B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODOLOGY

1. Description of Sampling Methodology

The goal is to conduct a national telephone survey of consumers on their experiences with consumer fraud. The potential respondent universe for this survey is all U.S. adults (18 years of age or older) residing in U.S. households in any of the 50 states or in the District of Columbia. The sample design will involve oversampling of selected minority groups (African Americans and Hispanics) that are likely to be at an elevated risk of becoming victims of consumer fraud. Following a stratified (geographically) sample design, a total of about 3,700 telephone interviews will be completed nationwide including around 400 interviews from each of the minority groups – African Americans and Hispanics. One hundred (100) telephone interviews will be completed in the pretest to test the survey instrument.

For the purpose of sampling, the target population will be stratified into 12 strata as described below in Table 1. Each of the four census regions (Northeast, Midwest, South, and West) will be sub-stratified into 3 strata – High Density African American stratum, High Density Hispanic stratum and the Rest (or the Other group). The twelve resulting strata will be defined in terms of complete counties. The high density strata will be formed based on the proportion of African American or the Hispanic adult population. As shown below in Table 1, for example, the High Density African American stratum in Census Region Northeast will consist of all counties where the proportion of African American adults among all adults in that county is at least 20%. Similarly, the High Density Hispanic stratum in Census Region West will consist of all counties in that region where the proportion of Hispanic adults among all adults in that county is at least 40%. The definition of High Density strata will, therefore, vary depending on the Census Region and the targeted minority group.

A dual frame (landline and cell) sample design will be used within each of the 12 strata and so the sample frame will consist of all residential telephone numbers – both cell and landline – in the United States. The selection of landline numbers will be based on list-assisted (1+) RDD (Random Digit Dialing) sampling of telephone numbers. The cell phone sample will be a simple random sample drawn from all dedicated exchanges for cell phones. For respondents reached on a landline phone, one respondent will be chosen at random from all eligible adults within a sampled household. For respondents reached on a cell phone, the person answering the call will be selected as the respondent if he or she is otherwise found eligible for this study (18 years of age or older). The goal will be to complete roughly 70 percent of the interviews from the cell phone frame while the remaining 30 percent will be from the landline frame. It may be noted that the actual number of completed landline and cell phone surveys for each stratum will depend on observed response rates and so they may not exactly match the corresponding targets. However, the goal will be to meet those targets to the extent possible by constant monitoring of the response rates and by optimally releasing the sample in a sequential manner throughout the data collection period.

Table 1: Planned Stratification and Sample Allocation

Strata	Description of Strata	Targeted number of Interviews
Northeast – High Density African American	Region: Northeast All counties where the proportion of African American adults is at least 20% of the county adult population	90
Northeast – High Density Hispanic	Region: Northeast All counties where the proportion of Hispanic adults is at least 20% of the county adult population	90
Northeast - Other	Region: Northeast All remaining counties that do not belong to High Density strata	483
Midwest – High Density African American	Region: Midwest All counties where the proportion of African American adults is at least 20% of the county adult population	180
Midwest – High Density Hispanic	Region: Midwest All counties where the proportion of Hispanic adults is at least 15% of the county adult population	25
Midwest - Other	Region: Midwest All remaining counties that do not belong to High Density strata	576
South – High Density African American	Region: South All counties where the proportion of African American adults is at least 40% of the county adult population	260
South – High Density Hispanic	Region: South All counties where the proportion of Hispanic adults is at least 40% of the county adult population	150
South - Other	Region: South All remaining counties that do not belong to High Density strata	978
West – High Density African American	Region: West All counties where the proportion of African American adults is at least 10% of the county adult population	50
West – High Density Hispanic	Region: West All counties where the proportion of Hispanic adults is at least 40% of the county adult population	270
West - Other	Region: West All remaining counties that do not belong to High Density strata	548

Total		3,700

The definition of the four census regions in terms of states is also given below.

Northeast: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont, New Jersey, New York, Pennsylvania.

Midwest: Illinois, Indiana, Michigan, Ohio, Wisconsin, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota.

South: Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia, Alabama, Kentucky, Mississippi, and Tennessee, Arkansas, Louisiana, Oklahoma, and Texas.

West: Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming, Alaska, California, Hawaii, Oregon, and Washington.

The proposed sample allocation in Table 1 above is derived taking into account the adult population in each stratum, the distribution of the adult population across African Americans, Hispanics and Others within each region, and the expected response rates for these groups. The proposed sample allocation is not proportional (to the total adult population) across strata because a proportional allocation scheme across the 12 strata is not likely to yield the minimum number (400) of interviews for the African Americans and Hispanics nationwide. The high density strata, therefore, are oversampled to some extent to meet those requirements.

2. Description of the Information Collection Procedures

Subject to OMB approval, the FTC has contracted with a survey firm to conduct a telephone survey of a random sample of adult respondents drawn from all 50 states and the District of Columbia.

Interviews will be offered in both English and Spanish.[1] As part of the initial screening of respondents, the contractor will use standard procedures to identify respondents who prefer to be interviewed in Spanish. For those who prefer a Spanish interview, a call-back with a Spanish-speaking interviewer will be scheduled.

As described above, a stratified sample design will be used and the sample frame will consist of both landline and cell phone numbers. Within each of the 12 strata, the sampling of landline and cell phones will be carried out separately from the respective sampling frames. The landline RDD sample of telephone numbers will be selected following the list-assisted telephone

1 Interviews will not be conducted in languages other than English and Spanish. Recent Federal Census Current Population Survey data finds that fewer than 5% of U.S. population age 5 and over do not speak either English at least "well" or Spanish. Therefore, the current protocol will allow us to include interviews with the overwhelming majority of U.S. residents. The contractor will document the number of potential respondents who speak neither English nor Spanish.

sampling method proposed by Casady and Lepkowski (1993). This procedure uses the Telcordia frame that is generated by appending all 10,000 four digit suffixes (0000 to 9999) to the area code-prefix combinations. In view of cost and operational efficiency, the truncated version of the Casady and Lepkowski (1993) method will be used and sample will be drawn from 100-banks containing at least one listed residential numbers (1+). For within household sampling, the next birthday method will be used to randomly select one eligible person from all eligible adults in each sampled household. Following the next birthday method, the interviewer asks to speak with the eligible person in the household who has the next birthday. This is much less intrusive than the purely random selection method or grid selection that requires enumeration of all household members to make a respondent selection.

The cell phone sample of telephone numbers will be drawn separately from the corresponding dedicated (to cell phones) telephone exchanges. For respondents reached on cell phones, there will not be any additional stage of sampling (like the within household sampling for landline sample). The person answering the call will be selected for the survey if he/she is found otherwise eligible. For both landline and cell phones, the geographic location of the respondent will be determined based on respondent's self-reported response to a question on location (like 'what is your zip-code?'). All respondents will be asked a series of questions to gather information about their use of telephone (cell only, landline only or dual user cell mostly and other dual users) for the purpose of weighting the sample data. It may be noted that due to continuous porting of numbers from landline to cell and cell to landline, some numbers from landline exchanges may turn out to be cell phones and conversely, some numbers sampled from the cell phone exchanges may actually be landline numbers. However, such numbers will be relatively small and the vast majority of landline and cell phone numbers will be from the corresponding frames. The survey will also find out from the respondents if the number called is actually a landline or a cell phone number.

Estimation procedure: Survey based estimates for this study will be weighted to minimize any potential bias that may be associated with unit level non-response. At the national level, the sampling error associated with estimates of proportions based on the total sample size of 3,700 is expected to be around 1.6% at 95 percent level of confidence ignoring any design effect. Under the assumption of a design effect of 1.2, for example, the sampling error associated with estimates of proportions based on the total sample size of 3,700 is expected to be around 1.8% at 95 percent level of confidence. For any subgroup of interest, the sampling error will depend on the sample size. Sample data will be weighted to generate unbiased estimates. Weighting adjustments will be carried out to correct for (i) probability of selection in the sample and (ii) non-response. Once the sampling weights are generated, weighted estimates can be produced for different unknown population parameters (means, proportions etc.) for the target population and also for population subgroups.

The weighting for this study will be done following the procedure described in Kennedy, Courtney (2007): Evaluating the Effects of Screening for Telephone Service in Dual Frame RDD Surveys, Public Opinion Quarterly, Special Issue 2007, Volume 71 / Number 5: 750-771. In studies dealing with both landline and cell phone samples, one approach is to screen for "cell only" respondents by asking respondents reached on the cell phones whether or not they also

have access to a landline and then interviewing all eligible persons from the landline sample whereas interviewing only "cell only" persons from the cell phone sample. The samples from such designs are stratified, with each frame constituting its own stratum. In this study, however, a dual-frame design is proposed where dual users (those with access to both landline and cell phones) can be interviewed in either sample. This will result in two estimates for the dual users based on the two samples (landline and cell). The two estimates for the dual users will then be combined and added to the estimates based on landline-only and cell-only population to generate the estimate for the whole population.

Composite Pre-weight: For the purpose of sample weighting, a total of 24 weighting adjustment classes will be formed by the cross-classification of the 12 strata (Table 1) and the two sample frames (landline and cell). Following Kennedy, Courtney (2007), the composite pre-weight will be generated within each weighting class. The weight assigned to the i^{th} respondent in the h^{th} weighting class (h=1, 2, …, 24) will be calculated as follows:

$$W_{(landline,hi)} = (N_{hl}/n_{hl})(1/RR_{hl})(n_{cwa}/n_{ll})(\lambda^{IDual}) \quad \text{for landline sample cases} \quad (1)$$

$$W_{(Cell,hi)} = (N_{hc}/n_{hc})(1/RR_{hc})(1-\lambda)^{IDual} \quad \text{for cellular sample cases} \quad (2)$$

where
N_{hl}: size of the landline RDD frame in weighting class h
n_{hl}: sample size from landline frame in weighting class h
RR_{hl}: response rate in weighting class h associated with landline frame
n_{cwa}: number of adults in the sampled household
n_{ll}: number of residential telephone landlines in sampled household
I^{Dual}: indicator variable with value 1 if the respondent is a dual user and value 0 otherwise
N_{hc}: size of the Cell RDD frame in weighting class h
n_{hc}: sample size from Cell frame in weighting class h
RR_{hc}: response rate in weighting class h associated with Cell frame
'λ' is the "mixing parameter" with a value between 0 and 1.

If roughly the same number of dual users are interviewed from both samples (landline and cell), then 0.5 will serve as a reasonable approximation to the optimal value for λ. For this study, we plan to use a value of 'λ' equal to the ratio of the number of dual users interviewed from the landline frame and the total number dual users interviewed from both frames.

It may be noted that equation (2) above for cellular sample cases doesn't include weighting adjustments for (i) number of adults and (ii) telephone lines. For cellular sample cases, as mentioned before, there is no within household random selection. The random selection can be made from all persons sharing a cell phone but the percentage of those sharing a cell phone is rather small and it will also require additional questionnaire time to try to capture such information. The person answering the call will be selected as the respondent if he or she is otherwise found eligible and hence no adjustment based on "number of eligible adults in the household" will be necessary. The information on the number of cell phones owned by a respondent could also be asked to make adjustments based on number of cell phones. However,

the percentage of respondents owning more than one cell phone is expected to be too low to have any significant impact on sampling weights. For landline sample cases, the values for (i) number of eligible adults (n_{cwa}) and (ii) number of residential telephone lines (n_{ll}) may have to be truncated to avoid extreme weights. The cut-off value for truncation will be determined after examining the distribution of these variables in the sample. It is anticipated that these values may be capped at 2 or 3.

The Response rates (RR_{hl} and RR_{hc} mentioned above in equations (1) and (2)) will be measured using the AAPOR (3) definition of response rate within each weighting class and calculated as follows:

RR = (number of completed interviews) / (estimated number of eligibles)

= (number of completed interviews) / (known eligibles + presumed eligibles)

It will be straightforward to find the number of completed interviews and the number of known eligibles. The estimation of the number of "presumed eligibles" will be done in the following way: In terms of eligibility, all sample records (irrespective of whether any contact/interview was obtained or not) may be divided into three groups: i) known eligibles (i.e., cases where the respondents, based on their responses to screening questions, were found eligible for the survey), ii) known ineligibles (i.e., cases where the respondents, based on their responses to screening questions, were found ineligible for the survey), and iii) eligibility unknown (i.e., cases where all screening questions could not be asked, as there was never any human contact or cases where respondents answered the screening questions with a "Don't Know" or "Refused" response and hence the eligibility is unknown).

Based on cases where the eligibility status is known (known eligible or known ineligible), the eligibility rate (ER) is computed as:

ER = (known eligibles) / (known eligibles + known ineligibles)

Thus, the ER is the proportion of eligibles found in the group of respondents for whom the eligibility could be established.

At the next step, the number of presumed eligibles is calculated as:

Presumed eligibles = ER × number of respondents in the eligibility unknown group

The basic assumption is that the eligibility rate among cases where eligibility could not be established is the same as the eligibility rate among cases where eligibility status was known.

Post-stratification weight: Once the two samples are combined using the composite weight (equations (1) and (2) above), a post-stratification weighting step will be carried out, following Kennedy (2007), to simultaneously rake the combined sample to (i) known characteristics of the target population (U.S. adults living in any of the 50 states and D.C.) and (ii) an estimated parameter for relative telephone usage (landline-only, cell-only, cell mostly, other dual users).

The target numbers for post-stratification weighting will be obtained from latest available Current Population Survey data. The target numbers for the relative telephone usage parameter

will be based on latest estimates from the National Health Interview Survey. After post-stratification weighting, the distribution of the final weights will be examined and trimming of extreme weights, if any, will be carried out if found necessary to minimize the effect of large weights on variance of estimates.

3. Methods to Maximize Response Rates/Reliability of Sample Data

In order to maximize response rates, several steps will be taken including (1) a call design that will ensure call attempts are made at different times of day and different days of the week to maximize contact rates, (2) conducting an extensive interviewer briefing prior to the field period that educates them about the content of the survey as well as how to handle reluctance and refusals, (3) having strong supervision that will ensure that high quality data are collected throughout the field period, (4) utilizing troubleshooting teams to attack specific data collection problems that may occur during the field period, and (5) customizing refusal aversion and conversion techniques. A seven-call design will be employed, i.e. up to seven calls will be made to establish human contact and complete the interview with the randomly selected person.

Besides taking steps to maximize the response rate, data will be collected to permit an exploration of whether those who are more difficult to reach and those who refuse to participate in the study differ significantly from those who do participate. A sample of 100 non-respondents to the main survey (those who refuse to participate and those with whom we did not have contact) will be asked to answer a small number of questions – primarily demographics. The responses of these people will then be compared to the characteristics of those who complete the survey to see if there are any differences between the two groups. Further, for each telephone number where an interview is completed, data on how many calls were needed before the interview was completed will be collected. Using th data, it will be possible to examine whether those who are more difficult to reach have different characteristics – and different experiences – than those who were reached more easily.

4. Testing of Procedures or Methods Undertaken

Staff will pretest the survey by sampling 100 respondents to ensure that all questions are easily understood. This pretest is also discussed in Part A above, and is part of the collection of information for which the FTC seeks OMB approval.

Staff and the contractor will conduct a number of different review steps during the pre-test, including--

- the review of open-ended responses to determine if additional pre-codes need to be added and, to ease live coding by the interviewers, the review of existing pre-codes to see if their wording corresponds with how respondents speak,

- the review of feedback from the special cognitive questions to evaluate understanding of question wording and to determine if it is obtaining desired information,

- data checks for responses out of expected bounds (e.g. number of experience or dollar values) that indicate misunderstanding of the question,

- data checks on do not know/refusal response which indicate an inability or unwillingness to answer certain topics, and

- data checks on the numbers and type of people reporting different fraud types against prior data to determine if question wording is impacting responses.

The FTC and Gallup will rework questions as appropriate and necessary to meet desired understanding of questions or delete questions as needed.

5. <u>Individuals Consulted on Statistical Aspect of the Surveys</u>

The study design has been prepared by Keith Anderson, Senior Economist, Bureau of Economics (202-326-3428). It has been reviewed internally by two attorneys with extensive experience with the Commission's efforts to fight fraud – Patti Poss, Senior Attorney (202-326-2431), and Monica Vaca (202-326-2245), Assistant Director, Division of Marketing Practices. The contractor (Gallup; Contacts: Camille Lloyd, Senior Consultant, 202-715-3188, and Manas Chattopadhyay, Chief Methodologist, 202-715-3179) has also extensively reviewed the survey instrument and design and is experienced in conducting statistically rigorous telephone surveys.

As noted in A. 8 b. above ("Consultation Outside the Agency"), the design of the 2011 survey on which the current survey is based was reviewed by the FTC's survey consultant, Manoj Hastak, Ph.D., Professor of Marketing at American University's Kogod College of Business Administration in Washington, DC. Moreover, as also noted and illustrated in A. 8.b above, the methodology of the earlier surveys has been well received by academia and government agencies.

Reference.

Robert J. Casady and James, M. Lepkowski (1993). Stratified Telephone Survey Designs. Survey Methodology, 19, 103-113.

Kennedy, Courtney (2007): Evaluating the Effects of Screening for Telephone Service in Dual Frame RDD Surveys, Public Opinion Quarterly, Special Issue 2007, Volume 71 / Number 5: 750-771.