SUPPORTING STATEMENT – PART B

B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Description of the Activity

The entire population of eligible facilities currently available in the National Industrial Security System (NISS) is used for the survey. The survey is divided into four phrases:

a. Planning: Ensure information such as email address, point of contact (POC) consolidated facilities is correct. Notify industry in advance to allow them time to gather the requirements.

b. Data Collection: Open module via web-based survey tool. Monitor the survey during the deployment period to correct any anomalies before the survey is closed.

c. Analysis: Clean up the data; separate responses and non-responses; perform missing imputation using K-NN algorithm; and perform regression to correct overestimated forecast.

d. Monitor and maintain oversight of projected vs actual requirements in order to meet the OMB +/-5% variant; ensure the forecasts remain in-line with PSI budget.

2. Procedures for the Collection of Information

a. Statistical methodologies for stratification and sample selection;

NA

b. Estimation procedures;

NA

c. Degree of accuracy needed for the Purpose discussed in the justification;

NA

d. Unusual problems requiring specialized sampling procedures; and

NA

e. Use of periodic or cyclical data collections to reduce respondent burden.

NA

3. Maximization of Response Rates, Non-response, and Reliability

DSS, to maximize response rates:

a. Notifies industry in advance to allow them time to gather the requirements.

b. Corrects or updates information such as email address, and the point of contact (POC) responsible for filling the survey.

c. Identifies consolidated facilities, a parent facility responsible for its subsidiary facilities.

d. Ensures POC information such as email address for consolidated facilities is correct.

e. Sends e-mail reminders to non-respondents weekly during the data collection period to ensure timely completion of the collection.

To deal with instances of non-response, a K-Nearest Neighborhood (K-NN) methodology is used:

a. Separate non-response from response populations based on facility level.

b. Calculate the distance based on the total employee population and total facility clearance population.

c. Calculate average of the requirements from these respondent facilities and assign the values to the non-response facility.

DSS maintains a high working relationship with the respondent field and has consistently experienced high response rates on data collection efforts.

4. Tests of Procedures

NA

5. Statistical Consultation and Information Analysis

Primary POC: Chris Pirch 571-305-6241, Statistician