

## Part B: Collections of Information Employing Statistical Methods

As has been noted in Part A, this clearance request covers a pilot study to inform a larger study of voice in mining workplaces. This Part B provides a discussion of potential statistical methods for collecting the data under a larger study, highlights areas where the pilot study will inform the design and implementation of the larger study, and discusses the metrics that DOL will use in comparing the proposed approaches that are being piloted under this study.

- 1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection method to be used. Data on the number of entities (e.g. establishments, State and local governmental units, households, or persons) in the universe and the corresponding sample are to be provided in tabular form. The tabulation must also include expected response rates for the collection as a whole. If the collection has been conducted before, provide the actual response rate achieved.**

### Population

Table B-1 provides an estimate of the population of miners in the United States for 2010 broken out by type of mining operation (coal or metal/non-metal) and type of mine (surface or underground).<sup>1</sup>

**Table B-1 – Population of Miners by Type of Mining and Type of Mine**

Category	Coal	Metal/Non-Metal
Underground	47,525	17,065
Surface	35,952	93,483
Total	0	0

The pilot study will be collecting data from coal miners only. Additionally, the two modes of implementation will be geographically focused. The in-person data collection will focus on two states: Alabama and Virginia. The phone/web survey recruited through advertising will be focused on the Charleston, WV area. Table B-2 provides a summary of the coal miner populations for these areas:

**Table B-2 – Populations of Miners in Areas for Pilot**

Mode and Location	Population of Coal Miners (NAICS 2121)
<i>In-Person Mode</i>	
Alabama	3,773
Virginia	4,575
<i>Phone/Web Recruited Mode</i>	
Charleston, WV	5,000 – 9,999 [a]

Source: Census Bureau, County Business Patterns, <http://www.census.gov/econ/cbp/>.

[a] This is based on the Charleston, WV metropolitan statistical area. Census has suppressed the actual value for that area. However, the note supplied by Census indicates that between 5,000 and 9,999 coal miners are employed in this area.

### Sample Size

The sample size for each stratum in a larger study of voice in mining will be determined based on statistical theory. DOL would select a sample size using statistical method for each stratum we want to obtain valid and reliable data for. For example, if DOL wished to obtain valid and reliable data for the four strata in Table B-1 (coal-underground, coal-surface, metal/non-metal underground, and

<sup>1</sup> The tabulation in Table B-1 excludes milling operations and contractors.

metal/non-metal-surface), then DOL would calculate an appropriate sample size for each stratum. Calculation of sample size is discussed under Question 2 below.

For the pilot study, DOL is targeting a sample of 125 miners for each mode. The selection of 125 is not based on a statistical calculation, but reflects the budget available to collect these data for the pilot. DOL balanced the number of respondents needed to perform some data analysis against the cost of collecting the data.

To assess the success of the different modes, DOL will use the **number of response attained** as an outcome measure. That is, neither mode is guaranteed to attain the targeted 125 responses. Thus, a good outcome measure for comparing the two modes will be the total responses attained.

#### Response Rate.

The response rate for this larger survey of voice in mining workplaces will be informed to some degree by the pilot study that we are seeking approval for. Specifically, the pilot will allow DOL to develop an initial estimate of the response rate for at least the in-person implementation mode. As noted in Part A, DOL is considering two potential modes: (a) implementing in-person through DOL grantee training events, and (b) recruiting miners to take the survey via phone or web. The in-person mode will involve DOL contractor staff handing out the survey at training events. Those that refuse to complete the survey will be counted as non-respondents. The other two modes we are considering, however, are less likely to allow for clean estimation of response rates. Both the web-based implementation and the phone implementation will involve recruiting respondents to take the survey. Thus, it will not be possible for DOL to determine the number that see the survey advertised and that refuse the survey.<sup>2</sup> Nevertheless, DOL can still calculate a ratio to reflect the number of responses relative to the potential population (see Table B-2 above).

To assess the two modes, DOL will have both the **response rate** from the in-person implementation and the **number of response relative to the population**<sup>3</sup> for the recruitment mode. We expect that the response rate for the in-person will be much larger than the value for the number of responses relative to the population for the recruitment approach. Thus, comparing the two values directly is not possible.

- 2. Describe the procedures for the collection, including: the statistical methodology for stratification and sample selection; the estimation procedure; the degree of accuracy needed for the purpose described in the justification; any unusual problems requiring specialized sampling procedures; and any use of periodic (less frequent than annual) data collection cycles to reduce burden.**

This section begins with a brief discussion of the methods that will be used to implement the pilot and then discusses how the pilot will influence the full-scale data collection for the topics identified in the question.

#### Procedures for Collecting the Data under the Pilot

***In-Person Mode Collected at Training Site.*** To implement the survey at training sites, DOL will have its contractor (ERG) make contact with sites in Alabama and Virginia. These two states were

---

<sup>2</sup> That is, not all of the population in Table B-2 will see the survey.

<sup>3</sup> Given that we do not know the number that will see the survey advertised, we cannot calculate a response rate and therefore do not refer to this as a response rate.

selected by MSHA due to both the size of the mining community and the ability of MSHA to secure cooperation with these states. ERG will work with the training sites to determine when training sessions will occur and will then travel to a few selected training sessions. Given the targeted value of approximately 125 responses, ERG expects to attend approximately 4-5 sessions (2-3 in each state). ERG will have the training coordinator make contact with miners taking the survey and inform them a survey will be conducted and ask trainees to arrive early. ERG will implement the survey before the training occurs since some of the subjects covered by the training are asked about in the survey.<sup>4</sup> The ERG staff person will hand out the survey with an envelope that it can be placed in when the respondent is done. Respondents will be asked to place the completed surveys in a box at the front of the room when they are done. The trainer will be asked to provide a quick encouragement to the miners to take the survey, but will leave the room prior to the survey being handed out. The ERG staff person will remain the room.

***Web and Phone Mode Recruited through Advertising.*** To implement this mode, DOL will place an ad in a newspaper in the Charleston, WV area. The ad will indicate that a survey is being implemented about coal workers' rights and will provide a web address and phone number. Workers that call the toll-free number will be connected to ERG to take the survey. The web address will direct the respondent to the online version of the instrument.

#### Stratification, Accuracy, and Sample Selection

The distinctions in Table B-1 (coal vs. metal/non-metal and underground vs. surface) are two of the major distinctions between mining operations. Other potential factors to consider in stratifying the population include:

- Union status—unionized mines may have higher levels of voice in the workplace and miners at those mines may have more access or familiarity with information on their rights.
- Size—there may be differences in the way voice works at larger mines compared to smaller ones.
- Geographic region—mines in the traditional coal mining region of the U.S. (Kentucky, West Virginia, Pennsylvania Virginia) may differ from mines in other regions. Furthermore, mines in the west of the country may differ from those in the east.

The pilot study will influence the factors used for stratification. The data that are collected under the pilot, albeit limited, should provide some indication of what factors are more important for stratifying the sample. Additionally, preliminary work performed by DOL's contractor to cognitively test the survey instrument indicated that union status may be a key factor, as well as mine size.

As noted under Question 1, DOL will determine a sample size for each stratum in order to collect valid and reliable data for that stratum. Calculating the sample size requires three pieces of information: the desired confidence interval, the variance of the estimate, and the desired level of precision:

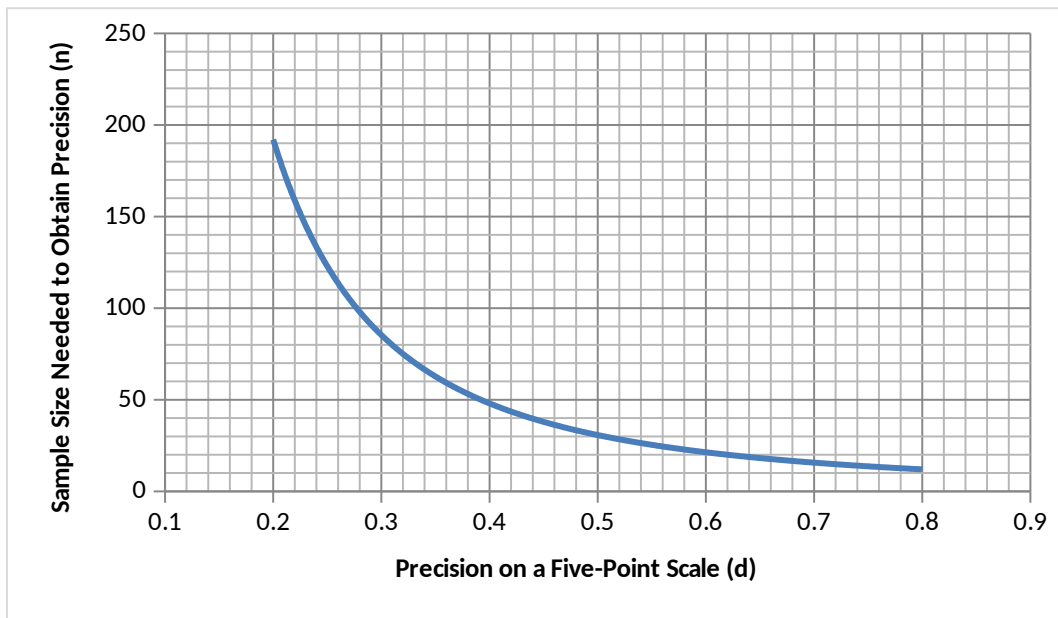
- **Confidence.** DOL will use a 95 percent confidence interval. The z-score from the standard normal distribution for a 95 percent confidence interval is 1.96, which is used in the sample size calculation.

---

<sup>4</sup> MSHA requires that all miners receive eight hours of refresher training annually. Thus, the implementation cannot take away from those eight hours

- **Variance.** The survey instrument asks a number of questions using a five point scale. McQuarrie (2006, Table 6.1)<sup>5</sup> indicates that a five-point scale will have a maximum variance of 2.0. The standard deviation (the square root of the variance) is used in the sample size calculation.
- **Precision.** The precision is the maximum amount (in units of the variable being estimated) that a researcher is willing to be away from a true estimate by taking a sample. Doubling the precision provides the length of the confidence interval being used in the study. In short, precision is a measure of accuracy with more precise estimates (smaller confidence intervals) requiring increasing larger sample sizes.

The basic formula for sample size is:  $n = \left(\frac{zS}{d}\right)^2$ , where  $n$  is the sample size being estimated,  $z$  is the standard normal score (1.96 for a two-sided 95 percent confidence interval),  $s$  is the standard deviation (square root of 2 in our case), and  $d$  is the desired precision. DOL has not selected a value for precision at this point. Figure B-1 provides a sketch of the relationship between sample size ( $n$ ) and precision ( $d$ ). For example, precision of 0.4 implies that our sample will result in an estimated mean value for variables measured on a five point scale being within 0.4 points (on the five point scale) of the “true” (population) value with 95 percent confidence. DOL has not selected a precision level yet for a larger study of voice in mining workplaces. The pilot study will assist DOL in setting a precision value by providing information on the likely distribution of responses for key questions in the survey.



**Figure B-1 – Relationship between Precision and Sample Size**

### Selection Methods

The selection method will depend on the implementation mode that is used in the larger survey. As noted in Part A, DOL is considering two potential modes: (a) in-person at DOL grantee training

<sup>5</sup> McQuarrie, Edward F., 2006. *The Market Research Toolbox*, Sage Publications Inc.

events and (b) recruiting miners to take the survey via phone or web. For the full-scale implementation, the in-person implementation will involve randomly selecting training events and then asking all miners that at that event to take the survey (i.e., a cluster design). The pilot study will provide information on how to best implement the survey on site. Additional work by ERG (not needing approval under this ICR) will provide background on the types of mines that access training events. The two other method (recruitment for phone or web) will not involve selection since the respondents are being recruited. However, recruitment will take place by posting advertisements in local media to recruit respondents. The pilot will assist us in determining the effectiveness of the recruitment process.

#### Estimation Procedure

Population parameters for a larger study will be estimated by appropriately weighting the sample responses. Sample weights would be calculated as the inverse of each unit's selection probability. The formulas used for this will depend on the selection process.

#### Unusual Problems Requiring Specialized Sampling Procedures

None are required.

#### Use of Periodic (Less Than Annual) Data Collection

DOL expects the larger survey of voice in mining to be done on an ongoing basis. The exact frequency has yet to be determined. In selecting a frequency, DOL will balance respondent burden with the need to collect data for program management. The pilot implementation that we are requesting clearance for will inform this decision. Specifically, the pilot will indicate the level of burden for respondents taking the self-administered version (at training events).

- 3. Describe the methods used to maximize response rates and to deal with nonresponse. The accuracy and reliability of the information collected must be shown to be adequate for the intended uses. For collections based on sampling, a special justification must be provided if they will not yield "reliable" data that can be generalized to the universe studied.**

The pilot study will provide information on the best approach to maximize response to the survey. The two primary reasons for doing the pilot prior to performing a larger study are (a) to determine the most effective ways to reach miners and (b) to determine the most effective approaches to obtain responses from miners that are contacted. Thus, the pilot we are seeking approval for will provide significant information on how best to ensure the data DOL collects under a larger study would be valid and reliable.

- 4. Describe any tests of procedures or methods to be undertaken. Tests are encouraged as effective means to refine collections, but if ten or more test respondents are involved OMB must give prior approval.**

In preparation for the pilot survey, DOL's contractor ERG and its subcontractor the National Opinion Research Center at the University of Chicago (NORC) performed nine cognitive tests with miners of the survey instrument. The cognitive tests focused on the understandability and relevancy of the questions. Additionally, the pilot that DOL is seeking approval for will provide additional

information on the feasibility of the implementation methods and will provide a more complete answer to this question when DOL files for approval for a larger study.

5. **Provide the name and telephone number of individuals consulted on the 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.**

DOL contracted with Eastern Research Group, Inc. (ERG) of Lexington, MA, and its consultants the National Opinion Research Center and University of Chicago, to design the survey instrument, develop the pilot plan, implement the pilot, and analyze the resulting data collected. The survey design team included the following individuals:

- Dr. Lou Nadeau, 781-674-7316; [lou.nadeau@erg.com](mailto:lou.nadeau@erg.com).
- Dr. Bernard Dugoni 773-256-6193, [dugoni-bernard@norc.uchicago.edu](mailto:dugoni-bernard@norc.uchicago.edu).