Independent Statistics and Analysis
U.S. Energy Information Administration

2020 Residential Energy Consumption Survey: Household Characteristics Technical Documentation Summary

June 2022

(revised March 2023)

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

Table of Figures

Table of Tables

The original version of this document (published June 2022) reflected preliminary 2020 RECS household characteristics estimates released in early 2022. We have revised this document to reflect updates to household characteristics data since that release. These updates include characteristics revisions based on a review of household energy billing data. Researchers who previously downloaded the preliminary microdata file or characteristics tables should consider this release to be the final, official characteristics data for the 2020 RECS.

Overview and History

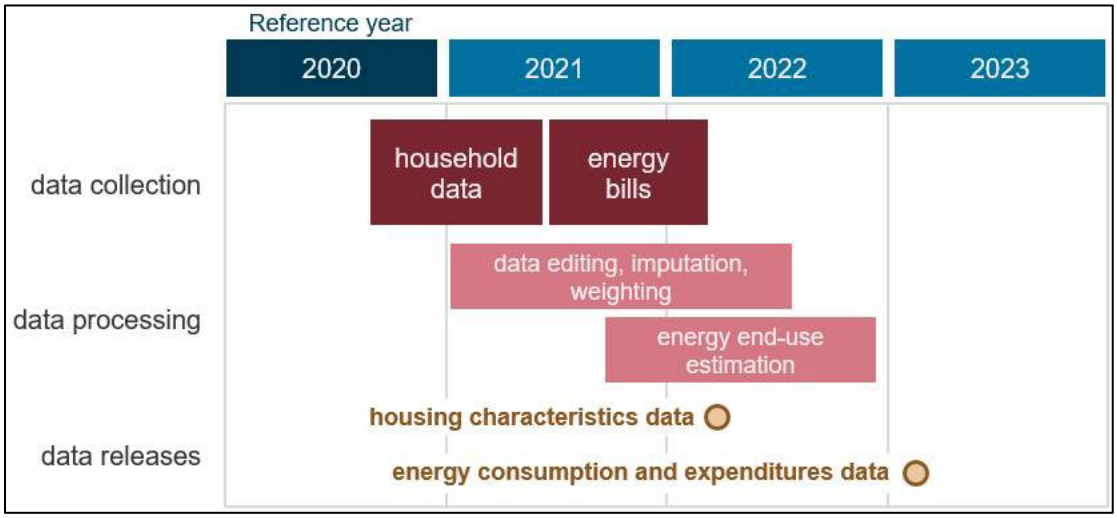
The *Residential Energy Consumption Survey* (RECS) is a periodic study we conduct that provides detailed information about energy usage in U.S. homes. RECS is a multiyear effort consisting of a Household Survey, data collection from household energy suppliers, and end-use consumption and expenditures 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. We use 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 our other products, but they are best suited for comparisons across different characteristics of homes within the residential sector.

We collaborated with IMG-Crown and RTI International to conduct the 2020 RECS Household Survey.

Figure 1. *2020 Residential Energy Consumption Survey* timeline



Survey Design Elements and Changes

We instituted a number of 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.
- We designed the 2020 RECS sample to meet precision requirements for 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.
- We selected housing units for the 2020 RECS 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. We augmented the frame information with supplemental data from the Decennial Census, the American Community Survey (ACS), and other sources to allow us to stratify the frame for better statistical efficiency and to represent 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, we conducted the study 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 could implement several other innovations, including increasing the sample size and moving to an unclustered sample design.
- We added new questions on emerging technologies and usage behavior, and we made some questionnaire changes as a result of the COVID-19 pandemic. The 2020 RECS added questions about solar capacity and installation, electric vehicle ownership and charging behavior, and

smart speakers. Just before data collection, in response to the COVID-19 pandemic, we added questions about working 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%. We collected the Household Survey data in two waves: the first wave was from September through November 2020, and the second wave was from January through April 2021.
- We conducted quality control checks and edits for all data 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. In addition, after we collected consumption and billing data from the ESS, we performed a comprehensive data editing process to reconcile the inconsistencies between respondent-reported characteristics data and the supplier-reported consumption data.
- To address item nonresponse, we used a hot-deck imputation method for the 2020 RECS. In this method, we match a recipient household that has a missing value for the variable being imputed to a similar donor household that has a response for that variable and use that value to replace the missing response. We imputed responses for about 250 Household Survey variables, and the median imputation rate was 2.9%.
- We used a new approach for housing unit control totals in weighting for the 2020 RECS To ensure the responding sample represented housing units at the national, census region, census division, and state levels, we implemented weighting adjustments. 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 used the Jackknife Repeated Replication method for variance estimation. Each RECS estimate has a corresponding relative standard error (RSE). We calculated the RSEs from the estimated variance using replicate weights, which were determined using the Jackknife Repeated Replication method for the 2020 RECS. RSEs are included 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.
- We conducted comprehensive nonresponse bias analysis. Based on the results from the nonresponse bias analysis, we 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.

[1] AAPOR Response Rate Calculator 4.1

Data Products and Revision Policy

Data products

We release 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
- Survey methods documentation

Although we release similar products across survey cycles, we change these products from one cycle to the next to adapt to changes in the residential energy sector and apply new dissemination methods and tools.

The following products are available on the RECS website:

- Data tables for housing characteristics, consumption, expenditures, and end use estimates
- Today in Energy articles
- Public-Use Microdata File and User Guide
- Webinars
- Topic-specific 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

We execute a series of survey data cleaning, editing, imputation, and coding steps to ensure RECS data and estimates meet EIA quality standards. We release preliminary RECS Household Characteristics results at the end of the RECS Household Survey phase. After this release, we perform 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 revisions to the preliminary Household Characteristics estimates, particularly for main-heating fuel and equipment and water-heating fuel. Differences between preliminary and final main space heating fuel estimates are discussed in Appendix A.

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. We do 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 we encourage RECS users to use caution when drawing analysis-based conclusions 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. Housing units on military installations are included. In addition, we benchmark 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

We selected housing units for the 2020 RECS using an Address Based Sample (ABS) design. The frame for this sample is a list of residential addresses, based on the U.S. Postal Service's (USPS) 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 allow us to stratify 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, we excluded drop-point addresses with four or more units from the sample frame. These units are 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 with fewer than four units, we used a sample substitution method.[3] In other words, if a drop-point address was selected, then we would select the nearest non-

[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

drop-point multifamily building with the same number of units as the substitution. A unit in the substitute building would 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 to make hiring and deploying in-person interviewers efficient and cost effective. For the 2020 RECS, introducing an entirely self-administered web and mail design meant that we could use an unclustered design, 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, we designed the 2020 RECS sample to meet the residential energy consumption precision requirements outlined in Table 1, where the fuel consumption data are collected from the ESS.

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

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%	-	-	-

Data source: U.S. Energy Information Administration

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

To achieve these statistical requirements and to improve the precision of other key energy-use metrics, we 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 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 we applied the same optimization process to satisfy the requirements at each geographic level. Once we allocated this initial sample, we applied the same approach to ensure the precision requirements for each energy fuel source were met.

Originally, we specified proposed precision requirements for propane at the national and census region levels. These requirements were dropped due to the need for oversampling and significant increase in sample size. We 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, we used data from the 2009 RECS and the 2015 RECS. Because these two datasets did not have complete data in every state, we derived the estimates using a pooled sample size from both RECS datasets where the combined sample total was 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, we allocated the remaining 7,429 households to each state in proportion to the occupied housing unit distribution of the 2017 ACS.

In addition, to achieve the number of completed cases allocated for each state, we selected additional sample addresses to account for losses due to ineligibility and nonresponse during data collection. We determined the starting sample we deployed 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, we estimated the yield rates 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 ACS. Table 2 contains 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. Within each state, the variables listed below were sorted as implicit stratification variables. Next, we used the Chromy's minimum replacement technique (Chromy, 1979) 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)

[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.

Table 2. Expected completed cases, assumed yield rate, and allocated starting sample, 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 available in English and 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 we conduct the RECS, we review the content and lessons learned from the previous cycle and revise the questionnaire appropriately. 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 as a result of the COVID-19 pandemic
- Information on power outages lasting longer than 24 hours

To improve response quality, we updated questions on square footage, space heating, and air conditioning. We also made minor adjustments 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 voluntary and conducted in two waves: the first wave was September to November 2020, and the second wave was 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-phase, 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 data collection, we determined that, based on response rates in Phase 1, we needed the additional Phase 3 sample to meet the targeted completed cases in certain states. To field Phase 3 efficiently, we decided to release the Phase 3 cases concurrently with Phase 2. This approach contributed to both schedule efficiencies and budget optimization.

An additional benefit to the phased approach was it allowed us 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 levels of formality on 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, neither the increased incentive nor the postcard color design made an impact on the web response rate, so we selected the colorful postcard and the $10 incentive for Phase 2 and Phase 3.

Contact materials

For each phase, we sent a sample of addresses up to six mailings over approximately six weeks:

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

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 we used the corresponding sum of weights in each disposition category.

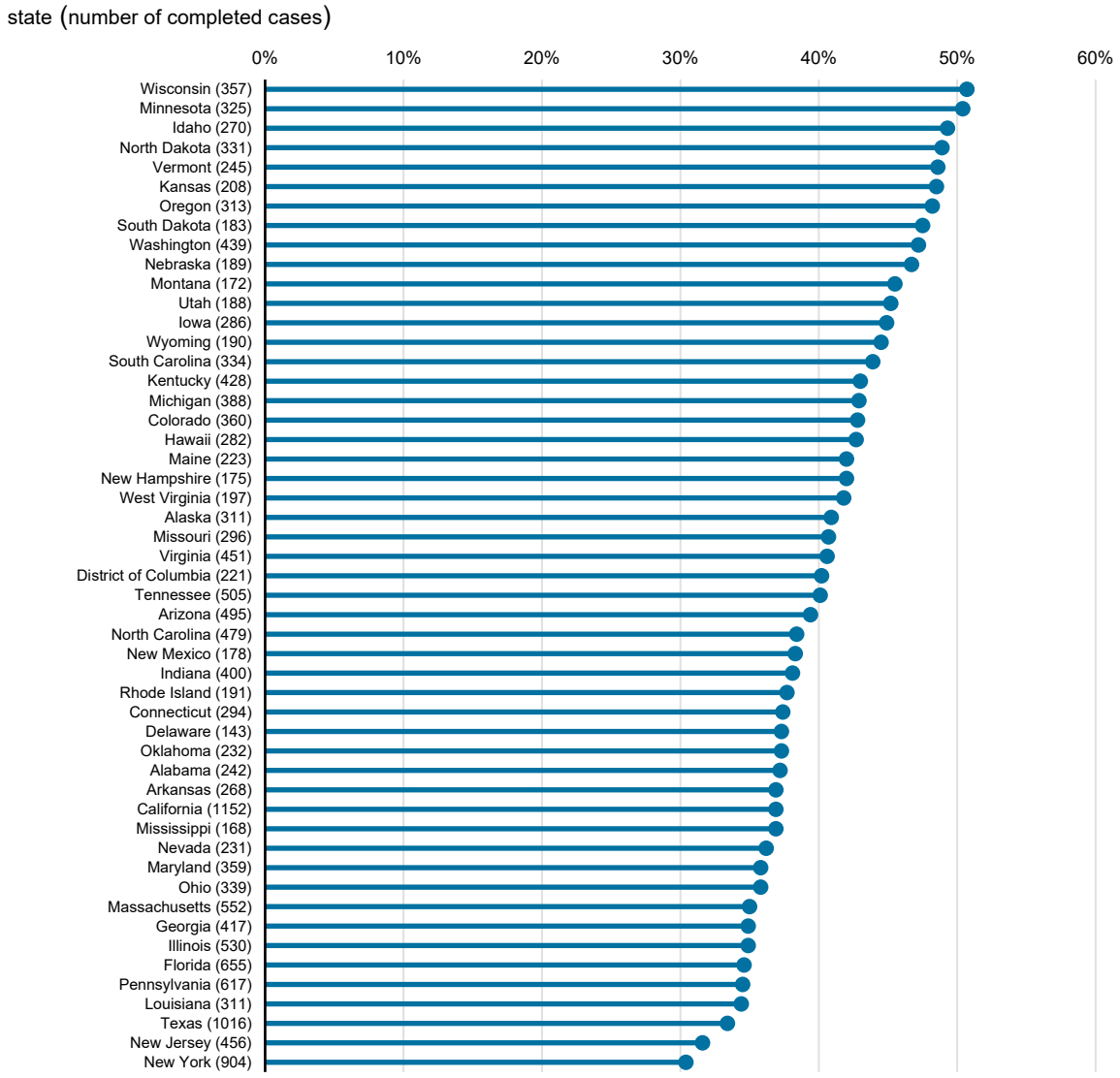
The response rate for 2020 RECS was lower than that of the 2015 RECS, which had an unweighted response rate of 51.2% (weighted response rate was 50.8%). We expected a lower response rate because we administered the 2015 RECS using both in-person and self-administered modes, and the 2020 RECS was entirely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

Sampled households in Wisconsin (50.7%), Minnesota (50.4%), and Idaho (49.3%) had the highest response rates. The states with the lowest response rates were New York (30.4%), New Jersey (31.6%), and Texas (33.4%). The unweighted response rate and weighted response rate are the same for each state because each case within a state has the same sample selection probability (Figure 2).

[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 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.

Figure 2. Unweighted and weighted self-response rates by state, 2020 RECS



We 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. We 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, we compared demographic variables 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.

- Analysis of how well the groups represented the general population 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 should use the relative standard errors (RSE) to determine if two estimates are statistically different from one another.

Editing and data quality

We employed several 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 available, which led to a categorical follow-up. In addition, range checks were available in the web instrument for numeric responses, which reduced the probability of a typographical error. Some questions used pictures as a guide in both the web and mail questionnaires, after pretesting indicated that the images 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, we thoroughly reviewed the data for inconsistent responses, numeric-response outliers, and write-in responses when a respondent chose *other* as a response. If the review indicated that a response was incorrect, it was either changed to a valid response, using deductive reasoning, or changed it to missing and then imputed.

Beginning with the 2015 survey cycle, we introduced an additional quality control step for the RECS Household Survey phase. We reviewed characteristics responses for inconsistencies with billing and fuel delivery data patterns reported during the ESS phase. For example, we changed a response of *electricity* for main heating fuel on the Household Survey to *natural gas* where ESS data indicated strong winter seasonal use in utility-reported natural gas bills for that household. We also reviewed cases where households reported solar photovoltaic (PV) generation to ensure ESS-provided electricity billing data included both the solar PV generation and the utility-generated consumption. These editing steps corrected some measurement errors in the Household Survey, resulting in more accurate main space-heating fuel, space-heating equipment, and water-heating fuel responses relative to the preliminary household data. In addition, these editing steps led to more consistent linkage between characteristics, annualized consumption and cost, and modeled end-use estimates. More detail about these changes is available in Appendix A.

Item imputation

Item nonresponse occurs when respondents do not know or refuse to answer a question in the survey or when we determine a response is invalid and remove it 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 we are imputing is matched with a similar donor case that has a response for that variable. We use the donor's response for that variable to replace the missing response for the recipient case. After imputation, final editing reviews ensured questionnaire skip patterns were maintained. For the 2020 RECS, we imputed all variables 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] where we use weights to match chosen donors to recipients. This method is the same imputation method we used for the 2015 RECS, except for the variables that measure square footage, which we imputed using the Predictive Mean Neighborhood (PMN)[9] hot-deck method in 2015.

We imputed responses for about 250 Household Survey variables, and the median imputation rate was 2.9%. Among the household survey variables included on the 2020 RECS public use file, about 240 variables were imputed, and the imputation rate ranged from 0% to 22.1%, with a median imputation rate of 2.7%. A total of 69% of the variables had a less than 5% imputation rate; and 87% of the variables had a less than 10% imputation rate. 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 in 2020. Without an interviewer present, respondents may have been more likely to leave a question blank if they were unsure how to answer. The 2020 RECS Household Survey also included more explicit *Don't Know* response options than in previous cycles.

Weather and Geographic Data

We gather 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 we assigned weights to the weather stations based on the horizontal and vertical distances between them and the RECS housing unit. We assigned the resulting weighted daily average temperatures to the RECS housing units, and then we calculated daily heating degree days (HDD) and cooling degree days (CDD) from the weighted temperatures, which are summed to yield

[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).

annualized HDD and CDD values. We also obtained thirty-year HDD and CDD averages[11] 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 also assigned to RECS housing units. We based these climate regions assignments 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.

Square Footage Data

Square footage data collection, editing, imputation, and data quality are available in the *Square Footage Technical Documentation* report.

Consumption and Expenditure Data

The Energy Supplier Survey (ESS) data collection and estimates produced from the RECS annualization and end-use modeling processes are available in the *Consumption and Expenditures Technical Documentation* report.

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, we weighted the sampled housing units to represent all housing units including those not in the sample. First, we calculated the base sampling weights, which are the reciprocal of the probability of selection for the RECS sample, for each sampled housing unit. We produced the final analysis weights (NWEIGHT) after applying various adjustments. In addition, we computed replicate weights for variance estimation purposes.

Similar to the weighting adjustment for the web and mail portions of the 2015 RECS, we calculated the 2020 RECS final analysis weights 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. We used the Generalized Exponential Model (GEM)[13] calibration method for the nonresponse and poststratification adjustments.

The last weighting adjustment, poststratification, improved the quality of the key 2020 RECS estimates by benchmarking them to other sources that we assume better represent the full population. The

[11] The most recent available data for the 30-year HDD and CDD averages covers the period between 1981 and 2010.
[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.

poststratification method used for 2020 RECS differed somewhat from previous RECS. We typically use 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. With no official 2020 ACS estimates, we 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 we used for 2020 RECS poststratification included state, housing unit type, and age of housing unit. We calculated the estimated control totals for housing unit type based on the proportional estimates of the 2019 ACS. We modeled the estimated control totals for age of housing unit based on the proportional estimates of the 2016 ACS to 2019 ACS.

The final analysis weight for each responding household was 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 non-sampled households.

Unlike 2015 RECS, which used the Balanced Repeated Replication (BRR) method for replicate weights, the 2020 RECS used the Jackknife method for variance estimation because Jackknife is more appropriate for a one-stage stratified sample. We constructed 60 Jackknife replicates.

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 is a measure of the precision of a particular 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. We estimated the standard errors 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 we collect that could identify respondents or their households is kept confidential and used only for statistical purposes. We use 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.

Appendix A. Comparing preliminary and final main heating fuel, main heating equipment, and water heating fuel estimates

To assist *Residential Energy Consumption Survey* (RECS) data users in determining the impact of the additional quality control step (see the *Editing and data quality* section), the following tables show national comparisons of main space heating fuel, main space heating equipment, and main water heating fuel estimates, relative standard errors (RSEs), and confidence intervals (CI) for the preliminary data release (June 2022) and the final data release (February 2023).

Table A1. Main space heating fuel in U.S. homes, 2020 RECS preliminary and final estimates

	Preliminary estimate (June 2022)			Final estimate (February 2023)		
Main heating fuel	**Estimate (million homes)**	**RSE**	**95% CI for estimate**	**Estimate (million homes)**	**RSE**	**95% CI for estimate**
Natural gas	56.25	0.97	(55.15, 57.34)	62.71	0.77	(61.75, 63.68)
Electricity	48.89	0.95	(47.96, 49.81)	42.57	0.92	(41.79, 43.35)
Fuel oil or kerosene	4.96	3.36	(4.63, 5.29)	4.93	3.50	(4.59, 5.28)
Propane	5.05	4.01	(4.65, 5.46)	5.21	3.76	(4.82, 5.60)
Wood	2.22	5.19	(1.99, 2.46)	2.25	4.88	(2.03, 2.47)
Some other fuel	Q	(-)	(-)	Q	(-)	(-)
Does not use heating equipment	6.10	4.53	(5.55, 6.65)	5.79	4.41	(5.28, 6.30)

Data source: U.S. Energy Information Administration, *Residential Energy Consumption Survey* (RECS)
Note: Q=Data withheld because either the relative standard error (RSE) was greater than 50% or fewer than 10 households were in the reporting sample. CI=Confidence interval.

Table A2. Main space heating equipment in U.S. homes, 2020 RECS preliminary and final estimates

	Preliminary estimate (June 2022)			Final estimate (February 2023)		
Main heating equipment	**Estimate (million homes)**	**RSE**	**95% CI for estimate**	**Estimate (million homes)**	**RSE**	**95% CI for estimate**
Central warm-air furnace	71.98	0.66	(71.02, 72.93)	74.42	0.58	(73.55, 75.29)
Heat pump	17.75	1.84	(17.10, 18.40)	16.13	1.83	(15.54, 16.72)
Steam or hot water system	9.32	2.85	(8.79, 9.85)	9.29	2.89	(8.75, 9.83)
Ductless heat pump (mini-split)	1.12	8.38	(0.94, 1.31)	1.06	8.94	(0.87, 1.25)
Built-in electric units	8.25	2.74	(7.80, 8.70)	7.65	2.83	(7.21, 8.08)
Built-in oil or gas room heater	3.04	5.53	(2.71, 3.38)	3.49	5.37	(3.12, 3.87)
Portable electric heaters	3.24	6.15	(2.84, 3.64)	3.03	6.15	(2.65, 3.40)

Heating stove burning wood	1.98	5.76	(1.75, 2.21)	1.94	5.46	(1.73, 2.15)
Some other equipment	0.74	10.02	(0.59, 0.89)	0.74	9.92	(0.59, 0. 89)
Does not use heating equipment	6.10	4.53	(5.55, 6.65)	5.79	4.41	(5.28, 6.30)

Data source: U.S. Energy Information Administration, *Residential Energy Consumption Survey* (RECS)
Note: Q=Data withheld because either the relative standard error (RSE) was greater than 50% or fewer than 10 households were in the reporting sample. CI=Confidence interval.

Table A3. Main water heating fuel in U.S. homes, 2020 RECS preliminary and final estimates

	Preliminary estimate (June 2022)			Final estimate (February 2023)		
Main water heating fuel	**Estimate (million homes)**	**RSE**	**95% CI for estimate**	**Estimate (million homes)**	**RSE**	**95% CI for estimate**
Natural gas	57.96	0.84	(56.98, 58.93)	59.33	0.75	(58.44, 60.23)
Electricity	58.24	0.82	(57.28, 59.19)	57.04	0.80	(56.13, 57.96)
Fuel oil or kerosene	4.28	4.49	(3.89, 4.66)	4.12	4.44	(3.76, 4.49)
Propane	2.64	5.20	(2.37, 2.92)	2.62	5.20	(2.35, 2.89)
Some other fuel	0.41	15.84	(0.28, 0.55)	0.41	16.08	(0.28, 0.54)

Data source: U.S. Energy Information Administration, *Residential Energy Consumption Survey* (RECS)
Note: Q=Data withheld because either the relative standard error (RSE) was greater than 50% or fewer than 10 households were in the reporting sample. CI=Confidence interval.

In addition to the changes at the national level (highlighted above), some notable differences exist at the state and regional levels as well.

Space heating fuels

Nationally, we estimate about 6.32 million fewer households use electricity for main space heating in the final data set. The number of households using natural gas for main space heating increased by about 6.46 million homes. This change was not evenly distributed across states—with Arkansas, Arizona, California, Colorado, Georgia, Iowa, Kansas, Louisiana, Mississippi, Missouri, New Mexico, Nevada, Oklahoma, South Carolina, South Dakota, Tennessee, and Texas increased between 7% and 9% for the number of homes using natural gas for space heating.

Space heating equipment

Regionally, the final data show a reduction in heat pumps used for main space heating in the South Census region, from 13.58 million heat pumps in the preliminary data to 12.52 million heat pumps in the final data. The largest decreases were in Arkansas, North Carolina, South Carolina, and Tennessee, which all had between a 4% and 5% decrease in the number of homes that use heat pumps for main space heating.

The estimate for the number of homes in the South Census region that use electric furnaces for main space heating also decreased, from 10.31 million homes in the preliminary data to 9.01 million homes in the final data.

There is a corresponding increase in the number of natural gas furnaces in the South Census region in the final data. In the preliminary data, 11.96 million homes used a natural gas furnace for main space heating compared with 14.36 million homes in the final data.

Water heating fuels

Nationally, 1.2 million fewer homes used electricity for their main water-heating fuel in the final data when compared to the preliminary data. The largest group of these homes were in California, where the preliminary data estimated that 2.78 million households used electricity for water heating compared with 2.48 million in the final data.