Supporting Statement Part B
U.S. DEPARTMENT OF ENERGY HYDROGEN PROGRAM KNOWLEDGE AND OPINIONS ASSESSMENT

B. Collections of Information Employing Statistical Methods

1. Respondent universe and selection methods

Potential respondent universe

The respondent universe for the general public survey will be non-institutionalized U.S. adults (ages 18 and over), approximately 223 million. The universe for students age 12-17 is approximately 25 million.

U.S. populations* (thousands) by survey age groups	
Adults (18 and over)	223,031
Students (12-17)	25,116
*Estimates for July 2008 from http://www.census.gov/population/projections/nation/detail/d2001_10.pdf	

The respondent universe for the state and local government personnel is defined as the primary representative or designee from each of the following:
- Three state government agencies (State Energy Offices, Departments of Transportation, and Departments of Environmental Protection) in all 50 States.
- Twelve largest cities and twelve largest counties in each of the four census regions.

The respondent universe for the end users is defined as representatives from the following three categories:
- Transportation.
- Energy users who require an uninterrupted power supply.
- Industrial end users who have large power requirements.

Data/respondent selection methods

The sampling method for the general public and students will be random digit telephone dialing (RDD) using the Genesys system (Marketing Systems Group, http://www.m-s-g.com/virtuallibrary/reflisting.aspx) for generating samples. This approach to public opinion surveys is standard and widely used. For example, it is being used by the Bureau of Transportation Statistics for their *Omnibus Household Survey* (see http://www.bts.gov/omnibus/household/general_methodology.html).

The targeted sample for state and local government agencies is composed of all 246 offices as noted in the previous section (i.e., State Energy Offices, State Departments of Transportation, State Departments of Environmental Protection, twelve largest cities in each of the four census regions, and twelve largest counties in each of the four census

regions). For each census region, if county and city governments are combined into a single government entity, only one call will be made to that office, and the next largest county in that census region will be selected for interviewing. An advance letter will be sent to a named person in each organization. Sampling will be accomplished by calling the appropriate office and requesting the person to whom the letter was addressed. The designated respondent (i.e., the addressee of the letter) may assign someone else in the office to represent the agency if need be.

End users are defined as businesses and industries with facilities in the United States which are potential commercial users of hydrogen and/or fuel cells. Businesses will be identified by North American Industry Classification System (NAICS) or Standard Industrial Classification (SIC) codes and stratified into three sectors. Component population numbers for each of the three sectors of potential end users are shown in the following table.

Populations and interview plans for the three sectors in the end user category				
Hydrogen usage sector	Number in component population*	Number in target population**	Number of completed interviews	Number of attempted interviews
Transportation (i.e., end users)	282,866	849	332	664
Businesses needing uninterrupted power supplies	891,817	2,675	334	668
Industrial sectors with large power requirements	344,188	1,033	334	668
Total	1,518,871	4,587	1,000	2,000
*Based on the Census Bureau's 2002 Economic Census. **Sampling was restricted to the largest 0.3% of the component populations.				

The sampling method for the end users is as follows:
- Lists of businesses meeting the above criteria will be purchased from Dun & Bradstreet (specifically the D&B Market Place database).
- Businesses in the sectors of transportation and businesses needing uninterrupted power supplies will be ranked by number of employees.
- Businesses in the category of industrial sector with large power requirements will be ranked by revenue.
- The largest 0.3% of businesses will be selected from each category and used to represent the subcategory in the strata because the largest businesses represent the greatest potential for hydrogen and fuel cell usage.

Expected response rate

Because of their technical nature, it is not reasonable to expect a high response rate for the hydrogen surveys. However, because they are intended to measure *changes* in

hydrogen awareness and understanding, non-response bias is less of a problem than in many other surveys, because non-response bias tends to cancel out of the cross-year differences computed to measure changes.

The response rate for the 2004 general public survey was 25%. The response rate for the 2004 student survey was 28%. The response rate for the 2004 state and local government survey was 96%. The response rate for the 2004 survey of end users was 30%. The response rate for end users may increase from the 2004 rate as the proportion of businesses actually using hydrogen and fuel cell technologies increases. Response rates for the other survey components are expected to be similar to their corresponding 2004 rates.

2. Procedure for collection of information

Statistical methodology for stratification and sample selection

The sampling for the general public and student surveys will be by list-assisted RDD using the Genesys system and CATI methods, as in the 2004 survey. The Genesys samples are implicitly stratified by Census Division and Metropolitan Statistical Area size. For purposes of statistical sampling, the general public and student target populations are essentially infinite. State and local officials compose a much smaller target population and will be explicitly identified. End users will be identified with Dun & Bradstreet's Market Place database, as in the 2004 survey.

Estimation procedure

Estimates and other statistical analyses will be computed using the Statistical Analysis System (SAS) including the SAS Surveymeans and Surveyfreq procedures. The analyses will properly account for the stratification as well as sampling weights. Sampling weights will be computed from selection probabilities adjusted for non-response, households with multiple telephone numbers, and, by iterative proportional fitting, post-stratification by age, sex, and region. In addition to standard summary statistics, exploratory analyses (e.g., contingency tables and correlations) will be performed to investigate possible relationships between responses to the various questions (e.g., performance on technical questions vs opinion questions, or differences attributable to responder age, sex, region, etc.). As with the 2004 survey, results of the current survey will be used as input to Hydrogen Program decisions regarding education and outreach.

Degree of accuracy needed for the purpose described in the justification

The 2004 data and analysis report, including estimates with standard errors and confidence intervals, is at http://www1.eere.energy.gov/hydrogenandfuelcells/hydrogen_publications.html (see links under "General/2004 Hydrogen Knowledge and Opinions Survey"). For example, the average score on the technical questions for the general public was 32.8% with 95% confidence limits of 31.3 and 34.4. The corresponding average score for state and local

officials was 65.8% with 95% confidence limits of 65.4 and 66.2. These and many other statistical results have been used in the Hydrogen Program. With the exception of the end user survey, they have been considered adequate but not excessive in statistical precision. With the exception of the end user survey, the survey will update the information collected in 2004 subject to the same accuracy requirements. For the end user survey, the sample size will be increased to about 1,000 so that it achieves the same accuracy as the survey of the general public.

Otherwise the sample sizes for the current survey will be the same as in the 2004 survey. Because the variance of a difference of two statistically independent estimates is the sum of the variances of the two estimates, cross-year differences will be estimated with standard errors of about $\sqrt{2}$ times the standard errors of the single-year estimates. This is also considered to be adequate but not excessive.

Unusual problems requiring specialized sampling procedures

None.

Use of periodic (less frequent than annual) data collection cycles to reduce burden

This survey is to be repeated every three or four years; it is not an annual survey.

3. Response rates and the issue of non-response

All surveys conducted in this information collection will employ CATI technology. All interviews will be conducted by professionals trained in interviewing techniques. Although a CATI interface will be used to facilitate sampling and dialing, potential respondents will interact only with the interviewers themselves, not the CATI interface (for example, respondents will not be asked to hold or push touch-tone keys).

Maximizing response rates

Procedures in this study for maximizing response rates will exceed usual standards for CATI surveys. Interviewers will make a minimum of 15 attempts to reach eligible households. Each call attempt will be a minimum of five rings. The CATI software will cycle the attempts in the following order: weekday day, weekday evening, Saturday day, and Sunday evening to maximize coverage of the residential population. Sample allocation and scheduling of interview sessions will assure a minimum of three attempts in each day part (e.g., weekday day, weekday evening, weekend). Lines that are busy will be called back a minimum of five times at 10-minute intervals. If the line is still busy after the fifth attempt, the number will be attempted again on different calling occasion(s). If the line is still busy after the fifth calling occasion, the CATI system will attempt to contact the phone company to ascertain whether the number is actually in service.

The surveys will be conducted to accommodate the special feature that relatively large proportions of "don't knows" and "no opinions" are expected. (Changes in the proportions of people having an opinion on these questions are of particular interest.) Respondents will be assured that relatively few people currently know the answers and that "don't know" and "no opinion" are perfectly acceptable answers. This will help to minimize item non-response rates. Opinion questions will be mixed with technical questions to minimize possible frustration with technical (hydrogen science) questions.

Methods for dealing with non-response

Statistical methods. A primary objective of the surveys is to measure changes in understanding and awareness about hydrogen and fuel cell technologies. Thus, statistics collected during the survey will be compared directly with the 2004 results. Although non-response bias is always a potential problem in CATI surveys, it is a less serious problem in this survey, because of the tendency for non-response biases to cancel out in differences. That is, the non-response bias in the 2004 baseline and current surveys are expected to be similar. When cross-year differences in estimates are computed, bias will tend to subtract out.

Another statistical technique that will be employed to compensate for non-response will be adjustment of sampling weights for differential response rates in various demographic categories. In addition to these statistical adjustments, the following CATI procedures will also be employed to reduce non-response.

Call backs. CATI surveys accurately handle large numbers of scheduled call back appointments. When a scheduled appointment time arrives, the CATI system finds the next available station and delivers the record as the next call. Ensuring that appointments are kept helps to maximize response rates (and to minimize imposition on study participants). The CATI system also allows for callbacks to reschedule interviews and to restart interrupted interviews (for example if a respondent wants to terminate an interview but to finish it later). Scheduled call backs can be either casual (general time) or definite (exact time) depending on the respondent's request. The CATI system also automatically handles callbacks for no-answer, busy, and answering machines. Call backs for busy signals are retried at several minute intervals; callbacks for no-answer and answering machines are scheduled to ensure coverage at different times of day.

Refusal Conversion. Another step that will be taken to maximize response rates will be to route each initial refusal to special survey staff trained and experienced at converting initial refusals to responses. How initial refusals are handled is decided on the basis of each particular case and the experience by the special staff member. A frequently successful technique that will be used is switching from an English-speaking to a Spanish-speaking interviewer.

4. Pretesting

The survey will be an update of the 2004 survey. Thus, the attached survey questionnaires have been thoroughly pretested for the most part. A few new questions have been added and these new questions will be tested prior to the survey being fielded.

5. Name and Telephone

Oak Ridge National Laboratory (ORNL) will oversee the survey and conduct the data analysis. ORNL contact: Richard L. Schmoyer, PhD, Statistics (1980), 865-946-1255, ric@ornl.gov.