

2009 RECS Sample Plan

Rachel Harter, Mike Kwanisai, Stephanie Eckman, Anna Wiencrot, Ned English, Katherine Dekker, Krishna Winfrey, NORC; Brenda Cox, SRA International, Inc.

Draft December 28, 2009

I. Overview

The Residential Energy Consumption Survey (RECS) is a periodic survey sponsored by the U.S. Energy Information Administration (EIA) that is used to estimate residential energy usage in all occupied housing units in the United States and in each census region and division. The 2009 RECS has been expanded to allow separate estimation capability for about 15 states selected based upon their population size and geographical location. As a consequence, the 2009 RECS will include a substantially larger sample than prior rounds of the survey, which will create an extraordinary opportunity for energy research. However, this level of funding may not be possible in the future, so the 2009 design must allow a return to the 2005 sample size levels. To do this, we will create a 2009 sample design that retains all primary sampling units (PSUs) and segments selected for the 2005 RECS, and supplements them with additional selections. Stratification of PSUs will be consistent with the 2005 stratification to the extent possible. Designing an expanded sample that both preserves the 2005 design and comprises a substantially larger sample for 2009 with appropriate probabilities of selection is an interesting challenge. This document summarizes the specific methods and design features of the 2009 RECS.

The larger sample size for the 2009 RECS was instituted so that it could meet more stringent precision requirements than in the past. The relative standard errors (RSEs) of average household consumption were set to be at least 0.01 at the national level, 0.02 at the census region level, 0.025 for census divisions (with Mountain Division split into Mountain North and Mountain South subdivision, each at 0.03), and 0.025 for the 15 separately reportable states. Fortunately, the target sample of 15,600 completed household interviews (assuming 80% response rate using AAPOR definition 3) is somewhat more than sufficient to meet the original precision criteria, enabling these more ambitious goals. The following sections discuss the specific geographic domains in the design and the PSU and housing unit allocations to support the criteria.

The frame of housing units for selection is based on the U.S. Postal Service's Delivery Sequence File (DSF) to the extent possible. In most geographic areas, the DSF should be as accurate as traditional listings for occupied housing units if not more so (INSERT REFERENCES FROM PROPOSAL). DSF frame building can be implemented with much less cost and time than

traditional listings and allows straightforward identification of new construction. Some areas where many of the DSF addresses are not city-style addresses (P.O. boxes and rural route boxes)¹ still require use of traditional listing, or updates to the 2005 traditionally generated listings. The specific methods used are discussed in Section VI Housing Unit Frame Construction.

Because the 2009 sample is so much larger than the 2005 sample, the requirement for an oversample of Low Income Home Energy Assistance Program (LIHEAP) households was no longer necessary. The sample should contain a sufficient number of LIHEAP households by chance, although without target precision requirements.

II. Geographic Domains

The 2005 RECS ensured geographic control of the sample allocated to each census division by subdividing them collectively into 19 geographic domains. The 19 domains were composed of one census division (New England), subdivisions of divisions, and the four states of New York, California, Florida, and Texas. Precision constraints were applied at the national, regional, and division levels to develop each division's sample allocation, which was then allocated to its geographic domains in proportion to their population. Specifically, estimates of average energy expenditure per household for geographic domains were to meet or exceed the following minimum levels of sample precision: Nation (RSE = .0125); Census Region (RSE = .0275); and, Census Division (RSE = .0450). The design was also required to ensure that state-level estimates could be constructed for New York, California, Texas, and Florida.

For the 2009 RECS, 29 geographic domains were defined. The domains include the 15 states that had precision constraints applied so that their sample allocation ensures separate estimation capability. These states are hereafter referred to as "reportable states". The remaining domains are composed of the remaining portions of the 2005 geographic domains, excluding the reportable states. The 2005 domains already had North and South Mountain Division in separate domains, so treating them as separate divisions did not involve a change in state alignment.

Determining the specific reportable states for separate estimation purposes was an iterative process as PSU allocation to divisions was being determined. After reviewing allocations for 12 initially defined reportable states, it appeared that a small number of additional states could be treated as reportable states with the targeted number of completed cases. EIA prioritized states to consider for reporting, primarily on the basis of population and climatic/geographical diversity. Iterations led to the geographic domain definitions shown in Table 1 of the Appendix, where the 15 reportable states are highlighted. The 2005 geographical domains are also shown for comparison.

¹ Only about 10% of HUs in the U.S. do not have city-style addresses.

Table 1 shows the selected states by census region and division, highlighting the geographical diversity. Tables 2, 3a, and 3b (see Appendix), respectively, show the selected states with the remaining states by climate zone, fuel type and average household energy consumption.

III. Primary Sampling Units

PSU Definition

The 2009 RECS retains the primary sampling unit definitions established for the 2005 RECS. Most PSUs consist of a single county or county-like area, using Census definitions for counties. Some independent cities, primarily in Virginia but also in Missouri, were combined with their surrounding county. Very small counties were combined with an adjacent county to achieve a minimum measure of size, which in 2005 was 2,000 occupied housing units. Altogether, there were 2,866 PSUs in the frame. The cities and counties that were consolidated to form PSUs are listed in Table 4 of the Appendix.

PSU Measure of Size

For purposes of PSU allocation and selection, the PSU measure of size (MOS) is defined as the number of residential addresses in the U.S. Postal Service Delivery Sequence File (DSF). More specifically, we are using the count of city-style addresses, rural route addresses, and those post office box addresses that are marked as the sole address for a residence². For this purpose, addresses in the file flagged as college dormitory addresses, seasonal homes, or vacant homes are excluded. Our definition of addresses included in the MOS determination is the strongest approximation of occupied housing units possible from the DSF. NORC receives a copy of the DSF monthly from Valassis (formerly known as ADVO). The measure of size (MOS) was derived from the file received in early August 2009. August was the latest month for which DSF data were available without delaying the PSU allocation and selection activities.

The decision to use DSF residential addresses as the MOS for PSU selection was made after comparing the DSF county-level counts with occupied housing unit counts available from Claritas and the American Community Survey (ACS). Nielsen sells Claritas annual updated demographic estimates and projections for Census-defined areas as a commercial product for use in market research (see <http://www.claritas.com/collateral/c2/data/update-demographics-fact-sheet-n8027.pdf>). After reviewing the distributions of the percent differences, we concluded that both the Claritas and DSF measures are comparable at the PSU level. While the ACS estimates would not be suitable due to their limited availability, they were considered because of EIA's use of ACS estimates at higher levels for benchmarking purposes. The DSF measure was chosen because of its comparability with the MOS at the segment level for most segments.

² The DSF file from Valassis contains a flag for P.O. Boxes that are the sole mailing address for a housing unit.

PSU Allocation to Domains

The initial allocation of sample PSUs to geographic domains led to the realization that sufficient sample would be available to report additional states separately. The additional sample after initial sample requirements could also be used to improve the precision at various geographic levels. The EIA prioritized the considerations, both increasing the number of states and adding sample to protect against estimation error in determining sample size requirements for precision.

- National RSE of consumption of 0.01, same as before
- Regional RSE of consumption of 0.02, same as before
- Decrease RSE of consumption from 0.03 to 0.025 for divisions, reportable states, and other domains except Mountain North and Mountain South, where .03 is sufficient
- No precision requirement for Alaska and Hawaii
- Include the following states as separately reportable in priority order:
 - NY, CA, FL, TX
 - MA, PA, IL, WI, MO, TN, CO
 - AZ
 - GA
 - VA
 - MI
- Add additional sample to improve division precision, as insurance against variability in our estimates
- Add WA as a reportable state

The allocation was thus determined iteratively until as many of the goals could be met as possible. For each iteration, the process for allocation was as follows:

1. Determine the minimum sample needed to support the precision requirements at the national, regional, divisional, and reportable geographical domain levels. This was accomplished in multiple steps:
 - a. First, the best estimates of average household consumption and the relative standard errors of the estimates were determined. For national, regional, and divisional levels, as well as the four states reported in 2005, the best source was the published 2005 RECS figures. For other states and geographical domains, the

source of average household consumption was state net consumption from 2006 State Energy Data System data divided by the Census Bureau's 2008 estimate of 2006 housing units; the relative standard errors were copied from the RECS division level. Although 2005 RECS estimates were prepared at the geographical domain level, the estimates were considered too variable due to small sample sizes.

- b. Second, the relative standard errors were multiplied by the average household consumption and corresponding sample sizes from 2005 to obtain an estimate of the population standard deviations.
 - c. Third, because the desired relative standard error is a function of the population standard deviation, average household consumption, and new sample size, the new minimum desired sample sizes could be derived for the reportable domains. Although not originally required, we imposed a .03 precision requirement on nonreportable domains (except Alaska and Hawaii) to obtain a minimum sample size.
 - d. We verified that the minimum needed for national estimates, when allocated to regions in proportion to population, did not exceed the sample needed for the region requirements. That is, the combination of regional samples would be more than sufficient for the national estimate. Similarly we verified that the combination of divisional samples would meet or exceed the proportionally allocated regional samples. The division level precision requirements allocated proportionately to domains did sometimes exceed the domain requirements. In this case we proportionately allocated the division sample to domains first, and then expanded the sample in any domains that required additional sample for domain level estimation. Note that most domains had more sample allocated than in 2005, sometimes substantially more.
2. In 2005, all reportable geographical domains had a minimum of 12 PSUs and their associated sample size allocations, with the exception of New England which had 11 PSUs one of which was a self representing PSU.³ For the 2009 sample allocation, a minimum of at least 12 PSUs was set for all reportable domains. In the 2009 allocation, the initial sample sizes in step one were first converted to an estimated number of PSUs by dividing by the 2005 average number of completed cases per PSU. (We calculated this both using an overall average and the geography-specific average. For the final allocation we used an overall average of 28 completed cases per PSU.) If the estimated number of PSUs for a reportable domain was less than 12, the PSU count was set to 12,

³ Four states—New York, Florida, Texas, and California were set aside as separate geographic domains to allow for separate estimation although no precision constraints were imposed on these states. These states had a minimum of ten PSUs assigned to them.

and the number of completed cases was increased accordingly, using the same average per PSU.

3. In the final allocation table (Table 5 of the Appendix), the samples at the higher levels were set to the sums of the sample sizes for the geographical domains within so that the sample summed to the same total regardless of the level. In this way the precision at higher aggregations was generally better than the requirement.

The final resulting allocation of completed cases and PSUs is as shown in Table 5. The final sample count is close to the desired 15,600 and the estimated number of PSUs is approximately 536. It should be noted that these are not necessarily unique PSUs, as some large PSUs may be considered to be selected multiple times. Proportionately more segments will be selected from these larger PSUs. Furthermore, we note that the proportional allocation at division level caused almost all nonreportable domains to have more sample than required for relative standard errors of 0.03.

PSU Stratification and Allocation

To the extent possible, the PSU strata definitions from 2005 were preserved for 2009. The variables used to define the strata are the same as before, with no updates. For this purpose only, even the measure of size is the same as in 2005. In this way, we kept PSUs in the same strata as before.

The 2005 RECS partitioned the U.S. into 19 geographically defined domains for sample allocation purposes with sample allocated within domains in proportion to their measure of size. Next, certainty PSUs were defined for each domain as those PSUs whose size measure exceeded 75 percent of the sampling interval being used in that domain for PSU selection. The Bronx was also defined to be a certainty PSU even though its size measure was not quite 75 percent of the skip interval. The remaining PSUs within each domain were stratified by population size and variables related to energy usage.

Except for the four states that were separate domains, the 2005 RECS strata were not defined by state within the 19 geographic domains, and the frame was not sorted by state prior to sample selection. Expanding the number of geographic domains by extracting the additional individual states to be oversampled meant splitting apart the 2005 strata originally defined for the associated 2005 geographic domains. Where an individual state was extracted from its original 2005 domain, each of its 2005 strata have at least two 2009 substrata – one for each of the newly extracted individual states and another for the remainder domain states. These substrata have smaller population sizes as a consequence of this splitting.

Table 6 in the Appendix shows the resulting 2009 substrata with their definitions. Note that the definitions vary across domains, and even the number of substrata varies by domain. These

substrata only apply to the noncertainty PSUs stratified for selection from the 2005 strata. The 2005 certainty PSUs may be considered as separate substrata within these domains.

Appendix Table 7 shows the number of frame PSUs in the 2009 substrata. Four substrata have no frame PSUs. A separate column has been added to show the number of 2005 certainty PSUs in each domain.

The first step in 2009 RECS sample selection was to use the 2009 measure of size at the PSU level to allocate the sample to each 2009 domain and then to proportionally distribute that sample to each substrata. Following the approach of the 2005 RECS, the 2005 certainty PSUs were each treated as if they were separate substrata within each domain. Then we translated the allocations for the 2009 substrata to hypothetical PSUs assuming 28 completed interviews per PSU. The resulting fractional PSU allocations are shown in Appendix Table 8. Some allocations were extremely small. For example, for substratum S5 of 2009 domain number 4, the New Jersey substratum (NJ) was initially allocated 0.576 PSUs.

As our next step, we defined 2009 certainty PSUs as any PSU whose size measure was 75 percent of the skip interval that would otherwise have been used for PSU sample selection from that substratum. The sample allocations for these certainties were removed from the substratum's total allocation, and then the result was rounded to yield an integer number to select from the noncertainty PSUs in each substratum. In rounding the allocation for each substratum's noncertainty PSUs, we first used these guidelines:

- At least two noncertainty PSUs had to be selected from each substratum. When there were only one or two noncertainty PSUs in the frame, all noncertainty PSUs were selected with certainty.
- Any rounded allocation that exceeded the number of noncertainty PSUs in the frame for that substratum led to all noncertainty PSUs being selected with certainty from that substratum.

For the remaining substrata within each domain, the rounding process began by dividing the expected total completed interviews allocated for each substratum by the rounded number of allocated PSUs, at first using the rounded down nearest integer and then using the rounded up nearest integer. The expected total number of completes per PSU was recorded under both scenarios. Any substratum whose rounded down allocation resulted in more than 32 completes per PSU was automatically rounded up instead. For the remaining substrata within the domain, the individual substratum allocations were rounded up or down to achieve about the same number of expected completes per PSU while maintaining the allocated number of PSUs for the entire domain.

Table 9 in the Appendix shows these domain-level allocations rounded to integers. The 75 percent rule for defining certainties led to 141 certainty PSUs, of which 27 were also certainties

for the 2005 RECS. An additional 25 PSUs entered with certainty either because their substratum had two or fewer noncertainty PSUs, or because their substratum had fewer PSUs than the rounded allocation specified. A total of 253 PSUs were allocated to noncertainty selections across all substrata and domains.

The substratum sample of 2009 certainty PSU selections is composed of all PSUs designated for certainty selection in 2009 regardless of whether they were sampled for the 2005 RECS. Of the 166 certainty PSUs, 27 were included in the 2005 RECS as certainty selections and 56 were sampled as noncertainty PSUs for the 2005 RECS. The 2009 RECS sample of noncertainty PSUs is composed of the union of two independent samples. Sample 1 is the 2005 RECS and its associated sample of 175 PSUs, of which 92 were defined to be noncertainty PSUs under the 2009 RECS. Sample 2 is a complementary sample of PSUs independently selected from each substratum, where the sample size is the difference between the substratum's total 2009 allocation for noncertainty PSUs minus the total number of such noncertainty PSUs that were sampled in the 2005 RECS.

Table 10 in the Appendix shows the distribution of 2005 sample PSUs in the new substrata split into those that are 2009 certainty PSUs and those that are 2009 noncertainty PSUs. When the 2005 noncertainty PSU total is subtracted from the rounded allocation of 2009 noncertainty PSUs, the result is the allocation of the complementary sample of 161 PSUs to be selected from the frame of 2009 RECS. Note that the two-per-stratum rule applies to the total sample, not the complementary sample alone. Thus, there may be substrata where 0 or 1 additional noncertainty PSUs are selected for the complementary sample.

Selection of Noncertainty PSUs for the Complementary Sample

In this section we discuss the selection of the complementary sample of noncertainty PSUs. The actual selection of the 161 PSUs was executed completely independently of the PSU selection for the 2005 sample and from a frame that consisted of all PSUs defined as NSRs in 2009, regardless of whether they were included in the 2005 RECS sample. The resultant complementary sample of non-self-representing PSUs contained nine PSUs that were previously selected for the 2005 RECS. By selecting an independent sample, the probabilities of selection can be determined more easily for estimation, as discussed in the weighting section below. These nine PSUs from the 2005 RECS also selected for the 2009 complementary sample of NSRs will be allocated sample sizes for segments and housing units that reflect their dual selection for the two independent samples. Altogether, the 2009 RECS sample contains 410 unique PSUs.

Prior to selection of the complementary sample, the non-self representing PSUs within each substratum were sorted by the same variables used to sort prior to PSU selection in 2005. The only difference is that the measure of size for selection was the updated measure of size from the DSF. Table 11 in the Appendix shows the sort variables for each substratum. Sorting ensures a proportional mix of sample across these variables, providing an implicit stratification.

Within each substratum, we determined the sampling interval and a random start. We selected the first PSU in the sorted list that contained the starting point within its cumulative MOS and then successively added the skip interval to the starting point and selected the first encountered PSU containing the updated value within its cumulative MOS. The process was repeated until the complete complementary sample had been selected for each substratum. Table 12 in the Appendix compares the counts of 2005 PSUs in the sample by domain, alongside the counts of the 2009 PSUs. The table also shows the number of PSUs each certainty PSU in the 2009 RECS represents in terms of an allocation of 28 completed interviews.

IV. New Construction

For the 2005 RECS, Secondary Sampling Units (SSUs) were selected within PSUs. The SSUs were sorted by a socioeconomic indicator prior to selection. The selected SSUs were canvassed for new construction by contacting government agencies for administrative records data. The segments within selected SSUs were explicitly stratified based on whether they contained at least 25 newly constructed HUs.

The new construction canvass likely led to varying levels of completeness and quality from the various agencies. Furthermore, the new construction canvass was extremely time consuming, in part because of the time involved in getting agencies to respond, and in part because of the need to handle resulting data in various forms. The schedule for the 2009 RECS did not permit this operation.

For the 2009 RECS, we skipped the SSU stage and selected segments directly from PSUs, but controlled for new construction in a different way. The vendor Valassis maintains a database of newly constructed homes since June 2005, roughly the time since the 2005 RECS. Thus the database is an excellent means of identifying new construction for the 2009 RECS. The database includes addresses that have not received mail in the prior year, thereby eliminating most inhabited homes and focusing on truly new construction. According to Valassis, all legal new construction has city-style addresses and is included in the DSF, even in rural areas.

As described in the Implicit Stratification section below, we used the counts of new construction within each segment, along with each segment's measure of size, to sort the segments prior to selection. In this way the proper proportion of new construction segments will be included in the sample.

V. Segments

Segment Definitions

We retained the same segment definitions for segments that were included in the 2005 RECS sample. We do not have segment definitions for 2005 frame segments that were not selected within sample PSUs, so new segments were defined. Furthermore, we defined new segments in newly selected PSUs.

Note that the 2009 segments were defined and selected within PSUs directly, without an intermediate SSU stage. The SSU stage in 2005 was used for the new construction canvass. Because there was insufficient time for a new construction canvass for 2009 RECS, and because we are using the new construction counts from Valassis, which we believe to be superior overall, we skipped the SSU stage.

For the 2005 RECS, segments were defined as blocks or aggregations of blocks, using Census definitions, with a minimum size of 50 occupied housing units.⁴ We followed the same guidelines for the new segments. There is no specified maximum size, in general.

Using large segments with systematic sampling throughout reduces the clustering effect. Larger segments do not pose a problem if the method of listing is to use the DSF file; however, large segments may be substantially more difficult to list in the field. In 2005, some large segments were “chunked” into smaller sub-segments with a single chunk sampled for listing. This is a common practice with traditionally listed segments. We replaced the 2005 chunks with the entire segments for 2009. As noted, the listings in urban segments can easily be updated with the DSF. Given the minimum HU size, the rural segments are probably smaller than those NORC has listed for the National Frame and used for other studies, so it is unlikely that we will have rural segments that truly necessitate chunking.

Special Problems

The sample design team encountered four challenges in implementing the segment definitions as presented in the segment frame file from 2005. We now suspect that the geographical definitions of the segments in the frame were not completely updated to reflect operational decisions that occurred after the 2005 segments were selected. To maintain the 2009 schedule, we temporarily set aside the affected segments and PSUs impacted by these issues so that the necessary manual reviews could be performed. Segment allocation and selection proceeded on the remaining, unaffected geographical areas and was resumed for these impacted areas once all geographical definitions were correctly determined. Therefore, the problem areas were on a schedule that lagged behind the majority. The four issues are described below.

1. Segments with Overlapping Boundaries

The 2005 frame construction effort featured a new construction canvass for stratifying segments. Based on the segment frame file, some new construction segments were overlapping ‘non-new

⁴ Some 2005 segments in new construction strata were defined with a minimum of ___ housing units.

construction' segments. An exhaustive review of the maps used by listers yielded the correct geographic definitions of both 2005 new construction and 'non-new construction' segments. A total of 62 segments were impacted by this issue.

2. Segments with a 2005 frame MOS of Zero

A number of the 2005 segments in the available frame file had a MOS of zero in 2005, which should not be possible. Manual review of the physical maps used by listers in the listing of the frames provided a more accurate geographical definition of the segment actually fielded, and the revised segments had a 2005 MOS that was in keeping with the methodology described in the frame construction documentation. A total of XX segments were impacted by this issue.

3. Chunking of Segments

As noted above, some segments were chunked in 2005. That is, oversized segments were subdivided into smaller geographies, and one of the sub-geographies was selected for listing. Unfortunately, the 2005 segment frame file did not include precise geographical definitions of the selected chunks. For 134 segments, NORC performed a manual review of the maps and listing sheets to determine the correct segment definitions. Ultimately the chunks were abandoned, and the entire segments were used for 2009. Since most segments are being updated with the DSF file, the large size is no longer an issue for listing.

4. Broomfield County, Colorado PSU

Broomfield is a prominent suburb and tier of the Denver metropolitan area in the State of Colorado. Broomfield is a part of the Denver-Aurora Metropolitan Statistical Area and the Denver-Aurora-Boulder Combined Statistical Area. The municipality of Broomfield was incorporated in 1961 in the southeastern corner of Boulder County. Over the next three decades, the city grew through annexations, many of which crossed the county line into adjacent counties. In 1998, the state of Colorado created a separate Broomfield county, integrating areas from four adjacent counties, with a three-year transition period. In late 2001, Broomfield County became the 64th and smallest county of Colorado.

At the time of the construction of the 2005 RECS frame, the data to support this new county as a unique PSU were not available widely. Thus, the 2005 frame used the older county definitions which put the Broomfield segments into adjacent counties as in the 2000 census. For consistency with the 2005 frame, the segments for Broomfield County, with their current MOS data, were separated into their predecessor counties. Because the original frame of PSUs did not have Broomfield listed separately, but our updated PSU MOS excluded the Broomfield portion of the adjacent counties, we adjusted the county MOS figures for the affected counties and determined that the allocation of segments was not affected.

Segment Measure of Size

Within PSUs, a segment measure of size was determined for use in defining and selecting segments from each sampled PSU. For this smaller level of geography, estimation error, geocoding imprecision associated with non-city style addresses, occasional side-of-street geocoding imprecision, and other anomalies become issues for concern when choosing the best data source for defining the segment MOS. The ACS cannot be considered for defining a segment MOS as estimates are unavailable at this level. Claritas estimates of occupied housing units are available at the block group level, but not for individual blocks. DSF addresses can also be aggregated to create a segment MOS, but the quality of the counts depends on the extent of non-city style addresses in the segment. P.O. box addresses (where the P.O. box is the only viable means for a household to get mail delivery) are geocoded at post office locations, and not in the specific blocks containing the corresponding housing units. This ZIP code centroid geocoding may not be a problem for most segments with primarily city-style addresses. However, some rural segments may have a substantial and non-ignorable proportion of addresses that are not city-style.

The selected segment measure of size is a hybrid system that incorporates multiple sources depending upon the incidence of city-style addresses. That is, the measure of size is based upon the likely method of listing if the segment is selected.

Within the selected PSUs, we matched the Census Bureau's Type of Enumeration Area (TEA) code to all blocks and tracts. A TEA code of 1 indicates that the Census treats that block as "urban" for the decennial and other studies, meaning they conduct block-canvassing and mail-outs. Other TEA codes indicate varying conditions that do not permit direct canvassing, especially the prevalence of PO Boxes. We used the TEA code to define a threshold of urbanicity within Census tracts as follows: if 95 percent or more of the population resided in TEA 1 blocks, we tentatively considered it urban. Tracts where less than 95 percent of the population resided in TEA 1 blocks were considered candidates for in-person updates, pending further investigation.

For urban tracts with 95% or more of their population in TEA 1 blocks, we tabulated the MOS for block groups or individual blocks using the DSF-based MOS process as defined at the PSU level. The DSF-based MOS was used for the majority of tracts in the selected PSUs. Tracts with less than 95% of the population on TEA 1 blocks used the Claritas MOS at the block group level. Based on our experience using the TEA code for other survey designs, we expected about 25% of the U.S. population to live in tracts that fall into this category. The TEA-based MOS as described here is the MOS used for allocation and selection of segments with probability proportional to size.

Segment Allocation to PSUs

Initially, the allocation of segments to PSUs was based on the assumed average of 28 completed cases per PSU, and an assumed average of four completed interviews per segment. This implies an average of seven segments per noncertainty PSU. However, we adapted this average as follows:

- Certainty PSUs in small strata would have far fewer than 28 completed cases per PSU, based on the original PSU allocation. For these PSUs we established a minimum number of expected completes at 16. The number of segments in these PSUs was based on the minimum number of completes and the average of four completes per segment. Thus, these PSUs from small strata had 4 segments.
- Certainty PSUs by virtue of the 75% rule (including the certainty PSUs from 2005) required an even number of segments, just in case EIA finds this helpful for variance estimation. The number of completes per PSU is generally greater than 28, in proportion to the size of the PSU. Thus the certainty PSUs may have more than seven segments, and in some cases many more.
- In strata where the PSU allocation exceeded the available frame count, the certainty PSUs have additional segments to accommodate the number of completes allocated for the stratum. Again, the certainty PSUs have an even number of segments.
- Most noncertainty PSUs have seven segments. Noncertainty PSUs selected for 2009 that were also in the 2005 RECS got a double allocation of segments.
- PSUs from the 2005 sample have at least as many segments as they had in 2005.

The number of completed cases per PSU will not be a uniform 28, nor will the number of completed cases per segment consistently be four, for a variety of reasons. First, the certainty allocations of PSUs are based on size, which may or may not be an even multiple of 28. Second, a proportional allocation of HUs to a rounded allocation of noncertainty PSUs or segments will cause some deviation. Third, the 28 per PSU and four per segment are expectations only, given the variability in the eligibility rates and response rates. Even so, at this stage we planned for approximately four completes per segment in determining the segment allocation. Table 13 in the Appendix shows the allocation of segments to PSUs.

Implicit Segment Stratification

Rather than define explicit strata within PSUs, we defined implicit stratification variables and sorted by those variables prior to systematic selection. This greatly simplified the allocation and selection process, while maintaining control over the variables used in the sort.

The first sort variable is the urban/rural status corresponding to the method of listing to be used. That is, segments for which the DSF can be used for listing, even if initially a Claritas MOS was

used, were classified as urban. The remaining segments were classified as rural. For PSUs where all segments are either urban or rural, this sort had no effect.

The second categorical sort variable is a socio-economic status (SES) variable defined as follows:

- For block groups, the lowest level for which census values are available, obtain the percentage of owned homes and the median home value. Also obtain the percentage of rental homes and the median rental value. Assign the block group values to the corresponding blocks, thereby obtaining variables for the segments. (Use weighted averages for segments that cross block groups, although this is likely to be rare.)
- Within each selected PSU, rank the segments by owned home value, and by rent value. Retain the two rankings.
- Define a SES variable as $(\% \text{ owned homes} \times \text{rank for owned home value}) + (\% \text{ rental units} \times \text{rank for rental value})$.
- Within each selected PSU, sort the segments by the SES variable, low to high. Divide the ordered list into thirds, and assign values 1, 2, 3 to each third for Low, Medium, and High SES, respectively. This categorical variable is used for sorting segments prior to selection.

The third sort variable was the percentage of HUs in the segment that were deemed new construction. For each segment we have a total HU count as the MOS, whether it was from Claritas or the DSF. We also have a new construction count from Valassis corresponding to the number of new homes on the DSF since June 2005. As described above, we expect this to be a reasonable measure of new construction, even in rural areas. We divided the new construction count by the total MOS for the segment, yielding a percentage of new construction. We believe that using this variable as the third sort prior to segment selection better ensures a representative proportion of new construction in the sample than the high/low explicit stratification employed at the SSU level in 2005.

Selection of Complementary Segments

For PSUs that are new in the 2009 Sample, the full allocation of segments was selected directly from each PSU's frame of segments. For PSUs that were also in the 2005 RECS sample, the segments in the 2005 RECS sample are automatically in the 2009 RECS. The 2005 segments were removed from the 2009 allocation, and the remaining allocation was selected from all segments in the PSU, including the 2005 segments. In this way the 2009 complementary selection was done independently of the 2005 sample. Having an independent selection simplifies the probabilities of selection for the weights. However, it is possible that a 2005 segment would be selected again, in which case it would get a double allocation of HUs; in fact,

three segments were reselected. The 2009 sample of segments is the union of the 2005 segments and the complementary 2009 segments. The total number of unique segments is 3895.

After sorting, the segments were selected systematically with probability proportional to size, where the measure of size was either the Claritas or the DSF measure of size as described above. The 2005 RECS had no certainty segments, so there was no need to set them aside first. Also, since there was no explicit stratification within PSUs, the allocation to strata was not necessary. Other than these simplifications, the process was the same as for PSUs. Any segment whose measure of size is at least 75% as large as the initial selection interval was automatically selected with certainty. After removing the certainty segments and their size from the allocation, a new selection interval was determined for the remaining allocation to the remainder of the frame, and systematic selection proceeded for the new segments.

VI. Housing Unit Frame Construction

The primary goal of the frame construction phase is to ensure that all occupied housing units in the selected segments have a chance to be selected for the 2009 RECS. We used multiple sources to construct an accurate and complete housing unit sampling frame. The procedures are slightly different in the two types of segments, those selected for the 2005 round and those newly selected for the 2009 round. We describe each of the procedures in detail below.

2005 Segments

1449 segments were selected and listed for the 2005 round of the RECS. As described above, all of these segments will be used again for the 2009 round. However, the housing unit frame for these segments is at this point out of date as listing was completed in the spring of 2005. To ensure that the RECS sampling frame represents all new construction in these segments, we will update the listings in all 2005 sample segments.

As discussed in our response to the 2009 RECS Task Order Request for Proposal, we feel that the best way to update the frame is with a commercially-available address database. These databases are certified by the United States Postal Service (USPS) to contain all residential delivery points known to USPS. NORC licenses this database from Valassis, known as the “ADVO” file or DSF file. Our research with this database, and with versions offered by other companies, has shown that the list offers very good coverage of housing units eligible for selection for in-person surveys in non-rural areas (Montaquila et al. 2009, O’Muircheartaigh et al. 2006, O’Muircheartaigh et al. 2007). In rural areas, the list tends to contain many post office boxes and rural route boxes which are not suitable for in-person surveys such as RECS.

Because USPS coverage is not expected to be uniform across the 2005 segments, we will divide the segments into two sets: those that require field listing because the database coverage is inadequate and those where the database coverage is adequate. The main purpose of updating the listings is to capture new construction. Most new construction has city-style addresses and will

be captured by the DSF, even if the rest of the DSF contains a lot of P.O. boxes or rural route addresses for a particular segment. However, to be conservative, we will make a preliminary urban/rural classification of 2005 segments using the Census-defined TEA code using the same method as described above for the segment MOS. That is, segments containing blocks with TEA code 1 are preliminarily considered urban. A second, manual review as described below may reclassify some rural segments as urban, or vice versa. Each of these two subsets, rural and urban, of the 2005 segments is discussed in more detail below.

Although the TEA indicator from the 2000 census is dated, it is a conservative indicator of the tracts that are not sufficiently covered by city-style addresses. It is conservative in the sense that newly constructed homes are more likely to have city-style addresses, so we might designate more tracts than necessary as requiring in-person listing, if selected. In addition to this first level identification of segments requiring a manual update in the field, we implemented three additional, nested checks to identify those segments requiring manual updating in the field. Based solely on the TEA code, over 25% of the segments were identified as rural, a far higher rate than had been proposed. If the first identification, based on the TEA code identified a segment as rural, we further assessed the ratio of DSF counts to the Claritas scaled count. If the ratio (DSF/Claritas scaled) was greater than or equal to 0.8, the segment was redefined as urban. Next, for the remaining 'rural' segments, we evaluated the ratio if the DSF count to 2000 census counts. If this ratio was greater than or equal to 0.9, the segment was identified as urban. Finally, the remaining rural segments were manually reviewed through satellite imagery and a review of the vacant housing unit counts to confirm the decision to manually update. These steps balanced the cost of manual updating, which was substantial given the dispersed nature of the segments and their relatively small size. Just under 15% of the 2005 segments were selected for manual, field-based updating, as shown in Table 14 in the Appendix. The urbanicity of all RECS geographies is illustrated in Figure 1 in the Appendix.

Following the final determination of segments requiring field based updating or listing, segments were mapped against Census layers outlining Native American Reservations. Six segments fell within tribal lands. Special outreach and contacting efforts were made to prior to the start of field listing.

Rural 2005 Segments

Some 2005 sample segments are not adequately covered by the address database. We updated the listings for these 215 segments in the field using experienced NORC listers and a process called enhanced listing. We provided the listers with the original 2005 listings and updated segment overview maps and block maps. The listers travel around the segment, following the same path as the 2005 listers, confirming 2005 listed units, updating addresses, and, if appropriate, noting the non-existence of a listed unit. They will add housing units missed by the initial listing, if any, and will also add newly constructed streets and housing units to the frame.⁵

⁵ NORC has found that dependent or enhanced listing, listing from a previous frame, is more efficient than traditional listing (O'Muirheartaigh, Eckman and Weiss 2003).

The new additions to the list can be flagged in 2009 listings available to EIA. The listers will have a copy of the DSF for the segment for quality checks and comparison.

All listing sheets were returned to a small group of centralized staff who were trained to provide feedback to listers, correct errors themselves, and seek resolution of more ambiguous recordings. Listing sheets were then double-keyed to reduce data entry error and checked by the centralized staff.

Non-Rural 2005 Segments

In the non-rural segments, which we expect will be the majority of the 2005 segments, one option is to use the DSF address database to identify missed and newly constructed housing units and add them to the frame. This process, which is quite similar to what we already do for the General Social Survey, the Survey of Consumer Finances, and other NORC area-probability household surveys, is EIA's preferred approach. Alternatively, we can replace the 2005 listings with the DSF addresses as for most of the 2009 segments. The approaches are discussed below, but both rely on the extraction of the DSF addresses for the 2005 segments.

From the national DSF address database (nearly 150 million delivery points) we subset the addresses that lie inside zip codes that are near the RECS 2005 non-rural segments. We geocoded all of these addresses to find those that lie inside the physical boundaries of the selected segments. "Geocoding" is the process of interpolating longitude and latitude coordinates based on a street range database, and so determines absolute locations for addresses. At NORC we use *MapMarker Plus* software for geocoding, which is based on *GDT/Teleatlas* street geometry. This is a high quality geocoding program that records its confidence in the resulting coordinates for each address. Any address that is flagged by the routine is manually checked. Generally geocoding is very accurate because the addresses in the DSF from Valassis are clean. Following geocoding, we then determined whether each address was inside or outside of the selected non-rural segments using their absolute locations.

After determining the addresses from the database that are inside the segments, we began to match these addresses to the housing unit listed in 2005 using a probabilistic matching routine that can handle "fuzzy" matches such as "Boulevard" vs. "BLVD." The intent was that the nonmatching addresses from the DSF be flagged as additions to the 2009 sampling frame and inserted after the 2005 listings so that a systematic sample of addresses will select the proper proportion of new construction. The DSF records that do match were dropped from the DSF portion, and the 2005 original listed address was to be retained on the updated 2009 frame.

The match process is not perfect for a variety of reasons. Vanity addresses such as roads with honorary corporate names are not handled well in the match, but that is not a major issue for residential addresses. If the manual listing has no street address ("NO Number") or other address anomalies, resources such as Google Maps or Google Earth combined with a review of the DSF can often solve the problem. The protocol includes watching for known situations that could cause match problems such as the exclusion of directional "NW" or "SW" designations in Washington, DC, or confusion between "ST"

and “RD.” While it is difficult to quantify a success rate, we expect from past experience that the match rate will be more than 90%. We note that the success rate depends heavily on the quality of the manual listings, and the 2005 RECS listings seem to be of good quality.

The listing process was begun, and it quickly became clear that the time involved for the manual reviews as described above were inconsistent with the time schedule for selecting the housing unit sample. The fallback position was to use the DSF listings in place of the 2005 listings. In this way the 2005 nonrural segments are treated the same as the new nonrural segments. As noted above, prior research has shown the DSF listings to be of quality at least as good as traditional listings. While this approach does not permit the identification of “new” housing units on the list, it should result in proportional representation of new construction in expectation. The segments that were matched using the specified protocol will be used to research the differences in approach.

Segments new in 2009

As discussed in previous sections, we selected additional segments for the 2009 RECS. The process for creating a housing unit frame for these segments is quite similar to the process used in the 2005 segments. Again, we classified the segments into two sets, based on the address database coverage, and produced frames in different ways. The Census TEA code was used for the first cut at urban/rural status for assigning a MOS, and then the second, manual check was applied. We were able to re-classify some of the ostensibly "rural" areas as urban.

Rural 2009 Segments

For those 260 segments where the DSF coverage is not adequate, we sent trained and experienced listers to collect addresses. Because we do not have a previous listing in these areas, and the coverage of the address database has been judged insufficient, these segments are being listed using traditional listing methods..

Listers were provided with segment overview maps and ‘zoomed’ block maps that defined the boundaries of the segments and described the natural navigation points – roads, geographic features such as rivers, etc. On the maps, listers were provided a start point in the northwest corner of the segment and were to proceed in recording all housing units in a systematic and choreographed method through the segment. At the completion of the listing operation, listers were instructed to open a sealed envelope that contained the DSF listing of residential units that corresponded to the segment definition. While in the segment, listers compared each match in address description between the DSF and newly completed listing to verify correctness of the actual unit. In addition, for each unit unique to either the field listing or the DSF list, the lister was instructed to locate the physical unit, determine whether it was within segment boundaries, and resolve any ambiguous matches.

These listings will be reviewed by centralized staff, corrected as needed, and double-keyed for quality, similar to the enhanced listings for 2005 segments. To verify the quality of the listings, the quality checks include reviewing the pattern of odd and even house numbers, verifying that all streets contain housing units (or have been specifically marked as not containing units), and

comparing listed counts to expected housing unit totals. Just as above, the listed addresses will be keyed and appended to the sampling database.

Non-rural 2009 Segments

In the non-rural segments, we will not do any field listing but will instead use the addresses from the DSF database that geocode inside the segment as a sampling frame. This is the same frame construction procedure NORC has used on many other area-probability household surveys. DSF addresses that are P.O. boxes, rural route boxes that are not city-style, and college dormitories will be excluded. The listings will be in delivery sequence order.

There are many features in the DSF that can be used to advantage. DSF addresses that are flagged as vacant (but not more than six months) or as secondary homes will be included because they have a chance of being eligible at the time of data collection. This may result in a discrepancy between the list size and the MOS defined above, but at this point in the process it is better to be more inclusive. The DSF also contains a flag for households that have requested that no junk mail be sent. We will ignore the flag but include the addresses in the frame because they are eligible for the study; we would exclude them from any mailings due to contract obligations. Another feature of the DSF is that we know how many distinct units are located at mail “drop points” such as multi-unit buildings or gated communities. The addresses are repeated sequentially, and we are able to assign "synthetic apartments" to each unit based on their delivery sequence order. We have used this procedure for other area probability surveys, including the General Social Survey (GSS) and the Survey of Consumer Finances (SCF), and the interviewers have been able to implement it successfully.

VII. Housing Unit Selection

Systematic Selection Throughout Segments

NORC’s software for selecting samples of housing units from area probability designs selects samples systematically throughout the segment. This method differs from the 2005 RECS in which small, ultimate clusters were selected, and all housing units in the ultimate cluster were selected. An advantage of the systematic approach is that it reduces the clustering of the sample.

Overlap Issues

The ultimate cluster approach is sometimes used by other organizations to prevent sample overlap from one round to the next. That is, by selecting a different ultimate cluster in 2009, the housing units would necessarily be different. NORC’s software allows for the 2005 housing units to be flagged so that they are not reselected. Both approaches are legitimate and artifacts of the software that the organizations use.

While the systematic approach could overlap with the ultimate cluster approach, the number of reselected housing units is likely to be small, with at most one housing unit reselected in a 2005 segment. The decision was made by EIA not to flag the 2005 sample HUs to prevent reselection.

NORC checked whether the 2009 RECS segments overlap with other NORC studies conducted in the field. For some studies there was no geographical overlap at all. Only 296 blocks (1.4% of blocks in RECS geographical areas) have any overlap with other studies. NORC management was satisfied that the potential for housing unit overlap was small and did not require special controls to prevent housing unit reselection.

Initial Sample Sizes

As the next step in the sampling process, we estimated the number of sampled listings that must be selected to yield the desired number of completed interviews. We also allocated the completes for each geographic domain in such a way that we mitigated unequal weighting effects associated with differences in size measures across stages of sampling, as discussed below. The steps in this process are outlined in the remainder of this section.

As a part of the early sample design process, we assigned the desired number of completed interviews (“completes”) to each 2009 geographic domain. PSUs and segments were selected with probability proportional to size, and completes were allocated to them based upon these PSU and segment size measures. Unlike traditional area probability samples based upon Census counts, the size measures for PSUs were not just an aggregated version of their segments’ size measures, so the two size measures may not be exactly equivalent to one another or exactly proportional in relation to one another. In selecting the segment sample and allocating the completes to segments, we decided to deal with any minor discrepancies between the PSU and segment size measures when selecting sample households as the actual count of frame listings is likely to differ from the segment size measure, too.

Translating completes into household listings requires estimates of eligibility rates and response rates. The 2005 RECS is the best source of data for this purpose but is limited in terms of sample sizes for the much finer geographic partitioning being used in the 2009 RECS. An examination of the distribution of sampled PSUs in the 2005 RECS led us to the conclusion that we could develop separate response and eligibility rates for the four states that served as separate 2005 geographic domains (CA, FL, NY, and TX), but for the remainder of the geographic domains we concluded that we can obtain reliable rates only at the division level.⁶ As an additional step, we examined eligibility rates for a recent comparable NORC national survey to see if we needed to fine-tune our eligibility rates, although for this other survey we were unable to estimate eligibility rates below the census region level. We also looked at recent CPS vacancy

⁶ This conclusion is in agreement with our approach in modeling variances for determining the sample allocations to each domain.

rates to see whether the RECS eligibility rates needed adjusting. Ultimately we decided that the estimates from these other surveys were not consistent and reliable enough to merit adjustment to the 2005 RECS rates. We may continue to explore the eligibility patterns after sample selection and throughout RECS data collection.

Using the 2005 RECS rates, then, we determined the number of listed housing units to be selected from each geographical domain to achieve our targeted allocation of completes. As a margin of safety, we plan to select 20% more sample units than we expect to need. The excess sample will be held in reserve and worked only as needed if our assumed rates prove faulty.

We will use NORC's proprietary software to select the sample of housing units. The housing unit listings will be sorted within segment prior to selection by old vs. new listing entry for segments that were in the rural 2005 sample, and otherwise in the order the listings were produced, either around the block or in DSF order. The software allocates the sample to each segment sampled across the domain. The probabilities of selection for each selected PSU and each selected segment within the domain will be input into the program. The program will probabilistically allocate sample listings to each segment in such a way to equalize the overall sampling weights across the geographic domain. The sample units and the associated probabilities of selection will be output.

For area probability surveys, the actual number of sampled listings can be expected to differ from the size measure used for PPS selection of the segment. If we proceeded and selected the same fixed number of sampled listings from each segment, we would not achieve the desired equal selection probabilities across segments within a PSU or across PSUs within a geographic domain and sample precision would be adversely affected. As a consequence, good survey practice calls for adjusting the fixed sample size planned for each segment to account for these differences between the PSU and segment size measures and the actual segment listing count. NORC's selection software adjusts for these differences in a way that equalizes the weights across PSUs and segments within any one sample draw.

While this is the general approach, we needed to make certain adjustments to ensure that the software selects the desired number of listed housing units from each PSU and segment. We checked each geographic domain to determine whether any PSU was assigned more than its proportional number of completed interviews. We do know that some segments were deliberately allocated more than their proportional number of completes. Where either event occurred, adjustments were needed to ensure that the PSU and/or the segment gets its required minimal sample size. To accomplish this, we will adjust the input probabilities so that the program oversamples these units as desired. That is, we will input pseudo-probabilities into NORC's software. This can be done by multiplying the unit's probability of selection by the ratio of the proportional allocation deserved divided by the desired sample allocation. Then, once the sample listings are selected, we will compute corrected probabilities of selection for weights.

Another problem to deal with is sample segments that were selected twice. The segment will only be input once to prevent selecting the same household twice. To deal with this problem we will multiply the segment's probability of selection by $\frac{1}{2}$ so the proper sample size is automatically allocated.

We anticipate that not all segments in all PSUs will have sampled listings prepared and ready for sample selection on January 7th when the sample needs to be selected. To deal with this eventuality for a domain, we would split the sample for the domain into two parts. Set 1 will be the segments with completed listings and Set 2 will be the remaining segments. Completed interviews will be allocated to the two sets based upon the PSU and segment size measures. Then the process described above will be implemented for Set 1 and Set 2 separately with the Set 2 sample not being selected till February 9. This approach would allow us to equalize weights within Set 1 and within Set 2, but not across the entire domain.

Replicates

The selected housing units will be assigned to replicates for sample management purposes. We expect to release the replicates to the field in three batches, with a possible fourth batch if preliminary assumptions prove problematic for meeting the sample targets. The statisticians will review the results of the first release or batch while the field is working the second batch. In that way the statisticians can adjust the number of replicates in the third batch as needed. It is possible that the third batch will include replicates from the reserve sample. A fourth batch will be released only if the second and third batches yield completed cases with lower eligibility rates or response rates than the first batch.

Not every segment will have sample in every replicate; the sample sizes per replicate are too small. Furthermore, it is inefficient to send interviewers to a single housing unit per segment per batch. Therefore, within each segment, the selected housing units will be randomly assigned to the reserve and non-reserve portions. To be conservative, we will assign less than the expected sample needed to the non-reserve portion, just in case not all of the non-reserve replicates are needed. Then all housing units in the reduced non-reserve portion will be assigned to a single replicate or batch of replicates. In this way all segments will be worked in the non-reserve portion, and the field work will be more economical. If reserve replicates are needed, the geographical distribution of those replicates will be less economical.

Because the housing unit sample will be drawn at two different points in time, some replicates, particularly the reserve replicates, will necessarily contain units from segments selected from both draws. Early replicates will have entire segments, however, so this is less of an issue.

New construction

It is extremely important to the EIA that new construction be proportionately represented in the 2009 RECS sample, but not oversampled. While it is not possible to explicitly identify and

control for new construction, the plan as described should represent new construction in right proportion in expectation.

Problems found in the field

With any method of listing, some listing errors will occur. The 2009 RECS is not using a missed housing unit procedure because (1) the listings are quite fresh, (2) the DSF sort order is not conducive to a half-open interval procedure, and (3) the 2005 missed HU procedure turned up almost no corrections. Nevertheless, we want to allow for corrections to erroneous listings such as a missed in-law apartment or an entire missed apartment complex. In such instances, the interviewers will be instructed to call the field manager for clarification or further investigation.

VII. Probabilities of Selection

PSU probabilities

The 2005 RECS PSUs were selected from fixed strata so that the two sampled PSUs represent the total stratum. Their probabilities, when combined with the conditional segment and HU selection probabilities, will yield overall probabilities that reflect the entire stratum.

The 2009 PSUs were selected independently from the 2005 segments, and this selection occurred within substrata which were nested within the 2005 strata. The substrata reflected the splitting required to reflect the finer geographic levels and greatly expanded sample size for the 2009 RECS.

Without adjustments, the two weighted samples will total to twice the national population since we are combining two national samples. The weights without adjustment will also yield distorted population estimates for 2009 substrata that contain one or more 2005 PSUs because the 2005 PSUs have weights that reflect the entire 2005 stratum, not just the 2009 substratum that contains it. Therefore, two different adjustments are needed to combine the two sets of PSUs. These adjustments are made to the weights, but we will need to adjust the input probabilities for the NORC sampling software which requires the computation of segment-level unconditional probabilities.

The first adjustment would be to adjust the PSU selection probabilities for 2005 PSUs so that they represent only the 2009 substratum. Presently the probabilities have the stratum MOS total as it existed in 2005 as the denominator and the PSU MOS times the number of sample PSUs as the numerator. Our adjustment will be to multiply the 2005 PSU probabilities by the ratio of the 2005 MOS stratum total divided by the 2005 MOS substratum total and the ratio of the number of 2005 PSUs in the substratum divided by the total possible PSUs in the substratum. (Note that 2 is the maximum number of 2005 PSUs that could have been selected from the substratum.)

Here we are using the MOS used in the 2005 sample selection. It is technically possible for the adjusted PSU selection probability for a 2005 PSU to be greater than 1. This might occur for a substratum with two 2005 PSUs in the substratum where one has a MOS that is more than 50% of the substratum PSU. Should this occur, we will treat this as an expected frequency of selection instead of a probability.

Next we need to combine the two independent samples together in such a way that allows for a disproportionate number of 2005 and 2009 PSUs within the substratum. That is, the same substratum population may be represented by different numbers of PSUs in the two samples. To combine the two samples, we will multiply the PSU probabilities for the 2009 PSUs by the ratio of (the total number of PSUs across both samples) divided by (the total number of 2009 PSUs). For the 2005 PSUs selected from that substratum we will multiply the 2005 probabilities by [the ratio of (the total number of PSUs across both samples) divided by (the total number of 2005 PSUs)]. If we were dealing with the weights, we would multiply the two sets of PSU weights (2005 vs. 2009) by the fraction of the total substratum PSUs found in that sample. For the probabilities, we need to do the reverse; that is, we need to divide the probabilities by these two fractions.

With these two adjustments, we have PSU probabilities that reflect a combined sample and can be used in weighting. For PSUs selected twice, the probability is the sum of the 2005 probability and the sum of the 2009 probability since the two samples were selected independently.

Segment probabilities

The conditional probabilities, after PSU selection, are essentially the measure of size for the segment times the segment allocation for the PSU. The unconditional probabilities are the conditional segment probabilities times the PSU probabilities. For PSUs that were selected for the first time in 2009, the segment probabilities are this simple.

Some PSUs were also in the 2005 sample. Within these PSUs, the segments might have been selected exclusively for 2005, or there might be a mixture of 2005 and 2009 selections. If all the segments in the PSU were selected in 2005, then the 2005 conditional segment probabilities (based on the 2000 census measure of size) are appropriate without further adjustment.

If there are both 2005 and 2009 segments in the PSU, then the probabilities and weights must be adjusted so that the weights do not sum to double the segment size. Since the PSU geography is the same as in 2005, there is no need to scale the 2005 segments for geographical reasons. However, the conditional probabilities were different, and the weights will need to be adjusted to account for the combination of the two samples.

HU probabilities

NORC's software for selecting a sample of housing units from a multistage cluster design tries to allocate the total desired sample size among segments in such a way that HU probabilities are equalized. The software requires the unconditional segment probabilities as discussed above. It also adjusts for the actual number of available lines in the frame, which may not match the MOS used in selecting the segments.

The 2009 RECS design is allocated to geographical domains in such a way that it is not necessarily desirable to have equal probabilities across all domains. Some domains are intended for separate reporting, and some are for supporting division level estimates, so they may have different variance requirements and sampling rates. Therefore, we plan to select the sample separately for each geographical domain to equalize the probabilities within domains.

IX. References

Montaquila, Jill; Valerie Hsu; J. Michael Brick; Ned English; and Colm O'Muircheartaigh. (2009) A Comparative Evaluation of Traditional Listing vs. Address-Based Sampling Frames: Matching with Field Investigation of Discrepancies. *2009 Proceedings of the American Statistical Association, Survey Research Methods Section* [CD ROM], Alexandria, VA: American Statistical Association.

O'Muircheartaigh, Colm; Ned English; and Stephanie Eckman. (2007) Predicting the Relative Quality of Alternative Sampling Frames. *2007 Proceedings of the American Statistical Association, Section on Survey Research Methods* [CD ROM], Alexandria, VA: American Statistical Association.

O'Muircheartaigh, Colm; Ned English; Stephanie Eckman; Heidi Upchurch; Erika Garcia Lopez; and James Lepkowski. (2006) Validating a Sampling Revolution: Benchmarking Address Lists Against Traditional Field Listing. *2006 Proceedings of the American Statistical Association, AAPOR Survey Research Methods Section* [CD ROM], Alexandria, VA: American Statistical Association.

[Reference list incomplete]

Appendix

Table 1. Geographical Domains for 2009 RECS Reportable States Highlighted in Yellow

Census Geography	2005 Domain	2009 Domain	States
Northeast Region			
New England Division			
	1	1	CT, ME, NH, RI, VT
		2	MA
Middle Atlantic Division			
	2	3	NY
	3	4	NJ
		5	PA
Midwest Region			
East North Central Division			
	4	6	IL
		7	IN, OH
	5	8	MI
		9	WI
West North Central Division			
	6	10	IA, MN, ND, SD
	7	11	KS, NE
		12	MO
South Region			
South Atlantic Division			
	8	13	VA
		14	DC, DE, MD, WV
	9	15	GA
		16	NC, SC
	10	17	FL
East South Central Division			
	11	18	AL, KY, MS
		19	TN
West South Central Division			
	12	20	AR, LA, OK
	13	21	TX

Appendix
Table 1. Geographical Domains for 2009 RECS, cont.
 Reportable States Highlighted in Yellow

Census Geography	2005 Domain	2009 Domain	States
West Region			
Mountain North Sub-Division			
	14	22	CO
		23	WY, ID, MT, UT
Mountain South Sub-Division			
	15	24	AZ
		25	NM, NV
Pacific Division			
	16	26	CA
	17	27	OR, WA
	18	28	HI
	19	29	AK

Appendix

Table 2. Frame PSUs by Geographic Domain and Climate Zone

Reportable States Highlighted in Yellow

Census Division	2005 Domain	2009 Domain	States	Climate Zone								Total	
				Subarctic	Very Cold	Cold	Mixed-Humid	Hot-Humid	Hot-Dry	Mixed-Dry	Marine		
New England	1	1	CT, ME,NH, RI, VT		1	52							53
		2	MA			14							14
Middle Atlantic	2	3	NY			54	8						62
		3	4	NJ			8	13					21
	5		PA			61	6						67
East North Central	4	6	IL			64	37						101
		7	IN, OH			141	39						180
	5	8	MI		9	73							82
9		WI		15	56							71	
West North Central	6	10	IA, MN, ND, SD		48	209							257
		7	11	KS, NE			81	66					147
	12		MO			26	84						110
South Atlantic	8	13	VA				97						97
		14	DC, DE, MD, WV			29	53						82
	9	15	GA				80	71					151
		16	NC, SC			6	121	18					145
10	17	FL					67					67	
East South Central	11	18	AL, KY, MS				211	55					266
		19	TN				95						95
West South Central	12	20	AR, LA, OK				143	66			2		211
	13	21	TX				13	138	43	20			214
Mountain	14	22	CO		9	39					2		50
		23	WY, ID, MT, UT		3	113				1			117
	15	24	AZ			3				10	2		15
25		NM, NV			23				8	11			42

Pacific	16	26	CA			6			26	8	16	56
	17	27	OR, WA			37					33	70
	18	28	HI					4				4
	19	29	AK	7	12							19
Total				7	97	1095	1066	419	88	45	49	2866

Source: U.S. Department of Energy, *High-Performance Home Technologies: Guide to Determining Climate Regions by County*

Appendix

Table 3a. Counts and Percentages of Housing Units by Geographic Domain and Fuel Type

Reportable States Highlighted in Yellow

Census Division	2005 Domain	2009 Domain	States	Total Housing Units	Counts of Housing Units by Fuel Type					
					Electric	Gas	Other Oils	% Electric	% Gas	% Other Oils
New England	1	1	CT, ME,NH, RI, VT	3,469,987	335,448	1,041,797	1,932,984	9.67%	30.02%	55.71%
		2	MA	2,708,108	349,346	1,332,389	980,335	12.90%	49.20%	36.20%
Middle Atlantic	2	3	NY	7,905,969	687,819	4,340,377	2,656,406	8.70%	54.90%	33.60%
		3	NJ	3,471,647	367,995	2,534,302	538,105	10.60%	73.00%	15.50%
	5	PA	5,451,386	975,798	3,009,165	1,253,819	17.90%	55.20%	23.00%	
East North Central	4	6	IL	5,196,936	660,011	4,453,774	20,788	12.70%	85.70%	0.40%
		7	IN, OH	7,792,017	1,625,007	5,733,933	233,237	20.85%	73.59%	2.99%
	5	8	MI	4,503,107	306,211	3,949,225	108,075	6.80%	87.70%	2.40%
		9	WI	2,533,518	319,223	1,960,943	134,276	12.60%	77.40%	5.30%
West North Central	6	10	IA, MN, ND, SD	4,256,207	681,344	3,268,773	157,760	16.01%	76.80%	3.71%
	7	11	KS, NE	1,980,945	382,770	1,544,106	9,376	19.32%	77.95%	0.47%
		12	MO	2,622,330	742,119	1,767,450	13,112	28.30%	67.40%	0.50%
South Atlantic	8	13	VA	3,226,878	1,510,179	1,303,659	316,234	46.80%	40.40%	9.80%
		14	DC, DE, MD, WV	3,838,835	1,340,532	1,943,758	440,790	34.92%	50.63%	11.48%
	9	15	GA	3,864,709	1,723,660	2,075,349	15,459	44.60%	53.70%	0.40%
		16	NC, SC	6,010,960	3,393,276	2,109,167	371,999	56.45%	35.09%	6.19%
	10	17	FL	8,504,557	7,713,633	595,319	42,523	90.70%	7.00%	0.50%
East South Central	11	18	AL, KY, MS	5,238,724	2,510,470	2,561,314	42,784	47.92%	48.89%	0.82%
		19	TN	2,681,661	1,448,097	1,142,388	26,817	54.00%	42.60%	1.00%
West South Central	12	20	AR, LA, OK	4,747,725	1,987,615	2,622,486	7,624	41.86%	55.24%	0.16%
	13	21	TX	9,224,352	4,981,150	4,132,510	9,224	54.00%	44.80%	0.10%

Appendix

Table 3a. Counts of Housing Units by Fuel Type, cont.

Reportable States Highlighted in Yellow

Census Division	2005 Domain	2009 Domain	States	Total Housing Units	Counts of Housing Units by Fuel Type					
					Electric	Gas	Other Oils	% Electric	% Gas	% Other Oils
Mountain	14	22	CO	2,091,502	328,366	1,698,300	4,183	15.70%	81.20%	0.20%
		23	WY, ID, MT, UT	2,184,690	401,597	1,621,191	35,057	18.38%	74.21%	1.60%
	15	24	AZ	2,596,351	1,456,553	1,054,119	2,596	56.10%	40.60%	0.10%
		25	NM, NV	1,913,034	447,820	1,371,889	11,476	23.41%	71.71%	0.60%
Pacific	16	26	CA	13,159,358	3,013,493	9,461,578	52,637	22.90%	71.90%	0.40%
	17	27	OR, WA	4,283,822	2,143,887	1,682,256	182,447	50.05%	39.27%	4.26%
	18	28	HI	499,276	175,745	20,970	0	35.20%	4.20%	0.00%
	19	29	AK	279,293	28,209	138,809	97,473	10.10%	49.70%	34.90%

Source: U.S. Census Bureau, 2005-2007 American Community Survey

Appendix

Table 3b. Net Household Energy Consumption by Fuel Type (in Trillion Btu)

Reportable States Highlighted in Yellow

Census Division	2005 Domain	2009 Domain	States	Net Housing Consumption	Net Housing Consumption by Fuel Type					
					Electric	Gas	Other fuels	% Electric	% Gas	% Other fuels
New England	1	1	CT, ME,NH, RI, VT	391.8	94.1	75.0	222.7	24.02%	19.14%	56.84%
		2	MA	294.9	68.7	116.2	110.0	23.30%	39.40%	37.30%
Middle Atlantic	2	3	NY	831.9	171.4	406.8	253.7	20.60%	48.90%	30.50%
		4	NJ	396.5	101.5	236.1	58.9	25.60%	59.55%	14.85%
	3	5	PA	564.7	186.3	240.8	137.6	32.99%	42.64%	24.37%
East North Central	4	6	IL*	643.5	163.9	438.9	40.7	25.47%	68.21%	6.32%
		7	IN**, OH	851.7	303.7	456.6	91.4	35.66%	53.61%	10.73%
	5	8	MI	525.6	120.7	336.5	68.4	22.96%	64.02%	13.01%
		9	WI	254.3	76.3	132.9	45.1	30.00%	52.26%	17.73%
West North Central	6	10	IA, MN, ND, SD	446.3	153.7	223.6	69.0	34.44%	50.10%	15.46%
		7	11	KS, NE	207.1	80.4	103.5	23.2	38.82%	49.98%
	12	MO	257.0	122.4	103.5	31.1	47.63%	40.27%	12.10%	
South Atlantic	8	13	VA	293.6	155.2	84.5	53.9	52.86%	28.78%	18.36%
		14	DC, DE, MD, WV	351.6	158.3	139.1	54.2	45.02%	39.56%	15.42%
	9	15	GA	330.5	191.8	114.7	24.0	58.03%	34.70%	7.26%
		16	NC, SC	444.2	292.3	85.9	66.0	65.80%	19.34%	14.86%
	10	17	FL	472.2	402.0	16.3	53.9	85.13%	3.45%	11.41%
East South Central	11	18	AL, KY, MS	428.4	270.7	111.9	45.8	63.19%	26.12%	10.69%
		19	TN	230.5	146.3	63.1	21.1	63.47%	27.38%	9.15%
West South Central	12	20	AR, LA, OK	393.1	230.8	134.5	27.8	58.71%	34.22%	7.07%
		13	21	TX	674.5	426.2	205.9	42.4	63.19%	30.53%

Appendix

Table 3b. Net Household Energy Consumption by Fuel Type (in Trillion Btu), cont.

Reportable States Highlighted in Yellow

Census Division	2005 Domain	2009 Domain	States	Net Housing Consumption	Net Housing Consumption by Fuel Type					
					Electric	Gas	Other fuels	% Electric	% Gas	% Other fuels
Mountain	14	22	CO	213.1	60.2	133.2	19.7	28.25%	62.51%	9.24%
		23	WY, ID, MT, UT	235.6	82.7	121.3	31.6	35.10%	51.49%	13.41%
	15	24	AZ	176.6	117.5	39.3	19.8	66.53%	22.25%	11.21%
		25	NM, NV	159.3	64.1	74.2	21.0	40.24%	46.58%	13.18%
Pacific	16	26	CA	878.8	304.2	498.5	76.1	34.62%	56.73%	8.66%
	17	27	OR, WA	354.5	186.8	126.0	41.7	52.69%	35.54%	11.76%
	18	28	HI	14.1	10.9	0.5	2.7	77.30%	3.55%	19.15%
	19	29	AK	39.3	7.2	19.9	12.2	18.32%	50.64%	31.04%

Source: Energy and Information Agency, 2007 SEDS data, <http://www.eia.doe.gov>

* The original SEDS data has other fuels totaling 45.9 for the state of Illinois. This does not yield the reported state net energy consumption of 643.5 trillion Btu considering the reported electricity and natural gas consumption of 163.9 and 438.9 trillion Btu respectively.

** The original SEDS data has other fuels totaling 33.4 for the state of Indiana. This value does not yield the reported state net energy consumption of 296.4 trillion Btu considering the reported electricity and natural gas consumption of 118.2 and 145.9 trillion Btu respectively.

Appendix

Table 4. Consolidated Counties and Cities in RECS PSUs

This table removed for confidentiality purposes.

Appendix

Table 5. Preliminary 2009 PSU Allocation by Geographic Domains

Census Division	2005 Domain	2009 Domain	States	Expected PSUs	Actual PSU Allocation				Expected Completed Cases
					2009 SR	2005 NSR	2009 NSR	Total	
New England	1	1	CT, ME, NH, RI, VT	19	5	3	7	15	530
		2	MA	22	10	0	2	12	608
Middle Atlantic	2	3	NY	33	12	3	5	20	925
	3	4	NJ	7	1	4	5	10	196
		5	PA	12	4	3	5	12	336
East North Central	4	6	IL	12	4	3	7	14	336
		7	IN, OH	17	5	8	5	18	472
	5	8	MI	12	5	1	5	11	336
		9	WI	12	5	4	3	12	336
West North Central	6	10	IA, MN, ND, SD	31	6	4	17	27	877
	7	11	KS, NE	15	5	1	7	12	406
		12	MO	29	7	2	11	20	807
South Atlantic	8	13	VA	12	5	1	6	12	336
		14	DC, DE, MD, WV	10	6	3	2	11	277
	9	15	GA	19	4	4	10	18	522
		16	NC, SC	15	2	4	9	15	427
	10	17	FL	44	17	3	8	28	1,219
East South Central	11	18	AL, KY, MS	17	2	8	8	18	467
		19	TN	12	3	1	7	11	336
West South Central	12	20	AR, LA, OK	12	2	7	4	13	340
	13	21	TX	48	12	9	10	31	1,340
Mountain	14	22	CO	12	7	1	3	11	336
		23	WY, ID, MT, UT	6	3	3	1	7	164
	15	24	AZ	12	2	1	3	6	336
		25	NM, NV	5	3	3	1	7	140
Pacific	16	26	CA	72	23	3	4	30	2,025
	17	27	OR, WA	17	6	4	5	15	488
	18	28	HI	2	1		1	2	54
	19	29	AK	1		1		1	29
Total				536	167	92	161	419	15,001

Appendix

Table 6. 2009 Substrata and Definitions

Census Division	2005 Domain	2009 Domain	States	Strata Definition						
				S1	S2	S3	S4	S5	S6	S7
New England	1	1	CT, ME, NH, RI, VT	150K+ HU, <\$44K INC	150K+ HU, \$44-51K INC	150K+ HU, \$51K+ INC	56-150K HU	<56K HU		
		2	MA							
Middle Atlantic	2	3	NY	90K+ HU, \$63K+ INC	90K+ HU, <\$63K INC	<90K HU				
		3	4	NJ	150K+ HU, <\$44K INC	150K+ HU, \$44-59K INC	150K+ HU, \$59K+ INC	80-150K HU	<80K HU	
	5		PA							
East North Central	4	6	IL	150K+ HU, 5900+ HDD	150K+ HU, <5900 HDD	28-150K HU, 70K+ HU, 6K+ HDD	28-150K HU, <70K HU, 6K+ HDD	28-150K HU, <6K HDD	<28K HU, 6K+ HDD	<28K HU, <6K HDD
		7	IN, OH							
	5	8	MI	70K+ HU, \$48K+ INC	70K+ HU, <\$48K+ INC	30-70K HU	<30K HU			
		9	WI							
West North Central	6	10	IA, MN, ND, SD	100K+ HU	15-100K HU	<15K HU				
		7	11	KS, NE	105K+ HU	15-105K HU	<15K HU			
	12		MO							
South Atlantic	8	13	VA	120K+ HU, <\$45K INC	120K+ HU, \$45K+ INC	31-120K HU	<31K HU			
		14	DC, DE, MD, WV							
	9	15	GA	200K+ HU	70-200K HU	20-70K HU, 3K+ HDD	20-70K HU, <3K HDD	< 20K HU		
		16	NC, SC							
	10	17	FL	80K+ HU, South Coast	80K+ HU, Central	80K+ HU, North	<80K HU			
East South Central	11	18	AL, KY, MS	200K+ HU	30-200K HU, 3500+ HDD	13-30K HU, 3500+ HDD	<13K HU, 3500+ HDD	28-200K HU, <3500 HDD	<28K HU, <3500 HDD	
		19	TN							
West South Central	12	20	AR, LA, OK	35K+ HU, 3300+ HDD	35K+ HU, <3300 HDD	<35K HU, 3300+ HDD	<35K HU, <3300 HDD			
	13	21	TX	North & West, 100K+ HU	North & West, East 20K-100K HU	North & West, West 20K-100K HU	North & West, <20K HU	Coastal & Southern, >35K HU	Coastal & Southern, <35K HU	

Appendix

Table 6. 2009 Substrata and Definitions, cont.

Census Division	2005 Domain	2009 Domain	States	Strata Definition						
				S1	S2	S3	S4	S5	S6	S7
Mountain	14	22*	CO	90K+ HU, 7500+ HDD	90K+ HU, <7500 HDD	25-90K HU	<25K HU			
		23	WY, ID, MT, UT							
	15	24**	AZ	100K+ HU	20-100K HU	<20K HU				
		25***	NM, NV							
Pacific	16	26	CA	100K+ HU, MidCoast/SF	100K+ HU, Central Valley	<100K HU, North/3500+ HDD	<100K HU, Central & South /<3500 HDD)			
	17	27	OR, WA	165K+ HU	70-165K HU	<70K HU				
	18	28	HI	None						
	19	29	AK	None						

HU=Housing Units, INC=Income, HDD=Heating degree days

* Note excluded for confidentiality purposes

** Note excluded for confidentiality purposes

*** Note excluded for confidentiality purposes

Appendix

Table 7. Frame PSUs in 2009 Substrata

Census Division	2005 Domain	2009 Domain	States	Strata Definition							2005 SR	Total	
				S1	S2	S3	S4	S5	S6	S7			
New England	1	1	CT, ME, NH, RI, VT	1	2	1	10	39					53
		2	MA	3	1	3	2	4				1	14
Middle Atlantic	2	3	NY	4	8	46						4	62
		3	NJ	1	6	4	4	6					21
	5	PA	2	2	3	11	49						67
East North Central	4	6	IL	3	0	5	8	3	33	48		1	101
		7	IN, OH	3	4	7	19	21	53	72		1	180
	5	8	MI	4	5	14	58					1	82
		9	WI	2	3	15	51						
West North Central	6	10	IA, MN, ND, SD	5	40	212							257
	7	11	KS, NE	3	13	131							147
		12	MO	2	22	86							
South Atlantic	8	13	VA	3	2	13	79						97
		14	DC, DE, MD, WV	2	4	15	61						82
	9	15	GA	4	3	11	17	116					151
		16	NC, SC	2	12	32	24	75					145
	10	17	FL	6	4	9	46					2	67
East South Central	11	18	AL, KY, MS	2	7	22	88	23	124				266
		19	TN	2	12	29	52	0	0				95
West South Central	12	20	AR, LA, OK	9	12	87	103						211
	13	21	TX	5	17	9	112	7	60			4	214
Mountain	14	22	CO	5	2	4	39						50
		23	WY, ID, MT, UT	0	3	15	99						117
	15	24	AZ	1	4	9						1	15
25		NM, NV	2	6	33						1	42	
Pacific	16	26	CA	6	6	19	15					10	56
	17	27	OR, WA	4	9	56						1	70

	18	28	HI	4								4
	19	29	AK	19								19
Total											27	2866

Appendix

Table 8. Unrounded Allocation by Stratum

Census Division	2005 Domain	2009 Domain	States	Strata Definition							2005 SR	Total		
				S1	S2	S3	S4	S5	S6	S7				
New England	1	1	CT, ME, NH, RI, VT	1.566	4.319	2.122	5.619	5.287					18.913	
		2	MA	5.854	2.529	6.179	1.400	0.787				4.979	21.729	
Middle Atlantic	2	3	NY	6.280	6.746	6.427						13.593	33.046	
		3	NJ	0.546	2.692	2.211	0.962	0.576					6.986	
	5	PA	2.807	0.986	1.717	3.218	3.273						12	
East North Central	4	6	IL	2.065	0	1.360	0.946	0.731	0.920	1.022		4.957	12	
		7	IN, OH	1.401	3.438	1.900	2.057	2.817	1.731	2.184		1.345	16.873	
	5	8	MI	3.256	2.096	2.180	2.216					2.252	12	
		9	WI	1.893	3.034	3.791	3.282						12	
West North Central	6	10	IA, MN, ND, SD	9.561	11.896	9.878							31.335	
		7	11	KS, NE	5.077	4.324	5.112							14.513
	12		MO	10.320	11.427	7.087								28.834
South Atlantic	8	13	VA	2.888	2.267	3.479	3.366						12	
		14	DC, DE, MD, WV	2.400	3.062	2.536	1.886							9.884
	9	15	GA	6.385	1.370	2.432	3.938	4.510						18.635
		16	NC, SC	2.044	4.456	3.683	2.946	2.133						15.262
	10	17	FL	8.343	9.213	8.848	7.827					9.303	43.535	
East South Central	11	18	AL, KY, MS	2.177	1.421	1.525	1.898	5.017	4.625					16.662
		19	TN	3.012	4.239	2.995	1.755	0	0					12
West South Central	12	20	AR, LA, OK	3.228	3.195	2.798	2.925							12.147
		13	21	TX	6.726	5.179	3.749	4.105	4.581	3.692			19.814	47.846
Mountain	14	22	CO	6.188	2.225	2.013	1.574							12
		23	WY, ID, MT, UT	0	1.920	2.138	1.794							5.853
	15	24	AZ	2.009	1.581	1.158						7.252	12	
		25	NM, NV	1.270	0.743	0.892						2.102	5.007	

Pacific	16	26	CA	6.137	6.951	3.552	3.448				52.224	72.313
	17	27	OR, WA	4.656	4.708	4.583					3.469	17.416
	18	28	HI	1.927								1.927
	19	29	AK	1.036								1.036
Total											121.28	535.75
											8	2

Appendix

Table 9. Distribution of 2009 RECS PSU Sample Allocation by Domain and Strata

Census Division	2005 Domain	2009 Domain	States	PSU Category	Strata							2005 SR	Total	
					S 1	S 2	S 3	S 4	S 5	S 6	S 7			
New England	1	1	CT, ME, NH, RI, VT	Total allocation	1	2	1	6	5				15	
				2009 SR	Certainty (by 0.75 rule)	1	2	1	1					5
					Certainty (Small strata)									
				2009 NSR	2005 NSR (also 2009 NSR)				1	2				
	2009 NSR Supplement					4	3					7		
	2	2	MA	Total allocation	3	1	3	2	2			1	12	
				2009 SR	Certainty (by 0.75 rule)	3	1	3	1				1	9
					Certainty (Small strata)				1					1
2009 NSR				2005 NSR (also 2009 NSR)										
	2009 NSR Supplement					2					2			
Middle Atlantic	2	3	NY	Total allocation	4	6	6					4	20	
				2009 SR	Certainty (by 0.75 rule)	3	4						4	11
					Certainty (Small strata)	1								1
				2009 NSR	2005 NSR (also 2009 NSR)		1	2						
	2009 NSR Supplement		1		4							5		
	3	4	NJ	Total allocation	1	3	2	2	2				10	
				2009 SR	Certainty (by 0.75 rule)									
					Certainty (Small strata)	1								1
				2009 NSR	2005 NSR (also 2009 NSR)		1	1	1	1				
		2009 NSR Supplement			2	1	1	1					5	
		5	PA	Total allocation	2	2	2	3	3					12
				2009 SR	Certainty (by 0.75 rule)	2								
Certainty (Small strata)						2								2
2009 NSR	2005 NSR (also 2009 NSR)					1	1	1					3	
	2009 NSR Supplement			1	2	2					5			

Appendix

Table 9. Distribution of 2009 RECS PSU Sample Allocation by Domain and Strata, cont.

Census Division	2005 Domain	2009 Domain	States	PSU Category		Strata							Total		
						S 1	S 2	S 3	S 4	S 5	S 6	S 7		2005 SR	
East North Central	4	6	IL	Total allocation		3		2	2	2	2	2	1	14	
				2009 SR	Certainty (by 0.75 rule)		1							1	2
					Certainty (Small strata)		2								2
				2009 NSR	2005 NSR (also 2009 NSR)				1	1	1				3
		2009 NSR Supplement				2	1	1	1	2			7		
		7	IN, OH	Total allocation		2	4	2	2	3	2	2	1	18	
				2009 SR	Certainty (by 0.75 rule)			3						1	4
					Certainty (Small strata)			1							1
	2009 NSR			2005 NSR (also 2009 NSR)		1		2	1	1	1	2		8	
		2009 NSR Supplement		1			1	2	1			5			
	5	8	MI	Total allocation		4	2	2	2				1	11	
				2009 SR	Certainty (by 0.75 rule)		2							1	3
					Certainty (Small strata)		2								2
				2009 NSR	2005 NSR (also 2009 NSR)			1							1
		2009 NSR Supplement				1	2	2					5		
		9	WI	Total allocation		2	3	4	3					12	
2009 SR				Certainty (by 0.75 rule)		2	1							3	
				Certainty (Small strata)			2							2	
2009 NSR	2005 NSR (also 2009 NSR)						2	2					4		
	2009 NSR Supplement				2	1					3				
West North Central	6	10	IA, MN, ND, SD	Total allocation		5	1	1	0					27	
				2009 SR	Certainty (by 0.75 rule)		5	1							6
					Certainty (Small strata)										
				2009 NSR	2005 NSR (also 2009 NSR)			2	2						4
2009 NSR Supplement			9		8						17				

Appendix

Table 9. Distribution of 2009 RECS PSU Sample Allocation by Domain and Strata, cont.

Census Division	2005 Domain	2009 Domain	States	PSU Category		Strata							2005 SR	Total	
						S 1	S 2	S 3	S 4	S 5	S 6	S 7			
West North Central, cont.	7	11	KS, NE	Total allocation		3	4	5							12
				2009 SR	Certainty (by 0.75 rule)	3	1								4
					Certainty (Small strata)										
				2009 NSR	2005 NSR (also 2009 NSR)		1								1
	2009 NSR Supplement		2		5							7			
	12	MO	Total allocation		2	1	1	7							20
			2009 SR	Certainty (by 0.75 rule)	2	5								7	
				Certainty (Small strata)											
2009 NSR			2005 NSR (also 2009 NSR)				2						2		
	2009 NSR Supplement		6	5							11				
South Atlantic	8	13	VA	Total allocation		3	2	4	3						12
				2009 SR	Certainty (by 0.75 rule)	2	1								3
					Certainty (Small strata)	1	1								2
				2009 NSR	2005 NSR (also 2009 NSR)				1						1
		2009 NSR Supplement				3	3						6		
		14	DC, DE, MD, WV	Total allocation		2	4	3	2						
	2009 SR			Certainty (by 0.75 rule)	2	2								4	
				Certainty (Small strata)		2								2	
	2009 NSR			2005 NSR (also 2009 NSR)				1	2					3	
		2009 NSR Supplement			2							2			
9	15	GA	Total allocation		4	2	3	4	5					18	
			2009 SR	Certainty (by 0.75 rule)	4								4		
				Certainty (Small strata)											
2009 NSR	2005 NSR (also 2009 NSR)		1	1	1	1	1				4				

					2009 NSR Supplement		1	2	3	4				10
--	--	--	--	--	---------------------	--	---	---	---	---	--	--	--	----

Appendix

Table 9. Distribution of 2009 RECS PSU Sample Allocation by Domain and Strata, cont.

Census Division	2005 Domain	2009 Domain	States	PSU Category		Strata								
						S 1	S2	S3	S4	S5	S 6	S7	2005 SR	Total
South Atlantic, cont.	9, cont.	16	NC, SC	Total allocation		2	4	4	3	2				15
				2009 SR	Certainty (by 0.75 rule)	2								2
					Certainty (Small strata)									
				2009 NSR	2005 NSR (also 2009 NSR)		1	1	1	1				4
	2009 NSR Supplement		3		3	2	1				9			
	10	17	FL	Total allocation		6	4	8	8				2	28
				2009 SR	Certainty (by 0.75 rule)	5	4	5					2	16
					Certainty (Small strata)	1								1
2009 NSR				2005 NSR (also 2009 NSR)			1	2					3	
	2009 NSR Supplement			2	6					8				
East South Central	11	18	AL, KY, MS	Total allocation		2	2	2	2	5	5		18	
				2009 SR	Certainty (by 0.75 rule)	2								2
					Certainty (Small strata)									
				2009 NSR	2005 NSR (also 2009 NSR)		1	1	2	2	2			8
	2009 NSR Supplement		1		1		3	3			8			
	19	TN	Total allocation		2	4	3	2					11	
			2009 SR	Certainty (by 0.75 rule)	2	1							3	
				Certainty (Small strata)										
2009 NSR			2005 NSR (also 2009 NSR)			1						1		
	2009 NSR Supplement		3	2	2					7				
West South Central	12	20	AR, LA, OK	Total allocation		4	3	3	3				13	
				2009 SR	Certainty (by 0.75 rule)	2							2	
					Certainty (Small strata)									
2009 NSR	2005 NSR (also 2009 NSR)	1	2	2	2					7				

Appendix

Table 9. Distribution of 2009 RECS PSU Sample Allocation by Domain and Strata, cont.

Census Division	2005 Domain	2009 Domain	States	PSU Category	Strata							2005 SR	Total	
					S1	S2	S3	S4	S5	S6	S7			
West South Central, cont.	13	21	TX	Total allocation	5	5	4	4	5	4		4	31	
				2009 SR	Certainty (by 0.75 rule)	5	1	1		1			4	12
					Certainty (Small strata)									
				2009 NSR	2005 NSR (also 2009 NSR)		1	2	2	2	2			9
2009 NSR Supplement		3	1		2	2	2			10				
Mountain	14	22	CO	Total allocation	5	2	2	2					11	
				2009 SR	Certainty (by 0.75 rule)	4	1						5	
					Certainty (Small strata)	1	1						2	
				2009 NSR	2005 NSR (also 2009 NSR)				1				1	
		2009 NSR Supplement				2	1				3			
		23	WY, ID, MT, UT	Total allocation		3	2	2					7	
				2009 SR	Certainty (by 0.75 rule)		1						1	
					Certainty (Small strata)		2						2	
	2009 NSR			2005 NSR (also 2009 NSR)			2	1				3		
		2009 NSR Supplement				1				1				
	15	24	AZ	Total allocation	1	2	2					1	6	
				2009 SR	Certainty (by 0.75 rule)	1						1	2	
					Certainty (Small strata)									
				2009 NSR	2005 NSR (also 2009 NSR)		1						1	
		2009 NSR Supplement			1	2					3			
		25	NM, NV	Total allocation	2	2	2					1	7	
2009 SR				Certainty (by 0.75 rule)	1						1	2		
				Certainty (Small strata)	1							1		
2009 NSR	2005 NSR (also 2009 NSR)				1	2					3			

Appendix

Table 9. Distribution of 2009 RECS PSU Sample Allocation by Domain and Strata, cont.

Census Division	2005 Domain	2009 Domain	States	PSU Category		Strata							2005 SR	Total			
						S1	S2	S3	S4	S5	S6	S7					
Pacific	16	26	CA	Total allocation		6	6	4	4					10	30		
				2009 SR	Certainty (by 0.75 rule)		4	5	1						10	20	
					Certainty (Small strata)		2	1								3	
				2009 NSR	2005 NSR (also 2009 NSR)				1	2							3
	2009 NSR Supplement					2	2							4			
	17	27	OR, WA	Total allocation		4	5	5						1	15		
				2009 SR	Certainty (by 0.75 rule)		4	1							1	6	
					Certainty (Small strata)												
				2009 NSR	2005 NSR (also 2009 NSR)				2	2							4
	2009 NSR Supplement					2	3							5			
	18	28	HI	Total allocation		2									2		
				2009 SR	Certainty (by 0.75 rule)		1									1	
					Certainty (Small strata)												
				2009 NSR	2005 NSR (also 2009 NSR)												
	2009 NSR Supplement		1											1			
	19	29	AK	Total allocation		1									1		
2009 SR				Certainty (by 0.75 rule)													
				Certainty (Small strata)													
2009 NSR				2005 NSR (also 2009 NSR)		1										1	
	2009 NSR Supplement																
Total	Total	All States	Total allocation		83	100	97	61	34	13	4		27	419			
			2009 SR	Certainty (by 0.75 rule)		65	35	11	2	1				27	141		
				Certainty (Small strata)		12		1							25		
			2009 NSR	2005 NSR (also 2009 NSR)		3	16	3	2	1	6	2			92		

--

		NSR)			0	3	2				
		2009 NSR Supplement	3	37	5	3	2	7	2		161
					6	5	1				

Appendix

Table 10. Distribution of 2005 Sample PSUs by 2009 Category, Domain, and Substrata

Census Division	2005 Domain	2009 Domain	States	2009 PSU Category	Strata Definition									
					S1	S2	S3	S4	S5	S6	S7	2005 SR	Total	
New England	1	1	CT, ME, NH, RI, VT	Certainty		1								1
				Non-Certainty				1	2					3
		2	MA	Certainty	2	1	2	1					1	7
Middle Atlantic	2	3	NY	Certainty	2	1						4	7	
				Non-Certainty		1	2						3	
	3	4	NJ	Certainty										
				Non-Certainty		1	1	1	1					4
		5	PA	Certainty	2	1								3
		Non-Certainty			1	1	1					3		
East North Central	4	6	IL	Certainty	1							1	2	
				Non-Certainty				1	1	1			3	
		7	IN, OH	Certainty		2						1	3	
			Non-Certainty	1		2	1	1	1	2			8	
	5	8	MI	Certainty	1							1	2	
				Non-Certainty		1							1	
9		WI	Certainty	1	1							2		
		Non-Certainty			2	2						4		
West North Central	6	10	IA, MN, ND, SD	Certainty	2								2	
				Non-Certainty		2	2						4	
	7	11	KS, NE	Certainty		1							1	
				Non-Certainty	2	1							3	
	12	MO	Certainty											
		Non-Certainty			2							2		
South Atlantic	8	13	VA	Certainty		1							1	
				Non-Certainty			1						1	
		14	DC, DE, MD, WV	Certainty	2	1		1	2				3	
			Non-Certainty											
	9	15	GA	Certainty	2								2	
				Non-Certainty		1	1	1	1				4	
		16	NC, SC	Certainty										
		Non-Certainty		1	1	1	1					4		
10	17	FL	Certainty	2	2	1					2	7		
		Non-Certainty				1	2					3		
East South Central	11	18	AL, KY, MS	Certainty		1	1	2	2	2			8	
				Non-Certainty	2	1							3	
West South Central	12	20	AR, LA, OK	Certainty			1						1	
				Non-Certainty	1	2	2	2					7	
	13	21	TX	Certainty	2	1						4	7	
		Non-Certainty		1	2	2	2	2				9		

Appendix

Table 10. Distribution of 2005 Sample PSUs by 2009 Category, Domain, and Substrata, cont.

Census Division	2005 Domain	2009 Domain	States	2009 PSU Category	Strata Definition							Total	
					S1	S2	S3	S4	S5	S6	S7		2005 SR
Mountain	14	22	CO	Certainty	2	1							3
		23	WY, ID, MT, UT	Non-Certainty				1					1
	15	24	AZ	Certainty		1						1	2
				Non-Certainty	1								1
		25	NM, NV	Certainty	1							1	2
				Non-Certainty		1	2						
Pacific	16	26	CA	Certainty	2	2	1					10	15
	17	27	OR, WA	Non-Certainty			1	2					3
				Certainty	2						1	3	
	18	28	HI	Non-Certainty		2	2					4	
	19	29	AK	Certainty	1								1
Non-Certainty													
TOTAL				Certainty	33	18	4	1				27	83
				Non-Certainty	3	16	30	23	12	6	2		92
				Total								175	

Appendix

Table 11. Sort variables by Substratum for PSU Selection

Census Division	2005 Domain	2009 Domain	States	Strata Definition							
				S1	S2	S3	S4	S5	S6	S7	
New England	1	1	CT, ME, NH, RI, VT	% SF	% NG	% NG	HDD	HDD			
		2	MA	% SF	% NG	% NG	HDD	HDD			
Middle Atlantic	2	3	NY	% NG	% NG	% NG					
		3	4	NJ	% NG	% NG	% NG	% NG	% Bulk		
	5		PA	% NG	% NG	% NG	% NG	% Bulk			
East North Central	4	6	IL	Med HH Inc	% NG	% NG	% NG	% NG	% NG	% NG	% NG
		7	IN, OH	Med HH Inc	% NG	% NG	% NG	% NG	% NG	% NG	% NG
	5	8	MI	% SF	% NG	% NG	% Bulk				
		9	WI	% SF	% NG	% NG	% Bulk				
West North Central	6	10	IA, MN, ND, SD	Occ. HU	HDD	HDD					
		7	11	KS, NE	HDD	% NG	% NG				
	12		MO	HDD	% NG	% NG					
South Atlantic	8	13	VA	% NG	% NG	% NG	% NG				
		14	DC, DE, MD, WV	% NG	% NG	% NG	% NG				
	9	15	GA	% NG	% EL	% NG	% EL	HDD			
		16	NC, SC	% NG	% EL	% NG	% EL	HDD			
	10	17	FL	% SF	% SF	% EL	CDD				
East South Central	11	18	AL, KY, MS	% NG	% EL	% EL	% EL	% EL	% EL	% EL	
		19	TN	% NG	% EL	% EL	% EL	% EL	% EL	% EL	
West South Central	12	20	AR, LA, OK	% NG	Occ. HU	% NG	% NG				
	13	21	TX	Med HH Inc	% NG	HDD	% NG	% NG	% NG		
Mountain	14	22	CO	Med HH Inc	HDD	HDD	HDD				
		23	WY, ID, MT, UT	Med HH Inc	HDD	HDD	HDD				
	15	24	AZ	HDD	HDD	HDD					
		25	NM, NV	HDD	HDD	HDD					
Pacific	16	26	CA	HDD	Med HH Inc	% NG	% NG				
	17	27	OR, WA	% EL	% EL	% EL					
	18	28	HI	None							
	19	29	AK	None							

HU=Occupied housing units, Med HU Inc=Median household income, HDD=Heating degree days, % SF=Percentage of single family housing units, % NG=Percentage of HU using natural gas as main heating fuel, % Bulk=Percentage of HU using bulk oil as main heating fuel, % EL=Percentage of HU using electricity as main heating fuel,

Appendix

Table 12. Distribution of 2005 PSU Sample, Newly Selected PSUs and Certainty PSUs by Domain

Census Division	2005 Domain	2009 Domain	States	2009 PSU Sample			2009 Certainty PSU sample	
				PSUs in 2005 sample	Newly Selected sample	Total	Certainty PSUs	Total PSUs represented *
New England	1	1	CT, ME, NH, RI, VT	4	10	14	5	8.963
		2	MA	7	5	12	10	20.941
Middle Atlantic	2	3	NY	10	10	20	12	24.626
		3	4	NJ	4	5	9	1
East North Central	4		5	PA	6	5	11	4
		6	IL	5	8	13	4	7.022
	5	7	IN, OH	11	7	18	5	4.783
		8	MI	3	8	11	5	5.508
West North Central	6	9	WI	6	6	12	5	4.927
		10	IA, MN, ND, SD	6	21	27	6	10.318
	7	11	KS, NE	1	11	12	4	6.015
12		MO	5	15	20	7	16.180	
South Atlantic	8	13	VA	2	10	12	5	5.156
		14	DC, DE, MD, WV	6	5	11	6	5.462
	9	15	GA	6	12	18	4	6.385
		16	NC, SC	4	11	15	2	2.044
East South Central	11	17	FL	10	18	28	17	33.012
		18	AL, KY, MS	8	9	17	2	2.177
West South Central	12	19	TN	4	7	11	3	3.882
		20	AR, LA, OK	8	4	12	2	1.638
Mountain	13	21	TX	16	13	29	12	29.829
		22	CO	4	7	11	7	8.413
	14	23	WY, ID, MT, UT	4	3	7	3	1.920
		24	AZ	3	3	6	2	9.261
Pacific	15	25	NM, NV	5	2	7	3	3.372
		26	CA	18	12	30	23	66.063
	17	OR, WA	7	7	14	6	8.938	
18	27	HI	1	1	2	1	1.413	
	19	28	AK	1		1		
Total				175	235	410	166	302.585

* Assuming an allocation of 28 completed interviews per PSU. The figures include the 27 certainty PSUs from 2005 RECS.

Appendix

Table 13. Allocation of Segments to Selected PSUs, With Revised Expected Completes

This table removed for confidentiality purposes.

Appendix

Table 14. Final Determination of Urban Segments vs. Those Needing Field Listing

Frame	Selected PSUs	Selected Segments	Urban Segments	Field Updated Segments	% Field Segments
2005	175	1449	1234	215	14.8
2009 Supp	266	2449	2180	269	11.0
	441	3898	3414	484	12.4

Figure 1. Urbanicity of RECS Geographies

This figure removed for confidentiality purposes.