Supporting Statement
State Energy Program Evaluation
OMB Control Number: XXXX-XXXX

B. Collections of Information Employing Statistical Methods

1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods to be used.

Overview of SEP Programmatic Activities and Sampling Plan

As discussed the overall objective of this study is to provide quantitative estimates, at the national level, of key program outcomes resulting from the State Energy Program (SEP) in 2008 and ARRA periods (2009-2011). The principal outcomes to be quantified are:

- Energy and demand savings;
- Renewable energy generation;
- Energy cost savings;
- Carbon emissions reductions; and
- Job impacts.

Over the course of the study period, SEP funded thousands of initiatives representing a wide range of strategies to reduce energy consumption including:

- Direct subsidy of energy-related capital improvements;
- Subsidy of capital improvements via reduced interest loans and guarantees;
- Technical training and support for facility managers and equipment vendors;
- Support for building code development and enforcement; and,
- Support for broad energy policy development.

Given the breadth of SEP and ARRA funded initiatives, evaluation of these activities requires a two-stage sampling approach. In the first stage, DOE selected the individual state-level initiatives, referred to as Programmatic Activities (PAs), to be assessed. In the second stage, DOE will select a sample of informants for each of the selected PAs.

For the first sampling stage, DOE grouped the PAs into strata that make up similar types of initiatives. The first grouping was by Broad program area categories (BPAC). Results of this study will be reported by BPAC Each BPAC was further divided into sub-categories, which varied by BPAC. For example, PAs that provide loans, grants and incentives for building retrofits constitute one BPAC, which is further divided into residential and non-residential sub-categories.

The groupings defined by BPAC-subcategory combination served as sampling strata. Once all programmatic activities were classified into BPACs and sub-categories, DOE selected a sample of PAs for analysis within each BPAC-subcategory stratum, as described further below.

The rationale for this approach reflects two methodological considerations. First, the variation in key outcome indicators will likely be smaller for groups of PAs that share the same types of initiatives and have similar operational systems than it will be for the population of PAs as a whole. Second, the research methods appropriate to the evaluation of an individual PA can vary by BPAC and subcategory. However, DOE will be able to apply standardized methods for quantifying key outcomes for PAs within a given BPAC. This has the added advantage of facilitating project management efforts, improving the efficiency of primary data collection and improving the efficiency of the data analysis and reporting efforts. That is, the PAs are stratified into homogeneous subgroups both to improve statistical sampling efficiency and to allow efficient execution of the evaluations of individual PAs. DOE details the implementation and results of the Stage 1 sampling below.

Quantification of outcome indicators for each sample PA will require collection of data and information from individuals involved in those initiatives, including: program participants, program administrators and staff, vendors who serve program participants, and observers of the targeted markets and policy organizations. This study does not include an evaluation of non-SEP Programs. Commonly, SEP-funded programs have other funding sources that may be external to the ultimate subgrantee. In order to assess the impact of SEP, DOE must also understand the relative impact of similar leveraged or cost-shared programs to SEP's funded efforts. The exact configuration of the study subjects will vary by BPAC and BPAC subcategory.

For example, for PAs that provide financial incentives, technical support, or training to facility owners to encourage energy-efficient capital improvements, DOE plans to interview samples of participants to characterize what measures they took and the influence of the program upon their decisions. DOE will apply probability sampling methods for selecting participants in this Stage 2. Those sampling methods are described in the section titled "Stage 2 Sampling within PAs" on Page 25 of this document. This approach will support proper extrapolation of the results to the PA level as well as variance calculations.

Other types of PAs, such as efforts to change state-level building codes or regulations in regard to renewable energy facilities, do not generate lists of participants. Moreover, the number of individuals engaged in these efforts is relatively small, and reliable assessment of program outcomes requires opinions from specific experts who have a detailed understanding of the associated processes, rather than random selection. Quantification of outcomes for these types of PAs will not make use of statistical sampling techniques in Stage 2. Once DOE estimates outcomes (i.e. energy and demand savings, emissions reductions, jobs) for each PA, DOE will expand PA level outcomes to the full BPAC group using conventional multi-staged sampling techniques (Stage 2 sample expansion techniques are described in the section titled "Estimation Procedure" on Page 23 of this document).

As stated above, quantification of SEP outcome indicators may require an understanding of the relative impact of similar leveraged or cost-shared programs to SEP-funded efforts. As such, some survey instruments collect data to build a narrative that establishes the veracity of leveraged or cost-shared impacts, leading to the ultimate quantification of the proportion of observed outcomes that are attributable to the SEP effort itself. In a given sequence of questions, some will establish the roles of the various contributing program efforts and help determine the way in which they interact, without immediately yielding a quantifiable result. Examples of such questions include those asked of Program Managers of Non-SEP Programs which run concurrent to SEP that are designed to clarify the decision-making process involved in designing the Non-SEP Program. At the heart of estimating the impacts of leveraged funding or cost-sharing is the determination of whether or not the presence of the sampled SEP program influenced funding decisions and/or program design. In most surveys, a final question or questions in an attribution-related sequence were specified to collect a quantitative indicator of fund leveraging or cost-sharing.

Figure 1 below depicts the process used to complete Stage 1, and our planned approach for Stage 2. Stage 1 is represented by the upper portion of the diagram, while Stage 2 is depicted by the lower portion of the diagram. Moving from the top of the diagram downward, the following lists the individual steps for each of these two general stages:

Stage 1– Selection of Sample of PAs

A. DOE developed the universe of PAs by merging the WINSAGA (A1) and PAGE (A2) databases. Within each database, SEP/ARRA grants awarded to states were listed under separate market titles. Within each database, each record corresponded to an individual grant, or "Market Title." A state could have multiple market titles. DOE reviewed the information available for each market title, and determined whether to treat the entire Market Title as a single Programmatic Activity, or as 2 or more distinct Programmatic Activities. In rare instances, multiple small Market titles within the same state with similar types of activities were combined into a single Programmatic Activity for purposes of the sampling frame. From the information provided, DOE identified the funding amounts associated with each PA in the final sampling frame, and assigned each one to a Broad program area category and Subcategory. Thus the sampling frame (A3) consisted of a set of PAs characterized at a minimum by BPAC and subcategory (stratum) as well as by state and funding level.

   Following guidance from DOE, DOE restricted the study universe to a subset of all the BPAC-subcategory strata, together accounting for over 80 percent of each study period's total budget. As described further below, the restriction mostly eliminated smaller strata in terms of program budget. This restriction was imposed to allow more effective use of evaluation resources, by reducing the number of different types of evaluations that needed to be developed and the number of different BPACs for which stand-alone estimates will be produced

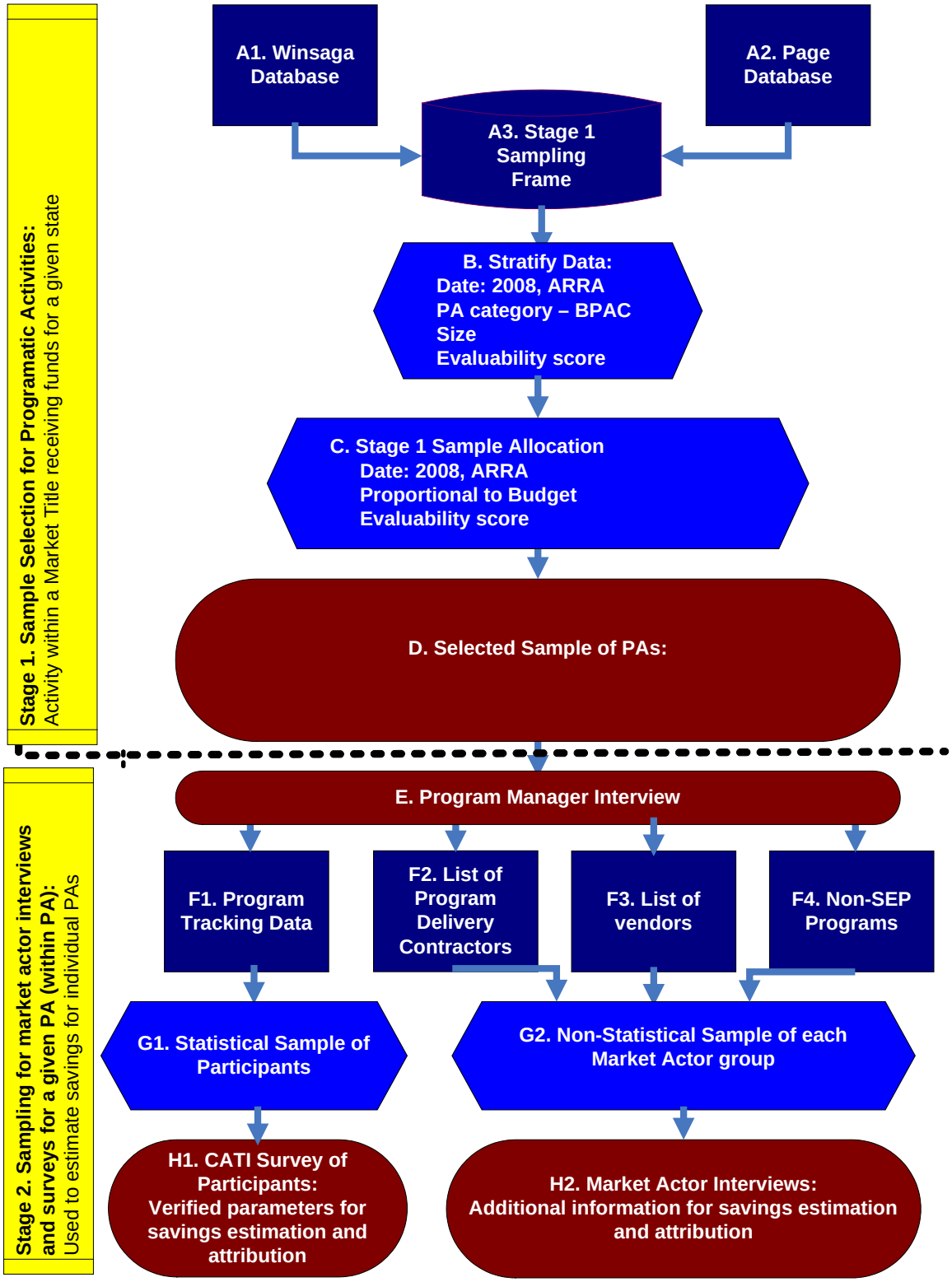
B. As indicated, the included programmatic activities for each study period (Program Year 2008 and ARRA period) were stratified by BPAC and subcategory. As described further below, DOE further stratified the included PAs by size (program budget) and by likelihood of successful evaluation.

C. For each study period, the total number of PAs to be evaluated was specified in consultation with DOE, based on preliminary budgetary assessment, review of the data bases, and DOE direction on the relative effort for the two study periods. Although funding in PY 2008 was much lower than during the ARRA period, the complexity of the 2008 assessment is actually greater because there is substantially more diversity in the activities funded. The Stage 1 targets were set at 53 PAs for PY 2008 and 29 for ARRA (PY 2009-2011). DOE allocated the Stage 1 target total for each study period to fine sampling cells defined by BPAC, subcategory, high-level size category, and evaluability category, as described further below. Thus, DOE sets Stage 1 sample targets for each fine sampling cell.

D. The relative emphasis placed on PY 2008 and the ARRA period was determined by several factors. The mix of project types, total funding available, and amount allocated per programmatic activity in PY 2008 are much more similar to the conditions expected in future years than what occurred during the ARRA period. Accordingly, the findings from an examination of PY 2008 efforts are expected to be especially helpful in understanding the kinds of activities that are likely to be important to future SEP efforts and to help inform decisions regarding Program operations and the best use of available funds. Another factor influencing the allocation of evaluation funds among the two study periods is that the large majority of funding was concentrated in a few large program areas during the ARRA years, in contrast to PY 2008 when Program funds were distributed over a wider range of activities. This means that a greater number of program areas had to be studied in PY 2008 than in the ARRA years to account for the same percentage of total Program expenditures. Within the broad program areas sampled, the number of programmatic activities to be studied per area is only slightly larger for PY 2008 than for the ARRA period.

E. For each fine sampling cell, DOE randomly selected individual PAs to be evaluated from the frame, according to the Stage 1 targets set for that cell in the sample design and allocation.

Figure 1: Overview of Sampling and Data Collection



Stage 2 – Selection of Sample Program Participants and Observers

In Stage 2, DOE will conduct evaluations for each of the PAs selected in Stage 1. Boxes E through H of Figure 1 outline the steps used to obtain the necessary data and provide quantitative estimates of program outcomes.

F. DOE will initiate primary research at each PA through interviews with PA managers. These interviews will serve as the first step in sampling for the remaining market actor groups.

G. One function of the Program Manager interview is to collect program tracking data. These data will serve as the basis for engineered savings estimates, and also provide the population data from which samples of program participants will be drawn. DOE will also use the program manager interviews to collect lists of the other program-related market actors, including program delivery contractors, vendors, and program managers of similar, non-SEP funded programs. DOE will use population data for each group of market actors to select a sample of survey and interview respondents for the respective survey or interview efforts. For PAs in which end-use participants will be surveyed, these will be selected using probability samples.

H. DOE will collect data from each of the program-related market actors selected.

I. Finally, DOE will use results from each of the research efforts to estimate savings and determine program attribution. For end-use participant samples, DOE will produce quantitative estimates such as BPAC-level aggregate program savings using sample weighting methods consistent with the sample design, as described in Section 2. DOE will choose respondents for the in-depth interviews based on their familiarity with the PA being evaluated. These responses will be aggregated qualitatively.

Detail on Key Sampling Concepts

Programmatic Activities (PAs). The State Energy Program (SEP) provides grants and technical support to the states and U.S. territories to enable them to carry out a wide variety of cost-shared energy efficiency and renewable energy activities. Through a structured planning process, the State Energy Offices and the U. S. Department of Energy work together to ensure that activities are designed to meet each state's unique energy needs while also addressing national goals, such as energy security. SEP provides money to each state and territory according to a formula that accounts for population and energy use. The formula is available at the following link (Click Here) and is also included as Attachment I "Allocation_SEP_Funds_To_States." In addition to these "Formula Grants," SEP "Special Project" funds are made available on a competitive basis to carry out specific types of energy efficiency and renewable energy activities. The resources provided by DOE typically are augmented by money and in-kind assistance from a number of sources, including other federal agencies, state and local governments, and the private sector.

For this evaluation, programmatic activities (PAs) conducted by State Energy Offices are the primary sampling unit. To be counted as part of SEP, a PA must be included in the State Plan submitted to SEP and supported, in part, by SEP funds. While it is not unusual for evaluators to

refer to a related set of activities (e.g., multiple energy audits) performed in a single year under a common administrative framework as a "program," such efforts are referred to in this document as programmatic activities (PAs)." The term "programmatic activities" is used here to refer to a related set of activities (e.g., multiple energy audits) performed in a single year under a common administrative framework. Typically, the programmatic activities designed and carried out by the states with SEP support involve a number of actions (e.g., multiple retrofits performed or loans given).

For program year (PY) 2008, the states' SEP efforts included several mandatory activities, such as establishing lighting efficiency standards for public buildings, promoting car and vanpools and public transportation, and establishing policies for energy-efficient government procurement practices. The states and territories also engaged in a broad range of optional activities, including holding workshops and training sessions on a variety of topics related to energy efficiency and renewable energy, providing energy audits and building retrofit services, offering technical assistance, supporting loan and grant programs, and encouraging the adoption of alternative energy technologies. The scope and variety of activities undertaken by the various states and territories in PY 2008 was extremely broad, and this reflects the diversity of conditions and needs found across the country and the efforts of participating states and territories to respond to them. A total of $33 million in SEP funding was made available during PY2008 to the states and territories.

Under the American Recovery and Reinvestment Act (ARRA) the amount of funding available to support the states' SEP activities increased dramatically and the mix of programmatic activities funded also changed considerably. A total of $3,069,000,000 in SEP funding was allocated to the ARRA funding period (2009-2011).

For the two study periods, a total of 1,025 PAs were identified across the states participating in SEP–450 PAs were in operation during PY2008 and 575 PAs were in operation during the ARRA funding period (PY 2009-2011). Because the amount of Program funding was very much greater during the ARRA period than in PY 2008 but the total numbers of programmatic activities were roughly similar, the average amount allocated per PA was nearly 25 times larger during the ARRA years. In contrast, the magnitude of funding per PA in future years is expected to be much closer to what was observed in PY 2008. This makes it likely that the findings from the study of PY 2008 activities will be more representative of the Program in future years and more helpful in informing decisions on future operations.

*Broad Program Area Categories (BPACs).*For this evaluation, the PAs conducted by the State Energy Offices are the primary Stage 1 sampling unit. The PAs are enumerated and classified by Broad Program Area Categories (BPACs). For this classification scheme, the BPAC provides a high level description of the PA's objectives and basic operations, for example: Building Retrofits, Renewable Energy Market Development Projects, Codes and Standards. These BPACs had been developed by DOE for previous administrative and evaluation applications, and it was decided to retain the basic BPAC structure for this study. Figure 2 shows the list of the BPACs developed by DOE.

Figure 2: List of Broad program area categories Developed by DOE

Building Retrofits
Renewable energy market development
Loans, grants, and incentives
Workshops, training, and education
Building codes and standards
Industrial retrofit support
Clean energy policy support
Traffic signals and controls
Carpools and vanpools
Technical assistance to building owners
Commercial, industrial, and agricultural audits
Residential energy audits
Government and institutional procurement
Energy efficiency rating and labeling
Tax incentives and credits
New construction and design

Based upon review of the SEP and ARRA program databases, DOE found that the BPACs did not provide sufficiently narrow definitions to support the sample stratification goals discussed above. DOE therefore developed a set of BPAC Subcategories that would support effective grouping of PAs by basic objectives and operating procedures.

BPAC Subcategories. One of the first tasks of this project was to conduct a thorough review of all SEP records for the study period in order to classify funded initiatives into PAs. DOE reviewed the database of all funding applications to develop workable definitions of the BPAC Subcategories and to identify the proper classification of each PA by BPAC and BPAC Subcategory. In cases where program descriptions contained in the database were unclear, DOE attempted to validate our understanding of the activities through brief conversations with regional and state-level officials.

One product of this process was the list of BPAC subcategories and their respective definitions. For example, the BPAC "Building Codes and Standards," was determined to consist of PAs in 3 subcategories:

- Building Code Development and Support
- Generalized Workshops and Demonstrations
- Training and Technical Assistance

Together these three subcategories contain all of the PAs in the Codes and Standards BPAC. Moreover, based on our review, DOE determined that within each of these subcategories, all the PAs subcategories could be evaluated using a similar set of methods.

The second key product of this classification process was the assignment of each PA and its budget to a unique BPAC Subcategory. This is a necessary step in the allocation of evaluation resources and sample points to the respective BPACs. As described above, the BPAC subcategory combinations defined a set of sampling strata for the Stage 1 sample of PAs.

2. Describe the procedures for the collection of information:

As described above, there are 2 stages of sample selection. The first stage is a sample of PAs. Each of the selected PAs is evaluated, meaning that quantitative outcome estimates are produced for each sampled PA. For certain types of PAs, the evaluation requires a second stage statistical sample of participants.

The required sampling and expansion elements are addressed first for the Stage 1 sample of PAs. The Stage 1 sample expansion translates the outcome estimates for the sampled PAs into program-level estimates for the study target, that is, for the universe of PAs included in the Stage 1 sampling frame.

The sampling and expansion elements for the Stage 2 statistical sample are then described. The sample expansion at this stage translates the findings for individual sampled participants within a PA into outcome estimates for that PA. The PA-level results are expanded to program-level results via the Stage 1 expansion. The combination of Stage 1 and Stage 2 results, and resulting accuracy of the primary estimates, are described at the end of the Stage 2 sampling discussion.

STAGE 1 SAMPLING AND EXPANSION: PA SAMPLE

Statistical methodology for stratification and sample selection

Sampling Objectives

The goal of this study, as noted in Section 1, is to provide quantitative estimates randomly selected from a representative sample, at the national level, of key program outcomes resulting from SEP in the 2008 and ARRA periods. This goal has been further specified by a number of policy objectives determined by DOE.

- Because funding for the study is limited, the study should address those BPACs that account for 80% of the program funding and randomly sample from them. In general, the smallest BPACs should be excluded. However, certain BPACs with relatively low funding but potentially high impacts should be included. Specifically, Codes and Standards, and Clean Energy Policy Support should be included even if not among the top 80% in funding. These small PAs represent 2.7% of funding in 2008 and 3.4% of funding during the ARRA period.
- Because the ARRA period was an abnormal program year, the study should provide separate estimates for the ARRA period and for the 2008 program year. The 2008 experience is expected to be more indicative of future program performance if spending priorities, funding levels and program categories do not change over the coming years. It is recognized that this choice means that the accuracy for each study period will be lower than if only one of the two periods

were studied using the same resources. However, a study that addressed ARRA only would have limited value to future program planning; a study that addressed program year 2008 only would not fulfill the objectives of the study funding source.

- More study resources are to be allocated to the 2008 program year, since this may be more indicative of future grant supported efforts. In addition, there is a greater variety of activity to be addressed in 2008, making a larger sample necessary to cover the range of activities of interest, even among the top 80%.
- The primary objective of the study is to produce outcome estimates that are as accurate as possible for the program study target as a whole, for each study period. At the same time, the study should produce separate estimates for each BPAC. It is recognized that the estimates for individual BPACs will not be as accurate as the estimate for the program study target as a whole.
- A policy decision affecting methodology made by DOE from the outset is that this study will be based on detailed evaluation of a sample of individual PAs. An alternative would be to apply a more generic assessment to all PAs. However, DOE and its stakeholders have an interest in evaluations based on direct information from participants where relevant, along with information from program managers and affected market actors. The costs of this approach limit the number of individual PAs that can be studied.

In developing the study plan, DOE attempted initially to explore the optimal balance between sampling more PAs with smaller Stage 2 samples in each, versus selecting fewer PAs with larger Stage 2 samples in each. However, there are substantial fixed costs associated with preparing evaluation methods for each BPAC. There are substantial costs also for evaluating each PA at any level of rigor. These PA base costs include understanding the PA structure, actions, and data sources, and interviews with program managers and key market actors. Moreover, for many Stage 1 sampling cells there are no Stage 2 participant samples. Thus, the primary determinants of the Stage 1 sample sizes are the base costs for BPAC and PA-level assessment. The Stage 2 participant sample sizes are set at levels that provide good accuracy for each PA. As described further in Section B.2.3, the Stage 2 participant samples are large enough that they contribute little uncertainty to the final results.

The Stage 1 samples are as large as the study funding permits. Section B.2.3 below shows that these samples are expected to be large enough to provide meaningful results. Specifically, the total number of PAs to be evaluated was set at 82, including 24 High-rigor[1] and 58 Medium-High-rigor[2] PAs, and a total sample size of 53 for PY2008 and 29 for ARRA. These numbers were

[1] High rigor evaluations require verification of savings through best practice methods, particularly methods recognized in the *California Evaluation Protocols*, DOE's *Impact Evaluation Framework for Technology Deployment Programs*, and the *International Performance Measurement and Verification Protocol*. These methods include on-site verification and/or performance monitoring of a sample number of projects supported by the program, whole building utility meter billing analysis, surveys of participants and nonparticipants, and combinations of building simulation modeling and other engineering analysis with the first two methods. In some cases, these verification methods may be mixed with less intensive approaches such as file review and telephone contact with program participants to increase sample size. Sample results are expanded to the population using statistical methods, such as ratio estimation or regression analysis.

[2] Medium-high rigor evaluations require verification of savings with individual participants, using less intensive data collection and analysis methods than those prescribed for high rigor. All input data may be collected through telephone contact with participants, supplemented by review of program documentation. These data are then combined with documented input assumptions and applied to standard engineering formulae to estimate savings for

determined based on an initial assessment of the distribution of funding by activity types, and the number of different types of evaluations that could be accommodated by the available budget.

Development of the Sampling Frame

The sampling frame for each study period (PY 2008 and ARRA) started as the largest BPAC-subcategory strata (in terms of program budget), that together account for at least 80 percent of non-administrative budget. That is, DOE defined a minimum funding size threshold such that the strata above this size total greater than but close to 80 percent of the total program budget. All strata that represent more than 3% of the total SEP program budget are included in the sampling frame. These strata alone account for 77.6% of total PY 2008 funding and 83% of SEP ARRA period funding. A few additional strata were then included for policy reasons despite being smaller than the size threshold. These additions increase the final sample frame to represent 80.3% of total funding for PY 2008 and 86.4% of SEP ARRA funding. The included strata define the population that will be represented for each study period. DOE refers to this population as the study universe. Within that universe the strata below the 3% funding threshold are as follows.

PY 2008 (2.7% of total funding)

- Building Codes and Standards – Development and Support; Targeted Training
- Loans Grants and Incentives – Generalized Workshops
- Technical Assistance – Generalized Workshops; Targeted Training

ARRA Period (3.4% of Total Funding)

- Building Codes and Standards – Codes; Generalized Workshops, Targeted Training
- Building Retrofits - Generalized Workshops; Targeted Training
- Loans Grants and Incentives – Generalized Workshops; Targeted Training

Rigor Level

After reviewing the activities in the course of the classification process DOE determined that High-rigor evaluations would be meaningful only for evaluation of building retrofit activities. These activities fall into two BPACS: (1) Building Retrofit and (2) Loans, Grants and Incentives. Under each of these 2 BPACS, there are Residential and Nonresidential building retrofit subcategories, for a total of 4 BPAC-subcategory cells. These 4 cells are assigned to High rigor evaluation. All other cells are assigned Medium-High rigor.

---

all or a sample of participants. On-site data collection, if used at all in medium rigor evaluations, will be applied either in exceptional cases, such as when a single project represents a large portion of potential savings for the PA, or where needed to support key assumptions used in the engineering-based assessments. Sample sizes will also be smaller in the medium-high rigor assessments.

Evaluability

Each PA was assigned an evaluability score indicating the chance of successfully completing an evaluation at the targeted rigor if DOE selects that PA. Evaluability scores were based on the following criteria:

a. *Match of actual program operations to the BPAC definition.* The DOE team developed detailed working definitions for each BPAC. If, upon selection and detailed review of activities, DOE found that a PA had been misclassified, it was evaluated consistent with its actual activity. Its expansion weight was based on the BPAC it was selected from.

b. *Progress in implementation.* In order to carry out high or medium-high rigor evaluations, the program needed to have resulted in a sufficient number of the targeted actions, such as completion of retrofit projects or installations of renewable energy equipment, for a sample to be drawn and tested. Prior to planning an evaluation of any sampled PA, DOE has established criteria to assess evaluability and the status of program or project completion will be assessed at that time. This would only apply to the 29 sampled ARRA-funded PAs (not the 53 PA from 2008) for which current and valid reporting guidance and progress tracking mechanisms exist to ensure accuracy of the program's or project's status. At this time, all funding has been obligated under ARRA and program or project completion will be considered as one of many variables in the evaluability assessment.

c. *Quality and availability of program records.* For high and medium high rigor evaluations, it will be necessary to contact participants in the program. In most cases DOE will need to be able to characterize the services that participants received from the program at the individual level. If such records were not available at the time of PA selection and could not, in the evaluator's judgment, be reconstructed within schedule and budget constraints, then the PA was dropped from the sample and a substitute selected.

While excluding a small number of eligible PAs potentially biases the results, the number of PAs and funding proportions they represent would be very small. Only 23 PAs were removed from the sample frame for this reason; 14 from PY2008 (5.2% of total budget) and 9 from the ARRA period (0.6% of total budget) and these were excluded from the State 1 sample frame. It is important to note that these eligible PAs represent those with a zero or low chance of successful Medium-High- or High-rigor evaluation due to limitations of available program data—but their exclusion says nothing about the effectiveness of those PAs themselves. Indeed, DOE will not claim any energy savings due to these programs and excluding them will result in less overall savings for the study universe and more reliable estimates for those PAs in the sample. DOE maintains that this is a reasonable approach and the value of accuracy for those included in the sample outweighs any potential bias which would be very small.

First Stage Sample Allocation

First stage sample allocation to BPAC-subcategory cells occurred in a few steps.

1. **Preliminary allocation.** Initially Stage 1 sample points (number of PAs to be selected) were allocated to BPAC-subcategory strata proportional to SEP program budget only. This process left some smaller strata, especially those included despite being below the minimum size threshold (at least 3% of total funding), with zero allocation.
2. **Forced allocations.** After reviewing the initial allocation strictly proportional to budget, some forced allocations were specified, to ensure the small strata that need to be covered would have some sample to allow for analysis. Forced allocations include two PAs in 2008 and four in the ARRA period. After establishing the allocations, however, the actual sampled PAs were selected randomly. These forced allocations will provide better stand-alone savings estimates for their associated BPACs without appreciably reducing the accuracy of the overall estimate of savings.
3. **Proportional allocation.** The strata that received forced allocations were set aside. The remainder of the total sample points for each period were allocated to the remaining strata proportional to size (program budget).
4. **Identification of certainty and non-certainty PAs.** Allocation proportional to size means that one sample PA is allocated for about every $850,000 of budget for PY2008, and for every $77 million of budget for ARRA.
    a. Any individual PA with budget above this amount is included with certainty in the first stage sample. The PAs so selected are called "first-pass certainty" PAs. In some cases, the budget for an individual PA would mean an allocation of two or more PAs. However, DOE selected a given PA only once.
    b. Once the large, first-pass certainty PAs have been identified, the remaining sample points are allocated to the remaining strata, proportional to the remaining size.
    c. DOE identifies a second set of certainty selections within this remainder sample, using the same approach as for the first pass. That is, all PAs with budget greater than the ratio of total remaining budget to remaining sample size are included with certainty. The PAs so selected are called "second-pass certainty" PAs.
    d. Once the first- and second-pass certainty PAs have been identified, the remaining Stage 1 sample points are allocated to the remaining strata, proportional to the remaining size. These allocations are referred to as the "non-certainty" or "remainder" sample.
5. **Assessment of achievability.** Once DOE identified the target numbers of certainty and non-certainty selections for each cell, DOE assessed whether there were cells whose targets were unlikely to be met based on evaluability.
    a. PAs included in the Stage 1 sampling frame all have evaluability scores either "high" or "moderate." DOE assumed that a "likely" evaluable PA has an 80 percent chance of being evaluated at the targeted High or Medium-High rigor

level, while a "possibly" evaluable PA has a 50 percent chance. Based on discussions with representatives from DOE, ORNL and the states who participated in the May 25th Network Committee Meeting, DOE believes these are conservative estimates.

b. The assumed success rates should be conservative for certainty PAs. Certainty PAs are high priority for successful completion because of their size. If after confirming with ORNL that the evaluation team is unable to complete evaluation of one of these PAs, DOE will substitute a smaller PA. However, this substitution will be a last resort.

c. The remainder sample was allocated to "likely evaluable" PAs at a higher rate than to "possibly evaluable" PAs. Specifically, the measure of size used to allocate the remainder sample was 2x the program budget for "likely evaluable" PAs, and 1x the program budget for "possibly evaluable" PAs. This procedure ensures that both levels of evaluability are covered by the sample, but that evaluation resources are devoted more heavily to the PAs that have a better chance of being evaluable.

d. Based on the assumed probabilities of successful evaluation at targeted rigor for likely and possible, DOE calculates the size of the oversample required to achieve the targeted sample sizes. With the assumed success probabilities, DOE needs a sample of five "likely" PAs to complete four evaluations successfully. DOE needs a sample of two "possible" PAs to complete one evaluation successfully.

e. If the total oversample required based on this calculation exceeds the number of PAs in the sample, DOE flags a potential shortfall. For most cells, the sample design does not have an anticipated shortfall. For these cases, unless the frequency of inadequate data availability is worse than projected in some cell, DOE expects to achieve these targeted sample sizes at the targeted rigor levels. There were three cells with some potential shortfall and they are discussed in the next section.

6. **Final targets.** After the iterative reallocation in steps 1-4, DOE reviewed the sample allocations and made some slight adjustments to be sure:

   a. Total samples after rounding still matched the targeted number by time period and by rigor level, and

   b. The iterative re-allocation of the remainder did not result in severe over- or under-allocation to any one stratum.

Because there was a large enough allocation to certain BPACs, DOE could afford to shift a few sample points to another BPAC to get a better stand-alone estimate for that BPAC, without appreciably reducing the accuracy of the overall estimate.

It is important to emphasize that the sampling objectives were maintained for sampling target estimates that were as fair as possible, and not slanted toward high savings PAs. The sample design did not include savings estimates for the proportion of the population excluded from the study universe and the study will make no claims of savings for the excluded 20% of spending omitted from the study plan.

Furthermore, regardless of how the allocations were set, selections were random, except for large PAs selected with certainty. These certainty selections represent only themselves. Random selections from a particular BPAC/subcategory combination represent only (the noncertainty portion of) that BPAC/subcategory.

The remainder of this section presents the results of each of the steps listed above.

Completed Sample Frame

As noted, our starting point for frame definition was to select the largest BPAC/Subcategory strata that sum to at least 80 percent of funding. To meet this condition, the minimum funding percentage for a BPAC/Subcategory stratum turned out to be 3 percent for each period. In addition, as noted, certain BPAC/Subcategory strata that are below the minimum size criterion were included for policy reasons to ensure adequate inclusion of important BPACs. The additional included strata are the following:

- Building Codes and Standards—this BPAC is anticipated to produce savings disproportionate to spending. Ideally, the evaluation team would include cells and allocate sample to them proportional to savings as opposed to spending, however, there is no general and consistent indicator of savings available to rely on.

- Subcategories of Workshops/Demonstrations and Training/Certification that are likely to be evaluable, if the other subcategories of the associated BPAC are included.

- Building Retrofit subcategories if not already included based on size.

The Stage 1 sampling frame represents 80.3 percent of SEP funding for PY2008, and 86.4 percent for the ARRA period. Figure 3 and Figure 4, respectively, display the included BPAC-subcategory strata for the two study periods, along with their program budgets, and number of available PAs (Stage 1 population counts). Also indicated in these tables is the program budget as a percentage of the total program budget in the study universe. Yellow highlighted rows are those BPAC/subcategory strata that are included for one of the three policy reasons noted above. As shown, 140 PAs are included in the Stage 1 sampling frame for PY2008 and 306 for the ARRA period.

Figure 3: Stage 1 Sampling Strata Included in the Covered Study Universe (PY2008)

BPAC	Subcat	Target Rigor	SEP Budget	Percent of Covered SEP Budget	Number of PAs
Building Codes and Standards	Building Code Development and Support	MH	393,875	1%	5
Building Codes and Standards	Generalized Workshops and Demonstrations (Participants maybe traceable)	MH	1,617,449	4%	6
Building Codes and Standards	Targeted Training and/or Certification (participants are traceable)	MH	677,485	2%	5
Building Codes and Standards	Technical Assistance to Building Owners	MH	2,972,522	7%	3
Building Retrofits	Building Retrofits: Nonresidential	H	903,728	2%	4
Building Retrofits	Building Retrofits: Residential	H	576,183	1%	4
Building Retrofits	Generalized Workshops and Demonstrations (Participants maybe traceable)	MH	2,589,997	6%	21
Building Retrofits	Targeted Training and/or Certification (participants are traceable)	MH	181,447	0%	5
Building Retrofits	Technical Assistance to Building Owners	MH	4,060,411	9%	11
Clean Energy Policy Support	Policy and Market Studies; Legislative Support	MH	5,539,183	13%	31
Loans, Grants and Incentives (Excl Retro)	Alternative Fuels, Ride Share and Traffic Optimization	MH	2,932,203	7%	10
Loans, Grants and Incentives (Excl Retro)	Generalized Workshops and Demonstrations (Participants maybe traceable)	MH	94,722	0%	2
Loans, Grants and Incentives (Excl Retro)	Technical Assistance to Building Owners	MH	5,062,979	12%	4
Loans, Grants and Incentives (Retro Only)	Building Retrofits: Nonresidential	H	9,392,550	21%	8
Loans, Grants and Incentives (Retro Only)	Building Retrofits: Residential	H	2,332,255	5%	2
Renewable Energy Market Development	Generalized Workshops and Demonstrations (Participants maybe traceable)	MH	2,704,155	6%	11
Technical Assistance	Generalized Workshops and Demonstrations (Participants maybe traceable)	MH	30,425	0%	1
Technical Assistance	Targeted Training and/or Certification (participants are traceable)	MH	95,880	0%	1
Technical Assistance	Technical Assistance to Building Owners	MH	1,800,693	4%	6
			43,958,143	100%	140

Figure 4: Stage 1 Sampling Strata Included in the Covered Study Universe (ARRA)

BPAC	Subcat	Target Rigor	SEP Budget	Percent of Covered SEP Budget	Number of PAs
Building Codes and Standards	Building Codes and Standards: Codes	MH	10,381,043	0%	10
Building Codes and Standards	Generalized Workshops and Demonstrations (Participants maybe traceable)	MH	19,223,610	1%	2
Building Codes and Standards	Targeted Training and/or Certification (participants are traceable)	MH	3,075,498	0%	7
Building Retrofits	Building Retrofits: Nonresidential	H	507,614,260	22%	70
Building Retrofits	Building Retrofits: Residential	H	91,550,234	4%	11
Building Retrofits	Generalized Workshops and Demonstrations (Participants maybe traceable)	MH	667,990	0%	1
Building Retrofits	Targeted Training and/or Certification (participants are traceable)	MH	25,637,692	1%	7
Loans, Grants and Incentives (Excl Retrofits and Projects)	Generalized Workshops and Demonstrations (Participants maybe traceable)	MH	4,247,962	0%	4
Loans, Grants and Incentives (Excl Retrofits and Projects)	Renewable Energy Market Development: Manufacturing	MH	251,957,503	11%	12
Loans, Grants and Incentives (Excl Retrofits and Projects)	Targeted Training and/or Certification (participants are traceable)	MH	9,548,163	0%	2
Loans, Grants and Incentives (Retrofits and Projects)	Building Retrofits: Nonresidential	H	488,804,472	22%	43
Loans, Grants and Incentives (Retrofits and Projects)	Building Retrofits: Residential	H	135,599,983	6%	13
Loans, Grants and Incentives (Retrofits and Projects)	Renewable Energy Market Development: Projects	MH	298,047,959	13%	55
Renewable Energy Market Development	Generalized Workshops and Demonstrations (Participants maybe traceable)	MH	2,072,750	0%	5
Renewable Energy Market Development	Renewable Energy Market Development: Manufacturing	MH	118,323,694	5%	9
Renewable Energy Market Development	Renewable Energy Market Development: Projects	MH	289,487,818	13%	51
Renewable Energy Market Development	Targeted Training and/or Certification (participants are traceable)	MH	14,718,684	1%	4
			2,270,959,315	100%	306

Stage 1 Sample Targets

The allocation procedure outlined above resulted in the sampling targets shown in Figure 5 and Figure 6 for the PY2008 and ARRA periods, respectively.

For reference, the figures show the allocation that would be assigned based strictly by allocating proportional to total budget (green highlighted columns), and also the allocations that would result from allocating strictly proportional to the number of PAs in the cell (red highlighted cells). Also shown is the total number allocated through the iterative process described above, in the blue highlighted cells, combining the certainty and non-certainty PAs.

The figures show a few cells with allocations of zero. These are cells initially included in the frame, but that were too small to receive an allocation with proportional allocation. These were all cells that were included in the frame to ensure some coverage of evaluable Workshops/Demonstrations and Training/Certification (subcategory) activities. DOE did not force allocations to these cells, because enough other activities in these subcategories were included.

There are a few cells (highlighted in yellow) where the final proposed allocation differs from the iteratively allocated targets (in blue).

- For PY2008, the iterative allocation (Steps 1-4 above) results in a target of 11 for Clean Energy Policy Support. This allocation would be 20 percent of the sample, for 13 percent of the budget and 22 percent of the number of PAs. In Step 6, DOE reduced this allocation to eight, and added one to "Building Codes and Standards/Targeted Training and/or Certification", one to "Loans, Grants and Incentives/Building Retrofits: Residential", and one to "Renewable Energy Market Development/Generalized Workshops and Demonstrations" (yellow highlighted cells).

- For ARRA, the rounding of cell targets from Steps 1-4 above resulted in a total of 28 selections instead of the targeted 29. In Step 6, DOE rounded down the allocation to "Loans, Grants, and Incentives/Renewable Energy Market Development: Manufacturing", and DOE added one each to "Loans, Grants and Incentives/Renewable Energy Market Development: Projects", and to "Renewable Energy Market Development/Renewable Energy Market Development: Projects" (yellow highlighted cells).

The figures also show that in most cases the proposed targets are within the range bracketed by allocation proportional to size and allocation proportional to number of PAs. Allocations less than proportional to size are mostly associated with large numbers of certainty selections.

Finally, the figures indicate that most of the targets are expected to be achievable based on the numbers available in each cell and the assumed evaluation success rates. That is, the likely shortfall is zero except for "Loans, Grants and Incentives/Technical Assistance to Building Owners", "Loans, Grants and Incentives/Building Retrofits: Nonresidential", and "Loans, Grants and Incentives/Building Retrofits: Residential" in PY2008. If it does turn out that some of the target sample sizes cannot be achieved for these or other strata, DOE will re-allocate sample to other strata as needed.

Figure 5: Stage 1 Sample Targets by BPAC/Subcategory (PY2008)

BPAC	Subcat	Target Rigor	Budget			Population # PAs			Iteratively Allocated Sample Size				Likely Shortfall	Final Target
			Budget	% Budget	Sample Proportional to Budget	Population # PAs	% Population # PAs	Sample Proportional to # PAs	Large/Certainty	Small/Noncertainty	Total	% Sample Total		
Building Codes and Standards	Building Code Development and Support	MH	393,875	1%	0	5	4%	2	0	1	1	2%	0	**1**
Building Codes and Standards	Generalized Workshops and Demonstrations (Participants maybe traceable)	MH	1,617,449	4%	2	6	4%	2	1	2	3	6%	0	**3**
Building Codes and Standards	Targeted Training and/or Certification (Participants are traceable)	MH	677,485	2%	1	5	4%	2	0	1	1	3%	0	**2**
Building Codes and Standards	Technical Assistance to Building Owners	MH	2,972,522	7%	4	3	2%	1	1	0	1	2%	0	**1**
Building Retrofits	Building Retrofits: Nonresidential	H	903,728	2%	1	4	3%	2	0	2	2	4%	0	**2**
Building Retrofits	Building Retrofits: Residential	H	576,183	1%	1	4	3%	2	0	2	2	4%	0	**2**
Building Retrofits	Generalized Workshops and Demonstrations (Participants maybe traceable)	MH	2,589,997	6%	3	21	15%	8	1	4	5	9%	0	**5**
Building Retrofits	Targeted Training and/or Certification (Participants are traceable)	MH	181,447	0%	0	5	4%	2	0	0	0	1%	0	**0**
Building Retrofits	Technical Assistance to Building Owners	MH	4,060,411	9%	5	11	8%	4	1	5	6	11%	0	**6**
Clean Energy Policy Support	Policy and Market Studies; Legislative Support	MH	5,539,183	13%	7	31	22%	12	2	9	11	20%	0	**8**
Loans, Grants and Incentives (Excl Retro)	Alternative Fuels, Ride Share and Traffic Optimization	MH	2,932,203	7%	4	10	7%	4	1	4	5	10%	0	**5**

BPAC	Subcat	Target Rigor	Budget			Population # PAs			Iteratively Allocated Sample Size				Likely Shortfall	Final Target
			Budget	% Budget	Sample Proportional to Budget	Population # PAs	% Population # PAs	Sample Proportional to # PAs	Large/Certainty	Small/Noncertainty	Total	% Sample Total		
Loans, Grants and Incentives (Excl Retro)	Generalized Workshops and Demonstrations (Participants maybe traceable)	MH	94,722	0%	0	2	1%	1	0	0	0	0%	0	**0**
Loans, Grants and Incentives (Excl Retro)	Technical Assistance to Building Owners	MH	5,062,979	12%	6	4	3%	2	2	1	3	5%	3	**3**
Loans, Grants and Incentives (Retro Only)	Building Retrofits: Nonresidential	H	9,392,550	21%	11	8	6%	3	3	1	4	8%	1	**4**
Loans, Grants and Incentives (Retro Only)	Building Retrofits: Residential	H	2,332,255	5%	3	2	1%	1	1	0	1	2%	3	**2**
Renewable Energy Market Development	Generalized Workshops and Demonstrations (Participants maybe traceable)	MH	2,704,155	6%	3	11	8%	4	2	3	5	9%	0	**6**
Technical Assistance	Generalized Workshops and Demonstrations (Participants maybe traceable)	MH	30,425	0%	0	1	1%	0	0	0	0	0%	0	**0**
Technical Assistance	Targeted Training and/or Certification (Participants are traceable)	MH	95,880	0%	0	1	1%	0	0	0	0	0%	0	**0**
Technical Assistance	Technical Assistance to Building Owners	MH	1,800,693	4%	2	6	4%	2	2	1	3	5%	0	**3**
		Total	**43,958,143**	**100%**	**53**	**140**	**100%**	**53**	**17**	**36**	**53**	**100%**	**6**	**53**
		MH	**30,753,427**	**70%**	**37**	**122**	**87%**	**46**	**13**	**31**	**44**	**82%**	**3**	**43**
		H	**13,204,716**	**30%**	**16**	**18**	**13%**	**7**	**4**	**5**	**9**	**18%**	**3**	**10**

Figure 6: Stage 1 Sample Targets by BPAC/Subcategory (ARRA)

BPAC	Subcat	Target Rigor	Budget			Population # PAs			Iteratively Allocated Sample Size				Likely Shortfall	Final Target
			Budget	% Budget	Sample Proportional to Budget	Population # PAs	% Population # PAs	Sample Proportional to # PAs	Large/Certainty	Small/Noncertainty	Total	% Sample Total		
Building Codes and Standards	Building Codes and Standards: Codes	MH	10,381,043	0%	0	10	3%	1	0	2	2.0	7%	0	**2**
Building Codes and Standards	Generalized Workshops and Demonstrations (Participants maybe traceable)	MH	19,223,610	1%	0	2	1%	0	0	1	1.0	3%	0	**1**
Building Codes and Standards	Targeted Training and/or Certification (participants are traceable)	MH	3,075,498	0%	0	7	2%	1	0	1	1.0	3%	0	**1**
Building Retrofits	Building Retrofits: Nonresidential	H	507,614,260	22%	6	70	23%	7	0	6	5.6	19%	0	**6**
Building Retrofits	Building Retrofits: Residential	H	91,550,234	4%	1	11	4%	1	0	2	2.0	7%	0	**2**
Building Retrofits	Generalized Workshops and Demonstrations (Participants maybe traceable)	MH	667,990	0%	0	1	0%	0	0	0	0.0	0%	0	**0**
Building Retrofits	Targeted Training and/or Certification (participants are traceable)	MH	25,637,692	1%	0	7	2%	1	0	0	0.3	1%	0	**0**
Loans, Grants and Incentives (Excl Retrofits and Projects)	Generalized Workshops and Demonstrations (Participants maybe traceable)	MH	4,247,962	0%	0	4	1%	0	0	0	0.0	0%	0	**0**
Loans, Grants and Incentives (Excl Retrofits and Projects)	Renewable Energy Market Development: Manufacturing	MH	251,957,503	11%	3	12	4%	1	0	3	2.8	10%	0	**2**
Loans, Grants and Incentives (Excl Retrofits and Projects)	Targeted Training and/or Certification (participants are traceable)	MH	9,548,163	0%	0	2	1%	0	0	0	0.1	0%	0	**0**

BPAC	Subcat	Target Rigor	Budget			Population # PAs			Iteratively Allocated Sample Size				Likely Shortfall	Final Target
			Budget	% Budget	Sample Proportional to Budget	Population # PAs	% Population # PAs	Sample Proportional to # PAs	Large/Certainty	Small/Noncertainty	Total	% Sample Total		
Loans, Grants and Incentives (Retrofits and Projects)	Building Retrofits: Nonresidential	H	488,804,472	22%	6	43	14%	4	1	4	4.9	17%	0	**5**
Loans, Grants and Incentives (Retrofits and Projects)	Building Retrofits: Residential	H	135,599,983	6%	2	13	4%	1	0	1	1.5	5%	0	**1**
Loans, Grants and Incentives (Retrofits and Projects)	Renewable Energy Market Development: Projects	MH	298,047,959	13%	4	55	18%	5	0	3	3.3	11%	0	**4**
Renewable Energy Market Development	Generalized Workshops and Demonstrations (Participants maybe traceable)	MH	2,072,750	0%	0	5	2%	0	0	0	0.0	0%	0	**0**
Renewable Energy Market Development	Renewable Energy Market Development: Manufacturing	MH	118,323,694	5%	2	9	3%	1	0	1	1.3	4%	0	**1**
Renewable Energy Market Development	Renewable Energy Market Development: Projects	MH	289,487,818	13%	4	51	17%	5	0	3	3.2	11%	0	**4**
Renewable Energy Market Development	Targeted Training and/or Certification (participants are traceable)	MH	14,718,684	1%	0	4	1%	0	0	0	0.2	1%	0	**0**
		Total	**2,270,959,315**	**100%**	**29**	**306**	**100%**	**29**	**1**	**28**	**29**	**100%**	**0**	**29**
		MH	**1,047,390,366**	**46%**	**13**	**169**	**55%**	**16**	**0**	**15**	**15**	**52%**	**0**	**15**
		H	**1,223,568,949**	**54%**	**16**	**137**	**45%**	**13**	**1**	**13**	**14**	**48%**	**0**	**14**

Implementing the PA Sample Design

Drawing the Stage 1 sample of PAs means selecting a simple random sample from each Stage 1 sampling cell, with the cell sample size specified by the design. For each selected PA, an evaluation will be conducted at the Rigor level specified by the design.

Misclassification and Multiple Classifications

In the course of evaluating a selected PA, the investigation may reveal that the PA was incorrectly classified at the frame development stage. In addition, many PAs are known even from the currently available data to include multiple categories of activity.

To deal with both misclassification and multiple categories, DOE distinguishes between the *sampling* category and the *analytic* category or reporting domain. PAs are assigned to BPACs and subcategories at the sample design and frame development stage based on the information available at that time. This assignment and the sample allocation determine each PA's probability of being included in the sample. That probability determines its sample expansion weight, and its stratum assignment for the calculation of ratio estimates and standard errors, described below.

For purposes of *analysis,* activities may be classified by information available at the design stage, or by information available only after collecting more information from the selected PAs. Information can be reported for all components of all PAs that include a certain type of activity, not just for the PAs assigned to a particular category for sampling. Thus, for example, to determine the total savings from all residential retrofits, as identified post-sampling, DOE would sum up the residential retrofit components in all sampling strata, each weighted by that stratum's expansion weight. This situation is analogous to stratifying buildings based on imperfect building type information. Each building may have multiple types of activities. A sample is stratified based on the best information available at the sampling stage to classify buildings by predominant activity type. During data collection, information may be obtained about the portions of the building corresponding to each activity type. Information can then be reported by domains corresponding to observed activity types. The weighting and stratification are based on the sampling information.

Estimation Procedures

BPAC-Specific Impact Calculations

For each selected PA, our evaluation will produce calculated impacts and error bounds. DOE will also have one or more measure of size (MOS) for each PA. At a minimum DOE will have the spending amount. DOE also may have more informative correlates of savings such as program-estimated impacts, or other activity measures such as number

of units or square feet affected. These other measures of size will vary by BPAC and subcategory.

From these results DOE will calculate a statistical ratio estimate for the BPAC for each of the key outcome metrics estimated from the PA sample. DOE will use the Combined rather than Stratified form of the ratio estimator, because the latter form has more bias when cell sample sizes are small as in this study. Specifically, DOE calculates the ratio $\hat{R}$ of the stratified estimate of population savings (or other metric) to the corresponding estimate of population measure of size from the same sample:

$$\hat{R} = \sum_k N_k \; y_{_k} / \sum_k N_k \, x_{_k}$$

Where

N_k = population number of individual PAs in cell k

$y_{_k}$ = sample mean outcome (e.g., energy savings) for cell k

$x_{_k}$ = sample mean MOS for cell k.

In applying this formula, sampling cells k are defined by BPAC, subcategory, certainty or remainder, and evaluability level.

This ratio is a form of unit savings estimate. For example, if the measure of size x is the number of square feet audited, and the outcome is savings, the ratio $\hat{R}$ is savings per square foot audited. DOE then calculates the population total savings (or other outcome) Y_{TOT} by multiplying the total measure of size X_{TOT} known from the database by this ratio (e.g., multiply savings per square foot by total square feet audited):

$$Y_{TOT} = \hat{R} \, X_{TOT}.$$

DOE will calculate the standard error of the ratio and the corresponding total savings estimate via statistical formulas for stratified ratio estimation, e.g. from Cochran (1977)[3].

Portfolio-Level Impact Calculations

The procedures described above will provide estimates of savings and other impacts for each BPAC. Total impacts for the study target—i.e., for the PAs covered by the Stage 1 sample frames-- will be calculated as the sum of the impacts by BPAC for each program year. However, as described further below, some adjustment parameters determined from high-rigor information collected only for 2008 may be used in calculations for ARRA as well.

[3] Cochran, W. G. 1977. *Sampling Techniques*. New York: John Wiley & Sons.

STAGE 2 SAMPLING WITHIN PAS

Statistical methodology for stratification and sample selection

As discussed in Section 1, for BPAC-Subcategories involving end-user projects, particularly the Building Retrofit Subcategories, DOE will select a second stage sample of individual projects or participants within each selected PA. For each selected project DOE will collect information via telephone survey. In addition, a subset of projects selected for telephone surveys will also be selected for onsite data collection. DOE will use this information to develop estimated energy savings and other outcomes for the selected project.

To design these second stage samples, the evaluation team will follow standard sampling procedures laid out in guidelines such as California's *Evaluation Framework*[4] for designing stratified samples to support ratio estimation. DOE will use the distribution of projects by tracking system savings and measure type as the basic guides to stratification.

- **Choice of Stage 2 sampling unit.** Generally, DOE will attempt to match the sampling unit to the purchase decision-making unit in order to capture and make best use of information on attribution of program influence on the quantity of measures, timing, and efficiency levels of equipment installed in direct relationship to the savings estimate. However, this is not always possible due to logistical, schedule, and tracking system problems. DOE developed a variety of methods to deal with this problem. For example, DOE often assesses attribution at the program level through large sample surveys of participants, surveys of vendors, sales and shipment data analysis, or combinations of the above.
- **Stage 2 Sample Frame.** The Stage 2 sample frame for each PA will consist of the database of facility owners or projects that have received support from the PA. DOE assumes for the purposes of this submittal that each project or participant record will contain some information that will be useful in sample stratification or selection. This information may consist of measures of size, such as project costs, estimates of energy savings based on engineering calculations, or more qualitative characterizations.

For strata designated for High Rigor evaluation, DOE will select a random subsample of telephone survey respondents to receive onsite visits. The onsite visits will verify by direct observation some of the physical information collected by phone, particularly for energy savings. For each project in the onsite subsample, DOE will produce verified values of savings estimates and other key parameters or outcomes based on this improved data.

Estimation procedure

[4] TecMarket Works, *The California Evaluation Framework*. San Francisco: California Public Utilities Commission. 2004. Chapter 13, Sampling.

Telephone Survey Results

The analyses of individual projects from the full Stage 2 sample yield a set of phone-based savings estimates that reflect findings concerning the actual quantity, efficiency features, operating environment, and operating patterns of the program measures installed for each project. DOE will use ratio estimation techniques for these project-level estimates of savings, along with a corresponding Measure of Size known for all units in the Stage 2 sample frames. Where available, the Measure of Size will typically be a program tracking system estimate of savings for the individual projects.

That is, similar to the Stage 1 expansion, DOE will calculate a ratio for each PA *a* as

$$R_a^{\wedge} = \sum_j N_j \; y_j \; / \sum_j N_j \; x_j$$

where

N_j = population number of individual projects in cell j of the PA
y_j = sample mean phone-based outcome (e.g., energy savings) for cell j of the PA
x_j = sample mean MOS for cell j of the PA.

The sampling cell j in this formula refers to whatever stratification cells are used for the particular PA, based on information available for the PA.

The total outcome y_{Ta} for the PA is then calculated as

$$y_{Ta} = R_a^{\wedge} \; x_{Ta}$$

where x_{Ta} is the total of the measure of size x for the PA.

A separate ratio expansion of this form is conducted for each PA, for several reasons.

1. The measures of size x available will vary quite a bit across PAs, even within a BPAC.
2. Even where PAs have a similar nominal Measure of Size, such as a tracking system estimate of energy savings, that nominal MOS is likely to be calculated inconsistently across different PAs. As a result, the relationship between the tracking energy savings x and the phone-survey-based estimate y will be different for each PA. At the same time, DOE has a large enough sample within each PA that PA-specific ratios can be estimated meaningfully.
3. The PA-level total x_{TOTa} is available only for the PAs in the Stage 1 sample. As a result, even if DOE produced a single ratio across all the sampled PAs for a given BPAC Subcategory, it could only be applied to the sampled PAs, not to the full population of PAs.

Onsite Data Collection

A subsample of the telephone sample will be drawn in order to conduct additional onsite data collection. This onsite subsample presents a different situation relative to the larger telephone sample, in that DOE will use the subsample to calculate a verification ratio not at the PA level, but at the BPAC-Subcategory level. That is, for each BPAC-Subcategory stratum that has an onsite sample, DOE will calculate a verification ratio:

$$R_v^{\wedge} = \sum_a \sum_j N_{ja} \; v__{ja} / \sum_j N_{ja} \; y__{ja}$$

where

N_{ja} = population number of individual projects in cell j of PA a

$v__{ja}$ = sample mean onsite-based outcome for cell j of PA a

$y__{ja}$ = sample mean phone-based outcome (e.g., energy savings) for cell j of PA a

Again, the cell j in this case refers to the stratification variables used within each PA a, and may vary by PA within the BPAC-subcategory stratum.

This verification ratio is calculated at the BPAC-Subcategory level rather than for individual PAs because the onsite sample size for individual PAs will be small. Moreover, for the verification ratio, the onsite outcome v and the phone-based outcome y are both calculated consistently across PAs. It is therefore reasonable to produce a single ratio across PAs.

Applying the Onsite Verification Factor

For each PA a, the PA-specific phone-based ratio $R_a^{\wedge}$ together with the known denominator total x_{Ta} provides an estimate of the outcome of interest y_{Ta} for that PA, as indicated above. These PA-level estimates are combined into stratum- or BPAC-level estimates using the Stage 1 ratio expansion described above, to produce stratum- or BPAC-level phone-based totals Y_{TOTp}.

For strata that have onsite subsamples, the Stage 1 aggregation is conducted at the stratum level, to produce stratum totals Y_{TOTp}. Each such total is then multiplied by the corresponding verification ratio based on the onsite subsample, to produce the final adjusted estimate:

$$Y_{TOTv} = R^{\wedge}_v \, Y_{TOTp.}$$

BPAC totals are then calculated as the sum of the Subcategory totals for each BPAC. For those outcomes for which the onsite subsample does not provide a verification adjustment, the phone-based estimate will be the final estimate.

Degree of accuracy and sample sizes needed for the purpose described in the justification

STAGE 1 AND STAGE 2 ACCURACY

Both the Stage 1 and Stage 2 sample expansions rely on ratio estimation. The relative precision of a ratio estimator is determined by the *error ratio.* The error ratio is the ratio-based equivalent of a coefficient of variation (CV). The CV measures the variability (standard deviation, or root-mean-square difference) of individual y values around their mean value, as a fraction of that mean value. Similarly, the error ratio measures the variability (root mean square difference) of individual y values from the ratio line y = Rx, as a fraction of the mean y value. Thus, to estimate the precision that can be achieved by the planned sample sizes, or conversely the sample sizes necessary to achieve a given precision level, it is necessary to know the error ratios for the sample components.

In practice, error ratios cannot be determined until after the data are collected. The sample design and projected precision are therefore based on error ratios assumed based on past experience with similar work.

The Evaluation Team has extensive experience in applying this kind of analysis to all of the types of measures and delivery mechanisms encompassed by SEP PAs. Based on this experience, DOE makes the following assumptions. DOE expects these assumptions to be on the conservative side in general.

Stage 1 error ratios, all BPACs: 1.0
Stage 2 error ratios, all Stage 2 CATI samples: 0.8
Stage 2 error ratios, onsite vs CATI: 0.8

For purposes of projecting the accuracy that will be achieved, the Stage 1 error ratio assumed is the variability of true PA-level savings within the BPAC or stratum. The observed variability will be higher, since it will include estimation error for each PA.

The Stage 1 error ratios reflect the variability across PAs in a given BPAC of the (true) savings per program dollar. DOE anticipates a fair amount of variability in the amount of leveraging of funds, and in the effectiveness of program structures and operations across the different PAs in the different states. DOE therefore assumes a relatively high value for the Stage 1 error ratio.

The Stage 2 CATI error ratios are based on experience from prior work with evaluations of individual programs. The assumption of an error ratio of 0.8 typically provides reasonable and adequate sample sizes for situations where the “y” variable is savings based on confirmed installation and operations, and the “x” variable is a tracking estimate of savings.

For the Stage 2 onsite vs CATI ratios, DOE assumes the onsite to CATI relationship has the same level of variability as that for the relationship between the phone outcome and PA-specific Measure of Size within each PA. On the one hand, DOE generally expects fairly close agreement between the outcomes based on the onsite and phone sources. On the other hand, the onsite:phone verification ratio will be estimated across all PAs for a particular BPAC-Subcategory cell. Thus, the variability around the overall ratio line for that cell may be greater than it would be within a single PA.

The relative standard error (RSE) of the estimate (standard error divided by the estimate) is related to the error ratio *er* by

$$RSE^2 = (1-n/N)er^2/n$$

where n is the sample size and N is the population size.

In the context of the 2-stage sampling used here, the total variance is the sum of the between-PA and within-PA contributions to variance. Since the within-PA population sizes are unknown, DOE assumes they are large enough to ignore the finite population correction factor (1-n/N) for the Stage 2 samples. This is a conservative assumption. For many of the PAs evaluated, the Stage 2 CATI sample will be a near census of the completed projects, and there will be minimal contribution to variance from this source.

The onsite sample is a subsample of the Stage 2 phone sample. As a result, the onsite:phone verification ratio $\hat{R}_v$ based on the onsite sample alone and the phone-based outcome to MOS ratios $\hat{R}_a$ are not strictly speaking independent. However, DOE assumes that the relationship between onsite- and phone-based outcomes is essentially independent of the relationship between phone-based outcomes and the MOS. DOE therefore treats the 3 sets of ratios—Stage 1 outcome per program dollar (R_1), Stage 2 phone-based outcome per PA MOS (R_a), and Stage 2 onsite: phone verification ratio (R_v)—as being independent. The combined effect of these errors is therefore the sum of these three terms' contributions to variance. DOE combines them using the approximation

$$RSE(Y_{TOT}) \simeq [RSE(R_1)^2 + RSE(R_a)^2 + RSE(R_v)^2]^{1/2}.$$

Thus, for those BPAC-Subcategory cells that have no Stage 2 statistical sample, the relative precision of the estimate is given by

$$RSE(Y_{TOT}) = RSE(R_1) = er_1[(1-n_1/N_1)/n_1]^{1/2}$$

where er_1, n_1, and N_1 denote the Stage 1 error ratio, sample size, and population count, respectively.

For those Stage 1strata that have a Stage 2 CATI sample but no onsite subsample, the relative precision of the estimate is given by

$$RSE(Y_{TOT}) \simeq [RSE(R_1)^2 + RSE(R_2)^2]^{1/2}$$
$$\simeq [(1-n_1/N_1)er_1^2/n_1 + er_2^2/n_2]^{1/2}$$

where er_2 and n_2 denote the Stage 2 error ratio and total sample size across all PAs. As noted, the FPC is ignored for the Stage 2 sample, but may turn out to be nontrivial.

For those Stage 1 strata that have a Stage 2 CATI sample and onsite subsample, the relative precision of the estimate is given by

$$RSE(Y_{TOT}) \simeq [RSE(R_1)^2 + RSE(R_a)^2 + RSE(R_v)^2]^{1/2}$$
$$\simeq [(1-n_1/N_1)er_1^2/n_1 + er_2^2/n_2 + er_v^2/n_v]^{1/2}$$

where er_v and n_v denote the onsite:phone verification error ratio and total sample size across all PAs.

Summing savings (or other outcomes) over BPAC-Subcategory strata s, the RSE of the sum is given by

$$RSE(\Sigma_s Y_{TOTs}) = [\Sigma_s f_s^2 RSE^2(Y_{TOTs})]^{1/2}$$

where $f_s = Y_{TOTs} / \Sigma_i Y_{TOTi}$ is the projected fraction of total savings that is in cell s.

The primary estimation goal of this study is to develop energy savings estimates for the study target as a whole—that is, for the included BPAC-Subcategories in total. To project the accuracy of this total, the additional information needed is the expected contribution of each of these primary cells to the total savings estimate. For purposes of this estimation, DOE assumes 3 different levels of savings per program dollar, in the proportions of 1:2:4. That is, the "high" savings per dollar BPAC-Subcategories have roughly twice the savings per dollar as the medium ones, and the "low" BPAC-Subcategories have savings per dollar roughly half as much as the medium ones.

With the assumptions and formulas above, the projected relative standard errors at the BPAC and total program level for each study period are shown in the tables below. As anticipated in the design of this study, the RSEs for individual BPACs are fairly wide, but the RSEs for the total are small enough for the study findings overall to be meaningful. The majority of the variability comes from the Stage 1 sampling.

Table 1: Projected Relative Standard Errors for Program Year 2008

2008					Relative Standard Error			
BPAC	**% Program Budget**	**Population # PAs**	**Sample # PAs**	**Projected % Savings**	**Stage 1 Only**	**Stage 2 CATI only**	**Stage 2 onsite: phone**	**Combined**
Building Codes and Standards	13%	19	7	18%	44%	10%	0%	45%
Building Retrofits	19%	45	15	13%	19%	9%	4%	22%
Clean Energy Policy Support	13%	31	8	16%	30%	0%	0%	30%
Loans, grants and Incentives (Excl Retro)	18%	16	8	16%	21%	8%	0%	23%
Loans, grants and Incentives (Retro Only)	27%	10	6	31%	31%	16%	12%	38%
Renewable Energy Market Development	6%	11	6	4%	28%	12%	0%	30%
Technical Assistance	4%	8	3	3%	38%	15%	0%	42%
TOTAL	**100%**	**140**	**53**	**100%**	**14%**	**6%**	**4%**	**16%**

Table 2: Projected Relative Standard Errors for ARRA

ARRA					Relative Standard Error			
BPAC	**% Program Budget**	**Population # PAs**	**Sample # PAs**	**Projected % Savings**	**Stage 1 Only**	**Stage 2 CATI only**	**Stage 2 onsite:phone**	**Combined**
Building Codes and Standards	1%	19	4	2%	44%	9%	0%	46%
Building Retrofits	28%	89	8	31%	35%	13%	17%	42%
Loans, Grants and Incentives (Excl Retrofits & Prjcts)	12%	18	2	7%	61%	0%	0%	61%
Loans, Grants and Incentives (Retrofits & Prjcts)	41%	111	10	39%	32%	12%	10%	36%
Renewable Energy Market Development	19%	69	5	20%	42%	13%	0%	44%
TOTAL	**100%**	**306**	**29**	**100%**	**19%**	**7%**	**7%**	**22%**

3. Describe methods to maximize response rates and to deal with issues of non-response.

Maximizing response rates

Based on previous experience, DOE anticipates that response rates for the surveys of probability samples planned will achieve the following response rates:

- Telephone surveys of residential customers/program participants in rebate programs: range between 30% and 64%, typically achieving response rates of roughly 45%.
- Telephone surveys of residential customers/participants in training and technical assistance programs: range between 29% and 70%, typically achieving response rates of roughly 45%
- Telephone surveys of commercial and industrial customers/program participants in rebate programs: range between 32% and 63%, typically achieving response rates of roughly 47%
- Telephone surveys of commercial and industrial customers/participants in training and technical assistance programs: range between 19% and 70%, typically achieving response rates of roughly 44%
- On-site surveys of residential customers/program participants in rebate programs : range between 50% and 80%, typically achieving response rates of roughly 75%%
- On-site surveys of commercial and industrial customers/program participants in rebate programs: range between 50% and 80%, typically achieving response rates of roughly 75%

DOE will employ a variety of best practices in order to minimize non-response bias. While the specific practices may vary among PAs and type of instrument, the general steps taken to increase response rates will include the following:

- **Conservative treatment of sample** –DOE will release sample in batches, with smaller Initial batches and larger later batches. Within each batch, DOE will make at least eight attempts to contact respondents calling at different times, over different days, leaving a minimum of three messages with a callback number. For residential surveys, DOE will ensure that each respondent is called over at least one weekend. DOE will place calls to both residential and non-residential respondents during hours most appropriate for reaching respondents in their respective time zones. For example, small contractors can typically be reached early in the morning (7am) or in the evening (7pm), with greater difficulty reaching them at other times of the day.
- **Call times** – For residential surveys, we will ensure that each respondent is called over at least one weekend. DOE will place calls to both residential and non-residential

respondents during hours most appropriate for reaching respondents in their respective time zones. For example, small contractors can typically be reached early in the morning (7am) or in the evening (7pm), with greater difficulty reaching them at other times of the day.

DOE expects the application of such techniques to yield response rates at the highest end of the scales described above.

Methods for dealing with non-response

In order to assess the presence and extent of non-response bias, DOE will contrast key parameters in the respondent group to the overall sample frame. For example, DOE will identify under-represented commercial and industrial segments by contrasting the proportion of each segment in the respondent pool relative to the overall sample frame. Other parameters within the sample frame used to identify the presence of non-response bias may include measure categories and company size. Where possible, DOE will also contrast parameters of the respondent pool's profile to secondary data sources. Use of secondary data sources will help examine whether non-response impacts the estimated outcomes as a result of regional differences between the sample frame across PAs and the overall population.

Once DOE characterizes the magnitude and potential direction of non-response bias on estimated outcomes, DOE will derive adjustment factors to the estimated outcomes. Secondary data will provide one source of possible non-response adjustments.

4. <u>Describe any tests of procedures or methods to be undertaken.</u>

Wherever possible, DOE adapted previously field-tested survey questions to develop each of the CATI surveys and in-depth interview guides. In addition to adapting individual survey questions, DOE developed the flow and skipping pattern of each instrument using previously tested questions. Therefore, while each guide is original in its entirety, they consist of a compilation of the research team's combined experience in fielding similar studies across a broad spectrum of research areas. DOE used this approach to limit the need for extensive field testing of each instrument thereby reducing response bias associated with the framing of questions.

Table 3 below provides an inventory of survey content adapted from previous research efforts in developing the CATI survey instruments. Due to the breadth of programmatic activities covered by this research effort and number of instruments required to obtain the necessary data, the table presents a summary of the information adapted from previous studies. The table shows the Market Actor (or group to be interviewed), the member of the research team who authored the contributing survey, the name of the previous study, and the general content areas addressed by adapted survey questions.

The table shows that DOE adapted much of the content from surveys used to evaluate WI focus on energy (KEMA), CA small business rebate programs (ITRON), the MI evaluation of electric and natural gas energy optimization (KEMA), and the CA indirect impact surveys (ODC). The WI study provided the general questionnaire sequence as well as specific questions used in a number of survey sections. For example, this study provided the basic framework for verifying general measure information in the tracking data, and capturing more specific engineering data. The study also provided the basic attribution sequence used throughout each of the CATI instruments. The CA study provided structure for technology specific questions used to obtain detailed measure data such specific measure properties and engineering values. This structure was used throughout the CATI surveys with specific content provided by the remaining contributing instruments.

Table 4 provides an inventory of the content adapted from previous studies to develop the in-depth interview guides. The table shows that the instruments required for the SEP evaluation spanned a broader range of market actors. Further, the specific information required from each market actor was often unique to this evaluation. However, the research team leveraged existing interview questions wherever possible.

Table 3: Survey content adapted from field-tested surveys for SEP/ARRA CATI surveys

	Question Sources		Survey Topic Area											
Market Actor(s)	Author	Previous Study Name	Intro & Finding Correct Respondent	Respondent Info & Role in Project	Initial Verification Inventory	Tech Specific	Attribution	Building Characteristics	Firmographics	On-site Recruiting	Housing & Demographics	Effect on Other Programs	Knowledge & Awareness Effects	Capacity Building
Participant; Training Recipient	KEMA, Inc.	Wisconsin Focus on Energy. Business Program Evaluation. Madison, WI:Wisconsin Department of Public Service	✓	✓	✓	✓	✓	✓	✓	✓				
Participant; Training Recipient	Itron, Inc.	Evaluation of California Statewide Small Business Rebate Programs. San Francisco:California Public Utilities Commission	✓	✓	✓	✓	✓	✓	✓	✓				
Participant; Training Recipient	KEMA, Inc.	Impact Evaluation of Electric and Natural Gas Energy Optimization Program. Clark Lake, MI.			✓	✓	✓	✓	✓	✓				
Participant; Training Recipient	TetraTech	Focus on Energy Efficiency Heating & Cooling Participant Survey. Focus on Energy, Public Service Commission of Wisconsin		✓				✓		✓	✓			
Participant	TetraTech	ACES Impact Evaluation Participant Survey. Focus on Energy. Focus on Energy, Public Service Commission of Wisconsin		✓				✓		✓	✓			
Participant; Training Recipient	Cadmus, Inc.	HIM PG&E and SDGE Clothes Washer Participant Survey. Residential Evaluation for CPUC		✓				✓		✓	✓			
Participant; Training Recipient	Summit Blue, Inc.	Partnership Program Participant Survey for Indirect Impacts Evaluation. CPUC		✓		✓		✓		✓	✓			
Participant; Training Recipient	Cadmus, Inc.	PG&E2000 Mass Markets Single Family Measures Participant Survey. Residential Evaluation for CPUC		✓		✓		✓						
Participant; Training Recipient	Cadmus, Inc.	SCE2501 Residential Energy Efficiency Incentive Program Survey. Residential Evaluation for CPUC		✓		✓		✓						
Participant; Training Recipient	Cadmus, Inc.	Appliance Rebate Program (ARP) Participant Survey. Residential Evaluation for CPUC		✓		✓		✓						
Participant; Training Recipient	Cadmus, Inc.	Residential Retrofit Evaluation Non-Participant Survey and Recruitment Effort. Residential Evaluation for CPUC		✓		✓		✓						
Participant	Cadmus, Inc.	PGE2000 Duct Test & Seal/Refrigerant Charge and Airflow Program. Residential Evaluation for CPUC		✓		✓		✓						
Participant; Training Recipient	Cadmus, Inc.	Single Family Water-Saving Devices Rebate Program Survey. Residential Impact Evaluation. CPUC. 2009.		✓		✓		✓						
Training Recipient	Opinion Dynamics Corporation	Indirect Impact Evaluation of the Statewide Energy Efficiency Education and Training Program. San Francisco: California Public Utilities Commission		✓	✓	✓	✓					✓	✓	✓

Table 4: Interview content adapted from field-tested surveys for SEP/ARRA in-depth interviews

Market Actor(s)	Question Sources		Interview Topic Area					
	Author	Previous Study Name	Intro & Finding Correct Respondent	Data Availability	Other Sources of Funding	Influence & Role in the Marketplace	Teacher Characteristics	Program Design (Logic Model)
Vendors	Tetra Tech	Adapted from qualitative questions asked in National Grid Residential High Efficiency Central Air Conditioning Program (Participating Contractor Interview Guide)				✓		
Vendors	Tetra Tech	Adapted from qualitative questions asked in the Sempra Utilities Non-Residential Process Study Staff Interview Template				✓		
Vendors	Tetra Tech	Adapted from qualitative questions asked of contractors in the Focus on Energy Efficient Heating and Cooling program				✓		
Program Delivery Contractor	Opinion Dynamics Corporation	Adapted from NYSERDA Program Implementation Interview Guide	✓					
Program Delivery Contractor	Opinion Dynamics Corporation	Adapted from NYSERDA Interview Guide for Business Partners, Implementers, and Contractors	✓					
Program Delivery Contractor	Opinion Dynamics Corporation	Adapted from FirstEnergy Pennsylvania Evaluation Program Staff Interview Guide	✓					
Program Delivery Contractor	Opinion Dynamics Corporation	Adapted from Black Hills Energy Comprehensive Evaluation Program Staff Interview Guide	✓					
Program Delivery Contractor	Opinion Dynamics Corporation	Residential Building Retrofit Vendors/Professionals In-Depth Interview Guide	✓					
Contractor; Program Manager	Tetra Tech	Adapted from qualitative questions asked in the Sempra Utilities Residential Process Study Staff Interview Template	✓					
Contractor; Program Manager	Tetra Tech	Adapted from CMS Energy Residential Program Manager Interview Guide	✓	✓				
Contractor; Program Manager	Tetra Tech	Adapted from Allegheny Pow er Comprehensive Evaluation Staff Interview Guide	✓	✓				
Program Manager	Tetra Tech	Adapted from qualitative questions asked in the NYSERDA Trade Ally Interview Guide for HPWES Contractors	✓					✓
Program Delivery Contractor	KEMA	Adapted from Xcel 2011 Business DSM internal review guide	✓		✓			
Program Manager; Vendors	KEMA	Adapted from qualitative questions asked of participating consultants and contractors in the WI Focus on Energy Home Performance w ith Energy Star In-depth Interview Guide	✓			✓		
Tech Assistance for Teachers/Trainers	Opinion Dynamics Corporation	Indirect Impact Evaluation of the Statew ide Education & Information Programs					✓	

5. **Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s) or other person(s) who will actually collect and/or analyze the information for the agency.**

- Miriam L. Goldberg, Ph.D.—Senior Vice President – Sustainable Use, KEMA, Inc.; 608-259-9152 x70211; miriam.goldberg@kema.com
- Mitchell Rosenberg—Vice President, KEMA, Inc.; 781-273-5700; mitch.rosenberg@kema.com

KEMA is the evaluation contractor and will coordinate data collection and analysis. Data collection will be carried out jointly by KEMA and its subcontractors.