SUPPLEMENTAL SUPPORTING STATEMENT B
OMB No. 1615-NEW
E-VERIFY DATA COLLECTIONS

Revised May 13, 2009

B.   Collection of Information Employing Statistical Methods

This section discusses the statistical methods that we will use for both the web survey of nonusers and the Arizona case studies. Section B1 describes the statistical methods that will be used for the web survey of nonusers. Although employers and their employees for the Arizona case studies will be sampled, we will not be able to generalize the results to the population studied. Please see Section B3 for information on the sample design and expected response rates for the interviews with Arizona employers and their employees and justification for the case study data collection.

B.1.   Respondent Universe and Sampling Methods for the Survey of Nonusers

The target population of this survey includes all employers who are not enrolled to E-Verify. Puerto Rico and other U.S. territories and the State of Arizona, which mandates the use of
E-Verify for all employers, are excluded from study. The domains of interest for the employer population are based on employer size classes within three industry sectors. The three industry sectors of interest are:

1. Employment agencies, temporary help services, and farm labor contractors;

2. Industries known to have relatively large percentages of undocumented workers;

3. All other industries.

Three industry sectors will be defined using the 2007 North American Industry Classification System (NAICS) codes. The size classes will be based on the number of employees (full-time and part-time) working in each company. The three size classes are:

1. Small (less than 15 employees),

2. Medium (15–99 employees), and

3. Large (100 or more employees).

In total, there are nine domains of interest established by three size classes within each of the three industry sectors.

Sampling Frame

The sampling frame will be MarketPlacePro, formerly known as the Dun's market Identifiers (DMI) register maintained by Dun & Bradstreet (D&B). DMI covers all of the U.S. economy and its coverage of most industries is quite complete. DMI, the single comprehensive publicly available database to provide coverage of business establishments, is updated monthly and its coverage of the target population is relatively complete. The records contain the following fields: a D-U-N-S number; North American Industry Classification System (NAICS) code or Standard Industrial Classification (SIC) code; Federal Information Processing Standards (FIPS) state code; Standard Metropolitan Statistical Area (SMSA) code; number of employees at the location; total number of employees for the entire organization; status indicator, i.e., single location, headquarters, or branch; a subsidiary indicator; D-U-N-S numbers of the domestic topmost firm, headquarters, and parent (if a subsidiary); and hierarchy and DIAS codes to identify its location within the corporate structure.

DMI provides the option of choosing alternative organizational levels. The DMI list includes both headquarters and branch level records. DMI defines a headquarters as a business establishment that has branches or divisions

reporting to it, and is financially responsible for those branches or divisions. We will include only the headquarters record for those employers with multiple branches. Therefore, the sampling units will be the single location companies (a business establishment with no branches or subsidiaries reporting to it) and the headquarters of the companies that have multiple branches. The headquarters record provides the total number of employees for the company, including the employees in the branches. It also provides the number of employees at that location.

Table B-1 shows the number of company records in the sampling frame by industry sectors and company employee size classes. Only the single location companies and headquarters of companies with multiple branches were used in this tabulation. That is, a company with a headquarters and multiple branches in different locations was included as a single unit in the tabulations. The number of employees for the headquarters refers to the total number of employees in the company, including the employees in the branches. The number of employees includes full-time and part-time employees.

Table B-1.   Number of employers in the universe, by industry sector and employee size in the sampling frame

| | Number of employees | | | |
Industry sector	Less than 15[1]	15-99[2]	100 or more	Total
1: Employment agencies, temporary help services, and farm labor contractors	42,983	5,230	1,766	49,979
2: Industries known to have relatively large percentages of undocumented workers	1,794,604	442,002	21,958	2,258,564
3: Other industries	6,876,356	834,904	121,837	7,833,097
Total	8,713,943	1,282,136	145,561	10,141,640

[1] Since the D&B's employee size includes owners/proprietors, the companies with an employee size of 1 are excluded.

[2] The employers with unknown employee size are included in size class 15–99.

Sample Design and Sample Size

The sample design will generate a national probability sample of employers that have not enrolled in E-Verify. The survey will utilize a stratified random sample design. The employers will be stratified on the basis of industry and number of employees. The employment agencies, temporary help services, and industries known to have relatively large percentages of undocumented workers will be oversampled. Larger employers will also be oversampled. However, all employers will be selected with equal probability within each industry by size stratum.

In total, a sample of 4,000 company records will be selected from the sampling frame. About 20 percent of the sampled companies are expected to be ineligible. The reasons for ineligibility include being out of business, having no employees, or being enrolled in E-Verify. The expected response rate is 70 percent. Thus, we expect to obtain a total of 2,250 completed surveys. In each industry by size domain, the target is, on average, to achieve 250 completed surveys (Table B-2). Note that the sample draw sizes displayed in Table 2, may be changed after we obtain updated frame counts (including the updated proportion of cases with unknown employee size) before we draw the sample.

Table B-2 shows the Census Bureau's 2006 County Business Patterns (CBP), the number of establishment estimates by industry sector and employment size. The CBP estimates do not include federal, state, and local government establishments whereas D&B includes them. In Table B-2, small size class had to be defined as less than 20 employees instead of less than 15 employees as they are defined in Table B-1.

Table B-2. County Business Patterns estimates of the number of establishments, by industry sector and employee size

	Employment size of the enterprise			
Industry sector	Less than 20	20-99	100 or more	Total
1: Employment agencies, temporary help services, and farm labor contractors	14,732	5,805	23,469	44,006
2: Industries known to have relatively large percentages of undocumented workers	1,432,322	180,168	200,168	1,812,658
3: Other industries	3,982,119	511,782	1,250,595	5,744,496
Total	5,429,173	697,755	1,474,232	7,601,160

B.2.   Procedures for the Collection of Information

The survey of nonusers will be administered via the web to facilitate collection and data analysis processes. As described in Section B.3, we will use a variety of techniques to achieve a 70 percent response rate. The Arizona case studies will be conducted via computer assisted personal interviewing (CAPI) application administered by experienced, trained field interviewers. Section A.3 describes the advantages of using CAPI.

Stratification and Sample Selection for the Survey of Nonusers

The sampling strata will be formed by three employee size classes within three industry sectors as described in Section B1. Three industry sectors will be defined based on the 2007 NAICS codes as shown in Table B-3.

The size classes, based on the total number of employees of each employer, will form a total of nine sampling strata.

Table B-3 Definition of industry sectors, by 2007 NAICS codes

Industry sector	2007 NAICS code	Description of the 2007 NAICS code
1: Employment agencies, temporary help services, and farm labor contractors	56131	Employment Placement Agencies and Executive Search Services
	56132	Temporary Help Services
	56133	Professional Employer Organizations
	115115	Farm Labor Contractors and Crew Leaders
2: Industries known to have relatively large percentages of undocumented workers	11 minus 115115	Agriculture, Forestry, Fishing and Hunting, excluding Farm Labor Contractors and Crew Leaders
	21	Mining
	23	Construction
	311	Food Manufacturing
	5617	Services to Buildings and Dwellings
	722	Food Services and Drinking Places
	812	Personal and Laundry Services
3: Other industries	All other NAICS codes	All other industries

The employers will be selected with equal probability within each size by industry stratum. The selection will be independent across the strata.

Expected Precision of the Estimates for the Nonuser Survey

As mentioned earlier, the target sample size for the survey of nonusers is a total of 2,250 completed surveys. In each of the nine strata, the target is, on average, to achieve 250 completed surveys. However, the number of completed surveys realized can vary across the strata and thus be lower or higher than 250 in a given stratum.

The overall target response rate for the survey is 70 percent. Therefore, to obtain 2,250 completed surveys, we need to contact about 3,200 eligible employers. We expect to find about 20 percent of the employers selected from the DMI frame as ineligible (including those companies that are out of business, have no employees, or are already enrolled to E-verify). Therefore, a sample of 4,000 employers is expected to be sufficient to obtain 2,250 completed surveys.

The population parameters of interest are mostly in the form of totals or proportions. For example, in the survey of nonusers, one such proportion can be the percentage of employers that have heard of E-Verify in a given industry by size domain. An estimate of percentage of nonuser employers, who are familiar with E-verify, in industry by size stratum h, $\hat{p}_h$ can be obtained as:

$$\hat{p}_h = \frac{100 \times \sum_{i \in S_h} w_{hi} y_{hi}}{\sum_{i \in S_h} w_{hi}}$$

where:

> S_h is the set of responding nonuser employers in stratum h;
>
> w_{hi} is the nonresponse adjusted sampling weight attached to responding nonuser employer i in stratum h; and
>
> y_{hi} is the indicator that nonuser employer i in stratum h is responded as familiar with E-verify.

Note that we recommend computing the survey estimates using the sampling weights as described in the above example. The sampling weights, if properly adjusted for nonresponse, can reduce potential nonresponse bias in the survey substantially.

A sample size yielding 250 completed surveys in an industry by size stratum should be sufficient to provide reasonable precision for estimates of proportions in that stratum. The sampling error for a 50 percent proportion obtained from a sample of 250 employers should not exceed 6.2 percent with a 95 percent confidence interval (sampling error is obtained by multiplying the expected standard error by 1.96). The percent sampling errors depend on the sample size and the magnitude of the population percentage to be estimated. For a given sample size, percent error is the largest for a 50 percent population proportion and decreases as proportion moves further away from the 50 percent/50 percent split. For example, for a population proportion of 20 percent (or 80 percent) with a sample size of

250, the sampling error will be less than 5 percent. The sampling errors will be smaller for estimates of proportions produced by overall industry sectors.

Sampling Weights and Estimation Procedures for Nonuser Survey

The sampling weights will be attached to every eligible employer record with a completed survey (1) to account for differential probabilities of selection, and (2) to reduce the potential bias resulting from nonresponse. Each sample employer with a completed survey will be assigned a final weight.

Initially, we will assign a base weight to each sample employer record as the reciprocal of its probability of its selection. The base weights will then be adjusted for nonresponse in order to reduce potential biases resulting from not obtaining a completed survey with every employer in the sample. These adjustments will be made by redistributing the weights of nonresponding employers to responding employers with similar propensities for response. A predictive model for response propensity will be developed to identify subgroups of population with differential response rates. These subgroups will then be used as nonresponse adjustment cells and a separate weight adjustment will be applied in each cell. The potential predictors that can be used in this modeling effort have to be known for both respondents and nonrespondents. These include industry sector, employee size, single location or headquarters status, census region, and MSA/non-MSA status.

If response propensity is independent of survey estimates within nonresponse adjustment cells, then nonresponse-adjusted weights yield unbiased estimates. There are several alternative methods of forming nonresponse adjustment cells to achieve this result. We plan to use Chi-Square Automatic Interaction Detector (CHAID) software[1] to guide us in forming the cells. CHAID partitions data into homogenous subsets with respect to response propensity. To accomplish this, it first merges values of the individual predictors, which are statistically homogeneous with respect to the response propensity and maintains all other heterogeneous values. It

---

[1] SPSS for Windows: CHAID, Release 6.0, User's Guide, Jay Magidson/SPSS Inc., 1993.

then selects the most significant predictor (with the smallest p-value) as the best predictor of response propensity and thus forms the first branch in the decision tree. It continues applying the same process within the subgroups (nodes) defined by the "best" predictor chosen in the preceding step. This process continues until no significant predictor is found or a specified (about 20) minimum node size is reached. The procedure is stepwise and creates a hierarchical tree-like structure.

Although nonresponse adjustment can reduce bias, at the same time it may increase the variance of estimates. Small adjustment cells and/or low response rates (or large nonresponse adjustment factors) may increase the variance and give rise to unstable estimates. In order to prevent an unduly large increase in variance and thereby an adverse effect on the mean square error of the estimates, we plan to limit the size of the smallest cell to a minimum and avoid large adjustment factors.

Variance Estimation

The estimates of standard errors in the nonuser survey can be obtained using a variance estimation software, such as SAS-callable SUDAAN or WesVar. SUDAAN provides variance estimation procedures using both Taylor series linearization method and replication methods. WesVar uses only replication methods. The replication method requires the development of a replication scheme and computation of the replicate weights. We propose to use SAS-callable SUDAAN with the Taylor linearization procedure, which requires less effort to obtain the standard errors of the survey estimates. The estimators in this survey are in the form of totals, means, and proportions. In Taylor linearization approach is appropriate to use with these types of estimators.

We do not anticipate any unusual problems requiring specialized sampling procedures.

Use of Periodic Data Collection Cycles to Reduce Burden

USCIS requires more frequent data collections to evaluate a growing program that has critical implications for immigration policy and reform. However, the last survey of nonusers was conducted in 1999. The last data collection for users of the E-Verify program was conducted in 2007; however, only a few of the respondents resided in Arizona, a state where E-Verify is now mandated.

B3. Methods to Maximize Response Rates and Deal with Issues of Non-Response

The techniques that will be used to achieve high response rates for the survey of nonusers are:

1. **Motivational material**

- Obtain letters of endorsement from one or more national professional employer organizations such as the National Chamber of Commerce, the National Small Business Association, the National Payroll Association, and the National Association of Manufacturers;

- Create a professional image for the study through a well designed and user-friendly website for the web survey of nonusers;

- Emphasize the importance of participation towards shaping future directions in a mandatory or a continued voluntary Federal immigration policy;

- Emphasize the steps that will be taken to ensure respondent confidentiality; and

- Use language appropriate for the target population.

2. **Aggressive followup.** One of the major factors that increases study response rates is the use of aggressive followup procedures to gain cooperation with the study. The web survey of nonusers therefore includes multiple contacts with selected respondents. More specifically, the data collection procedures for nonusers consist of the following steps:

- A personalized letter will be sent to all contact people followed by any letters of endorsement described above. This packet will be from the contractor for the survey of nonusers since nonusers would not necessarily be familiar with USCIS or E-Verify. The letter will stress both the importance of participation to future employment verification efforts and the fact that DHS will only use the information for research purposes.

- If the mailing results in a response that the address is no longer valid, a letter or email will be sent to the alternate contact person, if any.

- If no address or e-mail is provided for the contact person or if there is no alternate contact person for a non-valid e-mail address, phone interviewers will call the company to determine who is the correct contact person and, if possible, obtain the name and contact information for an alternate person who will be responsible for the study, if the primary contact person is not available.

- A reminder e-mail or letter will be sent to contact persons approximately one week after the initial contact.

- Approximately two weeks after the reminder email, phone interviewers will contact nonrespondents. Reasons for nonresponse will be requested and participation will be encouraged. Information on how to access the web survey site will be provided, if necessary.

- A second phone reminder will be made approximately two weeks after the first phone reminder. At that time, the interviewer will offer to complete the survey by phone if the respondent prefers to answer in this fashion.

- Approximately four weeks after the second phone reminder, a Federal Express packet will be sent to the remaining contacts who have not responded to any of the previous mail or e-mail correspondence or phone calls and who are not hard refusals.

(1). **Training.** All individuals who will be contacting potential respondents by phone or email and conducting telephone interviews will be trained in ways to optimize response. In addition to general survey procedures, they will be trained to respond to specific questions that are likely to be raised in this study.

(2). **Nonresponse conversion.** Experienced interviewers who are particularly skilled in nonresponse conversion will re-contact initial refusals. The major exception to this rule is for hard refusals.

(3). **Editing and Data cleaning.** A number of editing features will be built into the web survey. For example, if the respondent attempts to provide multiple answers to a question requiring a single response, the respondent will be asked to select only one response. Additional editing checks will be done subsequent to survey completion to check for completeness, inter-item consistency, extraneous remarks, and proper adherence to any skip instructions.

(4). **Pretesting.** A combination of focus groups and individual interviews has been used to obtain input on what factors are likely to motivate response to the surveys in the target populations,. In addition, lessons learned in the earlier data collections will be incorporated in the E-Verify data collections to improve respondent cooperation.

In addition to using the above procedures to increase response rates, for the Arizona case studies, an incentive of $25 will be offered to workers to complete the interviews. Based on our previous data collection experiences with similar workers, we expect a large number of them to be undocumented immigrants, who may fear their identity and status will be disclosed. This could occur especially since their co-workers may share with them that we are asking questions about their work status and experiences in obtaining employment. Since this population is difficult to locate, once they are found, it is especially important to be able to offer them tangible encouragement to participate in the study.

As mentioned above, another one of the most challenging aspects of achieving good response rates for the case studies is to locate workers who are no longer employed at the sampled companies. Therefore we will use the

employer's records and a tracing service (e.g., Peachtree. Accurint) to locate the most recent contact information. Additionally, we learned during the conduct of last year's case studies that the contractor's experienced field interviewers and supervisors were resourceful in searching the Internet for contact information and making discreet inquiries of neighbors and friends, etc. about how to reach employees. Finally, as E-Verify users, these employers have signed an MOU with the DHS and have agreed to cooperate with DHS and SSA designees' inquiries about the E-Verify program. Specifically, the MOU states the employer's responsibilities as follows:

> *The Employer agrees to cooperate with DHS and SSA in their compliance monitoring and evaluation of E-Verify, including by permitting DHS and SSA, upon reasonable notice, to review Forms I-9 and other employment records and to interview it and its employees regarding the Employer's use of E-Verify, and to respond in a timely and accurate manner to DHS requests for information relating to their participation in E-Verify.*

Nonresponse Bias Adjustments for the Survey of Nonusers

Please see section B2, Sample Weights and Estimation Procedures for a description of the approach to dealing with nonresponse bias.

Sampling and Justification for the Case Studies of Arizona Employers and Employees that Cannot be Generalized to the Population

For the case study portion of the evaluation, we expect to sample 540 employers having at least three tentative nonconfirmation findings within 3 months prior to sample selection. Based on our experiences in the fiscal year 2008 evaluation, this should yield a completed sample of approximately 100 employer cases. The sample will be stratified based on the number of employees and industry. Interviews and record reviews for employees with tentative nonconfirmation findings within 3 months prior to sample selection

will be conducted for each of the employers selected for the employer sample.

Additionally, a random sample of 20 employees will be selected from each sampled employer. For employers with 20 or fewer tentative nonconfirmation employees in the 3 months prior to review, all such workers will be selected for record review and employee interviews. For employers with more than 20 eligible employees, a random sample of 20 employees will be selected. We anticipate that approximately 2,000 workers will be sampled and that we will conduct 450 employee interviews. Table B-4 shows the universe, sample and response rates expected for each of the interviews to be conducted in Arizona.

Table B-4. Universe, sample size, and response rates for Arizona case studies

Collection	Universe*	Sample	Response rates	
			Number	Percent
EmployER interview	540	540	100	18
Worker Interview	50,000	2,000	450	22

*Universe of employers is defined as Arizona employers that have received at least 3 TNCs in the 3 months prior to sample selection. Data collected from case studies will not be generalized to the universe.

Based on our prior experience in which we used incentives and extensive follow-up procedures, we do not believe that it is feasible to obtain a sufficiently high response rate to permit inferences from the sample to the entire population. In the 2008 evaluation, we achieved an unweighted 37 percent response rate for employees due to the inability to locate the sampled employees. Employee contact information either was missing or incorrect and accurate updated information was unavailable from the employer, the tracing service, or neighbors. In a few cases, interviewers were fairly certain that the person they were trying to interview was the sampled employee, but the person denied that the identification was correct. Finally, a few workers refused to participate because they were afraid of employer retribution (i.e., they would be fired if their employer discovered they participated in the interview).

The purpose of the case studies is to examine in depth the procedures that employers and workers follow in the verification process, not to produce representative statistics. We are using sampling to ensure that a variety of employer/employee situations are examined, but do not require the statistics to be generalized in order to identify problems and potential solutions in the verification process. Also, we do not have a particular interest in providing statistics on the State of Arizona, but rather have chosen Arizona because it is the first state to fully implement a mandate that all employers use E-Verify. In this context, the case studies will help identify the problems and situations that would occur if/when using E-Verify is mandated in other states or in the entire nation. Statistics that are representative of Arizona may not necessarily be representative in other states, and thus our interest is in

identifying problems and solutions rather than providing statistics that can be generalized. A nationally representative sample of E-Verify users is planned for next year, and will be discussed in a separate OMB submission at that time.

B.4.   Tests of Procedures for Refining Data Collections

The web survey and the Arizona case study interview instruments submitted in this request for clearance are largely based on instruments used in last year's evaluations, though some changes have been made to accommodate the differences in programs and scope of the current studies. Since the instruments were effective last year, we have considerable evidence that the questions will again be effective this year. In addition, Westat conducted focus groups with nine participants with selected employers on the survey of non-users. Through that pretest, we identified minor issues involving the wording of particular questions, and have revised the instruments accordingly. We also conducted a stakeholders conference in Arizona to examine reactions of employers to the new mandate, and have used the information to further improve the case study instruments. After the CAPI programming of the case study instruments is completed, we will pretest the CAPI instruments with an E-Verify employer. The primary focus of that pretest will be on whether there are any difficulties with the CAPI programming, but data from that pretest will also provide one last additional test of the instruments.

B.5.   Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data

Sampling Statistician

Huseyin Goksel
Senior Statistician
Westat
1650 Research Blvd., RE 488
Rockville, MD 20850
301-251-4395
Huseyingoksel@westat.com

Data Collection

Denise Glover
Senior Study Director
Westat
1650 Research Blvd., TA 2128
Rockville, MD 20850
301-251-2269
deniseglover@westat.com

Joan Michie:
Senior Study Director
Westat
1650 Research Blvd., TA 2102
Rockville, MD 20841
301-294-2014

Data Analysis

Bradford Chaney
Senior Study Director
Westat
1650 Research Blvd., TA 2002
Rockville, MD 20850
301-294-3946
bradchaney@westat.com

Carolyn Shettle
Senior Study Director
Westat
1650 Research Blvd., TA 2058
Rockville, MD 20850
301-251-4324
carolynshettle@westat.com