B COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

B.1 Universe and Sampling Frame

Universe: The universe for this study is the population of all establishments with one or more employees under OSHA's purview, including private sector establishments in agriculture (with more than 10 employees only) and non-agriculture industries and public sector establishments in those states with OSHA-approved safety and health programs ("state-plan states").

Due to significant sampling frame inadequacies associated with agriculture and public sector establishments, OSHA has partitioned this universe into three segments for the purposes of this study. OSHA proposes to conduct a statistical survey, based on stratified random sampling, of those establishments in non-agricultural private sector industries. These establishments are classified in 3-digit NAICS industries 113000 through 813000. OSHA proposes to conduct case study interviews with 1) private sector establishments in the agriculture sector (NAICS 111000 and 112000) and with 2) government representatives in state-plan states. For both of these groups, adequate information to develop representative sampling frames is not available, thus precluding a statistical survey approach. A description of these case studies is provided in Section B.3 above.

OSHA is also proposing to conduct up to 50 site visits to employers who may be affected by a new standard. Collection of in-depth information via site visits is an essential element of OSHA rulemakings. These site visits will be primarily to small employers, and the Agency will seek employers across a wide range of industries. The primary goal of the information collected is not quantitative or statistical in nature; thus while the site visits may be described in some sense as a "survey," there is no accompanying statistical design to this collection of information. Site visits are described further in section B.8 below.

- **Private Sector, Non-Agricultural Establishments**

Because OSHA seeks accurate industry-specific estimates of current prevalence of baseline safety and health practices, we divided the industries within this universe into 20 industry groups Estimates of the prevalence of baseline safety and health management practices and other statistics will be generated for each industry group that meet the precision targets as described below. Within each industry group, stratified sampling based on establishment size will be conducted to improve the precision of the estimates. Estimates of the population establishment counts for each industry group are shown in Table B-1. Table B-2 shows the complete list of three-digit NAICS industries included in each industry group.

Sampling Frame: The sample frame for the study of these industries will be based on the Dun and Bradstreet (D&B) business establishment database. OSHA has used this frame in the past with good results, and it has proved to be efficient for establishment surveys. OSHA has not, at this time, acquired the D&B sampling frame data. The universe statistics, however, are believed to be similar to those derived from County Business Patterns data and shown in Table B-1. The

establishment-size strata are defined in terms of the total number of workers employed at each establishment. The boundaries for these strata are as follows:

Size class 1: 1-19 employees
Size class 2: 20-99 employees
Size class 3: 100-499 employees
Size class 4: 500+ employees

B.2 Sample Design for Statistical Survey

OSHA's objectives in designing the survey of private sector non-agricultural are as follows: for each industry group, OSHA wishes to develop estimates of the prevalence of baseline safety and health management practices in place among establishments within each industry group with 95 percent confidence intervals defined to equal the point estimator plus or minus a confidence-interval half-width 5.01 percentage points or less. If design effects are less than or equal to one, this objective can be achieved with a sample size of 400 or more for each group. The overall target sample for all industry groups is 8,000, as shown in Table B-1. The derivation of this relationship between the size of the sample and resultant precision of the estimates from the survey is shown in the appendix to this document.

Stratification: Within each industry group, the sample will be stratified by establishment employment size class. OSHA has stratified industry groups into four establishment size classes: 1-19, 20-99, 100-499 and 500 and more employees. Selecting independent samples within each group's stratum may potentially improve the precision of the estimates by reducing the variance of the estimates.

Table B-1
Safety and Health Management Practices Survey Sample Design: Non-Agricultural Private Sector

Group No.	Industry Group Name	Total Establishments [a]	Target Sample	Solicited Sample [b]
1	Utilities (NAICS 221)	16,674	400	691
2	Construction of Buildings (NAICS 236)	244,862	400	711
3	Heavy and Civil Engineering Construction (NAICS 237)	51,421	400	711
4	Specialty Trade Contractors (NAICS 238)	515,169	400	711
5	Primary Products Manufacturing (NAICS 311, 312, 313, 321, 322, 324, 325, 327, 331)	90,066	400	711
6	Final Products Manufacturing (NAICS 313 314, 315, 316, 323, 326, 332, 333, 334, 335, 336, 337, 339)	241,289	400	711
7	Wholesale Trade (NAICS 42)	434,464	400	711
8	Vehicle Sales and Service (NAICS 441, 447, 8111)	242,864	400	711
9	All Other Retail Trade (NAICS 44-45 except 441 and 447)	880,765	400	711
10	Transportation and Warehousing (NAICS 48-49)	219,806	400	711
11	Information (NAICS 51)	143,779	400	711
12	Professional and Financial Services (NAICS 52, 53, 54, 55, 561, 624, 813)	2,323,561	400	711
13	Educational Services (NAICS 61)	86,896	400	711
14	Ambulatory Health Care (NAICS 621)	547,183	400	711
15	Hospitals (NAICS 622)	7,352	400	711
16	Nursing Homes (NAICS 623)	75,606	400	711
17	Accommodations (NAICS 721)	63,903	400	711
18	Food and Drinking Places and Related Establishments (NAICS 713, 722)	642,236	400	711
19	Other Services and Non-manufacturing ((NAICS 71 except 713, 8112, 8113, 8114, 812)	51,572	400	711
20	Other Industries (NAICS 113, 114, 115, 211, 562)	24,339	400	711

Table B-1
Safety and Health Management Practices Survey Sample Design: Non-Agricultural Private Sector

Group No.	Industry Group Name	Total Establishments [a]	Target Sample	Solicited Sample [b]
	All Groups	**6,903,807**	**8,000**	**14,202**

[a] U.S. Census Bureau, County Business Patterns, 2007.
[b] Assumes 25 percent invalid sample frame listings (invalid address, out-of business, etc.) and 25 percent completion rate.
[c] The solicited sample in each cell is set to not exceed the total number of establishments in that cell.

Table B-2
3-Digit NAICS Non-Agricultural Industries Covered by the Safety and Health Management Practices Survey

Group No.	Industry Group Name	NAICS	Industry Title
1	Utilities (NAICS 221)		
		221000	Utilities
2	Construction of Buildings (NAICS 236)		
		236000	Construction of Buildings
3	Heavy and Civil Engineering Construction (NAICS 237)		
		237000	Heavy and Civil Engineering Construction
4	Specialty Trade Contractors (NAICS 238)		
		238000	Specialty Trade Contractors
5	Primary Products Manufacturing (NAICS 311, 312, 313, 321, 322, 324, 325, 327, 331)		
		311000	Food Manufacturing
		312000	Beverage and Tobacco Product Manufacturing
		321000	Wood Product Manufacturing
		322000	Paper Manufacturing
		324000	Petroleum and Coal Products Manufacturing
		325000	Chemical Manufacturing
		327000	Nonmetallic Mineral Product Manufacturing
		331000	Primary Metal Manufacturing
6	Final Products Manufacturing (NAICS 313 314, 315, 316, 323, 326, 332, 333, 334, 335, 336, 337, 339)		
		313000	Textile Mills
		314000	Textile Product Mills
		315000	Apparel Manufacturing
		316000	Leather and Allied Product Manufacturing
		323000	Printing and Related Support Activities
		326000	Plastics and Rubber Products Manufacturing
		332000	Fabricated Metal Product Manufacturing
		333000	Machinery Manufacturing
		334000	Computer and Electronic Product Manufacturing
		335000	Electrical Equipment, Appliance, and Component Manufacturing
		336000	Transportation Equipment Manufacturing
		337000	Furniture and Related Product Manufacturing
		339000	Miscellaneous Manufacturing
		423000	Merchant Wholesalers, Durable Goods
7	Wholesale Trade (NAICS 42)		
		424000	Merchant Wholesalers, Nondurable Goods

Table B-2
3-Digit NAICS Non-Agricultural Industries Covered by the Safety and Health Management Practices Survey

Group No.	Industry Group Name	NAICS	Industry Title
		425000	Wholesale Electronic Markets and Agents and Brokers
8	Vehicle Sales and Service (NAICS 441, 447, 8111)		
		441000	Motor Vehicle and Parts Dealers
		447000	Gasoline Stations
		811100	Automotive Repair and Maintenance
9	All Other Retail Trade (NAICS 44-45 except 441 and 447)		
		442000	Furniture and Home Furnishings Stores
		443000	Electronics and Appliance Stores
		444000	Building Material and Garden Equipment and Supplies
		445000	Food and Beverage Stores
		446000	Health and Personal Care Stores
		448000	Clothing and Clothing Accessories Stores
		451000	Sporting Goods, Hobby, Book, and Music Stores
		452000	General Merchandise Stores
		453000	Miscellaneous Store Retailers
		454000	Non-store Retailers
10	Transportation and Warehousing (NAICS 48-49)		
		481000	Air Transportation
		483000	Water Transportation
		484000	Truck Transportation
		485000	Transit and Ground Passenger Transportation
		486000	Pipeline Transportation
		487000	Scenic and Sightseeing Transportation
		488000	Support Activities for Transportation
		492000	Couriers and Messengers
		493000	Warehousing and Storage
11	Information (NAICS 51)		
		511000	Publishing Industries (except Internet)
		512000	Motion Picture and Sound Recording Industries
		515000	Broadcasting (except Internet)
		516000	Internet Publishing and Broadcasting
		517000	Telecommunications
		518000	Internet Service Providers, Web Search Portals, and Data Processing Services
		519000	Other Information Services
12	Professional and Financial Services (NAICS 52, 53, 54, 55, 561, 624, 813)		
		521000	Monetary Authorities - Central Bank
		522000	Credit Intermediation and Related Activities
		523000	Securities, Commodity Contracts, and Other Financial Investments and Related Activities
		524000	Insurance Carriers and Related Activities
		525000	Funds, Trusts, and Other Financial Vehicles
		531000	Real Estate
		532000	Rental and Leasing Services
		533000	Lessors of Nonfinancial Intangible Assets
		541000	Professional, Scientific, and Technical Services
		551000	Management of Companies and Enterprises
		561000	Administrative and Support Services

Table B-2
3-Digit NAICS Non-Agricultural Industries Covered by the Safety and Health Management Practices Survey

Group No.	Industry Group Name	NAICS	Industry Title
		624000	Social Assistance
		813000	Religious, Grantmaking, Civic, Professional, and Similar Organizations
13	Educational Services (NAICS 61)		
		611000	Educational Services
14	Ambulatory Health Care (NAICS 621)		
		621000	Ambulatory Health Care Services
15	Hospitals (NAICS 622)		
		622000	Hospitals
16	Nursing Homes (NAICS 623)		
		623000	Nursing and Residential Care Facilities
17	Accommodations (NAICS 721)		
		721000	Accommodation
18	Food and Drinking Places and Related Establishments (NAICS 713, 722)		
		713000	Amusement, Gambling, and Recreation Industries
		722000	Food Services and Drinking Places
19	Other Services and Non-manufacturing (NAICS 71 except 713, 8112, 8113, 8114, 812)		
		711000	Performing Arts, Spectator Sports, and Related Industries
		712000	Museums, Historical Sites, and Similar Institutions
		811200	Electronic and Precision Equipment Repair and Maintenance
		811300	Commercial and Industrial Machinery and Equipment (except Automotive and Electronic) Repair and Maintenance
		811400	Personal and Household Goods Repair and Maintenance
		812000	Personal and Laundry Services
20	Other Industries (NAICS 113, 114, 115, 211, 562)		
		113000	Forestry and Logging
		114000	Fishing, Hunting and Trapping
		115000	Support Activities for Agriculture and Forestry
		211000	Oil and Gas Extraction
		562000	Waste Management and Remediation Services

Sample Allocation: OSHA proposes allocating sample sizes among size strata within a given industry group based on the employee share of each stratum. The sample size for each stratum within an industry group has been developed by allocating the total group sample among the size strata in proportion to the stratum's share of the industry group's total number of employees in the universe. These sample targets are also found in Table B-1.

The overall number of solicitations required to achieve the sample targets shown in Table B-1 will depend on the incidence of invalid listings in the drawn sample and on the rate of nonresponse and other factors influencing completion rates among respondents. OSHA's previous experience with the D&B establishment database indicates that approximately 25 percent of the listings might be expected to be invalid. The Agency's previous survey efforts also indicate that a nonresponse rate of no more than 25 percent is likely for a survey such as this one. Table B-3 shows response rates obtained from recent OMB-approved surveys conducted OSHA's survey contractor, Eastern Research Group, Inc. (ERG).

Table B-3
Response Rates From Recent Surveys

Survey Title	Sponsoring Agency	Target Respondents	Response Rate
Single-Use Medical Device (SUD) Reuse and Reprocessing in Hospitals	FDA	Hospitals	79.4%
Personal Protective Equipment (PPE)	OSHA	Manufacturing Facilities	75%
Automated External Defibrillator Use in the Workplace	OSHA	General industry, except for construction	55% overall
			73% manufacturing

Conducted by Eastern Research Group, Inc. (ERG) under contract to the sponsoring Agency.

Eastern Research Group, Survey on the Reuse and Reprocessing of Single-Use Devices (SUDs) in U.S. Hospitals - Final Report. Prepared for Office of Policy, Planning, and Legislation Food and Drug Administration. Contract No. 223-98-8002, Task No. 19. August 5, 2002.

Eastern Research Group, PPE Cost Survey: Final Report. (Exhibit 14, OSHA Docket S-042: Costs of Personal Protective Equipment). Prepared for OSHA under Task Order 3, Base Year, DOL Contract No. J-9-F-9-0010. June 23, 1999.

Eastern Research Group, Automated External Defibrillator Use in Occupational Settings: Final Report. Prepared for OSHA under Task Order 30,Option-Year 2, Contract No. J-9-F-2-0030. September 30, 2006.

Combined, these figures suggest that the overall completion rate (percent of completed surveys among all respondents potentially solicited) of 56 percent or higher should be expected (100% x (100%-25%) x (100%-25%) = 56.25%). Thus, OSHA estimates that a solicited sample of 14,222 potential respondents is necessary to achieve the overall sample target of 8,000 for the statistical survey portion of this study.

Supplementary Sample

OSHA's estimates of the size of the initial solicited sample needed to achieve the stated accuracy targets are based on the best available information about the likely magnitude of sampling frame deficiencies and the efficacy of the proposed survey methods in obtaining the desired response rates. If, however, these response rates are not obtained for some industry groups or the magnitude of the sampling frame deficiencies is larger than expected, OSHA will consider, resources and schedule permitting, using a supplementary sample to increase the number of responses for those groups where the target sample sizes have not been obtained

At the time the initial sample is obtained, OSHA will draw an additional supplementary sample equal to 25 percent of the initial sample using the same sampling design and technique as used for the primary sample. Should the survey results show that further sampling is indicated, OSHA will use the same survey methodology for respondents solicited from the supplementary sample as used for those in initial sample.

B.3 Case Studies

- **Agricultural Establishments**

OSHA proposes to study the implementation of safety and health management practices among farming operations with more than 10 regular employees engaged in crop or animal production (NAICS 111000 and 112000, respectively). Due to the lack of a complete sampling frame for the establishments in this universe, OSHA proposes to conduct case study interviews of establishments in these two broad agricultural industries based on the Dun and Bradstreet (D&B) database of establishment listings.

Data from the 2007 Agricultural Census compiled by the U.S. Department of Agricultural (USDA) show that there are currently 1,051,889 and 1,152,903 farms in these industry classifications, respectively. Only a relatively small share of these farms, however, employ more than 10 employees on a regular basis. Based on the USDA data, OSHA estimates that only approximately 30,000 farms in NAICS 111000 and 13,000 farms in NAICS 112000 employ more than 10 employees.[1]

The D&B business establishment database includes listings for farming operations classified in these NAICS divisions and provide self-reported information on employee size. A comparison of the D&B coverage with the USDA estimates shows, however, that only a fraction (less than 50 percent) of the farms in the study universe are included in the D&B database. Further, a review of D&B listings suggests that the employment totals for some farms are derived from non-farm activities (e.g., farm stands, feed mill operations, etc.). Thus, the D&B establishment database is not representative of the study universe and, therefore, is unsuitable for use as a sampling frame for a statistical survey.

For this reason, OSHA proposes to conduct a case study review of relevant safety and health practices among establishments in the agriculture universe. OSHA plans to conduct 60 telephone interviews with farms selected at random from the D&B establishment listings. These interviews will use a telephone version of the same questionnaire to be used in the non-agriculture statistical survey. The case study interviews will be divided equally among small, mid-sized, and large farm operations in each of the two agricultural NAICS industries as shown in Table B-4.

Table B-4
Case Study Design for Agricultural Establishments

Industry	Number of Case Study Interviews			
	Farms with 11-19 employees	Farms with 20-99 employees	Farms with 100 or more employees	Total
Crop Production (NAICS 111000)	10	10	10	30
Animal Production (NAICS 112000)	10	10	10	30
Totals	**20**	**20**	**20**	**60**

- **State and Local Governments**

[1] OSHA has not included the use of migrant or other contract labor in the estimates of farms by employment size.

Public employees in state-plan states are also under the purview of OSHA safety and health rules. As such, OSHA proposes to study the use of safety and health management practices and other safety and health practices among agencies and departments operated by state and local governments in state-plan states. Public employees in the utility (e.g., municipal utility operations), education (e.g., public colleges), health care (e.g., publicly owned hospitals), and retail sectors (e.g., state-owned liquor stores) are covered by entities included in the D&B sampling frame for private sector, non-agricultural establishments, and, as such, will be covered by the statistical survey of establishments in that universe.

A range of governmental agencies and departments at the state and local level employ the remaining public employees. These include divisions of state, county, city, town, township, and special jurisdiction (e.g., unified school districts) governments. Because of the complexity of patterns in responsibility for the establishment and enforcement of safety and health practices in the public sector, OSHA proposes to interview knowledgeable state-level safety and health officials in each of the state-plan states to assess the prevalence of safety and health management practices in place among state and local governments.

OSHA proposes to conduct in-depth telephone interviews with one or more knowledgeable safety and health officials in each of the 25 states that have approved safety and health programs for public sector employees. Of these states, 20 have programs that cover both private and public sector activities, while four states have approved programs that cover public sector employees only. These states are shown in Table B-5 below. Case study interviews will be based on the private sector questionnaire, but will be less structured and more open-ended, given the ill-defined nature of responsibility for safety and health practices among governmental agencies and departments. OSHA will not consider safety and health practices among fire prevention and law enforcement personnel, given the extensive job-specific training given to these employees.

Table B-5
States with OSHA-Approved Safety and Health Programs

Alaska
Arizona
California
Connecticut (public sector employment only)
Hawaii
Illinois (public sector employment only)
Indiana
Iowa
Kentucky
Maryland
Michigan
Minnesota
Nevada
New Jersey (public sector employment only)
New Mexico
New York (public sector employment only)
North Carolina
Oregon

Table B-5
States with OSHA-Approved Safety and Health Programs

South Carolina
Tennessee
Utah
Vermont
Virginia
Washington
Wyoming

B.4 Information Collection – Statistical Survey

Data Collection: OSHA proposes to utilize a mixed-mode Internet and mail method to administer this survey and to collect the data. The data collection phase of the study will be initiated with an introductory letter from OSHA to the target respondents informing them of the study. Respondents will be invited to complete the survey electronically using a URL and password provided in the letter to access the on-line survey instrument. A copy of the instrument will also be provided for those respondents wishing to complete the questionnaire by hand. After two weeks, respondents will be mailed reminder postcards. After four weeks, those who have not responded will be contacted by telephone and urged to complete the survey. Finally, after six weeks, a final reminder postcard will be mailed to the remaining nonrespondents.

Sample Selection Methodology: The solicited sample of establishments within each stratum will be selected using sampling procedure described above. To facilitate the sampling, each establishment will be assigned a random index number, using a random number generator. The establishments in each stratum will then be arranged in ascending order according to their random index number. If S_{ij} is the size of the solicited sample in the ith group and jth class size, then those S_{ij} establishments with the smallest index numbers will be selected and included in the sample.[2]

Weighting and Estimation Procedures:

Weighting: Because OSHA is anticipating variable nonresponse rates by stratum, weighting procedures are necessary to produce overall estimates. The weights will be the inverse of the selection probabilities of the establishments. Each stratum will have a different weight because of the different actual nonresponse rates. The sampling weights are defined as follows:

(1) $$W_{ij} = \frac{U_{ij}}{S_{ij}}$$

where:

W_{ij}= the sampling weight for the ith industry group and jth establishment size class

[2] Alternatively, one could select the same number of establishments with the largest index numbers.

U_{ij} = the number of elements in the subset of the universe defined by the ith industry group and the jth size class
S_{ij} = the size of the solicited sample in the ith industry group and jth class size.

Adjustment Factor for Nonresponse: The adjustment factor for nonresponse is calculated by dividing the solicited sample size in each stratum by the actual number of responses from the corresponding stratum. This is written as:

(2) $$N_{ij} = \frac{S_{ij}}{R_{ij}}$$

where:

S_{ij} = the size of the solicited sample in the ith industry group and jth class size
R_{ij} = the actual (responded) sample size in the ith industry group and jth class size
N_{ij} = the nonresponse factor

Benchmarking

To facilitate comparison with published establishment estimates, OSHA will benchmark the survey industry group and stratum estimates to the totals reported in the 2007 County Business Patterns (CBP) database.[3] Benchmarking weights are calculated for each industry group and size stratum so that the benchmarked in-scope survey population estimates of the number of establishments equal the CBP-reported levels for that group and stratum. The benchmarking weights have the following property:

[3] $$B_{ij} = CBP_{ij} / U_{ij}$$

where B_{ij} is the benchmarking weight for industry group i and size stratum j, CBP_{ij} is the population of group i and stratum j as reported by County Business Patterns, and U_{ij} is the population of group i, stratum j in the sampling frame. From the previous discussion,

[4] $$U_{ij} = R_{ij} W_{ij} N_{ij}$$

where R_{ij} is the actual number of respondents in group i and size stratum j, and W_{ij} and N_{ij} are the sampling weight and nonresponse adjustment factor, respectively. Thus,

[5] $$B_{ij} = CBP_{ij} / (R_{ij} W_{ij} N_{ij})$$

[3] U.S. Census Bureau, *County Business Patterns, 2007.*

Point Estimation Procedures: The great majority of questions in this survey seek di- or polychotomous categorical responses. The estimation of the population statistics for these questions is discussed below. Three questions in the survey instrument, however, solicit verbatim numerical responses. An initial question asks about the number of employees at the respondent's facility. This information will be used, however, only to verify the employment size information for the respondent to be provided by the sampling frame database. Two other questions seek information about the annual average number of training hours and about the approximate number of hazards investigated in the previous 12 months, respectively. In neither case does OSHA intend to estimate population means from the survey data. Instead, verbatim responses to these questions will be used to categorize respondents in each industry estimation group according to the relative intensity of their activity in providing training or engaging in hazard investigation and mitigation. OSHA will then compare these ordinal measures with the attributes of respondents' safety and health systems and programs.

For the remaining questions, the estimator, $\hat{Y}_i$, for the industry group totals will take the form:

(63) $$\hat{Y}_i = \sum_j \sum_k W_{ij} N_{ij} Y_{ijk}$$

where

Y_{ijk} = response of the kth establishment in the ith industry group and the jth size class,
W_{ij} = sampling weight of the ith industry group and jth size class, and
N_{ij} = nonresponse adjustment of the ith industry group and jth size class.

Estimates of the mean for each group will be derived as follows:

(74) $$\bar{Y}_i = \frac{\sum_j \sum_k W_{ij} N_{ij} Y_{ijk}}{\sum_j W_{ij} N_{ij}}.$$

The mean for industry group i and stratum j is calculated as $\bar{Y}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk}$, where n_{ij} is the number of respondents in industry group i and stratum j. Thus, the equation for $\bar{Y}_i$ can be rewritten as

(85) $$\bar{Y}_i = \sum_j q_{ij} \bar{Y}_{ij}$$, where $$q_{ij} = \frac{n_{ij} W_{ij} N_{ij}}{\sum_j W_{ij} N_{ij}}.$$

Variance Estimation Methods: The point estimators presented above are, in general, nonlinear functions of the underlying random variables (e.g., the observations Y and the nonresponse

adjustment factors N). To account for these nonlinear effects, the following method based on balanced repeated replication will be used to construct the variance estimators used in this study.[4]

1. Using an appropriate random number generator, the set of sample units (regardless of response/nonresponse status) in each stratum will be partitioned into eight groups, referred to below as pseudo-primary sample units (pseudo PSUs). Each group will contain the same number of sample units, plus or minus one. Thus, within each of the eight original strata (four industry groups intersected with two size classes), step (1) will produce eight pseudo-PSUs, for a total of sixty-four pseudo-PSUs.

2. Within each stratum, the eight pseudo-PSUs constructed in step (1) will be assigned to four pseudo-strata, with two pseudo-PSUs per pseudo-stratum. Thus, steps (1) and (2) will result in a total of 32 pseudo-strata, with two pseudo-PSUs per pseudo-stratum.

3. In keeping with standard approaches, this work will use $k = 36$ half-sample replicates (due to the fact that 36 is the first multiple of four that is greater than 32, the number of pseudo-strata in this study). Inclusion and exclusion of a given pseudo-PSU within a given half-sample replicate will be based on the Hadamard matrix of order 36 provided in Appendix A.10 of Wolter (1985). Repeating notation and definitions from Wolter (1985, Chapter 3), for half-sample replicate α and for pseudo-PSU u within pseudo-stratum h, define the selection indicators $\delta_{h1\alpha}$ = {1 if pseudo-PSU (h,1) is selected for the α-th half-sample; 0 otherwise} and $\delta_{h2\alpha}$ = 1 - $\delta_{h1\alpha}$.

4. For a given point estimator $\hat{\theta}$ computed from the full sample (or from the appropriate identified subpopulation units contained in the full sample), for $k = 36$ and for each α = 1, …, k, define $\hat{\theta}_{\alpha}$ be the corresponding point estimator based on the α-th half-sample. This is equivalent to using the previously developed expression for the $\hat{\theta}$, but with the previous weights w_{ij} replaced by $2\,\delta_{hu\alpha}\,w_{ij}$ for all sample establishments (i, j) contained in pseudo-PSU u within pseudo-stratum h; and with the nonresponse adjustment factors N_{ij} now also based only on data from pseudo-PSU u within pseudo-stratum h that contains the specified sample establishment (i, j).

5. Following expression (3.4.1) of Wolter (1985), compute the variance estimator,

(96) $$v_k(\hat{\theta}) = \sum_{\alpha=1}^{k} (\hat{\theta}_{\alpha} - \hat{\theta})^2 / k$$

and the associated standard errors $se(\hat{\theta}) = \{v_k(\hat{\theta})\}^{1/2}$.

6. In addition, an approximate 95% confidence interval for the underlying parameter θ is $\hat{\theta} \pm (2.04) se(\hat{\theta})$ where the multiplier 2.04 represents the approximate 0.975 quantile of a t distribution on 31 degrees of freedom.

[4] For general background on balanced repeated replication, see, for example, K.M. Wolter, *Introduction to Variance Estimation* (New York, NY, 1985), Chapter 3.

Analysis of Possible Nonresponse Bias: OSHA will conduct an analysis of potential nonresponse bias in the survey estimates. Using standard procedures, OSHA will construct a logistic model of the propensity for survey completion based on both additional sampling frame variables as well as external parameters that can be unambiguously associated with the survey respondents.[5] The Agency has identified four exogenous variables that could provide information in assessing non-response bias: 1) employer size (either number of employees or revenues); 2) group or industry identification; 3) the state in which the establishment is located; and 4) whether the potential respondent is part of a multi-establishment company. Information about these variables are known for each respondent receiving a survey request. Industry identification (NAICS) may also serve as a proxy for occupational risk. There may also be variations within groups by sub-sector industry identification. Finally, twelve states require employers to have safety and health programs already and four others require employers to have safety and health committees. These existing regulatory requirements may also affect the liklihood of response.

The logistic function is given as,

$$(10) \quad P = \frac{e^{x\beta}}{(1+e^{x\beta})} = \frac{1}{(1+e^{-x\beta})}$$

where P in this context is the probability of a respondent completing the survey and x and β are the vectors of explanatory variables and their respective coefficients.

Given (10), the probability of survey nonresponse can be written as,

$$(11) \quad 1 - P = 1 - \frac{1}{(1+e^{-x\beta})} = \frac{1}{(1+e^{x\beta})}.$$

The odds of a positive survey response are, therefore,

$$(12) \quad \left(\frac{P}{1-P}\right) = e^{x\beta}$$

Taking the natural log of both sides, equation (12) becomes,

$$(13) \quad \ln\left(\frac{P}{1-P}\right) = x\beta.$$

[5] See Sharon L. Lohr, *Sampling Design and Analysis* (Pacific Grove, CA, 1999), Chapter 8, for a discussion of methods dealing with potential nonresponse bias. For a recent analysis of nonresponse in a large government survey using logistic regression techniques, see Katherine, G. Abraham, et. al., "Nonresponse in the American Time Use Survey," *Public Opinion Quarterly*, Vol 70, No. 5, Special Edition 2006, pp 676-703.

For the purposes of nonresponse analysis for this survey, the logit model (130) to be estimated can be specified as,

$$(141) \quad \ln\left(\frac{P_{ik}}{1 - P_{ik}}\right) = a_i + b_{1i}x_{1ik} + b_{2i}x_{2ik} + ... + c_{1i}y_{1ik} + c_{2i}y_{2ik} + ... + \varepsilon_{ki}$$

where P_{ki} is probability of respondent k in industry estimation group i responding to the survey; x_{ik} and y_{ik} are vectors of dichotomous variables for industry group i from the sampling frame and external databases, respectively, and corresponding to respondent k; a_i is the intercept; b_i and c_i are the associated coefficient vectors; and ε_{ki} is the error term.

Using maximum likelihood methods, OSHA will estimate the above logistic relationship and use the coefficients for those variables found to be significantly related to the probability of responding to the survey to estimate the predicted probability of responding for each respondent. Given this form of the regression model, the predicted probability of a positive survey response for a given potential respondent k can be calculated as:

$$(152) \quad \hat{P}_{ik} = \frac{1}{(1 + e^{-(\hat{a}_i + \hat{b}_{1i}x_{1ik} + \hat{b}_{2i}x_{2ik} + ... + \hat{c}_{1i}y_{1ik} + \hat{c}_{2i}y_{2ik} + ...)})}.$$

OSHA will use these probability estimates to recalculate the nonresponse adjusted weights to be applied to each respondent's responses. These weights can be expressed as follows:

$$(163) \quad \hat{N}_{ijk} = \alpha_{ij}\left(\frac{1}{\hat{P}_{ik}}\right)$$

where $\hat{N}_{ijk}$ is the adjusted weight for respondent k in industry group i and stratum j, $\hat{P}_{ik}$ is the estimated response probability derived from the logistic regression, and α_{ij} is the normalization factor. These factors are calculated to normalize the estimated response probabilities so that the set of non-response adjusted weights have the following property for each stratum within a given estimation group:

$$(174) \quad U_{ij} = \sum_k W_{ij}\hat{N}_{ijk} = \alpha_{ij}\sum_k \frac{W_{ij}}{\hat{P}_{ik}}$$

where k is summed over all respondents k in stratum j.

B.5 Methods To Maximize Response Rates

All reasonable steps will be taken to maximize response rates on this voluntary survey. Initially each target respondent will receive a letter from OSHA informing them of the study and requesting their participation. OSHA believes this step has had the effect of increasing response rates in previous surveys. Also, OMB's recently-issued draft survey design guidance

recommends sending advance letters to potential respondents as a means of improving response rates.[6] OSHA plans to include a copy of the survey instrument with a self-addressed stamped return envelope together with the introductory letter. This will allow respondents to gain a clear idea of the scope of the survey and the nature of the information request. Providing a copy of the questionnaire will provide respondents the option of completing the questionnaire by hand if they so desire.

The initial letter will provide the URL and password for accessing the questionnaire electronically. The web-based survey will allow respondents to complete the survey at their convenience and the survey software will permit respondents to "save" partially completed questionnaires and return later for completion.

As noted above, reminder postcards will be sent to potential respondents two weeks after the mailing of the initial survey invitation letter. After four weeks, telephone contact with each nonrespondent will be attempted to encourage completion of the questionnaire either electronically or by mail. Callers will attempt up to six calls to reach the potential respondents. A final reminder postcard will be mailed after six weeks to the remaining nonrespondents. OSHA's survey contractor will also implement a help line to answer both telephone and e-mail questions about the survey posed by respondents.

B.6 Tests Of Procedures

OSHA plans to administer a pretest of the survey instrument to two rounds of ten establishments each. These establishments will be selected based on a random sample by aggregated industry group and employment size of the target survey population. The following stratified simple random sample design will be used for each round of pilot testing.

Industry Group	Size Class	Pilot Test Sample Size
Construction	1-99 100+	1 1
Manufacturing and Utilities	1-99 100+	1 1
Wholesale and Retail Trade	1-99 100+	1 1
Finance, Insurance and real estate	1-99 100+	1 1
Service and related industries	1-99 100+	1 1

After the completion of the first round, the pretest results will be evaluated to ensure that the survey questions are clearly understandable, terms are well defined, and the survey format is logical. Any indicated changes to the instrument will then be tested with the second round of pretest interviews.

~~OSHA plans to administer a pretest of the survey instrument to nine establishments. These establishments will be selected based on a random sample by aggregated industry group and~~

[6]Office of Management and Budget, *Draft Guidance on Designing Surveys for Information Collection*, December 6, 2004, pp 52-53.

employment size of the target survey population. The following stratified simple random sample design will be used for pilot testing.

Industry Group	Size Class	Pilot Test Sample Size
Construction	All sizes	1
Manufacturing and Utilities	1-99 100+	1 1
Wholesale and Retail Trade	1-99 100+	1 1
Finance, Insurance and real estate	1-99 100+	1 1
Service and related industries	1-99 100+	1 1

Pretest results will be evaluated to ensure that the survey questions are clearly understandable, terms are well defined, and the survey format is logical.

B.7 Expert Review

The statistical aspects of the survey design have been reviewed by:

Chester Fenton, Ph.D. (ABD)
Eastern Research Group, Inc.
110 Hartwell Ave.
Lexington, MA 02421
781-674-7326

Aylin Sertkaya, Ph.D.
Eastern Research Group, Inc.
110 Hartwell Ave.
Lexington, MA 02421
781-674-7227

The data will be collected and processed by:

Chester Fenton, Ph.D. (ABD)
Eastern Research Group, Inc.
110 Hartwell Ave.
Lexington, MA 02141
781-674-7326

The data will be analyzed by:

Chester Fenton, Ph.D. (ABD)
Eastern Research Group, Inc.
110 Hartwell Ave.
Lexington, MA 02141

781-674-7326

Paul Bolon
Occupational Safety and Health Administration
200 Constitution Ave. N.W.
Washington, D.C. 20210

B.8. Site visits

In addition to the statistical survey (Baseline Safety and Health Practices) described above-- which also includes "case studies" throughout two industry sectors that could not be adequately sampled by the survey methodology--the Agency is proposing to conduct as many as 50 site visits to employers. These employers could potentially be affected by a new standard that could require a management program or system to address workplace hazards. Site visits would collect information on current employer practices (much like the information collected in the "case studies" and survey questionnaire itself), but also solicit information from employers on how they would comply with such a regulation, what time or costs would be required to do so, and what regulatory alternatives might be. Site visit reports capture much richer, or in-depth, detail about employer conditions and ideas than the survey instrument and also reflect variations of employer size and industry sector. These site visits would be conducted either by OSHA personnel or a contractor under the agency's direction. Most of the site visits would be target small employers and require one to two hours of an owner or proprietor's time. Site visits usually are conducted by a visit to the employer's place of business, but could be conducted over the phone. Site visits provide the Agency with a quality of information about standards that cannot, in general, be captured by a checklist or simple quantification. These site visits are a one-time effort, which could take up a year to complete, and are not a recurring collection of information. Employers who participate could opt to remain anonymous, but reports of the site visits would be added to either the rulemaking docket. Site visits to employers are conducted or managed by OSHA's Office of Regulatory Analysis and are separate from enforcement functions. Employer participation is purely voluntary. Typically employers are provided an opportunity to examine site visit reports and correct any errors. Site visits of this nature accompany virtually every Agency rulemaking effort. Site visits would be targeted to industry sectors where the Agency expects that management safety and health programs are less likely to be found. The Agency believes that programs are more frequently found in industry sectors with high hazards and among larger employers. Site visits would therefore be targeted to small employers in all sectors and in sectors with relatively lower injury and illness rates.

APPENDIX
Sample Size and Precision

The following discussion shows the mathematical relationships underlying the association between OSHA's precision objectives detailed in Section B.2 and the target sample sizes shown in Table B-1. The basic relationships relating sample size and statistical confidence intervals can be derived as follows: Let $Y = \{ y_1, y_2, y_3, ..., y_n \}$ be a simple random sample selected with replacement from the population with mean μ and variance σ^2. Then, the sample mean, $\bar{y}$, is expressed as

(1) $$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

where y_i is the value of the *i*th element of the sample and *n* is the size of the sample. Further, the standard deviation of the sample, s, is defined as

(2) $$s = \sqrt{\frac{1}{n - 1} \sum_{i=1}^{n} (y_i - \bar{y})^2}$$

And, it can be shown that $\bar{y}$ and s^2 are unbiased estimators of μ and σ^2, respectively. If a series of samples is taken from the population, it can also be shown that the resulting probability distribution of the sample means will have an expected value of μ and a variance s^2, such that

(3) $$s^2 = \frac{\sigma^2}{n}$$

With the Central Limit Theorem, it can be shown that as *n* increases, the sample mean, $\bar{y}$, will have a distribution that approaches a normal distribution with mean μ and a standard deviation $\frac{\sigma}{\sqrt{n}}$. Alternatively, the relationship

(4) $$z = \frac{\bar{y} - \mu}{s}$$

is approximately normal, with mean zero and standard deviation of one for a sufficiently large sample. Further, equation (4) can be rewritten using (3) as

(5) $$z = \frac{(\bar{y} - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

Values from the normal distribution will define the range or confidence interval around the mean within which the values of a normally distributed random variable may be expected to lie with a given probability, $1 - \alpha$, such that

(6) $$P\left(-z_{\alpha/2} \le \frac{(\bar{y} - \mu)}{\frac{\sigma}{\sqrt{n}}} \le z_{\alpha/2} \right) = 1 - \alpha$$

If the population standard deviation σ were known, the upper and the lower confidence limits (UCL and LCL) for the population mean, μ, are then defined as follows:

(7a) $$UCL = \bar{y} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

(7b) $$LCL = \bar{y} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Equations (6) and (7) show the basic relationships that relate the confidence interval, the level of precision, and the sample size. To determine the necessary sample size, the sample designer must, therefore, specify two parameters: the confidence level (the value of α) and the desired precision, *e* (i.e., the maximum value of $\bar{y} - \mu$). Then, for a maximum error of $e = \bar{y} - \mu$ and a desired confidence level of α, the minimum desired sample size becomes

(8) $$n \ge \frac{z_{\alpha/2}^2 \sigma^2}{e^2}$$

The requisite sample size, however, also depends on σ^2, the population variance, which is generally not known. Thus, two approaches are typically used to estimate the value of σ^2. First, the sample designer may have *a priori* information about the population distribution that permits the approximation of σ^2. This may come from previous surveys or published studies. More commonly, a worst case assumption about the response frequencies is used as follows: Suppose a key survey question asks respondents about an important attribute of their manufacturing practice. Let *p* be the probability that this attribute is present for the respondents in the survey universe. A random variable with a value of 1 if the attribute is present and 0 if not, therefore, is binomially distributed with an expected value of *p* and a variance of $\sigma^2 = (1 - p)p$. This variance is maximized where $p = 0.5$. Thus, substituting $0.5^2 = 0.25$ for σ^2 in (8) will ensure that the sample size is large enough to meet the precision and confidence level objectives.

For example, tables of the normal distribution show than a normally distributed random variable with mean zero and variance 1 will have values between ± 1.96 with a 95 percent probability. Thus, a sample size of *n*, where

(9) $$n = \frac{0.25 * 1.96^2}{e^2}$$

will result in a sample sufficient to achieve a precision of at least *e* at the 95 percent confidence level. That is, $|\bar{y} - \mu| \le e$ with a probability of 95 percent. Equation (9) shows that a sample size of 400 is sufficient to obtain estimates with an accuracy of 5.01% with a probability of 95%.