U.S. Energy Information Administration

Office of Energy Statistics

Office of Energy Consumption and Efficiency Statistics

Supporting Statement for Survey Clearance

Residential Energy Consumption Survey, EIA-457, Forms A-G

OMB No. 1905-0092

Statistical Methodology

Part B

Original Date: April 2015

Residential Energy Consumption Survey Supporting Statement, Part B

Table of Contents

COLLECTING INFORMATION BY STATISTICAL METHODS

B-1. Description of the Survey Universe

The survey universe for the 2015 RECS will be the same as the 2009 RECS, which is the population of all housing units occupied as a primary residence in the 50 states and the District of Columbia.

The definition of an occupied housing unit is the same as that used by the U.S. Census Bureau, which excludes housing such as military barracks, dormitories, and nursing homes. Non-barracks housing units on military installations are within scope, but are excluded from this sample frame.

The size of the universe will be interpolated from estimates of the latest American Community Survey published by the Census Bureau.

B-2. Sampling Methodology and Estimating Procedures

Introduction

EIA proposes a stratified three-stage sample design for the Household Survey. The design will satisfy all of the study objectives and increase efficiency compared to the 2005 RECS, which had a similar sample size. At the first stage, EIA will select 200 primary sampling units (PSUs) from the Census Bureau's Public Use Microdata Areas (PUMAs). At the second stage, EIA will select segments defined by census block groups (CBGs). At the third stage, EIA will construct a Housing Unit (HU) sampling frame based on (1) address based sampling (ABS) mailing list only; (2) ABS supplemented by RTI's frame-linking procedure, Check for Housing Units Missed (CHUM); or (3) traditional field enumeration. Samples will be selected from the HU sampling frame in each segment. The sample for the Rental Agent Survey will come from the Household Survey. The responding HUs from the Household Survey that are occupied by renters or where owners indicate that some or all utility or bulk fuel energy bills are not paid directly by the household will be selected for the Rental Agent Survey.

Some key features of the proposed design include:

- Target precision levels for average household energy consumption at national, Census region, and Census division levels
- Using census PUMAs as PSUs instead of county-type PSUs to avoid complexity of selecting certainty PSUs, to reduce unequal weighting effects that might be induced if county-type PSUs are used, and to improve PSU sample selection by using PUMA-level energy consumption data from the American Community Survey (ACS)

- Within PSUs, using larger Census Block Groups (CBG) as segments, instead of one or more census blocks, to reduce clustering effect and geocoding error
- Expecting an improvement in the design efficiency from the 2005 RECS and the 2009 RECS by selecting 200 PUMA-type PSUs and 4 CBG segments per PSU (resulting in a total of 800 segments)
- Compared to the 2005 RECS, expecting to achieve higher precision (low relative standard error, RSE) at the national, Census region, and Census division levels for HU energy consumption outcomes because of improved design efficiency
- Using a hybrid method of constructing HU sampling frames to obtain high coverage while controlling costs. The frames will be constructed using a combination of ABS, CHUM, and traditional field enumeration.

PSU Selection

At the first stage of sample selection, EIA will select a stratified sample of 200 PSUs. EIA proposes using PUMA-type PSUs instead of county-type PSUs for three reasons: (1) each PUMA has at least 24,000 occupied HUs, far surpassing the minimum size requirement, but with far less variability in the number of HUs than counties; (2) large counties are divided into multiple PUMAs, so EIA can avoid the complexity of selecting certainty PSUs and include more unique PSUs in the sample; and (3) ACS microdata contain rich information related to household energy consumption at the PUMA level, which could be used in PSU stratification to enhance the sample design.

The first-level of PSU stratification divides the United States into Census Divisions, with the Mountain Division split into North and South. Some Census divisions are further divided into subdivisions that group state(s) at the second-level stratification. There are 19 geographical domains after the second-level stratification as shown in the Table B1. The 19 geographical domains ensure a nationally representative 2015 RECS sample.

Table B1.   First and Second Level Stratification for PSU Sample Selection

Census Division	Geographical Domain	States
New England	1	CT, ME, NH, RI VT, MA
Middle Atlantic	2	NY
	3	NJ, PA
East North Central	4	IL, IN, OH
	5	MI, WI
West North Central	6	IA, MN, ND, SD
	7	KS, NE, MO
South Atlantic	8	VA, DC, DE, MD, WV
	9	GA, NC, SC
	10	FL
East South Central	11	AL, KY, MS, TN
West South Central	12	AR, LA, OK
	13	TX
North Mountain	14	CO, WY, ID, MT, UT
South Mountain	15	AZ, NM, NV
Pacific	16	CA
	17	OR, WA
	18	HI
	19	AK

Within each geographical domain, EIA will implicitly stratify by sorting the frame by the within-domain stratification variables before selecting the PSU sample. EIA plans to use 2009 RECS household energy consumption data and apply the decision tree analysis method to identify the most important characteristics or combinations of characteristics and form within-domain strata. Ten PUMA-level household characteristics that are correlated with energy consumption will be used as inputs to the decision tree analysis. After decision tree models are finalized, EIA will create a within-domain stratification variable for each domain corresponding to the terminal nodes from the tree models.

EIA will use Chromy's minimum replacement technique to select PSUs in the probability proportional to size (PPS) selection scheme within each geographical domain. The estimated number of occupied HUs based on the 2013 1-year ACS data is used as the size of measure in

the PPS selection. EIA will select a total of 200 PSU, and an independent sample will be drawn from each geographical domain.

Within each of the 19 geographic domains, a near epsem sample will be achieved by selecting PSU and segment samples with the PPS method, and allocating the same number of completed interviews per segment. PSUs will be allocated to the 19 geographical domains to satisfy all of EIA's target precision requirements. Specifically, the target RSEs for average household energy consumption are:

- National: 1.25%
- Census Region: 2.75%
- Census Division: 4.50%

The PSU sample will be allocated to 19 geographical domains using an optimization procedure so that the precision requirements are met, while minimizing the UWE nationally.

Table B2 below shows the proposed PSU allocation.

Table B2. Proposed 2015 RECS PSU Allocation

Division	RECS Domain	States	PUMA Frame Count	Occupied Units (Census 2010)	Occupied Units (2013 ACS)	Proportional PUMA Allocation[1]	Optimized PUMA Allocation[1]	Expected Number of Completed Interviews
New England	1	CT, MA, ME, NH, RI, VT	109	5,664,396	5,602,713	10	9	180
Middle Atlantic	2	NY	145	7,317,755	7,219,356	13	11	220
	3	NJ, PA	165	8,233,264	8,115,033	14	13	260
East North Central	4	IL, IN, OH	231	11,942,561	11,846,561	20	19	380
	5	MI, WI	108	6,152,276	6,121,890	11	9	180
	6	IA, MN, ND,	76	3,912,277	3,985,867	7	8	160

West North Central	7	KS, MO, NE	83	4,208,837	4,207,161	7	8	160
South Atlantic	8	DC, DE, MD, VA, WV	124	6,585,304	6,566,918	11	10	200
	9	GA, NC, SC	180	9,131,920	9,099,434	16	14	280
	10	FL	151	7,420,802	7,211,584	13	12	240
East South Central	11	AL, KY, MS, TN	138	7,213,076	7,109,313	12	12	240
West South Central	12	AR, LA, OK	82	4,335,894	4,301,325	7	7	140
	13	TX	212	8,922,933	9,110,853	15	14	280
Mountain	14	CO, ID, MT, UT, WY	90	4,066,454	4,121,055	7	8	160
	15	AZ, NM, NV	90	4,178,635	4,156,887	7	9	180
Pacific	16	CA	265	12,577,498	12,650,592	22	26	520
	17	OR, WA	87	4,139,014	4,168,356	7	9	180
	18	HI	10	455,338	450,120	1	1	20
	19	AK	5	258,058	246,015	—	1	20
Total			2,351	116,716,292	116,291,033	200	200	4,000

[1]The proportional allocation and optimized allocation were calculated using estimated occupied HUs based on the 2010 Census.

Segment selection

Compared to using smaller segments (block or aggregations of blocks) as in previous RECS, using larger segments (CBGs) will reduce the clustering effect and increase the precision of estimates. It also has the secondary benefit of reducing geocoding errors because CBGs have significantly fewer geocoding errors than Census blocks. Furthermore, the CBG-level housing characteristics from the ACS microdata can be used in the design to select a representative segment sample within a PSU.

Although no explicit stratification will be applied, implicit stratification will be used to spread out segment samples over important characteristics that have a large impact on energy usage. Three segment-level characteristics will be considered in the implicit stratification:

- Percent renter occupied
- Year housing unit built
- Median income

The segment sampling frame will be sorted by these implicit stratification variables before a PPS segment sample is selected within each PSU. We will apply the PPS selection method using Chromy's minimum replacement technique. We will select 4 segments per PSU, or a total of 800 segments. Similar to the PSU PPS selection, the estimated number of occupied HUs based on the 2013 5-year ACS data for each segment will be used as the size measure.

Housing unit frame construction and selection

An ABS list sampling frame consists of addresses from a vendor that qualifies for updates to its proprietary address list from the U.S. Postal Service's Computerized Delivery Sequence (CDS) file of active mail delivery points. The vendor can also update using the U.S. Postal Service's No-Stat file, which contains more than 9 million primarily rural and vacant mailing addresses that supplement the CDS file.

It is important to note that HU undercoverage will exist for in-person surveys using ABS frames, particularly in rural areas. The vast majority of this undercoverage is the result of non-city–style addresses like post office boxes, highway contract routes, and rural route boxes. Area probability samples can suffer additional coverage problems (over- and undercoverage) from geocoding error. To account for this undercoverage, EIA will employ field enumeration and CHUM, which is a frame-linking procedure used in the field by interviewers instead of by listers.

CHUM is a procedure designed to give each HU missing from the ABS frame a known probability of selection, theoretically providing 100% coverage of dwelling units. It will be used in segments where the ABS is adequate but not complete. Experienced interviewers will be trained to conduct this procedure in the field.

EIA will estimate the expected ABS coverage rate for each sampled CBG using the ratio of the number of city-style mailing addresses on the ABS list to the estimated number of HUs in the CBG from the Census or ACS. EIA will divide selected CBGs into three groups:

- Group 1 (CDS listings only): ABS coverage rate of 95% or more
- Group 2 (CHUM): ABS coverage rate between 50% and 94%
- Group 3 (Field Enumeration): ABS coverage rate below 50%

EIA estimates that there will be 560 CBGs or 70% of total CBGs in Group 1, 200 CBGs or 25% of total CBGs in Group 2, and 40 CBGs or 5% of total CBGs in Group 3. For the 560 CBGs in Group 1, the CDS plus No-Stat list will be used. For the 200 CBGs in Group 2, the CDS plus No-Stat list will be supplemented with the CHUM procedure. For the 40 CBGs in Group 3, EIA will conduct counting and listing (field enumeration).

After frame construction, addresses will be selected using a systematic sample. Assuming an 80% eligibility rate (occupied primary HUs) and an 80% response rate, EIA will select eight HUs to complete approximately five interviews per segment.

Estimation Procedures

Each interviewed sample observation (that is, a housing unit for which an interview is completed) will be assigned a weight. The sum of the weights over all sample observations will equal the number of occupied residential housing units as derived from the American Community Survey. The weight for a particular sample unit will be the number of occupied housing units in the U.S. population represented by that sampled unit. Based upon the RECS sample results, the weights will be used to produce estimates of population values such as total energy consumption, average expenditures, percent of housing units with dishwashers, and so forth.

For each interviewed housing unit, the weight will reflect the probability of selection for that housing unit and additional adjustments to correct for potential biases arising from the failure to contact all sample housing units and the failure to list all housing units in the sample area. Initially, each sample observation will be assigned a base weight that equals the inverse of the probability of selection for the housing unit. The base weights will be adjusted upward to correct for non-interviews by spreading the effects of non-interviews over the interviewed sample of households in the final cluster.

In addition, four ratio adjustments will be made: 1) reduce the sampling variation in the estimates of the number of housing units by main space-heating fuel resulting from sampling of PSUs during the first stage of the sample design; 2) adjust sample estimates to ACS estimates of the number of households by Census Divisions; 3) correct for the type of household (single-person male, single-person female, and all other types); this will correct for the lower response rate among single-person households; 4) readjust the sample estimates to ACS estimates using ratio

adjustment.

Precision of Estimates

Assuming 20 completed interviews per PSU, Table B3 displays expected sample sizes and RSEs for energy consumption per household (million BTUs) for the 2015 RECS. Estimates from the 2009 RECS are displayed for comparison. As shown in the table, all target precision requirements will be met with the proposed 2015 RECS sample allocation.

Table B3. Expected Completes and Estimated Relative Standard Errors for the 2015 RECS

	Completes		Energy Consumption per HH (million BTU)	
	2009	**2015**	**2009 RSE**	**2015 RSE**
National	12,083	4,000	0.0064	0.0103
Census Region and Division				
Northeast	2,266	660	0.0131	0.0223
New England	938	180	0.0205	0.0432
Middle Atlantic	1,328	480	0.0178	0.0273
Midwest	2,843	880	0.0117	0.0194
East North Central	1,150	560	0.0149	0.0197
West North Central	1,693	320	0.0209	0.0444
South	4,090	1,380	0.0106	0.0169
South Atlantic	2,246	720	0.0227	0.0370
East South Central	614	240	0.0304	0.0449
West South Central	1,230	420	0.0269	0.0425
West	2,884	1,080	0.0181	0.0273
Mountain	812	340	0.0268	0.0382
Pacific	2,072	740	0.0260	0.0400
Region Average	N/A	N/A	0.0134	0.0215
Division Average	N/A	N/A	0.0230	0.0375

NOTE: Target precision requirements: 4.5% at the division level; 2.75% at the region level, and 1.25% at the national level.

B-3. Maximizing the Response Rate

Household Survey

EIA and its contractor will employ methodologies similar to the 2009 RECS to obtain a minimum response rate of about 80 percent. This includes a multi-wave, multi-contact approach using a post card, advance letter, monetary and non- monetary incentives, a Spanish language personal interview questionnaire, brochures, and refusal conversion letters.

Prior to the in-person contact, a pre-notification postcard and then a pre-notification letter will be sent to sampled HUs. The letter from the Director of the Energy Consumption Division at EIA will briefly describe the purposes of the survey, stress its importance, and announce the impending arrival of the interviewer. This letter, as well as the post card, will mention the monetary and nonmonetary incentives. Nonresponse letters have been developed for use with households who are concerned with confidentiality or say that they are too busy.

Rental Agent Survey

If a Rental Agent Survey case is spawned as a result of the Household Survey interview, the interviewer will attempt to contact the Rental Agent and administer the questionnaire on-site. If this is not feasible, the interviewer will contact the rental agent and administer the CAPI questionnaire by telephone. Similar nonresponse techniques and letters are proposed for the Rental Agent Survey protocol. The content of the rental agent survey is limited to items most susceptible to household reporting error.

Energy Supplier Survey

Billing data will be requested under mandatory authority for a 20-month period. Consumption and expenditure data are critical to estimating energy demand within the residential sector; therefore significant time during Household Survey interviewer training is devoted to collecting accurate supplier information such as company names, addresses, and account numbers. Trainings include specific modules for scanning samples of supplier bills.

For the 2015 RECS ESS, EIA and its contractor will employ similar methods for contacting energy suppliers as were used in the 2009 RECS ESS. This includes utilizing existing EIA contact information (e.g., 2012 CBECS ESS, EIA electricity and natural gas supply surveys), precontacting suppliers to alert them of the data collection, and identifying key respondents within their organizations. All suppliers will then receive an official data request letter from EIA. This letter will include instructions on accessing the survey Website and inform suppliers of the mandatory requirement for this phase of the RECS. Nonresponse follow-up procedures will include reminder phone calls and letters, as well as late notice phone calls and letters from contractor staff and EIA.

Data Quality Control

For the Household and Rental Agent Surveys, EIA and its data collection contractors will employ real-time and post-collection procedures to detect and correct critical data errors. Real-time edits are programmed into the CAPI instrument to alert respondents and interviewers of inconsistent and outlier responses, or incorrect keying into the instrument. Additional systematic edits will be run via SAS during post-collection to compare household survey and rental agent responses and review inconsistencies and outliers. Households will not be revisited or recontacted to resolve any edit failures.

For the Energy Supplier Surveys, EIA and its contractor will use real-time and early detection edit procedures for energy suppliers submitting household billing records. This includes outlier range checks for consumption and cost data, missing fields, and gaps in billing records. Suppliers will be recontacted to address the most critical edit failures.

Unit and Item Nonresponse Adjustments

Two types of non-response adjustment will be used: one for non-interviews (unit nonresponse) and the other for item non-response. The adjustments for non-interviews consists of adjusting the household weights to correct for potential bias resulting from the failure to contact all sample units and/or the failure to list all housing units within a sample area.

Item non-response occurs when respondents do not know the answer, refuse to answer a question, cannot provide data within the quality requirements of the question, or the interviewer makes an error. These non-responses will be imputed by one of three methods: regression, "hot decking," and deductive procedures. Regression procedures will be used for fuel consumption and expenditures.

The "hot deck" procedure will be used for imputing such items as household income, availability of natural gas, year house was constructed, and age of respondent. This procedure requires sorting the file of households by variables related to the missing item. A household is then randomly selected from a pool of households which have the same values on the selected variables. This "donor" household supplies the value for the missing variable in the household with the missing item.

Deductive procedures are used primarily to fill in missing data on what fuels are used in the home, which fuels are used for which end-use, and what method is used to pay for the fuels. Information on these items is obtained from the household interview, rental agent survey, and fuel supplier surveys. Deductive procedures are also used when information is missing or when the information from different sources does not agree.

Nonresponse Bias studies

There are three methods that EIA will use to assess nonresponse bias in the 2015 RECS: validation with external benchmark data, validation with internal subdata, and reviewing the impact of weighting adjustments given the 2015 sample design decisions and post-stratification data options

EIA will review trends between RECS estimates and other federal surveys. Before unit non-response adjustments and final post-stratification adjustments are done, key estimates will be compared for a net estimate of bias at the lowest level of aggregation possible. Minimally, we will compare the number of occupied housing units, main space heating fuel, age of home, household income with the American Community Survey (ACS) and American Housing Survey (AHS). EIA will not compare the ACS and AHS data on energy costs as their method is incompatible with ours—self reported estimates of previous month's electricity and gas bills—and RECS data are used in some way to edit and adjust those to an annual basis. EIA collects actual energy costs for the reference year directly from a householder's energy suppliers. We would compare trends from RECS with EIA's supply surveys for the whole residential sector.

For within-survey evaluations of bias, EIA will focus on a̶ two critical areas: square footage and nonresponse by income. The RECS method of capturing the square footage on sampled units has changed over time, other federal surveys only ask for self-reports, and square footage is a key independent variable in the consumption model for energy end uses. We will evaluate the relationship between these two methods—self reported and measured—to understand the relationship between the RECS data and the AHS self-reports, the Survey of Construction and commercial survey trends. Nonresponse by income subgroups will be evaluated for any nonresponse impact on the subsample for the energy assistance module, on energy usage and behavior.

B-4. Tests of the Procedures

As noted previously, EIA is making updates to the RECS Household questionnaire to cover changes in technology (e.g., TV peripherals), new energy efficiency programs, and to improve question, interviewer and respondent performance. EIA relied on stakeholder consultation, survey methodologist review, crowd sourcing, and cognitive interviews in drafting new and revised questions. We also incorporate lessons learned from prior rounds, interviewer debriefings, analysis of edit failure rates and sources of those failures. Additionally, EIA and its contractor will conduct extensive review of the data collection instruments and post-collection procedures in advance of the 2015 RECS launch. As part of these procedures, detailed variable and edit specifications are being developed.

In January and February 2015, RTI conducted cognitive interviews and pretesting for the 2015 RECS. The cognitive interviews focused on draft questions about new technologies, revisions to questions that were outdated or known to be problematic, and comparisons of alternate response options that may improve the accuracy of respondents answers and data quality. There were two

round of cognitive interviews with each round consisting of fifteen participants in three different locations.

RTI also conducted three rounds of crowdsourcing using *Amazon's Mechanical Turk (MTurk)*, which rapidly collects large volumes of pretest data. Each round consisted of 19 respondents. The first round asked about water heating, the second asked about lighting, and the third round asked about home structure.

The pretesting provided specific recommendations for questions in these areas:

- Cooking behavior
- Appliance ownership and usage (especially appliances other than stoves, ovens, washers and dryers)
- Air conditioning usage and behavior (temperature control, etc.)
- Television equipment and usage
- Lighting

B-5. Statistical Consultation

The principal EIA official directing the RECS sample redesign is Hiroaki Minato, who can be reached at (202) 586-6040 or by email at *Hiroaki.Minato@eia.gov.* The principal EIA official directing the 2015 RECS is James (Chip) Berry, who can be reached at (202) 586-5543 or by e-mail at *james.berry@eia.gov.*

The agency point-of-contact regarding this clearance request is Alethea Jennings, who can be reached at (202) 586-5879 or by e-mail at Alethea.jennings@eia.gov.