Independent Statistics & Analysis
U.S. Energy Information Administration

# 2020 Residential Energy Consumption Survey: Household Characteristics Technical Documentation Summary

June 2022

*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.

# Contents


Figures ..... 3

Tables ..... 3

Overview and History ..... 4

Survey Design Elements and Changes ..... 5

Data Products and Revision Policy ..... 6

- Data products ..... 6
  - Within-cycle data revisions ..... 7
  - Methodological and content changes across survey cycles ..... 7

Frame and Sample Design ..... 7

- Frame sources and coverage ..... 7
- Sample allocation and sample selection ..... 8
  - Completed-case sample size allocation ..... 9
  - Sample selection ..... 10

Household Survey ..... 11

- Questionnaire design ..... 11
- Data collection methods ..... 12
  - Phased approach ..... 12
  - Contact materials ..... 13
- Response rate and nonresponse bias ..... 13
- Editing and data quality ..... 15
- Item imputation ..... 15

Weather and Geographic Data ..... 16

Weighting and Sampling Error ..... 17

- Relative standard errors ..... 17

Confidentiality of Information ..... 18

# Figures

Figure 1. *2020 Residential Energy Consumption Survey* timeline ..... 4

Figure 2. 2020 RECS unweighted and weighted self-response rate by state..... 14

# Tables

Table 1. 2020 RECS relative standard error (RSE) requirements for average fuel consumption..... 8

Table 2. Expected completed cases, assumed yield rate, and allocated starting sample for 2020 RECS... 10

# Overview and History

The *Residential Energy Consumption Survey* (RECS), is a periodic study conducted by the U.S. Energy Information Administration (EIA) that provides detailed information about energy usage in U.S. homes. RECS is a multiyear effort consisting of a Household Survey, data collection from household energy suppliers, and end-use consumption and expenditures estimation. The 2020 RECS is the 15th iteration of the study.

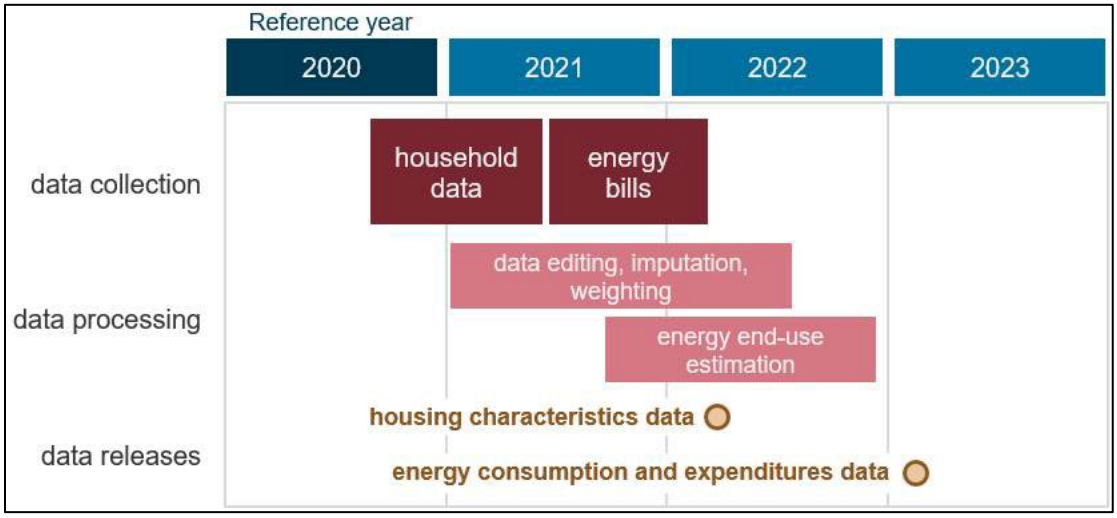
The Household Survey, a voluntary survey, collects data on energy-related characteristics and usage patterns at the national and sub-national (for example, state) level from a representative sample of housing units. The mandatory Energy Supplier Survey (ESS) collects data on how much electricity, natural gas, propane and liquefied petroleum gas (LPG), and fuel oil and kerosene were consumed in the sampled housing units during the reference year. It also collects data on actual dollar amounts spent on these energy sources. EIA uses engineering-based models to produce consumption and expenditure estimates for heating, cooling, refrigeration, and other end uses in all housing units occupied as a primary residence in the United States using the data collected from the Household Survey and ESS. Figure 1 outlines the 2020 RECS timeline.

The scope and purpose of RECS differ slightly from similar EIA products that report residential energy data. RECS samples homes occupied as a primary residence, which excludes secondary homes, vacant homes, military barracks, and common areas in apartment buildings. As a result, RECS estimates do not represent sector-level totals defined in other EIA products, but they are best suited for comparisons across different characteristics of homes within the residential sector.

The June 2022 version of this document includes information on the methodology for the Household Survey. Additional documentation will be published to coincide with the release of consumption and expenditure data in 2023.

The 2020 RECS Household Survey was conducted in collaboration with IMG-Crown and RTI International.

Figure 1. *2020 Residential Energy Consumption Survey* timeline



## Survey Design Elements and Changes

EIA instituted the following survey design revisions, content changes, and variable updates for the 2020 RECS:

- **The target population for the 2020 RECS is all occupied housing units in the 50 states and the District of Columbia (DC) that are used as primary residences.** Vacant homes, seasonal housing units, and group quarters (such as dormitories, nursing homes, prisons, and military barracks) are excluded. Housing units located on military installations are included.
- **The 2020 RECS sample was designed to meet precision requirements on energy consumption for all 50 states and DC, with an expected yield of 18,000 to 20,000 completed RECS questionnaires from sampled households.** For the first time in the program's history, estimates will be available for all 50 states and DC. The larger responding sample size also yields more precise estimates for key topics and for emerging technologies, such as electric vehicles.
- **Housing units for the 2020 RECS were selected using an Address Based Sample (ABS) design.** The frame for this sample was a list of residential addresses, based on the U.S. Postal Service's (USPS) Computerized Delivery Sequence file of active mail delivery points. The frame information was augmented with supplemental data from the Decennial Census, the American Community Survey (ACS), and other sources to enable stratification of the frame for better statistical efficiency and representation of the population of eligible housing units. Non-residential addresses were removed from the frame and procedures were implemented to account for special situations, such as accounting for non-deliverable and drop-point addresses.
- **The 2020 RECS introduced a completely self-administered design via web and paper questionnaire.** Before the 2020 RECS, all iterations of the study were conducted either through in-person interviews with trained interviewers at the sampled households or with a combination of in-person and self-administered modes. By eliminating interviewing staff for the 2020 RECS, the program was able to implement a number of other innovations, including increasing the sample size and moving to an unclustered sample design.
- **New questions were added on emerging technologies and usage behavior, and questionnaire changes were made in response to the COVID-19 pandemic.** The 2020 RECS added questions about solar capacity and installation, electric vehicle ownership and charging behavior, and the use of smart speakers. Just before data collection, in response to the COVID-19 pandemic, questions were added about regular telework or work from home, and whether anyone in the household was participating in K-12 distance learning or online college courses.
- **A total of 18,496 respondents completed the Household Survey: 73% by web and 27% by paper.** The total unweighted response rate (AAPOR 3)[1] was 38.6%, and the total weighted response rate was 37.9%. Household Survey data collection was conducted in two waves: the first wave was fielded September to November 2020, and the second wave was fielded January to April 2021.
- **All data underwent a series of quality control checks and edits to validate the sampled addresses of the responding households and to identify and resolve data inconsistencies.** This process included identifying outliers to numeric items, identifying and resolving logical inconsistencies, and recoding write-in responses into established response categories.
- **To address item nonresponse, a hot-deck imputation method was used for the 2020 RECS.** In this method, a recipient case that has a missing value for the variable being imputed is matched with a similar donor case that has a response. The donor's value for that variable is used to

[1] AAPOR Response Rate Calculator 4.1

replace the missing value for the recipient case. About 250 Household Survey variables were imputed, and the median imputation rate was 2.9%.

- **A new approach for housing unit control totals in weighting was used for the 2020 RECS.** Weighting adjustments were implemented to ensure the responding sample was representative of housing units at the national, census region, census division, and state levels. These weighting adjustments included ineligibility, nonresponse, and poststratification. RECS typically uses control totals from the current year's ACS data for poststratification. Because 2020 ACS one-year data were not published, 2020 RECS used a combination of 2020 Decennial Census counts and 2019 ACS estimates to calculate the control totals.
- **The 2020 RECS sample utilized the Jackknife Repeated Replication method for variance estimation.** Each RECS estimate has a corresponding relative standard error (RSE). RSEs were calculated from the estimated variance using replicate weights, which were determined using the Jackknife Repeated Replication method for the 2020 RECS. RSEs are shown as a separate tab in each published Excel table. Estimates greater than zero with a corresponding RSE of 0.00 indicate a variable used as a control total in poststratification.
- **EIA conducted comprehensive nonresponse bias analysis.** Based on the results from the nonresponse bias analysis, EIA identified no major concerns with the data quality of key 2020 RECS estimates, indicating that the final weighted 2020 RECS estimates are not significantly different from the target population parameters.

# Data Products and Revision Policy

## Data products

EIA releases a variety of RECS products across survey cycles tailored to a wide range of data users. These products include detailed tables of household energy use estimates across key geographic, structural, and demographic variables; topic-specific articles and reports; data-user webinars; microdata files; and survey methods documentation. Although similar products are released across survey cycles, changes are made from one cycle to the next to adapt to changes in the residential energy sector and leverage new dissemination methods and tools.

The following products are available on the 2020 RECS website:

- Household Survey data tables
- Today in Energy articles
- Public-Use Microdata File and User Guide
- Webinars
- Special-topic infographics and articles

RECS products from previous cycles are available on each survey cycle's *Data* page or in the archived *Analysis and Projections* page.

## Revisions and changes across cycles

### *Within-cycle data revisions*

EIA executes a series of survey data cleaning, editing, imputation, and coding steps to ensure RECS data and estimates meet EIA quality standards. EIA releases preliminary RECS Household Characteristics results at the end of the RECS Household Survey phase. After this release, EIA performs additional quality control steps to reconcile Household Characteristics data with energy billing data collected as part of the RECS Energy Supplier Survey. This additional quality control process may result in some revisions, usually minor, to the preliminary Household Characteristics estimates. EIA releases final Household Characteristics estimates concurrently with the release of Consumption and Expenditures estimates, usually one year after the preliminary release.

### *Methodological and content changes across survey cycles*

RECS is a cross-sectional study, with updates to questionnaire content, statistical methods, and dissemination strategies from the previous cycle. EIA does not currently conduct a longitudinal, household energy demand study. Each RECS, however, shares content and design elements across survey cycles. The unit of analysis for every RECS cycle is the occupied, primary housing unit. The sample is designed, using geographic and other stratification methods, using an address-based housing unit frame. Most survey questions are carried forward from one cycle to the next. Although RECS users are encouraged to use caution when drawing conclusions based on analysis across RECS cycles, many comparisons are valid and statistically sound.

Major changes to methods or questionnaire content from one cycle to the next are highlighted in Technical Documentation reports, special-topic reports (for example, the end-use modeling changes for 2015), and survey form specifications.

# Frame and Sample Design

The target population for the 2020 RECS is all occupied housing units in the 50 states and the District of Columbia (DC) that are used as primary residences. Vacant homes, seasonal housing units, and group quarters (such as dormitories, nursing homes, prisons, and military barracks) are excluded. But housing units on military installations are included. In addition, EIA benchmarks to occupied housing unit totals from the ACS. RECS uses the U.S. Census Bureau's definition of a housing unit, which is a single-family home, a unit in a multifamily building, or a mobile home.

## Frame sources and coverage

Housing units for the 2020 RECS were selected using an Address Based Sample (ABS) design. The frame for this sample is a list of residential addresses, based on the USPS's Computerized Delivery Sequence (CDS) file of active mail delivery points. The frame file is enhanced with supplemental data from the Decennial Census, the ACS, and other sources to enable stratification of the frame for better statistical efficiency and better representation of the population of eligible housing units. The following types of addresses were removed: nonresidential addresses, PO Box addresses that are not the only-way-to-get-

mail (OWGM),[2] and drop point addresses that have a frame indication of more than four units. Overall, the expected population coverage of the RECS ABS frame is about 99.6% of the target population of housing units.

Drop points are addresses that serve multiple housing units (drop units) without a unit identifier. These addresses posed logistical challenges for the 2020 RECS web- and mail-based data collection because the contact protocol included multiple mailings. Because the mailings went to a residential address, not a specific respondent name, subsequent mailings at a drop point address could be received by respondents from different housing units. Using a frame indicator for number of units, drop point addresses with four or more units were excluded from the sample frame. These units represent a small percentage (about 0.4%) of the overall drop points, and they are more likely to be group quarters that are out of scope for RECS. For the drop point addresses consisting of fewer than four units, a sample substitution method was used.[3] In other words, if a drop point address was selected, then the nearest non-drop point multifamily building with the same number of units would be selected as the substitution, and a unit in the substitute building would then be randomly selected as a sampling unit.

## Sample allocation and sample selection

Previous RECS cycles used clustered sample designs by grouping housing units into clustered geographies in order to make hiring and deployment of in-person interviewers efficient and cost-effective. For the 2020 RECS, the introduction of an entirely self-administered web and mail design meant that an unclustered design could be utilized, since geographic proximity was no longer necessary for efficient interviewer assignments. The design was a single-stage sampling with explicit and implicit stratification. In addition, the 2020 RECS sample was designed to meet the residential energy consumption precision requirements stated in Table 1, where the fuel consumption data are collected from the ESS.

Table 1. 2020 RECS relative standard error (RSE) requirements for average fuel consumption

| Geography | All fuel total[a] | Electricity | Natural gas | Fuel oil |
|---|---|---|---|---|
| United States | 1% | 1% | 1% | 3% |
| Midwest, South, and West regions | 2% | 2% | 2% | - |
| Northeast region | 2% | 2% | 2% | 4% |
| Census divisions | 3% | 3% | 3% | - |
| 50-states and District of Columbia | 4% | - | - | - |

[a] Includes electricity, natural gas, fuel oil, and propane

To achieve these statistical requirements, as well as to improve the precision of other key energy-use metrics, EIA targeted an estimated 18,000 to 20,000 completed cases. Compared with previous cycles of RECS, the sample allocation for 2020 was more complex due to additional geographic and fuel precision

[2] OWGM PO Boxes represent addresses at which a resident's only acceptable form of postal delivery is through a PO Box address. See McMichael, J., & Brown, D. (2018). PO Boxes on Address Based Sampling (ABS) frame: Under- or over-coverage or both? American Association for Public Opinion Research, Denver, CO.
[3] Amaya, A.E (2017). RTI International's Address-Based Sampling Atlas: Drop points. Research Triangle Park, NC: RTI Press. RTI Press Publication No. OP-0047-1712

requirements. However, the sample selection was simpler with a one-stage sample design, compared with a multistage sample design in the previous RECS cycles.

### *Completed-case sample size allocation*

When allocating the completed-case sample size, no single formula could optimize and satisfy all the precision requirements simultaneously. Therefore, the allocation process was done using a bottom-up approach, by optimally satisfying one requirement at a time. The first step was to calculate the minimum sample size needed to meet the precision requirement for total energy consumption (in British thermal units) at the state level, then check to see if the requirements at higher geographic levels, such as the division level, regional level, or the national level, were also satisfied. If they were not, then the same optimization process was applied to satisfy the requirements at each geographic level. Once this initial sample size was allocated, the same approach was then applied to ensure the precision requirements for each energy fuel source were met.

Originally, proposed precision requirements for propane were specified at the national and census region levels. These requirements were dropped due to the need for oversampling and significant increase in sample size. EIA determined that acceptable precision levels (between 4% and 6% at the region level) and improved quality over previous cycles could be attained for 2020 RECS propane estimates without meeting the original regional precision requirements. The estimation of the RSEs with the final allocated, completed-case sample size required prior estimation of the means, standard errors, and design effects. For the estimates of the means and standard errors used in the RSE formula, data from the 2009 RECS and the 2015 RECS were used. However, because these two datasets did not have complete data in every state, the estimates were derived using a pooled sample size from both RECS datasets where the combined sample total is at least 30 cases. If the combined sample size was less than 30 cases, then the estimates were derived using the average of two different modeled estimates. The design effect, which is an unequal weighting effect in this case, was 1.05 to account for eligibility and nonresponse adjustments to the equal design weights within each state.

As a result of the allocation process, the minimum sample needed to meet the precision requirements was approximately 10,571 households. However, as mentioned earlier, 18,000 households was the target sample for completed cases; therefore, the remaining difference of 7,429 households was allocated to each state in proportion according to the occupied housing unit distribution of the 2017 ACS.

In addition, to achieve the number of completed cases allocated for each state, additional sample addresses were selected to account for losses due to ineligibility and nonresponse during data collection. This starting sample for deployment was determined based on the number of expected completed cases and the corresponding assumed yield rate in each state. The yield rate is the proportion of starting sample cases that result in a complete, eligible questionnaire based on previous or external information. For 2020 RECS, the yield rates were estimated based on either data from the 2015 RECS and the National Pilot[4] or modeled from the self-response rate of the 2013−2017 five-year

[4] The National Pilot was a study with a nationally representative sample conducted in 2015 focusing on testing the feasibility of the self-administered modes.

ACS. See Table 2 on the allocated starting sample for each state calculated from the expected completed cases and the yield rate.

### *Sample selection*

For sample selection, the frame was stratified explicitly by state, and then within each state, the following variables were sorted as implicit stratification variables. The Chromy's minimum replacement technique (Chromy, 1979) was then used to select housing units systematically within each state:

- International Energy Conservation Code (county level, climate zone from U.S. Department of Energy)
- Multifamily dwelling unit indicator (address level, from CDS file)
- Rural-Urban Commuting Area code (census track level, from USDA)
- Zip code
- Carrier route (for mail delivery)
- Walk sequence (mail delivery sort order within carrier route)
- Zip+4 (for addresses that do not have a walk sequence)

Table 2. Expected completed cases, assumed yield rate, and allocated starting sample for 2020 RECS

| State or district | Expected completed cases | Assumed yield rate | Allocated starting sample |
|---|---|---|---|
| Alabama | 268 | 0.377 | 712 |
| Alaska | 211 | 0.384 | 549 |
| Arizona | 506 | 0.355 | 1,424 |
| Arkansas | 243 | 0.379 | 642 |
| California | 1,172 | 0.355 | 3,299 |
| Colorado | 321 | 0.356 | 901 |
| Connecticut | 315 | 0.375 | 841 |
| Delaware | 120 | 0.370 | 324 |
| District of Columbia | 194 | 0.323 | 600 |
| Florida | 676 | 0.383 | 1,763 |
| Georgia | 430 | 0.335 | 1,284 |
| Hawaii | 275 | 0.377 | 730 |
| Idaho | 234 | 0.399 | 587 |
| Illinois | 505 | 0.345 | 1,465 |
| Indiana | 355 | 0.379 | 938 |
| Iowa | 249 | 0.497 | 501 |
| Kansas | 183 | 0.397 | 461 |
| Kentucky | 430 | 0.397 | 1,083 |
| Louisiana | 234 | 0.360 | 650 |
| Maine | 196 | 0.386 | 507 |
| Maryland | 321 | 0.368 | 873 |
| Massachusetts | 553 | 0.378 | 1,465 |

| | | | |
|---|---|---|---|
| Michigan | 366 | 0.440 | 832 |
| Minnesota | 330 | 0.481 | 685 |
| Mississippi | 188 | 0.365 | 515 |
| Missouri | 330 | 0.421 | 783 |
| Montana | 168 | 0.403 | 417 |
| Nebraska | 177 | 0.402 | 440 |
| Nevada | 243 | 0.354 | 686 |
| New Hampshire | 179 | 0.398 | 450 |
| New Jersey | 475 | 0.310 | 1,534 |
| New Mexico | 182 | 0.339 | 537 |
| New York | 997 | 0.313 | 3,190 |
| North Carolina | 444 | 0.328 | 1,353 |
| North Dakota | 292 | 0.398 | 734 |
| Ohio | 405 | 0.396 | 1,023 |
| Oklahoma | 238 | 0.395 | 602 |
| Oregon | 309 | 0.444 | 697 |
| Pennsylvania | 637 | 0.356 | 1,788 |
| Rhode Island | 202 | 0.373 | 542 |
| South Carolina | 276 | 0.333 | 829 |
| South Dakota | 167 | 0.408 | 410 |
| Tennessee | 509 | 0.414 | 1,228 |
| Texas | 1,033 | 0.334 | 3,094 |
| Utah | 193 | 0.440 | 439 |
| Vermont | 217 | 0.390 | 556 |
| Virginia | 425 | 0.361 | 1,177 |
| Washington | 405 | 0.408 | 993 |
| West Virginia | 170 | 0.400 | 425 |
| Wisconsin | 320 | 0.425 | 753 |
| Wyoming | 133 | 0.393 | 338 |
| **Total** | **18,001** | | **48,649** |

# Household Survey

## Questionnaire design

The 2020 RECS Household Survey was designed to be entirely self-administered using either a web or paper questionnaire. Both questionnaires were also translated into Spanish. The 2020 RECS questionnaire specification is available on the EIA website and consists of the following topical sections:

- Your home
- Appliances
- Electronics

- Space heating
- Air conditioning
- Thermostats and temperatures
- Water heating
- Lighting
- Energy bills
- Household characteristics
- Energy assistance
- Energy supplier information

Each time EIA conducts the RECS, there is a review of the content and lessons learned from the previous cycle, and the questionnaire is revised as appropriate. The content revisions typically include adding or dropping questions to account for household technology changes or to improve response quality. For the 2020 RECS, new questions included information on:

- All-electric plug-in and hybrid plug-in vehicles and charging
- Household solar capacity
- Smart speakers
- Teleworking and online education at home due to the COVID-19 pandemic
- Information on power outages lasting longer than 24 hours

To improve response quality, EIA updated questions on square footage, space heating, and air conditioning. Minor adjustments were also made to the wording of questions to acknowledge potential changes in household energy consumption and behaviors related to the COVID-19 pandemic.

## Data collection methods

The RECS Household Survey was conducted on a voluntary basis with respondents in two waves: the first wave was fielded September to November 2020, and the second wave was fielded January to April 2021. A total of 18,496 eligible respondents completed the survey: 72.8% (13,469) responded via web questionnaire, and 27.2% (5,027) responded via paper questionnaire. Of the respondents that completed the web questionnaire, 73.4% (9,886) used a desktop or laptop computer, 22.2% (2,995) used mobile phones, and 4.4% (587) used tablets. Based on survey timing paradata collected without any data transformation, the web survey took an average of about 35 minutes to complete, with a median time of 32 minutes. In contrast, for the 2015 RECS, 5,686 households completed the Household Survey using a combination of in-person personal interviews, web questionnaires, and paper questionnaires.

### *Phased approach*

The 2020 RECS was planned as a three-phased, responsive-design approach. Phase 1 consisted of 20% of the initial starting sample, Phase 2 consisted of 80% of the initial sample, and an optional Phase 3 provided additional sample to address potential precision and representativeness issues. During the course of data collection, EIA determined that, based on response rates in Phase 1, the additional Phase 3 sample would be necessary to meet the target completes in select states. To field Phase 3 efficiently, the decision was made to release the Phase 3 cases concurrently with those in Phase 2. This approach contributed to both schedule efficiencies and budget optimization.

An additional benefit to the phased approach was it allowed EIA to conduct a series of experiments during Phase 1 and implement those findings during subsequent phases. Phase 1 included experiments to test the effectiveness of two different levels of formality for the RECS postcards (using a less formal version with colors rather than a more formal black-and-white version) and optimal incentive amounts to maximize web response (an additional $10 for web response versus an additional $20 for web response). Based on the experimental results from Phase 1, there was no significant difference between the submission rates of the differing postcard designs or the incentive amounts; therefore, the colorful version of the postcard and the $10 incentive options were selected for Phase 2 and Phase 3.

### *Contact materials*

For each phase, sample addresses were sent up to six mailings over the course of approximately six weeks.

- **Prenotice postcard**: sent to all cases
- **First invitation**: sent to all cases
- **Thank you or reminder postcard:** sent to all cases
- **Second invitation**: sent to remaining eligible or open cases
- **Reminder letter**: sent to remaining eligible or open cases
- **Third invitation**: sent to remaining eligible or open cases

## Response rate and nonresponse bias

The overall unweighted response rate for the 2020 RECS Household Survey is 38.6%, and the weighted response rate is 37.9%. The unweighted response rate was calculated using the American Association for Public Opinion Research (AAPOR) formula 3 (AAPOR, 2020):

$$RR = I/ (I+R+E)$$

where *I* is the number of complete interviews,[5] *R* is the number of refusal and eligible incompletes, and *E* is the number of eligible cases estimated from cases with unknown eligibility. The *E* was determined based on models for 2020 RECS.[6]

The weighted response rate was calculated using the same concept, except now the corresponding sum of weights were used in each disposition category.

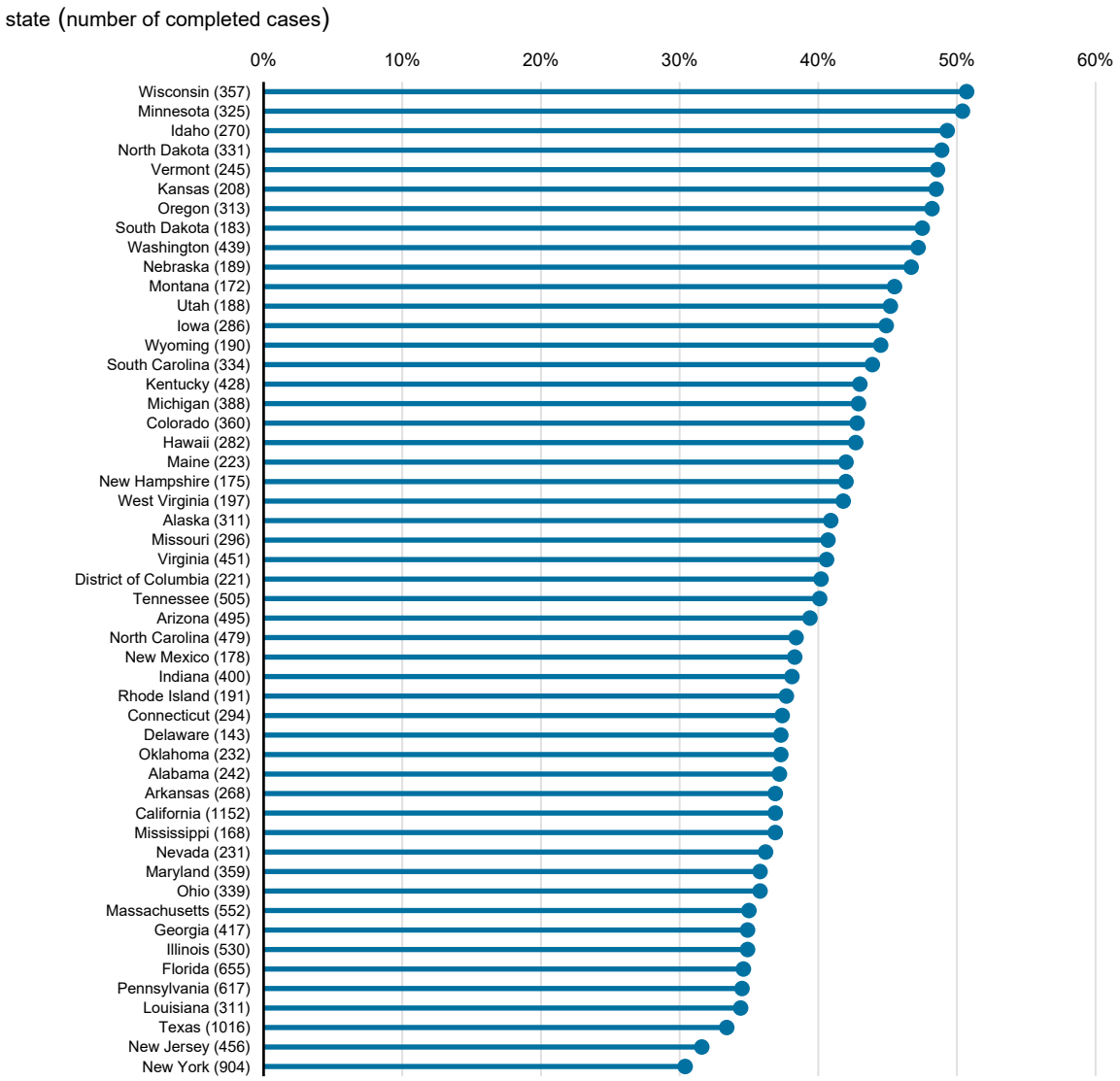
The response rate for 2020 RECS is lower than that of the 2015 RECS, which had unweighted response rate of 51.2% (weighted response rate was 50.8%). This lower response rate was expected, since the 2015 RECS was administered using both in-person and self-administered modes, and the 2020 RECS was entirely self-administered. In-person modes achieve higher response rates than self-administered modes because interviewers can build rapport with respondents in person.

Sampled households in Wisconsin, Minnesota, and Idaho responded at the highest response rates: 50.7%, 50.4%, and 49.3%, respectively. The states with the lowest response rates are New York (30.4%), New Jersey (31.6%), and Texas (33.4%). The unweighted response rate and weighted response rate are

[5] Completed interviews include interviews where the respondent did not answer all questions in the survey. The respondent must have answered at least 7 out of 10 key RECS questions in order for the interview to be considered complete. Partially completed interviews that did not meet that definition were defined as eligible incompletes.
[6] See the Weighting and Sampling Error section for more details.

the same for each state because each case within a state has the same sample selection probability. See the response rate chart by state in Figure 2.

**Figure 2. 2020 RECS unweighted and weighted self-response rate by state**




EIA conducted a comprehensive nonresponse bias study to understand how representative 2020 RECS respondents were of the general population or if any subpopulations were underrepresented in the responding sample. The study compared response rates by sample subgroup and compared the estimates of key frame variables between the respondent and nonrespondent groups. Differences in any comparisons could indicate potential nonresponse bias. In addition, demographic variables were compared to the ACS estimates to assess potential differences between the types of households responding to the RECS and the general household population of the United States.

The nonresponse bias study reached the following conclusions about the Household Survey:

- Unit response rates varied across different subgroups such as housing type, census regions, and urban or rural classification.
- Although statistical tests resulted in statistically significant differences for some characteristics variables, the differences were small.
- Representativeness analyses found that key 2020 RECS estimates are statistically similar to ACS and American Housing Survey (AHS) estimates. Most comparisons were not statistically different; for variables that were significantly different, the differences were small.
- The potential for nonresponse bias in RECS was reduced by applying weighting adjustments, which are described in the Weighting and Sampling Error section.
- When comparing estimates within the 2020 RECS or estimates from previous RECS studies, data users are encouraged to use the relative standard errors (RSE) to determine if two estimates are statistically different from one another.

## Editing and data quality

EIA employed many strategies to analyze and improve data quality in the 2020 RECS Household Survey. For numeric questions where an accurate response was important for modeling the energy use in the household, such as square footage and year of construction, an explicit *Don't Know* response was offered, which led to a categorical follow-up. In addition, range checks were available in the web instrument for numeric responses to reduce the probability that the respondent made a typographical error. Pictures were used as guides for some questions in both the web and mail questionnaires after pretesting indicated that they improved response quality.

All completed surveys went through a validation process to ensure that the correct sampled households responded and that key questions were answered.

After the validation process, the data were thoroughly reviewed for inconsistent responses, numeric-response outliers, and write-in responses when *other* was chosen as an answer to a question. If the review indicated that a response was incorrect, it was either changed to a different valid response using deductive reasoning or changed to indicate that it was missing and then imputed.

## Item imputation

Item nonresponse occurs when respondents do not know or refuse to answer a question in the survey or when a response is determined to be invalid and removed during editing. Item imputation is the process of filling in the missing responses using a statistical model to produce a complete dataset and to reduce the bias associated with item nonresponse.

The 2020 RECS used the hot-deck imputation method. In this method, a recipient case that has a missing value for the variable being imputed is matched with a similar donor case that has a response. The donor's value for that variable is used to replace the missing value for the recipient case. After imputation, final editing reviews ensured questionnaire skip patterns were maintained. For the 2020

RECS, all variables were imputed using the Cyclical Tree-Based (CTB)[7] hot-deck method. This method uses classification trees to group recipients and potential donors and uses a weighted sequential hot-deck imputation procedure[8] in which weights are used to match chosen donors to recipients. This method is the same imputation method that was utilized for the 2015 RECS, with the exception of the variables measuring square footage, which were imputed using the Predictive Mean Neighborhood (PMN)[9] hot-deck method in 2015.

Among the household survey variables in the 2020 RECS public use files, about 240 variables were imputed, and the imputation rate ranged from 0% to 22.1%, with a median of 2.7%. A total of 69% of the variables had an imputation rate of less than 5%; and a total of 87% of the variables had an imputation rate of less than 10%. The median imputation rate was higher than that of the 2015 RECS due to the change from partially in-person data collection in 2015 to entirely self-administered modes in 2020. Without an interviewer present, respondents may have been more likely to leave a question blank if they were unsure of a response. The 2020 RECS Household Survey also included more explicit *Don't Know* response options than previous cycles.

## Weather and Geographic Data

EIA gathers weather and certain geographic indicators from other government agencies to complete the characteristics profile of sampled housing units. The daily average temperature, calculated as the average of the daily minimum and maximum temperature, is available for a number of weather stations within the United States from Climate Data Online (CDO),[10] part of the National Centers for Environmental Information (NCEI). Each sampled RECS housing unit was associated with its nearby weather stations, and then weights were assigned to the weather stations based on the horizontal and vertical distances between them and the RECS housing unit. The RECS housing units were assigned the resulting weighted daily average temperatures, and EIA then calculated daily heating degree days (HDD) and cooling degree days (CDD) from the weighted temperatures, which are summed to yield annualized HDD and CDD values. Thirty-year HDD and CDD averages[11] were also obtained from the CDO data; however, because these normal values were pre-calculated, weighting them was not possible, so they reflect the average weather of the nearest station. Building America climate regions and International Energy Conservation Code (IECC) climate zones are assigned to RECS housing units. These climate regions are assigned based on a housing unit's county, but each county's designation is ultimately based on its typically observed values of annual HDD, annual CDD, and average humidity conditions throughout a year.

[7] Creel, D. V., & Krotki, K. (2006). Creating imputation classes using classification tree methodology. In *Proceedings of the Survey Research Methods Section, American Statistical Association, Joint Statistical Meeting 2006,* pp. 2884–2887.
[8] Cox, B. G. (1980). The weighted sequential hot-deck imputation procedure. In *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp.721−726.
[9] Singh, A., Grau, E., & Folsom, R. (2004). Imputation and unbiased estimation: Use of centered predictive mean neighborhoods method. In *Proceedings of the 2004 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Toronto, Ontario, Canada* (pp. 4351-4358). Alexandria, VA: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/srms/proceedings/]
[10] Formerly known as the National Climatic Data Center (NCDC).
[11] The most recent available data for the 30-year HDD and CDD averages covers the period between 1981 and 2010.

# Weighting and Sampling Error

The 2020 RECS used a single-stage probability design to select a sample of households that represents the housing unit population in the United States. To produce population estimates, the sampled housing units were weighted to represent all housing units including those not in the sample. Base sampling weights, which are the reciprocal of the probability of selection for the RECS sample, were first calculated for each sampled housing unit. The final analysis weights (NWEIGHT) were then produced after applying various adjustments. In addition, replicate weights were computed for variance estimation purposes.

Similar to the weighting adjustment for the web and mail portion of the 2015 RECS, the 2020 RECS final analysis weights were calculated by applying eligibility, unit nonresponse, and poststratification adjustments to the base weights. The eligibility adjustment consisted of two components: an adjustment to unoccupied housing units via a latent-variable technique[12] to predict the probability that a housing unit is occupied or unoccupied and an adjustment to not-primary housing units via a logistic regression model to predict the probability that a housing unit is primary or not-primary. The Generalized Exponential Model (GEM)[13] calibration method was used for the nonresponse and poststratification adjustments.

The last weighting adjustment, poststratification, was intended to improve the quality of the key 2020 RECS estimates by benchmarking them to other sources that are assumed to have better representation of the full population. The poststratification method used for 2020 RECS differed somewhat from previous RECS. EIA typically uses American Community Survey (ACS) estimates from the U.S. Census Bureau as population control totals. However, the Census Bureau did not release ACS estimates for 2020 as a result of the COVID-19 pandemic's impact on data collection efforts. Without official 2020 ACS estimates, EIA used an alternative strategy to develop control totals for 2020 RECS poststratification. This strategy used a combination of 2020 Decennial Census estimates and housing unit occupancy rates from the 2019 ACS. The derived control totals used for 2020 RECS poststratification included state, housing unit type, and age of housing unit. The estimated control totals for housing unit type were calculated based on the proportional estimates of the 2019 ACS. The estimated control totals for age of housing unit was modeled based on the proportional estimates of the 2016 ACS to 2019 ACS.

The final analysis weight for each responding household is the number of households in the population that the observation represents. For example, if the analysis weight for a household is 5,000, that household represents itself and 4,999 other non-sampled households.

Unlike 2015 RECS, which used the Balanced Repeated Replication (BRR) method for replicate weights, the 2020 RECS uses the Jackknife method for variance estimation because Jackknife is more appropriate for a one-stage stratified sample. Sixty Jackknife replicates were constructed.

## Relative standard errors

Estimates from a sample survey like RECS are not exact; they are statistical estimates with some associated sampling error—the result of generating estimates based on a sample rather than conducting a census of the entire population. The standard error provides a measure of the precision of a particular

[12] Biemer, P., Murphy, J., & Kott, P. (2016). Estimating mail or web survey eligibility for undeliverable addresses: A latent class analysis approach. In *JSM Proceedings*, pp. 1166–1172. American Statistical Association.
[13] Folsom, R. E., & Singh, A. C. (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. In *Proceedings of the American Statistical Association, Survey Research Methods Section*, pp. 598−603. Alexandria, VA: American Statistical Association.

statistic for a characteristic based on how variable it is in the population and a given sample size. Standard errors are used with survey statistics to measure sampling error, construct confidence intervals, or perform hypothesis tests. As mentioned above, the standard errors were estimated using the Jackknife method with a coefficient of 0.983 (59/60 replicates).

The relative standard error (RSE) measures how large the standard error is relative to the corresponding statistic; the larger the RSE, the less precise the survey statistic. The RSE is expressed as a percentage and is calculated as (standard error/statistic) x 100.

## Confidentiality of Information

The 2018 Confidential Information Protection and Statistical Efficiency Act (CIPSEA) protects 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. EIA applies disclosure protection measures before releasing the public use data files. These measures include removing localized geographic information such as addresses and top coding certain variables.[14] These disclosure steps mask the data so that the public cannot identify a sampled housing unit or its occupants.

[14] See *How to Use the 2020 RECS Microdata File* for a complete list of top-coded variables.