

## **SUPPLEMENTAL SUPPORTING STATEMENT B**

### **Survey of Recently Naturalized Citizens**

*(File No. 52)*

**OMB No. 1615-NEW**

#### **B. Collection of Information Employing Statistical Methods.**

The USCIS or its designated contractor(s), may employ statistical methods in this survey. As new surveys are developed which will use statistical methods, responses to questions 1 through 5 of this section will be provided, along with specific information for that particular activity, including survey instruments. Upon completion of the analysis of data, results of the survey activities will be provided.

#### **Survey of Recently Naturalized Citizens**

##### **1. Respondent Universe and Sampling Methods**

The sample design will generate a national probability sample of new U.S. citizens who completed the naturalization process in FY 2008 (October 2007 – September 2008). This study utilizes a two-stage sampling plan: Stage One selects a sample of geographic areas, or Primary Sampling Units (PSUs). Stage Two draws a sample of individuals from the sampled PSUs.

##### ***Sampling Frame***

The Contractor, Abt SRBI, will use the automated case tracking systems and the Central Index System to construct the sampling frame. The sampling frame will contain the address of the individual at the time of naturalization and a high percentage will also contain a telephone number. Some individuals will, however, move locally or will move outside their county of residence at time of naturalization. Also, it will not be possible to identify a telephone number for some sample individuals. In this situation a multi-modality survey design works as the best approach to attaining a high response rate, and making use of mail, telephone, and in-person modes of data collection.

A two stage sample design will be deployed for the Survey of Recently Naturalized Citizens.

### Stage One

FY 2008 naturalized citizens appear to be distributed in roughly the same geographic distribution as the U.S. population. However, some naturalized citizen county/region of birth groups are likely to exhibit some geographic distributions that differ from both the U.S. population and from the total naturalized citizen population. This necessitates a sufficiently large number of PSUs to capture this geographic variability. Stage one will proceed as follows:

1. Determine the count of individuals by county using the sampling frame of FY 2008 naturalized citizens
2. Remove counties with zero, or a very small number of individuals, from the sampling frame.
3. Examine the population size of remaining counties and combine each county that contains only a small number of individuals with a geographically adjacent county. (In other words, the PSUs will either be individual counties or county groups.)
4. Stratify the PSUs by Census Region and other variables that are available in the FY 2008 Naturalized Citizen sampling frame.
5. Identify “certainty” PSUs, that is, PSU’s that will be certain to be selected into the sample because of the size of the FY 2008 naturalized citizen population;
6. Draw probability sample proportional to size (PPS) sample of noncertainty PSUs. The measure of size for the PPS sample will be total FY 2008 naturalized citizens in the PSU or a composite measure of size that takes into account the count of individuals by the twelve country/region of birth domains listed in the table A below.

The first stage sample will consist of 100 PSUs.

### Stage Two

At the second stage of sampling the contractor will draw a sample of around 7,150 FY 2008 naturalized citizens. With 100 PSUs, the average sample size per PSU will equal approximately

72 individuals. Within each PSU the contractor will draw a stratified random sample of individuals. The stratification variables will include the twelve country/region of birth groups, and other characteristics of FY 2008 naturalized citizens, such as age group and gender, that are available in the sampling frame data base.

Although an overall 70% response rate is anticipated, the actual achieved response rate will be unknown until data collection begins. Also, the response rate may vary by country/region of birth group. The best way to deal with these uncertainties is to divide the sample within each PSU into replicates that reflect the within-PSU stratification, and to release sample on a controlled replicate-by-replicate basis. Once a replicate is released all of the individuals in that replicate will be contacted to participate in the survey. Quota sampling techniques that destroy the probabilistic nature of the sample will not be used.

### **Sample Size**

Country and region of birth will make up the most important domains for this sample design. Other potential domains will include major category of immigrant admission, length of stay in the U.S. prior to naturalization, age group, gender, occupational group, race/ethnicity, and marital status. The following primary stratification and sample sizes appear below:

Table A. Recently Naturalized Citizens Sample Size by Domain

| Twelve Country/Region of Birth Domains | FY 2007 Population Size | Percent of Total FY 2007 Population | Proposed Initial Sample Size |
|--|-------------------------|-------------------------------------|------------------------------|
| Mexico                                 | 122,258                 | 18.5                                | 1,144                        |
| India                                  | 46,871                  | 7.1                                 | 572                          |
| Philippines                            | 38,830                  | 5.9                                 | 572                          |
| China                                  | 33,134                  | 5.0                                 | 572                          |
| Vietnam                                | 27,921                  | 4.2                                 | 572                          |
| Dominican Republic                     | 20,645                  | 3.1                                 | 572                          |
| Africa                                 | 41,652                  | 6.3                                 | 572                          |
| Balance of Asia                        | 92,041                  | 13.9                                | 572                          |
| Europe                                 | 86,742                  | 13.1                                | 572                          |
| Balance of North America               | 98,260                  | 14.9                                | 572                          |
| South America                          | 48,133                  | 7.3                                 | 572                          |
| Oceania/Other/                         | 3,990                   | 0.6                                 | 286                          |

|         |         |     |              |
|---------|---------|-----|--------------|
| Unknown |         |     |              |
| Total   | 660,477 | 100 | <b>7,150</b> |

## 2. Procedures for the Collection of Information

Achieving the desired response rate of 70% will require multiple modes of survey administration: telephone, mail and in-person. The strategy employed will maximize telephone interviews, minimize field data collection costs, and avoid issues that can arise with mixed-mode data collection.

Survey administration will take place in order of the following procedures:

- Sample file preparation and respondent location
- Telephone Survey Administration phase: 15 call design
- Follow-up Abridged Mail Survey
- In-person Location and Follow-up

To maximize efficiency, the contractor will perform an initial round of locating work to verify contact information. This will allow for the best use of our telephone efforts. Following verification respondents will be sent an advanced contact letter that will briefly introduce the study and the participation incentive to the study subject. This letter is included as Appendix A. A week after the letter is sent, telephone recruitment will begin utilizing an intense, 15-call design. The contractor will use a multi-lingual project team with bi-lingual interviewers in Spanish and the following languages: Hindi, Gujarati, Tagalog, Mandarin, Cantonese and Vietnamese. Third-party interpreters will be utilized for other, less common languages.

The contractor will conduct in-person field follow-up on cases where they have made a final determination of incorrect telephone numbers and/or after they have not been able to reach the respondent or confirm the number is correct. Field-located respondents will conduct the interview by telephone, using a cell phone provided by the on-site interviewer.

## 3. Methods to Maximize Response Rates and Deal with Non-response

In the context of this study, "response rate" is defined as the number of adults who complete the interview divided by the number of adults sampled from the FY 2007 database of naturalized citizens. The response rate expected to be 70%.

Several procedures to maximize response rates will be employed. 1) Approximately one week prior to the initiation of phone interviews, letters explaining the intent of the survey will be mailed to households within the study area. . (See appendix A for advance letter ) 2) Interviewers will make multiple attempts to contact a respondent. 3) Interviews will be conducted during various times of the day and seven days a week to increase the likelihood of finding the respondent at home. 4) Respondents will be provided with the option of scheduling the interview at the time that is convenient for them. 5) For soft-refusals, "interview converters" who have extensive training in telephone interviewing and converting non-responders will be used to increase the response rate.

In households where there is no answer, a minimum of 12 call attempts will be made to each respondent in the sample. The attempts will be made according to the contractor's standard day, evening, weekend rotation scheme. If, after twelve calls, a respondent is not interviewed, the supervisor will evaluate the calling strategies and will assign the case to another interviewer if it appears that the respondent is reachable (i.e., someone has answered the telephone at the current number during the attempts and verified that it is an appropriate number for reaching the subject). A total of 15 call attempts will be made per household.

Hard to reach respondents will be mailed a hard-copy, abridged version of the survey in an attempt to recruit them into the study. The abridged survey will contain critical items. This will establish the legitimacy of the study by introducing the respondent to the content of the survey, and improving the overall response rate to the study. The mailed abridged surveys will inform respondents that they may phone- in to complete to entire survey if they wish, will offer to provide an interpreter or bilingual interviewer and will provide a toll-free number for the respondent to phone in.

The contractor will conduct in-person field follow-up on cases where they have made a final determination of incorrect telephone numbers and/or after they have not been able to reach the respondent or confirm the number is correct. Field-located respondents will conduct the interview by telephone, using a cell phone provided by the on-site interviewer.

The contractor will use a multi-lingual project team, including bi-lingual interviewers in Spanish and the following languages: Hindi, Gujarati, Tagalog, Mandarin, Cantonese and Vietnamese. Third-party interpreters will be used for other, less common languages.

#### Nonresponse Bias Analysis

The AAPOR response rate for the Survey of Naturalized Citizens is expected to be around 70%. We therefore plan to conduct a nonresponse bias assessment by using two main techniques. First, we will compare respondents and nonrespondents on the wide range of variables that are available in the FY 2008 sampling frame data base, including age group, date of naturalization, gender, country of birth, Census Region, degree of urbanicity of the county of residence. Second, the survey design will involve three sequential modes of data collection: telephone, mail, and in-person. Individuals who are interviewed in-person are likely to be more similar to individuals that we never interview. We will therefore compare individuals interviewed by telephone or mail with those interviewed in-person in terms of substantive questionnaire variables.

#### Weighting Methodology

After the interview file has been cleaned and edited we will develop final weights for each completed interview. First, we will calculate a base sampling weight equal to the reciprocal of the probability of selection of the FY 2008 naturalized citizen in a released replicate. The base sampling weight has two components – the selection probability of the PSU and the selection probability of a FY 2008 naturalized citizen in a released replicate within a PSU. Then at the PSU level we will make an adjustment to the base sampling weight to account for unit nonresponse. This will yield a nonresponse adjusted base sampling weight. We will then assemble population control totals for FY 2008 naturalized citizens using the sampling frame

data base. The weighting variables will include Census Region and country/region of birth as well as other variables identified in the nonresponse bias assessment as exhibiting differences between respondents and nonrespondents. To bring the weighted distribution of completed interviews in alignment with the control totals for the selected variables, we will use raking ratio estimation (also known as raking).

#### Standard Errors

Once the interview file has been cleaned and weights are calculated, we will calculate standard errors for key questionnaire estimates. Because we will use a two-stage sample design with unequal weights, we plan to use software such as SUDAAN to obtain valid standard errors. SUDAAN will take into account stratification, clustering, and unequal weighting in the variance estimation process. This will ensure that we do not provide underestimates of the actual magnitudes of the standard errors.

#### **4. Tests of Procedures or Methods to be Undertaken**

Cognitive testing with eight (8) individuals has been conducted on the survey instrument to uncover major sources of response error. Significant revision and reduction of the questionnaire resulted from that testing.

The contractor will conduct further testing of the instrument prior to the initiation of the full study. This will include pilot testing of 25 individuals in multiple languages under conditions similar to actual survey administration, including foreign language administration. These tests are undertaken to uncover programs in CATI programming and other minor wording or interviewer training issues. Only minor modifications are expected from this testing.

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

1) Sampling Statistician:

Mike Battaglia, VP  
Survey Sampling and Methodology Division (SSM)  
Abt Associates Inc., 55 Wheeler St., Cambridge, MA 02138  
(V) 617-349-2425, (F) 617-386-8317  
[mike\\_battaglia@abtassoc.com](mailto:mike_battaglia@abtassoc.com)

2) Data Collection

Chintan Turakhia  
Senior Vice President  
Abt SRBI  
275 7<sup>th</sup> Avenue, Suite 2700  
New York, New York 10001  
Phone 212.779.7700 Fax 212.779.7785  
[c.turakhia@srbi.com](mailto:c.turakhia@srbi.com)

3) Data Analysis

Chintan Turakhia  
Senior Vice President  
Abt SRBI  
275 7<sup>th</sup> Avenue, Suite 2700  
New York, New York 10001  
Phone 212.779.7700 Fax 212.779.7785  
[c.turakhia@srbi.com](mailto:c.turakhia@srbi.com)

Kelly Daley, Ph.D.  
Senior Analyst  
Abt SRBI  
640 N. LaSalle St. Suite 640  
Chicago, Illinois 60654  
312-529-9703  
[k.daley@srbi.com](mailto:k.daley@srbi.com)

John Mollenkopf, Ph.D.  
Center for Urban Research  
CUNY Graduate Center  
365 Fifth Avenue Room 6202  
New York, NY 10016