B. STATISTICAL METHODS

B-1. Description of the Survey Universe

The survey universe for the 2009 RECS will be the same as the 2005 and 2001 RECS, which is the population of all housing units occupied as a primary residence in the 50 states and the District of Columbia. The frame population is based on updates to the 2000 Census.

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

The size of the universe will be interpolated from estimates of the latest Current Population Surveys published by the Census Bureau.

B-2. Sampling Methodology and Estimating Procedures

Introduction. The RECS is a national stratified multistage area probability sample survey that has been conducted since 1978 with minimal alterations to its original sampling methodology. As a result of a design optimization for the 2005 RECS, however, the sample frame was redesigned to granulate primary sampling units (PSUs) down to the level of counties instead of metropolitan statistical areas (MSAs). Moreover, a two-PSUs-per-stratum strategy was adopted prior to second stage stratification and subsequent formation and selection of intermediate areas, listing segments, and the ultimate housing units that comprised the final sample. The 2009 RECS sampling methodology will utilize these essential features of the RECS 2005 frame redesign, making updates to segments selected in 2005 and selecting additional PSUs and segments.

The target number of completed, in-scope housing unit surveys will be 15,400. The purpose of this increase is two-fold:

1. Reduce the sampling error of key statistics for Census Divisions and selected states.

   The 2009 design will restrict the sample precision for average energy consumption per household to within the following RSEs: Nation: RSE=.01; Census Regions: RSE=.02; Census Divisions and individual states: RSE=.03.

2. Increase the number of states for which EIA can publish reliable estimates.

   In 2005, state estimates were made for the four most populous states only: NY, FL, TX, and CA. Preliminary plans for the 2009 RECS are to add estimates for the following eight states: MA, MO, IL, CO, TN, NC, PA, and WI. The additional states are chosen based on geography, climate, and population.

2005 RECS Sampling Methodology. In 2005, the housing units (HHs) in the survey universe were first divided into 19 geographically contiguous census domains, each domain consisting of one or more states. The table below identifies these domains, the state(s) that comprised them, and the corresponding census region and division. For 2009 the number and content of the domains will be revised based on the final list of targeted state estimates. Based on the preliminary selection of states, there will be 27 domains.

Table B1. 2005 RECS Sampling Domains

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

Within each census domain, the housing units were divided into primary sampling units (PSUs) that are either counties or groups of counties. Small counties were combined with other contiguous counties until county groups of at least 2000 occupied housing units are formed as PSUs. City-county pairs were also merged to make PSUs more consistent with other PSUs in the census domain, and Virginia independent cities were merged with surrounding counties to form their PSUs. These PSUs vary in size from just over 2000 to nearly one million. The total number of PSUs thus constructed was 2,910.

The PSUs within each of the 19 geographic domains were then stratified based on PSU size (number of HHs per PSU). Strata where PSUs are larger than a target size were considered as self-representing. The PSUs in other, non-self-representing groups were further stratified by one or more of the following variables: PSU size; weather (heating degree days and cooling degree days); main heating fuel shares; housing unit type shares; and median income. This process stratified all PSUs into 102 strata. There were 27 self-representing (SR) strata that were selected with certainty; 73 non-self-representing (NSR) strata where two PSUs were selected from each; and Hawaii and Alaska where only one PSU was selected from each. The PSUs were selected from each stratum using probability proportional to size, with the count of occupied HHs per PSU from the 2000 census as measure of size (MOS). This resulted in 175 PSUs. The distribution of these 175 PSUs by census division is given in the following table:

Table B2. 2005 RECS PSU Distribution

Census Division	Total Number of PSUs	Self-Representing PSUs	Non Self_Representing PSUs
New England	11	1	10
Middle Atlantic	20	4	16
East North Central	25	3	22
West North Central	12	0	12
South Atlantic	28	2	26
East South Central	12	0	12
West South Central	24	4	20
Mountain	16	2	14
Pacific	27	11	16
United States	175	27	148

Next, the selected PSUs were divided into secondary sampling units (SSUs). The SSUs were defined using 2000 Census definition for blocks, block groups, and census tracts, and stratified using 2000 Census tract data on energy-related characteristics prior to selection. SSUs were then randomly selected within the strata. The total number of SSUs selected was 1,300. The distribution of these SSUs by census division is presented in the table below.

Table B3. 2005 RECS SSU Distribution

Census Division	Number of SSUs
New England	106
Middle Atlantic	156
East North Central	180
West North Central	104
South Atlantic	212
East South Central	102
West South Central	138
Mountain	104
Pacific	198
United States	1,300

Detailed field listings were then created for each of the sampled SSUs, and a penultimate cluster of 50 housing units were randomly selected from each. Finally, an ultimate cluster (which averaged from three to four housing units) was systematically randomly sampled from each of the penultimate clusters. The ultimate clusters constituted the assignments given to the interviewers.

2009 Sampling Methodology Updates. The 2009 RECS sample design will incorporate the 2005 RECS sample geographical units and augment them with a supplementary sample of approximately 125 independently selected PSUs. Selection probabilities will be adjusted based on 2005 data so that the desired 2009 probabilities are attained for the new allocation.

New to the 2009 design is the use of the United States Postal Service Delivery Sequence File (DSF) to create listings for the supplemental PSUs and to update the 2005 listings for new construction. Where the DSF is not adequate, such as in rural areas, traditional listing methods will be used.

Estimation Procedures. Each interviewed sample observation (that is, a housing unit for which an interview is completed) will be assigned a weight. The sum of the weights over all sample observations will equal the number of occupied residential housing units as derived from the American Community Survey. The weight for a particular sample unit will be the number of housing units in the U.S. population

represented by that sampled unit. Based upon the RECS sample results, the weights will be used to produce estimates of population values such as total energy consumption, average expenditures, percent of housing units with dishwashers, and so forth.

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

In addition, four ratio adjustments will be made: 1) reduce the sampling variation in the estimates of the number of housing units by main space-heating fuel resulting from sampling of PSUs during the first stage of the sample design; 2) adjust sample estimates to ACS estimates of the number of households by the nine Census Divisions and, the ACS estimates for the 12 states selected; 3) correct for the type of household (single-person male, single-person female, and all other types); this will correct for the lower response rate among single-person households; 4) readjust the sample estimates to ACS estimates using ratio adjustment.

Precision of Estimates. The 2005 RECS produced estimates of energy consumption and expenditures at the national level that both have a relative standard error (RSE) of 1.2 percent. This precision is considered to be sufficient for the purposes of the survey. The RSEs for household energy consumption and expenditures for the 2005 RECS by census division are given in the table below.

Table B4. 2005 RECS Precision of Estimates

Census Division	Average Household Energy Consumption	Average Household Energy Expenditures
New England	3.9	3.5
Middle Atlantic	2.8	4.3
East North Central	2.9	2.1
West North Central	5.8	6.1
South Atlantic	4.1	3.0
East South Central	3.6	2.3
West South Central	4.0	3.6
Mountain	1.6	2.8
Pacific	4.3	4.0
United States	1.2	1.2

The RECS was designed to produce estimates of energy consumption and expenditures at a minimum level of precision in each census division. With the almost three-fold increase in target sample size for the 2009 RECS, the new sample design will meet the required RSEs as stated in the Introduction.

B-3. Maximizing the Response Rate

Household Survey (Form EIA-457A) EIA and its contractor will employ similar methodology that was used in the 2005 RECS to obtain a minimum response rate of 80 percent. This includes a multi-wave, multi-contact approach using a post card, advance letter, monetary and non monetary incentives, a Spanish language personal interview questionnaire, brochures, and refusal conversion letters.

Prior to the initial contact, a pre-notification postcard will be sent followed by a notification letter. Recent studies have shown that households look at postcards more than letters and will be aware of the survey when the letter arrives—especially if there is a mention of a small gift. Discounted express mail will be used to deliver the advance letter. If the mail should return, and a household doesn't have a viable address,

the interviewer will be aware of the advance letter status prior to contacting the household and deliver a letter in person before initial contact. The letter from the Director of the Energy Consumption Division at EIA will briefly describe the purposes of the survey, stress its importance, announce the impending arrival of the interviewer, and ask if they would have a bill for each energy source ready. This letter, as well as the post card, will mention the monetary and nonmonetary incentives. Refusal conversion letters have been developed for use with households who are concerned with confidentiality or say that they are too busy.

Rental Agent Survey (Form EIA-457C) Some households that rent do not directly pay some or all of their energy bills. In many of these cases, bills are paid through rent payments. The Rental Agent survey is a vehicle used to obtain more accurate data about these respondents' home energy characteristics. The information from the rental agent is used as a basis for changing data in the household record in those cases where it has been judged that the rental agent is more knowledgeable than the household.

For the 2009 RECS, approximately 550 rental agents will be contacted, representing about 1100 households. While still in the field the interviewer will attempt to contact the Rental Agent and administer the questionnaire on-site. If this is not feasible, the interviewer will contact the rental agent and administer the CAPI questionnaire by telephone.

Energy Supplier Survey (Forms EIA-457D through G) Billing data will be requested under mandatory authority for a 17-month period or whatever shorter period is maintained in computer files. If requested, copies of the authorization form signed by households will be mailed to each energy supplier named by the respondent. Consumption and expenditure data are the most important data collected on the RECS, therefore significant time during interviewer training is devoted to collecting accurate supplier billing information, including specific training modules for scanning samples of supplier bills. Obtaining copies of bills from the respondents at the time of the personal interview increases the quality of company names, addresses, and account numbers as well as provides baseline consumption information in cases where supplier data cannot be obtained.

The mandatory requirement for all energy supplier surveys is noted on each supplier survey form. Energy suppliers may submit computer printouts of the requested data as a convenience to reduce their reporting burden. Data are to be collected from fuel oil and LPG (bottled gas) dealers over the telephone following preliminary contacts by mail.. Many of the energy suppliers will be asked to provide information for more than one household.

Editing and Respondent Call-Back Procedures

EIA and its data collection contractors are revising post-collection procedures to emphasize early detection of critical item edit failures. These procedures include flagging cases that require immediate call backs to respondents. This should reduce the number of "missing", "don't know", or "refusal" responses for this cycle of the RECS as compared to previous rounds.

Nonresponse Adjustments

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

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

The "hot deck" procedure will be used for imputing such items as household income, availability of natural gas, year house was constructed, and age of respondent. This procedure requires sorting the file of

households by variables related to the missing item. A household is then randomly selected from a pool of households which have the same values on the selected variables. This "donor" household supplies the value for the missing variable in the household with the missing item.

Random selection procedures will be used primarily to assign dates when those responses are missing, and to impute for missing values that are conditional on other values (for example, the number of rooms air conditioned by central systems, conditional on the total number of rooms).

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

Nonresponse Bias studies

There are three methods that EIA will use to assess nonresponse bias in the 2009 RECS: benchmarking, studying variation within the respondent dataset, and reviewing the impact of weighting adjustments given the 2009 sample frame decision and post-stratification data options (CPS-derived, ACS, decennial counts, etc.). EIA sees this as an opportunity to assure that the address-based sampling methodology is valid for this study, that estimates for evaluating HHS energy programs are sufficiently accurate for their report to Congress, and that we provide sound estimates for new sample domains (states) where energy policy is chiefly written.

EIA will review trends between RECS estimates and other federal survey benchmarks across as many as three survey cycles: 2001, 2005 and 2009. Before final post-stratification adjustments are done, key estimates will be compared for a net estimate of bias at the lowest level of aggregation possible. Minimally, we will compare the number of occupied housing units, main space heating fuel, household square footage, age of home, household income, participation in social programs such as energy assistance, SNAP, etc from the American Community Survey (ACS) and American Housing Survey (AHS). EIA will not compare the ACS and AHS data on energy costs as their method is incompatible with ours—self reported estimates of previous month's electricity and gas bills—and RECS data are used in some way to edit and adjust those to an annual basis. EIA collects actual energy costs for the reference year directly from a householder's energy suppliers. We would compare trends from RECS with EIA's supply surveys for the whole residential sector.

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

With the introduction of the postal frame delivery sequence file (DSF) as a sampling frame for RECS, we intend to evaluate DSF coverage against the traditional area frame method. It was our intention to use the DSF as a 'new construction' update to the 2005 RECS area frame, but the cost and schedule impact of the matching operation was prohibitive. Because we retained the 2005 segments in the 2009 design, we can evaluate the change between matched segments as well as the full survey. We have several options for final post-stratification now (typically adjusting the number of occupied housing units across the sample domains): the DSF, the ACS and the CPS derived OHUs.

B-4. Tests of the Procedures

EIA received permission from OMB to conduct more than nine cognitive interviews for selected items on the 2009 RECS. The results of this testing were used to construct the final version of Section K, which contains questions specifically used for LIHEAP analysis. Less than nine cognitive interviews were conducted on other items in the 2009 RECS household questionnaire, mainly those related to use of electronic devices in the home.

As noted previously, many changes to the RECS Household questionnaire address changes in technology (e.g., TV peripherals), new energy efficiency programs, and question clarity. EIA relied on expert consultation, survey methodologist review, and the limited cognitive interviews in drafting new and revised questions. Additionally, EIA and its contractor are conducting extensive quality review of the data collection instruments and post-collection procedures in advance of the 2009 RECS launch. As part of these procedures, detailed variable and edit specifications are being developed. Much of this pre-survey effort was either shortened or overlooked for the 2005 RECS resulting in a lengthy post-collection phase and delayed release of public data products.

B-5. Statistical Consultation

The principal EIA official directing the RECS sample redesign is Eugene Burns, who can be reached at (202) 586-1385 or by email at *eugene.burns@eia.doe.gov.* The principal EIA official directing the 2009 RECS is James (Chip) Berry, who can be reached at (202) 586-5543 or by e-mail at *james.berry@eia.doe.gov.*

The agency point-of-contact regarding this clearance request is Grace Sutherland, who can be reached at (202) 586-6264 or by e-mail at *Grace.Sutherland@eia.doe.gov.*

APPENDIX A: DATA PRODUCTS FROM THE 2005 RESIDENTIAL ENERGY CONSUMPTION SURVEY

Note: All 2005 RECS data products were released electronically. No hard copy publications are available.

Detailed Data Tables

Household Energy Characteristics:
http://www.eia.doe.gov/emeu/recs/recs2005/hc2005_tables/detailed_tables2005.html

Household Consumption and Expenditures:
http://www.eia.doe.gov/emeu/recs/recs2005/c&e/detailed_tables2005c&e.html

Public-Use Microdata Files
http://www.eia.doe.gov/emeu/recs/recspubuse05/pubuse05.html

Historical Data Tables and Methodological Reports are available at the following link under the "Special Topics" and "Methodology" labels.
http://www.eia.doe.gov/emeu/recs/contents.html