PART B OF THE SUPPORTING STATEMENT

1. QUESTIONNAIRE OBJECTIVES, KEY VARIABLES, AND OTHER PRELIMINARIES

1(a) Questionnaire Objectives

The primary objectives of the survey are to gather and update information about coalbed methane (CBM) extraction facilities to determine if an effluent guidelines rulemaking is required to control the discharge of pollutants into surface waters of the United States and to publicly owned treatment works (POTWs). More specifically, the survey will provide EPA with preliminary economic, technical, and environmental data needed to quantify any adverse environmental impacts of the discharges of produced waters from coalbed methane extraction facilities, evaluate the effectiveness of treatment technologies, determine the incremental pollutant removals, and develop compliance costs and economic impacts for a range of possible wastewater management options.

1(b) Key Variables

EPA is focusing on the availability and affordability of CBM produced water treatment technology options in advance of a decision on whether to initiate an effluent guidelines rulemaking. EPA plans to use the survey to collect information specific to individual CBM projects as the economics and environmental impacts of CBM production are dependent on the location of the CBM development and the surrounding ecosystem. EPA will use this information to assess the statutory factors, particularly technological and economic achievability of available controls, production processes, potential environmental impacts, and wastewater treatment residuals disposal practices. EPA is using the following key variables as they relate to the potential technology options and beneficial use practices for this industrial sector.

- Produced Water Management Practices and Pollutant Discharges;
- Capital and Operating and Maintenance Costs for Produced Water Management Practices;
- Economic Size of CBM Operators and Projects;
- CBM Reserves Estimates and Production by Basin and Future Timing of New CBM Production; and
- Potential Economic Impacts of Different Produced Water Management Practices on CBM Operators and CBM Exploration and Production.

Please see Part A, Section 4(b)(i) of this Information Collection Request for detailed information on the data items for the survey.

1 (c) Terminology

For purposes of this data collection, EPA has defined two terms: operator and project. This section describes both in more detail because they are key concepts in the sample design.

An *operator* is the firm or division (if a profit center) that is responsible for management and the day-to-day operation of a project. This operator is generally a working-interest owner or a company under contract to the working interest owner(s). The working-interest owner bears the costs of exploration, development, and operation of the property and, in return, is entitled to a share of the mineral production from the property or to a share of the proceeds there from. The sample design assumes that an operator is located in only one basin. For example, if an operator has projects in three basins, EPA has counted it three times: once in each basin.

A *project* is defined to be comprised of a well, group of wells, lease, group of leases, or recognized unit operated as an economic unit when making production decisions. (EPA recognizes that industry has multiple definitions for the term "project.") One reason that EPA is interested in project-level information is because many projects handle the produced water in a single water management system. For purposes of estimating sample sizes for the information collection request (ICR), EPA has assumed that a project consists of three or fewer wells.

As described in the next section, operators would identify their projects and the number of wells in each in their responses to a screener questionnaire. EPA would then use this information to refine the sample design.

2. STATISTICAL APPROACH FOR THE SURVEY

The statistical approach considers the target population, the available information in the sample frame, the sample design, and sources of error. The following sections describe each component in more detail.

2(a) Target Population

The principal task in the development of a sample survey design is establishing a clear, concise description of the target population. The definition should clearly identify every element of the target population so that all non-population elements can be excluded. For the screener, EPA would define the target population to be all operators in the U.S. The screener then would collect information from the operators about each project, because the target population for the detailed questionnaire would be all projects producing CBM in 2007.

2(b) Sample Frame

A sample frame is a list or set of procedures for identifying all elements of a target population. In addition to listing population elements, sample frames also contain additional information, such as addresses and key characteristics of the population that will be used to draw samples. Sample frames are essential to the quality of surveys because sample elements are drawn from them.

(i) Screener

For the *screener*, the target population is all operators in the U.S. that produced CBM in 2007. EPA has created a sample frame of operators in the U.S. using licensed database information on historic well production from HPDI, Inc. HPDI, Inc. compiles information from nearly all of the oil and gas producing states and provides detailed data in a consistent format to clients accessed through a web-based query system. This information includes well identification information (such as API number, lease name and number, well name and number, operator name, location, basin designation, field, and reservoir/producing formation), historic production information (including summary information on first production, last production, cumulative production, and last 12 months production as well as detailed information on year-by-year production), status information (active/inactive), and operator contact information (where available). EPA has supplemented the sample frame with information publicly available from States. As a result, EPA considers that it has created a relatively complete sample frame of operators in the U.S. It identifies 484 operators maintaining 45,369 wells that were active CBM producers in the U.S. as of mid-2007.

(ii) Detailed Questionnaire

For the *detailed questionnaire*, EPA would use the information that it had gathered from the screener to create a sample frame identifying the projects. As explained in Section 2(c), EPA estimates, at most, 15,251 projects would be identified from the screener.

2(c) Sample Design

EPA would collect information from the industry using a two-phase sample design. In the first phase, EPA would collect information about the projects from all operators. In the second phase, EPA would select a subset of projects for the detailed questionnaire. Based upon worse case assumptions, EPA estimates that the sample would consist of 650 projects, which would then represent an estimated 15,251 projects in the total population. This section describes the sample design and selection for the screener and detailed questionnaire.

(i) Screener

The screener would be sent to all operators with more than two wells. To minimize burden for the screener, EPA has taken three approaches. First, for operators with one or two wells, EPA would attempt to complete the screener using publicly available information and contacts with industry associations. Only as a last resort would EPA contact the operators themselves. EPA expects that any additional information required from the operators could be obtained through a brief telephone call, instead of a written response to the complete screener. Second, EPA has only requested information necessary to statistically select the projects for the detailed questionnaire. Third, by applying recent questionnaire formatting research, EPA has designed and formatted the questions and response categories to be easy to read and interpret.

(ii) Detailed Questionnaire

The detailed questionnaire would be sent to a sample (subset) of projects identified from the screener responses. By requiring only a subset to respond to the detailed questionnaire, the survey burden would be greatly reduced. EPA would select projects in a manner that would be statistically representative of all projects in the target population. EPA would supplement the statistical sample with a small judgment sample of hand-picked projects that would be used for qualitative review. In an effort to reduce burden EPA may supplement the data with information from treatment vendors or by requesting follow-up information from questionnaire respondents with treatment technologies of interest (see DCN 5767 for an example request letter). For example, the detailed questionnaires may contain summary information about a potential treatment technology for which EPA may require additional design and cost information for the costing analyses. EPA may then need to request more detailed information about this specific technology.

Stratification

For the statistical sample, EPA would *stratify* the projects, primarily, by basin and business size. Stratification is performed by selecting one or more characteristics of interest and dividing the members of the population into the strata based on those characteristics. The sample frame created from the screener responses would identify these characteristics or provide a basis to reasonably assign characteristics to each population member (i.e., project). Stratified sampling consists of selecting a sample from within each stratum, then combining them to constitute the total sample. There are several benefits that result from stratifying the population, including:

- Ensuring that the sample contains representatives from every stratum;
- Improving the precision of parameter estimates;
- Allowing important parameters to be estimated at the stratum level; and
- Allowing certain subpopulations of particular interest to be sampled at a greater rate than others.

To select systems to receive questionnaires, EPA intends to use the following stratification variables:

- *Basin.* EPA's screener sample frame identifies the basin for each operator. EPA intends to stratify by basin to capture geological and regional differences in the industry. Because strata with small sample sizes can lead to inefficient sample designs, EPA has combined a few small basins within the same state. The following sample design estimates and descriptions refer to the combined basins as if they were a single basin. EPA has combined basins within the following three states:
  - *Alabama*: Alabama/Florida/Mississippi Salt Dome (two operators with a total of six wells) has been combined with Cahaba (two operators with 313 wells).
  - *Texas*: Fort Worth (one operator with one well) has been combined with Permian (one operator with six wells) and Texas Gulf Coast (one operator with two wells).

- *Wyoming*: Big Horn (one operator with one well) has been combined with Wind River (six operators with 20 wells).

- *Business size (Small or Non-Small)*. The screener would collect this information by asking if the parent company qualifies as a small business under the Small Business Administration definitions. By stratifying by business size, EPA would collect information that would allow for national estimates about specific impacts to small businesses.

In addition, EPA would further stratify certain basins by:

- *State* for the Powder River Basin (PRB). This basin has the largest number of wells for any basin over a wide geographical area, and thus, EPA considers a subdivision by state (Montana and Wyoming) would be appropriate. This subdivision could easily be incorporated into the sample design because the state location is available in the screener sample frame for each operator and well. The following sample design estimates and descriptions incorporate PRB/MT and PRB/WY as if they were separate basins.

- *Discharge Category* for two basins: PRB/WY and Appalachian. The screener would collect information about discharge practices that EPA would combine into two categories: *zero* and *other*:
  - The *zero* category would include any project that specified *only* these practices: re-injection hauled or trucked off-site, land applied such as irrigation and dust suppression. (In the screener, EPA has chosen to specify the practices separately to lessen any confusion.) Operators that only practice land application for a project where there is a discharge to a surface water would select the fourth check box ("Disposed or discharged by some other method") which would put this project in the other category.
  - The *other* category would include any project that disposes or discharges by some other method, possibly in combination with zero discharge practices.
  - For completeness, the screener includes a third possibility: no water produced by the project. For any such responses, EPA would contact the operator for more information.

Population Estimates for Each Stratum

Table B-1 shows the number of operators, wells, and estimated projects within each basin. EPA estimated the number of operators, wells, projects, small businesses, and discharge status as follows:

- *Operator*. EPA identified 484 operators in its screener sample frame. The screener would ask that the respondent correct the operator name, if appropriate.

- *Wells*. EPA identified 45,369 wells in its screener sample frame. The screener would ask for the number of wells in each project for two reasons. First, to assess

completeness of the responses and the quality of the screener sample frame, EPA intends to compare the number of wells reported by each operator to the number in the screener sample frame. EPA would contact an operator if there was a substantial difference between EPA's previously compiled information compared to the operator's response. Second, to create a more efficient sample design, EPA would statistically select the project with probability proportional to size. Size would be determined by number of wells and volume of gas production.

- *Projects*. EPA estimates that the industry has 15,251 projects. EPA derived this estimate as follows. For each operator with one, two, or three wells, EPA assumed that the operator had one project (i.e., 202 operators have an estimated 202 projects). For each operator with four or five wells, EPA assumed that the operator had two projects (i.e., 32 operators have an estimated 64 projects). For remaining operators (i.e., those with six or more wells), EPA estimated the number of their projects in each basin by dividing by three and rounding upwards to the nearest integer. For example, EPA identified four operators in the Arkla basin. One operator has one well, which EPA assumed is a single project. Three other operators have 26 wells that EPA combined into an estimated 9 projects (i.e., 26/3=8.67 rounded to 9). In total, Arkla has an estimated 10 (i.e., 1+9) projects. EPA used this calculation procedure for all basins to obtain the estimated number of 15,251 projects. EPA expects that this is an overestimate because it has visited projects with many more than three wells. However, the assumption provides a reasonable upper bound (i.e., worse case) on the sample sizes used to estimate the burden of the survey.

- *Small Businesses*. EPA estimated that 178 operators are small businesses. For purposes of developing estimates for the sample design, EPA assumed that operators with only one or two wells are small businesses. EPA recognizes that some operators in this category are large companies (e.g., one is a large Canadian company), and operators with more than two wells might be a small business. Consequently, EPA intends to use the screener responses to revise its small business classifications.

- *Discharge Status*. (Not shown in table because it applies only to two basins). For purposes of the burden estimates for the ICR, EPA estimated the discharge status using its professional judgment based upon its site visits. For the Appalachian basin, EPA estimated that 60 percent (i.e., 892 projects) would have some form of zero discharge, and the remainder would have other types of discharge. For the Power River Basin in Wyoming, EPA assumed that 50 percent (3219 projects) would have zero discharge. EPA would use the screener responses to revise its allocations to the two discharge categories before selecting the sample for the detailed questionnaire.

Table B-1 Numbers of Operators, Wells, Projects, and Small Businesses

Basin(s)	Number of Operators	Number of Wells	Estimated Number of Projects	Estimated Number of Small Businesses
Alabama/Florida/Mississippi Salt Dome & Cahaba	4	319	108	1
Anadarko	68	1,214	433	44
Appalachian	47	4,421	1,487	20
Arkla	4	27	10	1
Arkoma	55	1,420	492	23
Big Horn & Wind River	7	21	10	5
Black Warrior	13	4,771	1,593	4
Cherokee/Forest City	49	2,311	786	22
Ft. Worth, Permian, & Texas Gulf Coast	3	9	4	2
Green River	20	287	103	10
Illinois	21	141	54	8
Powder River, MT	3	877	293	1
Powder River, WY	79	19,295	6,440	11
Raton	10	2,736	914	1
Salina	3	3	3	3
San Juan	76	6,546	2,191	13
Sweetgrass Arch	1	1	1	1
Uinta-Piceance	21	970	329	8
Total	**484**	**45,369**	**15,251**	**178**

Precision Estimates Used to Derive the Estimated Sample Size

The precision of the detailed questionnaire estimates depends on both the sample design and the sample size, that is, the number of projects that would be selected. One measure of precision is the width of the confidence interval for the estimate. Confidence intervals provide a range of values for a particular estimate that would be likely if the study were repeated an infinite number of times. Thus, when using 95 percent confidence intervals, 95 percent of such intervals would include the true value, if we could take an infinite number of samples.

The binomial distribution is often used as the basis of sample designs, and can be used to estimate precision. The binomial distribution applies to situations where there are only two outcomes (yes or no) to a dichotomous question such as “Does the project have zero discharge?” The presence or absence of the attribute for a particular project is a dichotomous, or binary, variable. The binomial distribution models these data, based on the notion of obtaining national estimates of the percentage or proportion of projects in the target population (or a subset of the target population) that have a particular attribute. The binomial distribution also provides

estimates of the variance that is used to calculate the confidence intervals. Because a proportion of 0.5 (or 50 percent) results in the largest possible variance for the binomial distribution, EPA assumed that the probability of one outcome would be 0.5 (e.g., zero discharge occurs at 50 percent of the plants). In other words, if the population value is any value other than 50 percent, the survey estimate will be more precise – in statistical expectation – than it would be if the population value is 50 percent.

Because EPA is developing a national rule, it is primarily concerned with the precision of the overall estimates. Consequently, in estimating the overall sample size, EPA assumed more stringent requirements for overall estimates than basin estimates. First, EPA assumed that the sample (unadjusted for non-response) would be expected, with 95 percent confidence, to yield sufficient data to estimate the value of an unknown proportion to within ±0.05 of its true value for the target population (i.e., projects). This precision target will hold when the proportion's true (unknown) value is equal to 0.5, and even greater precision is expected when the true value of the proportion is not equal to 0.5. EPA then allocated the sample among the different strata. For strata with five or fewer projects, EPA assumed that the sample would include all of them. For the remaining strata with more than five projects, EPA applied another precision target for purposes of estimating the sample size. EPA assumed that the sample (unadjusted for non-response) would be expected, with 90 percent confidence, to yield sufficient data to estimate the value of an unknown proportion to within ±0.15 of its true value for the target population of all projects. EPA then adjusted the stratum estimates upwards by 10 percent for potential non-response. (This assumption of a 10 percent non-response rate is based upon a typical effluent guidelines questionnaire.) As a result of these calculations, EPA estimated that a statistical sample size of 610 projects would be appropriate. In addition, EPA intends to further adjust the sample to include 40 additional projects for a judgment sample. Typically, in EPA's experience with surveys for effluent guidelines, we identify facilities for which additional information would be useful, but were not captured into the statistical sample. For example, industry may identify a project with a unique treatment system for which we need data to evaluate its performance for the CBM Study. For this reason, we are including a judgment sample. We will develop our national estimates based upon the statistical sample with the non-response adjustment, and will use the data from the judgment sample separately in a qualitative manner. Table B-2 summarizes the steps that EPA used to estimate the sample size.

Table B-2 Steps In Sample Size Estimation

Estimated Number for:	Sample Size
National Estimates	375
After Allocation to Strata	556
After Strata Adjusted for 10% Non-response	610
With addition of Judgment Sample	650

Actual Sample Design

The previous section described EPA's procedure for estimating the sample size used to calculate the burden estimates for the ICR. However, if EPA were to use that procedure in actually selecting the sample, it would result in statistically inefficient estimates. In order to reduce the variability associated with the estimates, EPA intends to use the information from the

screener responses to create a more efficient sample design. In this improved design, EPA still intends to maintain at least an overall precision target that would be expected, with 95 percent confidence, to yield sufficient data to estimate the value of an unknown proportion to within ±0.05 of its true value for the target population (i.e., projects). Within this precision target, the sample design would select projects with probability proportional to size (PPS) within each stratum. The size of the project would be determined from the screener responses for the number of wells and the total gas production. Because strata with small population sizes are statistically inefficient, EPA also intends to evaluate whether collapsing any strata would improve the estimates while ensuring that sufficient information would be collected within the basins, small businesses, and the other strata.

2(d) Sources of Error

In developing the sample design, as described previously, EPA considered the estimated precision targets for data collected from the target population. EPA also considered potential error that could be associated with estimates calculated from the collected data, due to sources associated with sampling, such as response rates, as well as non-sampling sources of error, such as processing error.

(i) Response Rates

In developing the sample design, EPA considered both unit (questionnaire) and item (question) non-response. EPA expects that the response rate would be relatively high for this mandatory survey effort. The survey would be conducted under the authority of Section 308 of the Clean Water Act. The cover letters and instructions for the screener and detailed questionnaire would explain the legal authority, responsibility to respond, reasons for the questionnaire, and penalty for non-response. EPA would use reminder letters and/or telephone calls to remind respondents of the duty to respond under Section 308 of the Clean Water Act. If possible, EPA would seek the endorsement of the major trade associations, which would be expected to increase the response rate from its members. EPA recognizes that some non-response is unavoidable, and in past survey efforts, EPA has waived the duty to respond in extreme and rare cases (e.g., natural disasters) which also might occur for this survey effort. However, for the screener, timely and complete responses would be particularly important because it would be used to create the sample frame for the detailed questionnaire. If an operator is unable (or unwilling) to respond within the short timeframe, EPA might estimate the responses from publicly available information so that it would have a complete list of projects to draw the sample for the detailed questionnaire.

Prior to distributing the detailed questionnaires (units), EPA would adjust the initial sample sizes to help ensure that the effective sample sizes (i.e., respondents) would be sufficient for precision requirements. EPA would adjust the statistical sample size for an estimated non-response rate of 10 percent. (This assumption of a 10 percent non-response rate is based upon a typical effluent guidelines questionnaire.) In addition to increasing the initial sample size, EPA would strive to improve the response rate by reminder letters and/or telephone calls. Furthermore, after receiving the responses, EPA intends to adjust the detailed questionnaire weights for any non-response and to review publicly available information (e.g., State databases) in order to determine if non-respondents appear to have different characteristics than

respondents. EPA would examine these characteristics both for the entire industry and for subgroups in the analyses. For any differences, EPA intends to determine the major causes, and to incorporate appropriate adjustments for bias. (Bias is the difference between the expected value of an estimate and the true value of a parameter or quantity being estimated. If the data collection process generates estimates that are consistently (or on average) above or consistently below the true value, the data collection process is biased.)

To minimize item non-response, EPA's subject matter experts have worked closely with industry to develop questions that would be easy to understand with clearly defined and familiar terms; are formatted in a logical sequence; and would request data that are readily available within the industry. In this manner, EPA expects to minimize inaccurate or incomplete response of the questions that can occur due to misunderstanding or misinterpretation of questions and the unintentional skipping of questions by respondents. Additionally, EPA would operate an e-mail helpline and website to assist respondents with the screener and detailed questionnaire. After receipt of the completed detailed questionnaires, EPA intends to conduct extensive follow-up with respondents for any item non-response. If necessary, EPA would impute responses to key questions in our analyses.

(ii) Processing Errors

Processing errors can occur when questionnaire responses are coded, edited, and entered into the database. The design and implementation of the questionnaire database would employ a number of quality assurance techniques to reduce the frequency of such errors. These techniques would include the following:

- Double-entry keypunch verification on critical questions;
- Computerized comparison of selected responses to detect inconsistencies and illogical responses;
- Computerized analyses to screen for out-of-range and inconsistent numerical values; and
- Computerized analyses to detect missing numerical data and missing units.

3. PRETESTS AND PILOT TESTS

EPA does not intend to pre-test the questionnaire. For more than 30 years, EPA's Engineering and Analysis Division has conducted surveys of numerous industrial sectors to collect information to support regulation development activities in the effluent guidelines program. While EPA develops different questionnaires for each industry, there are common elements for all industries. The questionnaires collect the same basic data such as information about processes, treatment, and financial status. Thus, when EPA develops a questionnaire for a particular industry, it generally tailors the questions for specific terms and processes used by that industry. In past years, EPA has relied predominantly on active participation by trade groups in

reviewing the questionnaires. In EPA's experience, such collaboration generally tends to better reflect the industry at large than pre-tests. For this reason, EPA considers additional review through the pre-test process to be unnecessary for this industry.

4. COLLECTION METHODS AND FOLLOW-UP

Please See Part A, Section 5(b) of this ICR for this information.

5. DATA PREPARATION AND ANALYSIS

5(a) Data Preparation

Upon receipt of completed questionnaires, EPA and its contractors would review the questionnaires for completeness and accuracy and enter data codes to prepare the questionnaires for data entry. Follow-up calls would be performed as needed to clarify inconsistencies in responses. The coded questionnaire responses would be entered into a database, with double key-entry of critical questions. Once the data are entered into a database, numerous manual and electronic QA activities would be performed and the results would be provided to engineering and economic staff for further resolution and documentation. This database would then be used to perform data analyses.

5(b) Analysis

The questionnaire objectives include identifying available and affordable CBM produced water treatment technology options in advance of a decision on whether to initiate a an effluent guidelines rulemaking. Before deciding to initiate a rulemaking EPA will use the survey responses to help answer the following questions:

- What are the observed and potential impacts of CBM produced water discharges on aquatic environments and communities, riparian zones, and other wetlands?

- What is the range of pollutant concentrations and CBM produced water flow rate for the major CBM basins?

- What are the current and potential industry treatment technologies and beneficial use options for CBM produced water?

- How effectively do these treatment technologies and beneficial use practices reduce the discharge of pollutants?

- What is the range of incremental annualized compliance costs associated with these technologies and practices? How do these costs differ between existing and new sources?

- What is the demonstrated use and economic affordability (*e.g.*, production losses, firm failures, employment impacts resulting from production losses and firm failures,

impacts on small businesses) of these technologies across the different CBM basins?

- What are the types of non-water quality environmental impacts (including energy impacts) associated with the current industry treatment technologies and beneficial use practices for CBM produced water?

The objectives of the information collection would be achieved by the statistically-designed sample survey because the resulting inferences and analyses would be as statistically unbiased and as precise as is practicable. EPA would apply sample weights derived from the statistical sample design and adjust for non-response to the data during statistical analysis. Weighting the data would allow inferences to be made about all eligible projects, including those that did not respond to the questionnaires. Another advantage is that weighted estimates would have smaller variances than unweighted estimates. EPA also would evaluate whether estimates could be improved by post-stratifying the detailed questionnaire data using the gas production data collected from the screener. EPA would use accepted statistical methods for survey statistics, such as those described in *Sampling Techniques* (Cochran, 1977) and *Survey Sampling* (Kish, 1965). EPA would use the data from the judgment sample separately in a qualitative manner.

See Part A, Section 2(b) of this Information Collection Request for a detailed discussion of the technical and economic analyses.

EPA intends to use the following contractors to assist in conducting this survey:

Sample Frame Preparation and Analysis:
PG Environmental, LLC
447B Carlisle Drive
Herndon, VA 20170

Statistical Design and Analysis:
Battelle
505 King Avenue
Columbus, OH 43201

Economic Impact Analysis:
Eastern Research Group
14555 Avion Parkway
Suite 200
Chantilly, VA 20151

Advanced Resources International, Inc.
4501 Fairfax Drive, Suite 910
Arlington, VA 22203

6. REFERENCES

Cochran, W.G. (1977). *Sampling Techniques*. New York: Wiley.

Dillman, D. (2000). *Mail and Internet Surveys: The Tailored Design Method*. New York: Wiley.

Israel, G. (1992) "Sampling Issues: Nonresponse," University of Florida, IFAS Extension Electronic Document. Available at: http://edis.ifas.ufl.edu/PD006.

Kish, L. (1965). *Survey Sampling*. New York: Wiley.

Morrison, R. (2007). "Towards the Development of Establishment Survey Questionnaire Design Guidelines at the U.S. Census Bureau." *Proceedings of the Third International Conference on Establishment Surveys (ICES-III), Montreal, Quebec, Canada, June 18-21, 2007* (pp. 662-673). Alexandria, Virginia: American Statistical Association.