS sample, were first calculated for each sampled housing unit. The base weights were adjusted to account for survey nonresponse and ratio adjustments were used to ensure that the RECS weights add up to Census Bureau estimates of the number of occupied housing units for 2015. The variable **NWEIGHT** in the data file represents the *final sampling weight*, accounting for different probabilities of selection and rates of response, and being adjusted for the Census Bureau housing unit estimates. NWEIGHT is the number of households in the population that the observation represents. For example, if NWEIGHT for a household is 10,000, that household represents itself and 9,999 other non-sampled households. More details about the sample design can be found in the *RECS 2015 Technical Documentation – Summary*.

Sampling error

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

Standard errors are used in conjunction with survey estimates to measure sampling error, construct confidence intervals, or perform hypothesis tests. 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. An RSE is shown for each estimate in the RECS tables.

Fay's balanced repeated replication (BRR) method of estimating standard error

RECS uses Fay's method of the balanced repeated replication (BRR) technique for estimating standard errors. This method uses replicate weights to repeatedly estimate the statistic of interest and calculate the differences between these estimates and the full-sample estimate.

See Fay (1989), Heeringa, West, and Berglund (2010), Judkins (1990), Lee and Forthofer (2006), Roa and Shao (1999), Rust (1985), and Wolter (2007) for technical details.

If θ is a population parameter of interest, let $\hat{\theta}$ be the estimate from the full sample for θ. Let $\widehat{\theta_r}$ be the estimate from the r-th replicate subsample by using replicate weights and let ε be the Fay coefficient, $0 \leq \varepsilon < 1$. The variance of $\hat{\theta}$ is estimated by:

$$\hat{V}(\tilde{\theta}) = \frac{1}{R(1-\varepsilon)^2} \sum_{r=1}^{R} (\widehat{\theta_r} - \hat{\theta})^2$$

For the 2015 RECS, R=96 (the number of replicate subsamples) and ε = 0.5. The formula for calculating the RSE is:

$$\left(\frac{\sqrt{\hat{V}(\hat{\theta})}}{\hat{\theta}} \right) X\ 100$$

Examples: Using final weights (NWEIGHT) and replicate weights to calculate estimates and RSEs

The following instructions are examples for calculating any RECS estimate using the final weights (NWEIGHT) and the associated RSE using the replicate weights (BRRWT1 – BRRWT96). We have provided instructions for Excel users and users with access to SAS/STAT and R. Software packages like SAS/STAT, R, Stata, SUDAAN, and WesVar can process replicate weights to calculate RSEs. Note that while Excel can be used to calculate point estimates, it cannot process replicate weights to calculate RSEs for RECS or other complex sample designs with varying probabilities of selection. EIA recommends calculating standard errors or RSEs in conjunction with estimates to account for sampling error.

For Excel Users (estimates only, no RSEs)

Excel Example 1: Calculate the frequency of households that used natural gas as their main space heating fuel (Table HC6.1)

A simple count of households can be estimated using the sum of NWEIGHTS for a specified subset of cases within the RECS data file. For this example, filter the file for all cases where natural gas space heating was used as the main heating fuel (FUELHEAT= 1). There are 2,790 cases with FUELHEAT = 1. By adding the NWEIGHT column for these 2,790 cases, the estimated number of households that used natural gas as main heating fuel was approximately 57,667,485. This is equal to 49% of all homes, or 57.7 million/118.2 million (the sum of NWEIGHT for all cases in RECS.)

Table HC6.1 Space heating in U.S. homes by housing unit type, 2015[1]

Release date: February 2017
Revised date: May 2018

	Number of housing units (million)					
		Housing unit type				
	Total U.S.[2]	Single-family detached	Single-family attached	Apartment (2- to 4-unit building)	Apartment (5 or more unit building)	Mobile home
All homes	118.2	73.9	7.0	9.4	21.1	6.8
Space heating equipment						
Use space heating equipment	113.1	72.1	6.7	8.9	18.9	6.5
Have space heating equipment but do not use it	3.6	1.2	0.2	0.3	1.6	Q
Do not have space heating equipment	1.6	0.5	Q	Q	0.6	Q
Main heating fuel and equipment						
Natural gas	57.7	40.2	4.2	4.6	7.3	1.4

Excel Example 2: Calculate energy intensity by the number of household members in the South (Table CE1.4)

To find the energy intensity by the number of household members, first filter the microdata file for households in the South (REGIONC = 3). There should be 2,010 cases. In a new column, calculate the weighted number of household members (NHSLDMEM × NWEIGHT) for each case and sum the column to get 112,316,570. In a separate column, calculate the weighted total fuel consumption (TOTALBTU × NWEIGHT) for each case and sum the column to get 3,063,515,106,263. Divide the sum of the weighted total fuel consumption by the sum of the weighted number of household members. The result should be 27,275 thousand Btu or 27.3 million Btu as shown in Table CE1.4.

Release date: May 2018

Table CE1.4 Summary annual household site consumption and expenditures in the South—totals and intensities, 2015

	Number of housing units (million)	Site energy consumption[1]				Energy expenditures[1]			
	Total South[2]	Total (trillion Btu)	Per household (million Btu)	Per household member (million Btu)	Per square foot (thousand Btu)	Total (billion dollars)	Per household (dollars)	Per household member (dollars)	Per square foot (dollars)
All homes	44.4	3,064	68.9	27.3	35.6	85.19	1,917	758	0.99

SAS Example 1: Calculate the frequency and RSE of households that used natural gas as their main space heating fuel (Table HC6.1)

Create a new variable to flag the records we are interested in - households that used natural gas as their main space heating fuel. This new variable NG_MAINSPACEHEAT is equal to 1 if the household used natural gas as their main space heating fuel, and 0 otherwise.

```
DATA RECS15;
   SET RECS2015_PUBLIC_V3;
   IF FUELHEAT=1 THEN NG_MAINSPACEHEAT =1; ELSE
   NG_MAINSPACEHEAT =0;
RUN;
```

Use the variable NWEIGHT in the WEIGHT statement and the variable NG_MAINSPACEHEAT in the TABLES statement in PROC SURVEYFREQ. To get the sampling error (RSE) associated with the estimate, we can use PROC SURVEYFREQ to process the replicate weights.

```
PROC SURVEYFREQ DATA=RECS15 VARMETHOD=BRR(FAY);
   REPWEIGHTS BRRWT1-BRRWT96;
   WEIGHT NWEIGHT;
   TABLES NG_MAINSPACEHEAT;
RUN;
```

The estimated number of households that used natural gas as their main space heating fuel is 57,667,485. The standard deviation of the frequency is 1,317,409 and the calculation for the RSE is: (1,317,409 / 57,667,485)*100 = 2.3. This means that the sampling error is about 2.3% of the estimate, relatively small.

Table of NG_MAINSPACEHEAT

NG_MAINSPACEHEAT	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent
0	2896	60540765	1317409	51.2153	1.1145
1	2790	57667485	1317409	48.7847	1.1145
Total	5686	118208250	0.03206	100.000	

SAS Example 2: Calculate total and average space heating consumption by region, and associated RSEs, for households that used natural gas as their main space heating fuel (Table CE6.1)

To calculate total and average consumption for a specified subset of households in SAS, use the SURVEYMEANS procedure. For this example, use TOTALBTUSPH in the VAR statement, and the newly created variable NG_MAINSPACEHEAT in the DOMAIN statement. For a further breakout of consumption, add a second dimension to the DOMAIN statement. For this example, Census region (REGIONC) is added. The WEIGHT and REPWEIGHT variables are the same as the PROC SURVEYFREQ example above. Use the options *sum*, *clsum*, *mean*, and *clm* to request the sum, confidence interval for the sum, mean, and confidence limit of the mean, respectively, of the variable TOTALBTUSPH.

```
PROC SURVEYMEANS DATA=RECS15 VARMETHOD=BRR(FAY) SUM CLSUM MEAN CLM;
   REPWEIGHTS BRRWT1-BRRWT96;
   WEIGHT NWEIGHT;
   DOMAIN NG_MAINSPACEHEAT * REGIONC;
   VAR TOTALBTUSPH;
RUN;
```

The table of output shows the space heating consumption by region (in thousand Btu), the standard deviation (error), and confidence intervals for the average and total space heating consumption. Note that the estimates for NG_MAINSPACEHEAT = 0 reflect consumption for homes that do not use natural gas as a main space heating fuel.

The SURVEYMEANS Procedure

Statistics for NG_MAINSPACEHEAT*REGIONC Domains

NG_MAINSPACEHEAT	REGIONC	Variable	Mean	Std Error of Mean	95% CL for Mean		Sum	Std Error of Sum	95% CL for Sum	
0	1	TOTALBTUSPH	46221	2321.684403	41612.4336	50829.4478	443960349372	39738164073	3.65081E11	5.2284E11
	2	TOTALBTUSPH	25860	2283.203530	21327.5160	30391.7624	198437074973	26948164531	1.44945E11	2.51929E11
	3	TOTALBTUSPH	14266	937.545802	12404.9002	16126.9276	438898083779	34067874314	3.71274E11	5.06522E11
	4	TOTALBTUSPH	10119	608.617474	8910.9052	11327.0975	126451680280	9548225009	1.07499E11	1.45405E11
1	1	TOTALBTUSPH	57884	1807.611661	54295.4807	61471.6423	659951339848	48547089646	5.63586E11	7.56317E11
	2	TOTALBTUSPH	63107	1301.144730	60523.7512	65689.2550	1.1799726E12	44347592091	1.09194E12	1.268E12
	3	TOTALBTUSPH	37291	2000.805273	33319.2855	41262.4197	510007091585	36888037275	4.36785E11	5.83229E11
	4	TOTALBTUSPH	27906	1320.163371	25285.9195	30526.9267	387663319748	28997740116	3.30103E11	4.45223E11

The first set of columns below shows the average space heating consumption, the standard error, and 95% confidence limits. The average space heating consumption for homes that use natural gas as their main space heating fuel in the northeast is 57.9 million Btu. The RSE for the average is (1,807.6 / 57,884)*100 = 3.1%. The lower 95% confidence limit is 54.3 million Btu and the upper 95% confidence limit is 61.5 million Btu. This means that if the sample were repeatedly taken and the confidence intervals were constructed from each sample, then 95% of the time, those confidence intervals would cover the true population mean.

The SURVEYMEANS Procedure

Statistics for NG_MAINSPACEHEAT*REGIONC Do

NG_MAINSPACEHEAT	REGIONC	Variable	Mean	Std Error of Mean	95% CL for Mean	
0	1	TOTALBTUSPH	46221	2321.684403	41612.4336	50829.4478
	2	TOTALBTUSPH	25860	2283.203530	21327.5160	30391.7624
	3	TOTALBTUSPH	14266	937.545802	12404.9002	16126.9276
	4	TOTALBTUSPH	10119	608.617474	8910.9052	11327.0975
1	1	TOTALBTUSPH	57884	1807.611661	54295.4807	61471.6423
	2	TOTALBTUSPH	63107	1301.144730	60523.7512	65689.2550
	3	TOTALBTUSPH	37291	2000.805273	33319.2855	41262.4197
	4	TOTALBTUSPH	27906	1320.163371	25285.9195	30526.9267

The second set of columns below shows the total space heating consumption, the standard error, and 95% confidence limits. For Northeast homes that use natural gas as their main space heating fuel, this results in a total space heating consumption of 0.660 quadrillion Btu, an RSE of 7.4%, and a 95% confidence interval

of 0.564 quadrillion Btu to 0.756 quadrillion Btu.

NG_MAINSPACEHEAT	REGIONC	Variable	Sum	Std Error of Sum	95% CL for Sum	
0	1	TOTALBTUSPH	443960349372	39738164073	3.65081E11	5.2284E11
	2	TOTALBTUSPH	198437074973	26948164531	1.44945E11	2.51929E11
	3	TOTALBTUSPH	438898083779	34067874314	3.71274E11	5.06522E11
	4	TOTALBTUSPH	126451680280	9548225009	1.07499E11	1.45405E11
1	1	TOTALBTUSPH	659951339848	48547089646	5.63586E11	7.56317E11
	2	TOTALBTUSPH	1.1799726E12	44347592091	1.09194E12	1.268E12
	3	TOTALBTUSPH	510007091585	36888037275	4.36785E11	5.83229E11
	4	TOTALBTUSPH	387663319748	28997740116	3.30103E11	4.45223E11

SAS Example 3: Calculate energy intensity by the number of household members by region and U.S. (Table CE1.1)

To calculate the energy intensity in SAS, use the SURVEYMEANS procedure. For this example, use REGIONC in the DOMAIN statement and TOTALBTU and NHSLDMEM in the RATIO statement to calculate the intensity per household member. The WEIGHT and REPWEIGHT variables are the same as the examples above. Use the *sum* option to request the sums of both TOTALBTU and NHSLDMEM.

```
PROC SURVEYMEANS DATA=RECS15 VARMETHOD=BRR(FAY) SUM;
   REPWEIGHTS BRRWT1-BRRWT96;
   WEIGHT NWEIGHT;
   DOMAIN REGIONC;
   RATIO TOTALBTU/NHSLDMEM;
RUN;
```

To find the intensities, refer to the *Ratio Analysis* tables in the output. The first *Ratio Analysis* table shows the intensity for all U.S. homes. The national level of energy intensity per household member is 30,270 thousand Btu or 30.3 million Btu, as shown in Table CE1.1.

Ratio Analysis			
Numerator	Denominator	Ratio	Std Err
TOTALBTU	NHSLDMEM	30270	388.366666

The regional intensity per household member can be found in the *Domain Ratio in REGIONC* table. The total consumption per household member in the Midwest (REGIONC = 2) is 37,800 thousand Btu or 37.8 million Btu, as shown in Table CE1.1.

Domain Ratio in REGIONC				
REGIONC	Numerator	Denominator	Ratio	Std Err
1	TOTALBTU	NHSLDMEM	38067	733.288676
2	TOTALBTU	NHSLDMEM	37800	877.753024
3	TOTALBTU	NHSLDMEM	27276	729.528912
4	TOTALBTU	NHSLDMEM	22297	546.867192

Release date: May 2018

Table CE1.1 Summary annual household site consumption and expenditures in the U.S.—totals and intensities, 2015

	Number of housing units (million)	Site energy consumption[1]				Energy expenditures[1]			
	Total U.S.[2]	Total (trillion Btu)	Per household (million Btu)	Per household member (million Btu)	Per square foot (thousand Btu)	Total (billion dollars)	Per household (dollars)	Per household member (dollars)	Per square foot (dollars)
All homes	118.2	9,114	77.1	30.3	38.4	219.34	1,856	728	0.92
Census region and division									
Northeast	21.0	1,984	94.4	38.1	45.2	47.66	2,269	915	1.09
New England	5.6	547	97.3	40.3	44.5	14.31	2,541	1,054	1.16
Middle Atlantic	15.4	1,436	93.4	37.3	45.5	33.36	2,169	866	1.06
Midwest	26.4	2,486	94.3	37.8	41.4	46.42	1,760	706	0.77
East North Central	18.1	1,755	97.0	38.1	43.1	31.88	1,762	693	0.78

For R Users

First install the survey package (Lumley 2017):

```
install.packages("survey")
library(survey)
```

Save the CSV file and read into R:

```
RECS15 <- read.csv(file='< location where file is stored >', header=TRUE, sep=",")
```

R Example 1: Calculate the frequency and RSE of households that used natural gas as their main space heating fuel (Table HC6.1)

Create a new variable to flag the records we are interested in - households that used natural gas as their main space heating fuel. This new variable NG_MAINSPACEHEAT is equal to 1 if the household used natural gas as their main space heating fuel, and 0 otherwise. Convert it to a factor since it is a categorical variable.

```
RECS15$NG_MAINSPACEHEAT <- ifelse(RECS15$FUELHEAT == 1, 1, 0)
RECS15$NG_MAINSPACEHEAT <- as.factor(RECS15$NG_MAINSPACEHEAT)
```

Define the survey design so the replicate weights are taken into account using Fay's method with the Fay coefficient equal to 0.5.:

```
sampweights <- RECS15$NWEIGHT
brrwts <- RECS15[, grepl("^BRRWT", names(RECS15))]
des <- svrepdesign(weights=sampweights, repweights=brrwts, type="Fay", rho=0.5, mse=TRUE,
data=RECS15)
des
```

Use svytotal to sum the number of households by NG_MAINSPACEHEAT, using the survey design defined above.

```
svytotal(~NG_MAINSPACEHEAT, des)
```

The estimated total of households that used natural gas as their main space heating fuel is 57,667,485. The calculation for the RSE is: (1,317,409 / 57,667,485)*100 = 2.3. This means that the sampling error is about 2.3% of the estimate, relatively small.

```
                          total      SE
NG_MAINSPACEHEAT0 60540765 1317409
NG_MAINSPACEHEAT1 57667485 1317409
```

R Example 2: Calculate total and average space heating consumption by region, and associated RSEs, for households that used natural gas as their main space heating fuel (Table CE6.1)

To calculate average space heating consumption in R, use the svymean function on the variable TOTALBTUSPH. To group by NG_MAINSPACEHEAT and Census region (REGIONC), use svyby.

```
means <- svyby(~TOTALBTUSPH, by=~REGIONC+NG_MAINSPACEHEAT, des, svymean)
means
```

The output below shows the average consumption by region (in thousand Btu) and the standard error. The average consumption in the northeast Census region (REGIONC=1) is 57.9 million Btu. The RSE for the average is (1,807.6 / 57,883.6)*100 = 3.1%. Note that the estimates for NG_MAINSPACEHEAT = 0 reflect consumption for homes that do not use natural gas as a main space heating fuel.

```
    REGIONC NG_MAINSPACEHEAT TOTALBTUSPH         se
1.0       1                0    46220.94  2321.6844
2.0       2                0    25859.64  2283.2035
3.0       3                0    14265.91   937.5458
4.0       4                0    10119.00   608.6175
1.1       1                1    57883.56  1807.6117
2.1       2                1    63106.50  1301.1447
3.1       3                1    37290.85  2000.8053
4.1       4                1    27906.42  1320.1634
```

To compute 95% confidence intervals, use confint and specify the correct number of degrees of freedom. The number of degrees of freedom is equal to the number of replicate weights, which is 96.

```
confint(means, df=ncol(brrwts))
```

For the average main space heating consumption in the Northeast, the lower 95% confidence limit is 54.3 million Btu and the upper 95% confidence limit is 61.5 million Btu. This means that if the sample were repeatedly taken and the confidence intervals were constructed from each sample, then 95% of the time, those confidence intervals would cover the true population mean.

```
          2.5 %   97.5 %
1.0 41612.434 50829.45
2.0 21327.516 30391.76
3.0 12404.900 16126.93
4.0  8910.905 11327.10
1.1 54295.481 61471.64
2.1 60523.751 65689.25
3.1 33319.286 41262.42
4.1 25285.919 30526.93
```

To calculate the total space heating consumption, use svytotal with svyby. To calculate the 95% confidence intervals, use confint.

```
sums <- svyby(~TOTALBTUSPH, by=~REGIONC+NG_MAINSPACEHEAT, des, svytotal)
sums
confint(sums, df=ncol(brrwts))
```

The results below show the total space heating consumption, the standard error, and 95% confidence limits. For homes in the Northeast that use natural gas as their main space heating fuel, this results in a total space heating consumption of 0.660 quadrillion Btu, an RSE of 7.4% ((.4855 / 6.600)*100), and a 95% confidence interval of 0.564 quadrillion Btu to 0.756 quadrillion Btu.

```
    REGIONC NG_MAINSPACEHEAT  TOTALBTUSPH          se
1.0       1                 0 4.439603e+11 39738164049
2.0       2                 0 1.984371e+11 26948164560
3.0       3                 0 4.388981e+11 34067874265
4.0       4                 0 1.264517e+11  9548225021
1.1       1                 1 6.599513e+11 48547089559
2.1       2                 1 1.179973e+12 44347592085
3.1       3                 1 5.100071e+11 36888037263
4.1       4                 1 3.876633e+11 28997740137

                  2.5 %       97.5 %
 1.0 3.650807e+11 5.228400e+11
 2.0 1.449454e+11 2.519288e+11
 3.0 3.712739e+11 5.065223e+11
 4.0 1.074986e+11 1.454048e+11
 1.1 5.635861e+11 7.563166e+11
 2.1 1.091943e+12 1.268002e+12
 3.1 4.367849e+11 5.832293e+11
 4.1 3.301033e+11 4.452234e+11
```

R Example 3: Calculate energy intensity by the number of household members by region and U.S. (Table CE1.1)

To calculate the intensity for all U.S. homes, use svyratio to compute intensity per household member.

```
svyratio(~TOTALBTU, ~NHSLDMEM, des)
```

The national level of energy intensity per household member is 30,270 thousand Btu or 30.3 million Btu, as shown in Table CE1.1. The standard error is 388.4 thousand Btu and the RSE is (388.4 / 30,270.0)*100 = 1.28.

```
Ratios=
         NHSLDMEM
TOTALBTU 30269.98
SEs=
         [,1]
[1,] 388.3667
```

To calculate the regional energy intensity per household member, use svyratio with svyby.

```
svyby(~TOTALBTU, by=~REGIONC, denominator=~NHSLDMEM, des, svyratio)
```

The total consumption per household member in the Midwest (REGIONC = 2) is 37,800 thousand Btu or 37.8 million Btu, as shown in Table CE1.1.

	REGIONC	TOTALBTU/NHSLDMEM	se.TOTALBTU/NHSLDMEM
1	1	38067.28	733.2887
2	2	37800.34	877.7530
3	3	27275.72	729.5289
4	4	22296.68	546.8672

Notes to consider when using the microdata file and replicate weights

1. *Publication standards:* EIA does not publish RECS estimates where the RSE is higher than 50 or the count of households used for the calculation is less than 10 (indicated by a "Q" in the data tables). These are EIA's recommended guidelines for custom analysis using the public use microdata file.

2. *Imputation variables:* Most variables were imputed for "Don't Know" and "Refuse" responses. The "Z variables", also referred to as "imputation flags", are included in the public use microdata file. The imputation flag indicates whether the corresponding non-Z variable was based upon reported data (Z variable = 0) or was imputed (Z variable = 1). There are no corresponding "Z variables" for variables from the RECS questionnaire that were not imputed, variables where there was no missing data, and variables that are not from the questionnaire. EIA recommends using the imputed data, where available, to avoid biased estimation.

3. *Standardized coding:* Variables that were not imputed use the response codes -9 for "Don't Know" and -8 for "Refuse". Variables that are not asked of all respondents use the response code -2 for "Not Applicable". For example, if a respondent said they did not use any televisions at home (TVCOLOR = 0) then they were not asked what size of television is most used at home, thus TVSIZE1 = -2. Use caution when performing calculations on variables that may have -2, -8, or -9 responses.

4. *Indicator variables:* The microdata file contains variables to indicate the use of major fuels and specific end uses within each housing unit for 2015. These variables are derived from answers given by each respondent and indicate whether the respondent had access to and actually used the fuel and engaged in the end-use. All indicators are either a 0 or 1 for each combination of major fuel and end-use. For example, a respondent who says they heated their home with electricity in 2015 will have the derived variable ELWARM = 1. If a respondent says they have equipment but did not use it the corresponding indicator will be 0. As an example, a respondent in a cool climate might have air-conditioning equipment but did not use it in 2015. For this case, ELCOOL would be 0.

5. *Confidentiality:* The 2015 RECS was collected under the authority of the Confidential Information Protection and Statistical Efficiency Act (CIPSEA). EIA, project staff and its contractors and agents are personally accountable for protecting the identity of individual respondents. The following steps were taken to avoid disclosure of personally identifiable information on the public use microdata file.

- Local geographic identifiers of sampled housing units, such as zip codes, were removed.
- Building America Climate Regions with few sample cases ("Very Cold" and "Mixed- Dry") were combined with the most similar region.

- The variable indicating on-site wind generation (WIND) was removed due to too few responses.
- The variable HHAGE (age of the householder) was top-coded at 85.
- Weather and climate (HDD and CDD) values were inoculated with random errors. Adjustments were minor and will not result in significant differences than those estimates displayed in data tables.

References

Fay, R. E. (1989), "Theory and Application of Replicate Weighting for Variance Calculations," in Proceedings of the Survey Research Methods Section, 212–217, American Statistical Association.

Heeringa, S., West, B. T, & Berglund, P. A. (2010). *Applied survey data analysis*. Boca Raton, Fla.: CRC Press.

Judkins, D. R. (1990), "Fay's Method for Variance Estimation," Journal of Official Statistics, 6(3), 223–239.Lee, E. Sul, & Forthofer, R. N. (2006). *Analyzing complex survey data.* 2nd ed. Thousand Oaks, Calif.: Sage Publications.

Lumley, T. (2017) "survey: analysis of complex survey samples". R package version 3.32.

Rao, J. N. K. and Shao, J. (1999), "Modified Balanced Repeated Replication for Complex Survey Data," Biometrika, 86(2), 403–415.

Rust, K. (1985), "Variance Estimation for Complex Estimators in Sample Surveys," Journal of Official Statistics, 1(4), 381–397.

Wolter, K. M. (2007). *Introduction to Variance Estimation*, 2nd ed. Springer, New York.

The R code presented in this document was developed and tested in version 3.5.1.