

Department of Energy
Supporting Statement: ICR for DOE Impact Evaluation of HVAC, Water Heating and
Appliance R&D Program
OMB Control Number 1910-NEW

This supporting statement provides additional information regarding the Department of Energy (DOE) Office of Energy Efficiency and Renewable Energy (EERE) request for processing of the proposed information collection for the study, DOE Impact Evaluation of HVAC, Water Heating and Appliance R&D Program.

B. Collections of Information Employing Statistical Methods.

The information collection will gather opinions of experts in industry regarding the impact of DOE/EERE Building Technologies Office (BTO) investments on the development and diffusion of energy-efficient HVAC, water heating, and appliance technologies. Expert opinions are necessary in evaluation studies of R&D investments, to confirm the actual role of BTO in the financing and development of the index technologies, and to characterize counterfactual patterns of technology development and diffusion in the absence of BTO investments, thus enabling (by comparing these counterfactuals with actual observations) the estimation of the impact attributable to BTO.

Interviews with industry experts will be conducted for two technologies:

1. Advanced Refrigeration and Refrigerator/Freezer Technology
2. Alternative Refrigerants Research and the Heat Pump Design Model

The scope of these interviews will be limited to R&D performed within the past 15 years.

Separate interview guides will be developed and separate sets of respondents will be targeted for each of these two technologies. If any respondent has knowledge of more than one of these two technologies, we will select one technology for the interview; no interviewee will be asked to provide information on more than one technology. EERE will select the technology for the interview based on the applicability of one or more of the following criteria:

- 1) the technology that respondent has the greatest number of years of relevant experience since 2000;
- 2) The technology that respondent has been most actively involved with since 2000.

1. Respondent universe and selection methods

Qualified respondents must be industry experts who have knowledge in each of the following three areas:

- The relevant BTO investments and activities that use the two technologies being evaluated in this study;
- Technology development, commercialization, and market adoption and diffusion trends for these two technologies;

- Other factors that influenced those trends and, together with BTO activities and investments, contributed to the outcomes observed.

The universe of experts that meet these criteria is relatively small, heterogeneous, and difficult to define *ex ante*. Our general approach, explained in more detail below, will be to construct a sampling frame that is reasonably representative of the universe, given the best information we are able to gather before going into the field with our interview guides; to contact potential respondents in this frame according to a randomized process; and to then add respondent-referred contacts to the frame according to a process that will let us test and adjust for any bias that accepting such referrals may introduce.

Within each of the two technology areas, there will be subsets of respondents whose expertise is focused more on some aspects than others. For example, some respondents will be more familiar with DOE activities while others are more familiar with non-DOE contributing factors, and it is natural to expect perceptions and opinions about attribution to DOE to vary across these groups. Some of this variation may align with respondent characteristics that are possible to identify *ex ante*, and to the extent possible we will stratify the universe along these lines and select respondents from each strata.

For each of the two technologies, these strata would include at least the following:

1. Individuals who can be identified because of their direct involvement with BTO-funded research and development (R&D), such as engineers at national laboratories and private companies;
2. Representatives of companies manufacturing and marketing equipment related to the technology;
3. Individuals identified independently (through engineering journals, trade journals, conference proceedings, ACEEE Summer Papers and similar, etc.) as having relevant knowledge of the technology;
4. Current and former members of technical societies (ASHRAE; ASME), industry associations (e.g. Association of Home Appliance Manufacturers; Air-Conditioning, Heating, and Refrigeration Institute; National Oilheat Research Alliance; Oilheat Manufacturers Association), advocacy organizations (e.g., American Council for an Energy-Efficient Economy; California Institute for Energy and Environment), or state energy offices (see <http://www.naseo.org/members-states>), as pertains to the technology.

2. Sampling Methodology

We will build and manage our sampling frame for each of the two technologies according to the following steps:

1. Identify initial contacts in the above groups (and in other groups, as appropriate) to ensure a mix of perspectives for each technology. EERE seeks to obtain interviews from the following categories of respondents shown in Table B.1. As stated in Supporting

Statement A, the target number interviews for each technology is 40 (total of 80 across the two technologies). Based on an expectation of converting 50% of invitations to interviews, the target number of invitations for each group is twice the target number of respondents.

TABLE B.1

Category of Respondent	Target Number of Contacts Invited to Interview	Target Number of Respondents
Directly Involved	32	16
Use Technology to manufacture equipment	16	8
Possess Relevant Knowledge	16	8
Member of Technical Society	16	8
TOTALS	80	40

2. Augment the initial contact list with a respondent-driven or snowball sampling approach. These contacts will be used to augment the sampling frame, if needed, for hard-to-identify populations. We will ask each respondent to recommend other potential respondents in their professional networks. If certain groups appear underrepresented in our sample, we might ask specifically if the respondent can suggest someone in that group. Often, a respondent will refer us to someone they know, or know of, when they do not feel comfortable answering a certain question, or feel that they are not the most qualified to answer. The totality of experts compiled from Steps 1 & 2, will comprise our sampling frame. The respondent-driven approach is an appropriate sampling method in developing a frame for hard-to-identify populations.
3. Periodically reassess the stratification strategy and review response rates within each strata: Are there additional groups for which we should identify potential contacts? Do we need to focus on any one group to reach the target number of interviews? Has a potential contact been identified as having special knowledge that would be especially valuable to include in the study? Adjust the approach as appropriate.

This sampling methodology is appropriate for hard-to-reach populations that are familiar with one another when a sampling frame is not readily available or difficult to build.

3. Information collection procedures

Email invitations will be sent to prospective respondents for each technology, with a brief description of the project and its objectives. An interview guide, with background materials for the technology, will be sent once the respondent has agreed to participate and a time has been set

for the interview. Interviews will be conducted by conference call attended by two RTI staff, one to lead and one to take notes. If respondents prefer to provide responses on the interview guide (typing their responses into the Word document) and return to us by email attachment that is also acceptable; this approach may be supplemented with brief email or telephone follow-up.

To ensure that variation among responses reflects, as nearly as possible, only true differences of opinion regarding DOE's impact, we will provide respondents with timelines and tables summarizing relevant objective facts: (1) the actual observed outcomes most relevant for each of the technologies; (2) the factors that are most likely to have contributed to those outcomes (with as much detail as possible on organizations, activities, and dates). The data collection instrument will ask respondents to react to the information presented: Is it complete? Are there important factors missing that they would want to add?

4. Methods to maximize response rates

We will leverage mutual connections when possible to increase the likelihood of response. This may take the form of an email introduction from one respondent to another prospective respondent, or referencing in an email interview invitation the name of a mutual connection (with that person's permission) who recommended that we reach out.

Every effort will be made to ensure that a potential respondent has an opportunity to contribute their opinions and perspectives to the study in the way that is most convenient for them. As described in Section 2 for example, although data collection will most often take the form of telephone interviews, some respondents may prefer to provide responses offline on the interview guide (typing their responses into the Word document) and return to us by email attachment.

A follow-up email will be sent when an individual does not respond to our first interview invitation. If a contact has been identified as having special knowledge that would be especially valuable to include in the study, the email may include additional text explaining our interest in speaking with them specifically. The maximum number of contact attempts will be three.

5. Analysis of Results

Summary statistics will be calculated for quantitative responses: mean, median, quartiles, standard deviations. Where appropriate, standard statistical tests for differences of mean (t test) will be used to explore whether different respondent groups offered significantly different answers (e.g., national lab versus industry, or BTO-funded versus non-BTO-funded).

To test and adjust for any bias resulting from the respondent-driven sampling approach, we will also report adjusted means, with responses weighted by inverse propensity scores, so that the opinions of respondents who were less likely to be included in our sample are weighted more heavily. We would estimate a propensity score (i.e. probability of inclusion in the sample) for each potential respondent as a function of characteristics. These characteristics could include,

e.g., indicators (0/1, or dummy variables) for affiliation with companies, national labs, and industry associations; indicators of the source of the contact information (e.g., DOE-provided, RTI-identified, respondent-referral); and the number of times another respondent referred us to the potential respondent. We would then report results based on the weighted average of responses, where weights are given by the inverse of the respondent's propensity score. A respondent in our sample with a propensity score of 0.25 (i.e., this respondent had 1 chance in 4 of being included) would be included 4 times when calculating an adjusted average.

To flesh out this example, suppose that our respondent was identified by RTI, which found his name among the authors of a relevant article in the ACEEE Summer Study Proceedings archives, pre-1990; that he had worked in industry; that he had been affiliated with an industry association; and that he had retired more than 10 years ago. Suppose our sampling frame contains three other individuals who fit this description but from whom we did not receive responses. Then this one individual would have a propensity score of 0.25 and his responses would be given a weight of 4 when calculating an average. In comparison, suppose that 8 out of 10 contacts RTI was given by another respondent, being referred to each one exactly once, provided a response; then each of those respondents would receive a weight of 1.25. Suppose further that every individual to whom RTI was referred by more than one other respondent provided a response themselves; then each of those respondents would receive a weight of 1.0. There may be variability in the characteristics among the cases being weighted up to approximate the population and that may affect the accuracy of the findings. The weighted and unweighted estimates will be used to assess how close the estimates from the sample approximate the target population parameters. The analysis will be based solely on results from the interviews. Since there are no available estimates of the target population parameters, the evaluation will be based on how reasonable the results are based on the interviews.

The foregoing discussion is more heuristic than it is technically exact. In carrying out these procedures in our data, we would follow accepted best practice in the social science literature. For references to some key articles, see <http://www.respondentdrivensampling.org>. Note that the proposed methods cannot create nationally representative estimates, because the actual underlying population of informed experts cannot be completely characterized. The proposed methods endeavor to build a sampling frame that is as close as possible to representative, based on available information, and then weight responses to the frame. We will not claim that the frame is nationally representative; therefore we cannot claim that our results (weighted to the frame) are nationally representative.

To reiterate, weighted metrics will be reported *in addition to* unweighted metrics, as a robustness check against potential selection bias. Given the impossibility of perfectly defining the sampling universe *ex ante*, our weighting scheme must be recognized for what it is: a good-faith best effort to account for potential selection bias. If weighted and unweighted statistics are significantly different from one another, that fact will be discussed in our report.

As is customary, a range of robustness checks will be provided, including but not limited to the inverse propensity score weighting described above. For another example, it will certainly be appropriate to include a comparison of the variance of quantitative responses within and across chains of referrals. If there is correlation within groups defined by respondents having referred or been referred by others in the group, then standard errors and confidence intervals will be adjusted appropriately. The potential for bias caused by correlation within chains of referrals is more subtle. Bias could arise, for instance, if respondents who take an especially (un)favorable view of BTO contributions make a greater number of referrals or if their referrals are more likely to respond. This issue could be addressed within the propensity score weighting framework described above, or by a straightforward extension of it. Suppose that for every 1 standard deviation increase (decrease) in a constructed ‘favorability index’ for respondents, the average number of responses generated by referrals made by that respondent increases (decreases) by 20%. Then an adjusted average could be calculated in which a response that came from a referral made by someone whose favorability index was 1 standard deviation (s.d.) below the mean would be weighted by 1.2; referred by someone 2 s.d. below the mean, weighted by 1.44; referred by someone 1 s.d. above the mean, weighted by 0.83; referred by someone 2 s.d. above the mean, weighted by 0.69; etc. Although these measures provide some quantifiable measure on the variability of the responses, the method is not designed to produce estimates of the sampling error, frame coverage error, or model error.

5. Statistical Consultants

Troy Scott, PhD in economics, will oversee the data collection process at RTI.

Jake Bournazian, at the US Energy Information Administration, has advised us on this data collection. His contact information is below.

Jacob Bournazian
Survey Development Team Leader
Office of Survey Development and Statistical Integration
US Energy Information Administration
(202) 586 - 5562
jacob.bournazian@eia.gov