



*Independent Statistics & Analysis*

U.S. Energy Information  
Administration

---

# Supporting Statement B for Residential Energy Consumption Surveys

## **Part B: Collections of Information Employing Statistical Methods**

---

Form EIA-457A *Household Survey*

Form EIA-457D *Energy Supplier Survey: Household Propane Usage*

Form EIA-457E *Energy Supplier Survey: Household Electricity Usage*

Form EIA-457F *Energy Supplier Survey: Household Natural Gas Usage*

Form EIA-457G *Energy Supplier Survey: Household Fuel Oil/Kerosene Usage*

**OMB No. 1905-0092**

---

June 2020



# Table of Contents

---

Part B: Collections of Information Employing Statistical Methods.....	i
B.1. Respondent Universe.....	1
B.2. Statistical Methods.....	1
B.3. Maximizing Response Rates.....	7
B.4. Test Procedures and Form Consultations.....	9
B.5. Statistical Consultations.....	10

## B.1. Respondent Universe

The survey universe for the 2020 RECS will be 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, college dormitories, nursing homes, and prisons, as well as vacant and seasonal homes.

An address-based frame for the 2020 RECS was created from the U.S. Postal Service's Computerized Delivery Sequence (CDS) file of residential addresses that receive mail. A sample of housing units for the Household Survey was drawn from this frame file. The sample selection is described in section B.2. The list of energy suppliers that will complete the Energy Supplier Survey forms is obtained directly from the suppliers that are identified by respondents in their responses to the Household Survey.

## B.2. Statistical Methods

### 2020 RECS Survey Design Precision Requirements

The 2020 Residential Energy Consumption Survey (RECS) sample design is optimized to meet precision requirements for key household energy consumption metrics, including total average energy consumption and average consumption for individual fuels used in the residential sector. The precision requirements are based on relative standard errors (RSEs), which are the estimated standard errors of the means divided by the mean estimates, expressed as a percentage. Table 1 displays the precision requirements, by geographic domain, for the 2020 RECS sample design.

**Table 1. Relative Standard Error (RSE) Precision Requirements**

Geography	Average Household Consumption Outcome Variable	RSE of Mean
Total United States	Total energy consumption	1%
	Electricity consumption	1%
	Natural gas consumption	1%
	Fuel oil consumption	3%
4 Census Regions	Total energy consumption	2%
	Electricity consumption	2%
	Natural gas consumption	2%
Northeast Census Region	Fuel oil consumption <sup>1</sup>	4%
10 Divisions <sup>2</sup>	Total energy consumption	3%
	Electricity consumption	3%
	Natural gas consumption	3%
50 States and DC	Total energy consumption	4%

<sup>1</sup> A precision requirement for fuel oil consumption is only imposed for the Northeast as it is a highly regional fuel and used less frequently in other parts of the country.

<sup>2</sup> The divisions are the same as the standard Census divisions except that the Mountain division is split into Mountain South (AZ, NM, and NV) and Mountain North (CO, ID, MT, WY, and UT).

EIA set a target of 18,000 completed interviews. This target threshold is sufficient to meet the precision requirements at the various geographic levels listed in Table 1, and provides reasonable precision for other key metrics. As noted in Supporting Statement Part A, an additional 2,000 completes (for a total of 20,000 completes) may be needed to account for lower than expected response or coverage issues for certain subpopulation groups in the sample.

### **Housing Unit Frame**

The 2020 RECS uses an address-based sampling frame that starts with a vendor address list derived from the U.S. Postal Service's Computerized Delivery Sequence (CDS) file. The frame includes all mailable residential addresses in the 50 states and DC, with the exception of a small proportion of addresses that are non-active (No-Stat) addresses, and PO Box addresses that correspond to households that also have a street address for mail delivery.

#### *Other Way to Get Mail (OWGM) and Drop Point Addresses*

During the development of the 2020 RECS sampling frame, special attention was given to the households that only get mail via PO Box, known as "Only Way to Get Mail" (OWGM), and households receiving their mail at drop points. While each of these phenomena represent approximately 1% of the households nationally, they are not equally distributed across states. A minority of the states have a significant share, possibly impacting population coverage within these states. Given the 2020 RECS objective of producing estimates for all states, it was necessary to include all OWGM addresses and most drop point addresses in the housing unit frame.

For OWGM addresses, we will mail an invitation to the PO Box, but attempt to collect the physical address for responding households via a special question on the Household Survey form. Having a physical address for OWGM households will allow us to better link critical weather station data to household respondent data and to include these households in the Energy Supplier Survey data collection.

Drop points are single-mail receptacles shared by multiple housing units and there is a variable on the CDS file that indicates which postal service addresses are drop points. Drop points are found primarily in multifamily buildings with only 2 to 4 units, and only these units are included in the 2020 RECS frame. Drop points with 5 or more units are a small percentage of the overall drop points, and are likely to be institutional units that are out of scope for RECS. For these reasons, drop points with 5 or more units are excluded from the 2020 RECS frame.

Drop points create operational problems such that specific households cannot be targeted for mailing reminders and incentive payments. To eliminate the need for a special data collection protocol for sampled drop point housing units, EIA and its contractor substituted any selected drop point address with a non-drop point statistical (in most cases physical) neighbor. This method was chosen because most drop points are similar in building structure to their non-drop point neighbors<sup>3</sup>. Under this

---

<sup>3</sup> Amaya, A., LeClere, F., Fiorio, L., & English, N. 2014. "Improving the utility of the DSF address-based frame through ancillary information". *Field Methods*, 26(1), 70-86.

substitution approach, if a drop point unit was selected, a substitute unit in a non-drop point multi-unit building with the same number of units will be contacted instead.

### Determining expected completed interviews

The 2020 RECS uses a stratified sample by geographic area. The initial target of 18,000 completed interviews was allocated to the 50 states and DC using a “bottom up” statistical approach. First, we calculated the necessary number of completed interviews in each state to meet the state-level RSE requirement for total average energy consumption. Next, the process was repeated to determine the necessary number of completed interviews to meet the RSE requirement in each Census division, which was then allocated optimally<sup>4</sup> to the states within each division. The same process was repeated to determine the necessary number of completed interviews to meet the regional and national RSE requirements.

After the total average energy consumption requirements were met, sample sizes at the state level were checked for sufficiency to meet RSE requirements for those of average electricity and natural gas consumption. Of the two, meeting the natural gas precision requirements generally required a larger sample size because not all homes use natural gas and its use varies greatly by region. Therefore, the assumption was that the requirements for electricity should be met after satisfying the natural gas requirements. For natural gas, the first step was to take into account the proportion of housing units in each state that use natural gas based on prior RECS data. If the divisional, regional, and national requirements for natural gas were not met by the number of completed interviews necessary to meet the total average energy consumption requirements, the sample allocation was increased using the same process as done for the total average energy consumption requirements.

The general formula used for sample size determination at each geographical level is:

$$n = \frac{(UWE * \hat{\sigma}^2)}{\bar{x} * RSE^2}$$

Where *UWE* is the unequal weighting effect to account for nonresponse weighting, a *UWE* of 1.05 was assumed in the calculation;  $\hat{\sigma}^2$  is the variance estimate of an energy consumption value;  $\bar{x}$  is the mean of an energy consumption value; and *RSE* is a precision requirement.

Because the mean and variance estimates are different at each geographical level, the estimation approach for each fuel started with the state-level estimated means and variances from the combined data of 2009 RECS and the 2015 RECS. As these prior RECS studies were not designed to produce state-level estimates for all states, the sample sizes for the mean and variance estimation were pooled from 2009 RECS and 2015 RECS. For the variance estimation, if the pooled respondent sample size was at least 30, the pooled direct estimates were used. But if the pooled respondent sample size was less than 30, then the average of two different Empirical Bayes (EB) model estimates were used. Domain-level estimates were used in the EB models, because it was assumed that these estimates were more stable than the state-level estimates when sample size was small. A domain was defined as a group of

<sup>4</sup> Cochran, W.G. 1977. “Sampling Techniques”, third edition. New York: John Wiley & Sons.

congruent states in the 2009 RECS. The general EB model for the variance and mean estimators are the following:

$$\text{Mean estimator: } \bar{y}_{EBs} = \alpha_s \bar{y}_s + (1 - \alpha_s) \bar{y}_M$$

$$\text{Variance estimator: } \text{var}_{EBs}(\hat{y}) = \alpha_s \text{var}_s(\hat{y}) + (1 - \alpha_s) \text{var}_M(\hat{y})$$

Where S is denoted as state, M is denoted as domain, and  $\alpha_s$  is defined by:  $\alpha_s = \frac{n_s}{\sum_{i \in M} n_i}$  or

$$\alpha_s = \frac{SSB}{SSW + SSB}, \text{ with } n \text{ being the sample size of a state, } SSB \text{ being the weighted square error of}$$

consumption between the states in a domain, and  $SSW$  being the weighted square error of consumption for the samples within states.

Of the two EB models, one estimate for  $\alpha_s$  was based on sample size, and the other estimate for  $\alpha_s$  was based on consumption. The final EB estimate is the average of the two EB estimates. Of the 50 states and DC, only 7 states used the EB model for estimation.

After the effective minimum sample size was determined to meet all national, regional, divisional, and state-level RSE requirements, the remaining sample needed to meet the overall target of 18,000 completed interviews was allocated to states in proportion to their number of occupied housing units, based on 2017 American Community Survey (ACS) estimates. The additional sample allocation is required to achieve residential demand estimation objectives for the 2020 RECS beyond the explicit precision requirements stated in Table 2 above. For example, estimation for key energy-use metrics by housing type within states is achievable with sufficient sample sizes for single-family and multi-family units. Expected RSEs were analyzed for these subpopulations during the sample allocation phase and the design team considered these subpopulations during the implicit stratification phase described below.

### Determining the starting sample size

To achieve the target number of 18,000 completed interviews, a significantly larger *starting* sample size is needed to account for ineligibility and nonresponse. The necessary size for the starting sample was determined by dividing the number of completed interviews by the yield rate. The yield rate equals the original sample size minus the respondents who are out of scope. Based on the 2015 RECS and RECS National Pilot results, the expected yield rate for the 2020 RECS is about 37% nationally. However, it is assumed that the yield rates will differ somewhat across states. The state-level yield rates were estimated using either direct estimates from the 2015 RECS and the RECS National Pilot, or a regression model using Census Block Group (CBG) level estimates from the two surveys and the Census CBG-level ROAM (Response Outreach Area Mapper) data. Table 2 displays the expected completed interviews determined by the methodology above, and the starting samples for the 50 states and DC. The 2020 RECS will use a starting sample size of 48,649 housing units to yield 18,000 completed interviews.

Table 2. Expected completed interviews and starting samples for 50 states and DC

State	Minimum Expected Completed Interviews	Proportionally Added to the Minimum	Final Expected Completed Interviews	Expected yield rates	Starting Sample Size
AK	195	16	211	38.4%	549
AL	154	114	268	37.7%	712
AR	172	71	243	37.9%	642
AZ	348	158	506	35.5%	1424
CA	367	805	1172	35.5%	3299
CO	189	132	321	35.6%	901
CT	231	84	315	37.5%	841
DC	177	17	194	32.3%	600
DE	98	22	120	37.0%	324
FL	200	476	676	38.3%	1763
GA	198	232	430	33.5%	1284
HI	247	28	275	37.7%	730
IA	171	78	249	49.7%	501
ID	195	39	234	39.9%	587
IL	207	298	505	34.5%	1465
IN	197	158	355	37.9%	938
KS	113	70	183	39.7%	461
KY	323	107	430	39.7%	1083
LA	127	107	234	36.0%	650
MA	392	161	553	37.8%	1465
MD	184	137	321	36.8%	873
ME	163	33	196	38.6%	507
MI	123	243	366	44.0%	832
MN	196	134	330	48.1%	685
MO	182	148	330	42.1%	783
MS	120	68	188	36.5%	515
MT	142	26	168	40.3%	417
NC	199	245	444	32.8%	1353
ND	272	20	292	39.8%	734
NE	130	47	177	40.2%	440
NH	146	33	179	39.8%	450
NJ	276	199	475	31.0%	1534
NM	134	48	182	33.9%	537
NV	175	68	243	35.4%	686
NY	545	452	997	31.3%	3190
OH	116	289	405	39.6%	1023
OK	147	91	238	39.5%	602
OR	210	99	309	44.4%	697
PA	327	310	637	35.6%	1788
RI	177	25	202	37.3%	542



State	Minimum Expected Completed Interviews	Proportionally Added to the Minimum	Final Expected Completed Interviews	Expected yield rates	Starting Sample Size
SC	158	118	276	33.3%	829
SD	146	21	167	40.8%	410
TN	349	160	509	41.4%	1228
TX	438	595	1033	33.4%	3094
UT	133	60	193	44.0%	439
VA	232	193	425	36.1%	1177
VT	201	16	217	39.0%	556
WA	229	176	405	40.8%	993
WI	175	145	320	42.5%	753
WV	126	44	170	40.0%	425
WY	119	14	133	39.3%	338
<b>Total</b>	<b>10,571</b>	<b>7,430</b>	<b>18,001</b>		<b>48,649</b>

### Stratification Strategy

The frame will be stratified by state. Within each state, the frame will be sorted prior to systematic selection to provide an implicit stratification. The use of implicit stratification in combination with systematic selection ensures the selected sample is representative of the frame distribution for the variables used for sorting. In other words, the implicit stratification helps protect against a poorly representative sample by chance. The frame will be sorted by the following variables listed generally from highest to lowest:

- o Climate zone as defined at the county level by the International Energy Conservation Code<sup>5</sup>
- o Multi-family dwelling unit indicator defined at the address level on the CDS
- o Rural-Urban Commuting Area code defined at the Census tract level by the US Department of Agriculture<sup>6</sup>
- o ZIP code
- o Carrier Route on the CDS
- o Walk Sequence (mail delivery sort order within carrier route)
- o ZIP+4 (for addresses that do not have a walk sequence; otherwise this does nothing)

### Sample Selection

As discussed in Supporting Statement Part A, there will be three phases of data collection for the 2020 RECS Household Survey. Phase 1 will consist of 20% of the starting sample and Phase 2 will consist of the remaining 80% of the starting sample. The Phase 1 and Phase 2 starting sample will be selected

<sup>5</sup> [https://www.energy.gov/sites/prod/files/2015/10/f27/ba\\_climate\\_region\\_guide\\_7.3.pdf](https://www.energy.gov/sites/prod/files/2015/10/f27/ba_climate_region_guide_7.3.pdf)

<sup>6</sup> <https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/documentation/>

simultaneously prior to the start of Phase 1 data collection using Chromy's minimum replacement technique<sup>7</sup>. This is a systematic selection technique that selects sample units from successively ordered zones created on a sorted sampling frame.

Phase 3 is planned as a risk mitigation strategy, in case EIA does not meet the targeted 18,000 completed interviews or precision targets after Phase 1 and Phase 2. Additional sample will be released in Phase 3 based on the preliminary results from Phase 1 and Phase 2. Up to 2,000 completed interviews (for a total of up to 20,000 completed interviews) are planned for this phase, if necessary.

### **Weighting and Estimation Procedures**

Each completed interview will be assigned a final weight. The sum of the weights for all completed interviews will equal the number of occupied, primary housing units from the 2020 American Community Survey (ACS). The weight for a particular case is equivalent to the number of occupied primary housing units that interviewed household represents. The weighted 2020 RECS survey data will be used to produce a wide range of population estimates, such as total household energy consumption, average energy expenditures, percent of housing units with dishwashers, and so forth.

For each interviewed housing unit, the final 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 for ineligible and non-responding households. In addition, the weights will be adjusted to match ACS estimates for specific items such as the number of occupied housing units in each state.

## **B.3. Maximizing Response Rates**

The 2020 RECS will utilize robust contact, data collection, statistical analysis, and risk mitigation strategies to maximize the response rate and produce representative samples of key household energy subpopulations.

### **Maximizing Unit Response and Coverage in the Household Survey**

#### *Contact Strategy and Protocol*

The 2020 RECS Household Survey data collection relies on self-administered Web and paper modes, and a contact protocol known as "Choice+". The Choice+ protocol offers households a choice of responding by Web or paper questionnaire. Household respondents are offered a higher monetary incentive to respond by Web to encourage data collection via the Internet. The overall contact protocol calls for up to six mailings to sampled households over approximately seven weeks, including three postcards and three invitation letters (which also include the paper questionnaire).

---

<sup>7</sup> Chromy, J.R. 1979. "Sequential sample selection methods". In Proceedings of the American Statistical Association, Section on Survey Research Methods, pp. 401-406.

All contact materials include the same text in both English and Spanish. The Web questionnaire will be available in both English and Spanish, with the URLs for both versions included in all mailings. The paper questionnaire will be available in both English and Spanish, but households will need to call the RECS helpline to request the Spanish paper questionnaire. The RECS helpline will also be available throughout the data collection period for respondents to call with any questions or concerns.

### *Responsive Design*

The 2020 RECS Household Survey will use a responsive design with three phases of data collection. Phase 1, which will consist of 20% of the starting sample, will include two experiments related to contact materials and incentives. These experiments will test whether alternative strategies can increase response rates and/or improve coverage. Phase 1 will also allow for a full end-to-end test of the data collection, monitoring, and data management processes to ensure they are performing as expected prior to releasing the majority of the sample in Phase 2. After Phase 1 is completed, EIA will analyze the results of the experiments and apply the optimal strategy to all of the Phase 2 sample. The optional Phase 3 reserve sample will be deployed as a risk mitigation strategy in case the overall target of 18,000 completed interviews, or the precision targets, are not met after Phase 1 and Phase 2 data collection. This phase can address any shortfalls in the areas that have lower-than-expected yield rates.

Two experiments in Phase 1 will test the design of postcards used as part of the contact protocol and different incentive levels. For the postcard experiment, half of the Phase 1 sample will receive colorful postcards with graphics, similar to those used for the RECS National Pilot test in 2015. The other half of the Phase 1 sample will receive more official-looking postcards solely in black-and-white. For the incentive experiment, half of the Phase 1 sample will be promised an additional \$10 for completing the questionnaire via the web mode.

During Household Survey data collection, a robust dashboard of metrics will be used to monitor the data collection outcomes and data quality in real time. A variety of key metrics will be tracked, including comparisons with the frame and/or other external benchmarks such as those from the ACS to ensure data adequacy and sample representativeness to the population.

### **Item Nonresponse**

Item non-response occurs when respondents do not know the answer or refuse to answer a question, therefore, no data is provided to a request. Item nonresponse has been generally low for prior RECS Household Survey data collections as well as the self-administered RECS National Pilot. For the 2020 RECS Household Survey, item non-response will be corrected using the hot-deck imputation method, which will preserve the distribution of the outcome variables and variance structure in the data. Hot-deck imputation uses the non-missing values of a variable as donors to impute a missing item. It requires sorting the file of households by variables related to the missing item, and then imputes the missing item with the value of a household donor selected from a pool of households having the same values within an imputation class. The procedure will be done using the Cyclical Tree-Based Hot Deck (CTBHD) imputation system<sup>8</sup>.

---

<sup>8</sup> Martin, P., Wang, J., Frechtel, P., Sukasih, A., Lewis, K., Deng, G., & Kinyon, D. 2017. "Three-based hot deck imputation cycling —does cycling help?" Proceedings of the 2017 Joint Statistical Meeting, Baltimore, Maryland.

The CTBHD implements classification or regression tree analysis to select variables for construction of imputation classes and uses weighted sequential hot deck (Cox, 1980) to select donors within imputation class. Under CTBHD, variables with missing values are imputed sequentially, which accounts for questionnaire skip patterns caused by the relationship between a gate question and subsequent follow-up questions. In addition, the CTBHD implements a cycling process which can help stabilize imputed values.

### **Nonresponse Bias Analysis**

After Phase 1 and Phase 2 data collections are completed, EIA will assess whether characteristics of survey non-respondents differ from respondents to the 2020 RECS. EIA will compare respondent data with external benchmark data of other federal surveys, data from prior RECS studies, and analyze response using a Nonresponse Follow-up (NRFU) questionnaire. (If needed, the NRFU questionnaire will be submitted separately under EIA's generic clearance.) The NRFU questionnaire will be significantly shorter than the primary 2020 RECS questionnaire, collecting a limited set of key items such as housing unit type, main heating fuel, and number of household members. The responses from the NRFU respondents and 2020 RECS respondents will be compared to check for any meaningful differences between the two groups.

EIA will compare the distributions of a few key 2020 RECS estimates to those of the ACS, the American Housing Survey (AHS), and the 2015 RECS. Before unit non-response adjustments and final post-stratification adjustments are done, the estimates will be compared for a net estimate of bias at the lowest level of aggregation possible. Minimally, we will compare the type of occupied housing units, main space heating fuel, age of home, and household income data from 2020 RECS respondents with the ACS and AHS, and also be able to compare a more robust set of variables from the 2020 and 2015 RECS studies.

In addition, during data collection, a real-time dashboard accessible by all project staff will be used to continuously monitor nonresponse bias for some key variables.

### **Energy Supplier Survey**

Energy billing data for household respondents will be requested from the utilities and other energy suppliers identified on Household Survey questionnaires. EIA uses its mandatory data collection authority to collect information from the energy suppliers on Forms EIA-457D, E, F, and G. This collection of energy billing data is part as the Energy Supplier Surveys (ESS) of the RECS collection. The 2020 RECS ESS will occur after all phases of the household data collection are complete and employ similar methods for contacting energy suppliers as were used in the 2015 RECS ESS. These strategies include utilizing existing EIA contact information (e.g., 2018 CBECS ESS, 2015 RECS ESS, EIA electricity and natural gas supply surveys), advanced notification to suppliers to alert them of the data collection, and identifying key respondents within their organizations. All suppliers will then receive an official data request. This request will include instructions on accessing the survey website and submitting data, 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. The estimated number of respondents for the ESS is shown in Supporting Statement Part A.

## B.4. Test Procedures and Form Consultations

As part of determining the 2020 RECS questionnaire content, EIA consulted with stakeholders, reviewed lessons learned from prior rounds, and conducted pretesting activities. In general, changes to the questionnaire content from the 2015 RECS focused on:

- o adding questions to reflect current household energy-related technologies, behaviors, and emerging topics,
- o revising questions to improve response quality, and
- o removing questions that were outdated or had not performed well in prior rounds.

Under *EIA-882T: Generic Clearance of Questionnaire Testing, Evaluating, and Research, OMB 1905-0186 (Expiration 4/30/2022)*, EIA's survey contractor conducted cognitive interviews and online pretesting for the 2020 RECS Household Survey. This pretesting took place across two rounds during February and March 2020. In total, 30 in-person cognitive interviews and 720 online pretests were completed. The focus of this pretesting effort was on potential new and revised questions. A variety of topical areas were covered in the pretests, including square footage, cooking appliances, heating and cooling equipment, TVs and computers, and lighting. After the pretesting was completed, EIA's survey contractor provided a comprehensive report of the results, including recommendations about specific questions. EIA finalized the 2020 RECS household questionnaire content based on these recommendations.

## B.5. Statistical Consultations

The principal EIA official directing the RECS sample design is Katie Lewis, who can be reached at (202) 586-5138 or by email at [Katie.Lewis@eia.gov](mailto:Katie.Lewis@eia.gov). The principal EIA official directing the 2020 RECS is James (Chip) Berry, who can be reached at (202) 586-5543 or by e-mail at [james.berry@eia.gov](mailto:james.berry@eia.gov).