**B.      COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS**

####      1.      Description of Sampling Methodology

An overview of the information collection procedures for the study is given in part A, Specification 2.  The universe of analysis for the study is all adult members of households who have credit histories.[1]  Essentially all such individuals appear in the databases kept by Experian, Equifax, and TransUnion.  Without financial incentives, we would expect the overall response rate to be the same as in the second pilot study.  The overall response rate was 3%, and the response rates for each of six FICO groups were the following:

**Table 1**

| Fico Group | Population Share | Response Rate without Financial Incentives |
|---|---|---|
| Under 600 | 0.12 | 0.018 |
| 600-649 | 0.12 | 0.008 |
| 650-699 | 0.15 | 0.018 |
| 700-749 | 0.2 | 0.026 |
| 750-799 | 0.28 | 0.044 |
| 800 plus | 0.13 | 0.053 |

The information collection procedure in this study differs from that of the second pilot study in that we will provide financial incentives for participation (Specification 3).  We do not have a reliable estimate for the response rate under incentives.  In terms of a potential non-response bias, the information collection procedure in this study will allow us to obtain a great deal of information about individuals who choose not to participate despite the incentives (Specifications 2 and 3 below).

####      2.      Description of the Information Collection Procedures

The purpose of the study is to assess the accuracy of the credit reports of individuals in the universe of analysis, which is all adult members of households who have credit histories.  To accomplish this, we plan to obtain a nationally representative sample of 1000 individuals as study participants.  The two-step process we describe below is designed not only to achieve this representative sample, but also to meet four other important objectives:

A. no CRA can identify who is participating in the study,
B. the FTC's Bureau of Economics never obtains personal identifying information on either study participants or study non-respondents,
C. individuals who choose not to participate in the study are anonymous to the Bureau of Economics and its contractors, as the identified non-respondents are identified only by abstract ID numbers and not by personal identifying information, and
D. only the study contractor obtains *sensitive* personal identifying information and only from study participants, i.e., only from those who have given express permission for the contractor's possession of such information.

---

[1]      An individual has a credit history if he or she has at least one credit card or at least one reported installment debt (or both).

The first step in the study is to have one of the CRAs draw a random sample of 200,000 individuals from its database. This list constitutes the individuals *selected for possible contact* ("SPC list"). This list contains a person's name, address, ZIP code, FICO score, and a unique identifying number ("abstract ID"). The CRA does not report this list to anyone; instead, it creates two sublists, a *BE sublist* (to be sent to the Bureau of Economics) and a *mailing sublist* (to be sent to the FTC's mailing contractor). These lists contain different informational items. The BE sublist contains, for each of the 200,000 individuals, an abstract ID, FICO score and ZIP code; thus, it does not contain any personal identifying information. The mailing sublist contains, again for each individual, the abstract ID, name, address, and ZIP code. This list does not contain any *sensitive* personal identifying information, while it does contain identifying information of a type that could be found in publically available telephone books.

In a second step, BE staff sorts the records in the BE sublist by FICO score and assigns each record to one of the six FICO groups listed in Table 1 above. BE staff then randomly selects about 6,000 records from the BE sublist (further described in detail below). These records constitute the individuals initially selected for contact, called *selected for contact in round 1* or the *SC(1) list*. We further take the ID numbers from SC(1) to form a related list, called *ID(1)*. The list ID(1) contains only ID numbers, which are sent to the mailing contractor for the purpose of soliciting the implicitly identified consumers.[2] The individuals who agree to participate in the study reply only to the *study contractor*, and participants conduct all further business with that contractor. As part of the registration procedure, participants send their assigned ID numbers (contained in their solicitation letters) to the study contractor, and the study contractor will share these ID numbers with BE staff. Table 2 summarizes the main elements of the procedure so far:

**Table 2**

| SPC List (internal to the CRA) 200,000 individuals chosen at random | |
| --- | --- |
| **BE Sublist** 200,000 Abstract IDs, FICO scores, ZIP codes | **Mailing Contractor Sublist** 200,000 names, Abstract IDs, addresses, ZIP codes |
| **SC(1) List** About 6,000 Abstract IDs, FICO scores, ZIP codes | **ID(1) List** About 6,000 Abstract IDs only |

As learned from the second pilot study, the information collection procedure anticipates that the study participants obtained via SC(1) and ID(1) may not have the desired distribution of FICO scores. As necessary, BE staff will draw additional records from the BE sublist to create SC(2) and ID(2) in the same manner as SC(1) and ID(1) were created.[3] This process of

---

[2]     The mailing contractor never learns the participation decision of anyone who has been solicited. A solicitation consists of two mailings to each individual. Multiple mailings comprise part of the strategy for addressing non-response bias. (See Specification 3.)

[3]     We use the observed response rates from the initial solicitations to construct further solicitations via the SC(2) and ID(2) lists. This is explained in greater detail below.

solicitation and response is repeated through progressive rounds until, among other things, the study participants have a distribution of FICO scores close to the desired distribution of scores (more on this below). The *full SC list* is simply the union of SC(1), SC(2), etc. Again, this final list contains no personal identifying information.

### Further Role of the Study Contractor

By the above procedure, individuals who choose not to participate are anonymous to the Bureau of Economics and its contractors. Yet, the study contractor can perform a detailed analysis of the redacted credit reports of non-respondents. We explain how this is possible. The study contractor will receive the ID numbers of the participants, as they respond to the contractor, and will also receive the final SC list from BE staff (containing the ID numbers of all potential respondents; i.e., all who were sent invitation letters). The study contractor is thus able to classify all individuals on the final SC list into participants and non-respondents. The contractor will then obtain the credit reports of all non-respondents by sending their ID numbers to the CRA, which has the master list of all 200,000 names and associated IDs. These credit reports will be redacted by the CRA to remove any personal identifying information.[4] Upon obtaining these redacted reports, the study contractor will assist BE staff in determining whether the class of non-respondents (a certain subset of the final SC list) have significantly different credit histories from study participants. (See also Specification 3.)

From the above procedure it is evident that while the study contractor obtains detailed information about the credit histories of non-respondents, neither BE staff nor its contractors obtain personal identifying information about these individuals. (The mailing contractor never learns the participation decision of anyone whom it solicits.) Further, while the CRA does learn the identity of *non-respondents* (expected in a range of 5,000 - 10,000 names), this conveys no meaningful information to the CRA and has no negative implications for any consumer. Finally, the ultimate set of *participants* remains unknown to the CRA, since it does not know the SC list. The CRA knows only that the participants, whoever they might be, comprise about 1,000 people out of some 190,000 - 195,000 residual names on the SPC list (i.e., the original SPC list minus all of the identified non-respondents).

### Further Detail on the Construction of the Various SC lists

We further describe the construction of SC(1), SC(2), etc., and the final SC list. The goal of the procedure is to obtain a nationally representative sample of approximately 1000 study participants whose distribution of FICO scores matches the national distribution of FICO scores. Here we consider two facts gleaned from the second pilot study. First, if we contact individuals at random and provide no financial incentives, we should expect 3% of the people contacted to participate in the study (the unconditional response rate from the second pilot study). Second, we know that the response rate conditional on FICO group varies from about .8% to 5.3% across the 6 groups, with lower response rates from individuals with lower FICO scores (Table 1 above). Thus, we need to send substantially more invitation letters to the groups with the relatively lower FICO scores in order to obtain a distribution of scores in the sample that matches the distribution of scores in the population.

---

[4]     We know from the CRAs that such redaction is feasible.

If we were to provide no financial incentives, we would need to contact about 50,000 people from the BE sublist in order to obtain 1,000 participants (using the results of pilot 2):

**Table 3**

| FICO Group | Population Share | Number of Contacts | Response Rate without Financial Incentives | Number of Participants | Sample Share |
|---|---|---|---|---|---|
| Under 600 | 0.12 | 6857 | 0.018 | 120 | 0.12 |
| 600-649 | 0.12 | 16,000 | 0.008 | 120 | 0.12 |
| 650-699 | 0.15 | 8333 | 0.018 | 150 | 0.15 |
| 700-749 | 0.2 | 7843 | 0.026 | 200 | 0.2 |
| 750-799 | 0.28 | 6374 | 0.044 | 280 | 0.28 |
| 800 plus | 0.13 | 2449 | 0.053 | 130 | 0.13 |
| **Total** | | **47,857** | | **1000** | |

The sum over all FICO groups of (population share x response rate) is 3%, the unconditional response rate.

The procedure just described would likely provide a representative distribution of FICO scores, but it is important to improve upon the average 3% response rate.[5] We will therefore provide financial incentives for participation (Specification 3).

To take into account that we do not know how much a financial incentive would improve responsiveness, we allow for the multiple rounds of solicitations highlighted above, i.e., SC(1), SC(2), et cetera. In the first round, we select ID numbers from the BE sublist, conditional on FICO score, as enumerated in Table 4 below. These figures presently assume the unconditional response rate may increase to 25%, while further taking into account the variation in response rates conditional on FICO groups. The individuals are selected for contact in round 1 and comprise the SC(1) list defined above. This scenario yields about 6,000 contacts for SC(1):

**Table 4**

| FICO Group | Population Share | Number of Contacts | Response Rate with Financial Incentives | Number of Participants | Sample Share |
|---|---|---|---|---|---|
| Under 600 | 0.12 | 823 | 0.146 | 120 | 0.12 |
| 600-649 | 0.12 | 1920 | 0.063 | 120 | 0.12 |
| 650-699 | 0.15 | 1000 | 0.150 | 150 | 0.15 |
| 700-749 | 0.2 | 941 | 0.213 | 200 | 0.2 |
| 750-799 | 0.28 | 765 | 0.366 | 280 | 0.28 |
| 800 plus | 0.13 | 294 | 0.442 | 130 | 0.13 |
| **Total** | | **5743** | | **1000** | |

The sum over all FICO groups of (population share x response rate) is 25%, an unconditional response rate under incentives.

We recognize that the participants obtained from SC(1) may not have the desired distribution of FICO scores, and BE staff will draw additional records from the BE sublist, as

---

[5] As one illustration of non-response, it may be that people who have gone through bankruptcy could be more reluctant than others to participate in the study (i.e., an embarrassment /trauma factor). It is known that the reporting of bankruptcies and their eventual disposition is a potential source of error. Hence, it is possible that our sample, although representative on FICO scores (inclusive of low scores) could have an under-representation of people who have gone through bankruptcy (and thus potentially deficient on a source of error). See also related discussion in Specification 3 regarding the class of non-respondents.

necessary.  These individuals are selected for contact in round 2 and comprise SC(2) defined above.  In constructing this further list, BE staff will be able to use the observed response rates by FICO group from the SC(1) solicitation.  The process is repeated until the distribution of FICO scores in the full sample of study participants is close to the desired distribution.  Again, the final SC list is simply the aggregation of SC(1), SC(2), et cetera.[6]

### Statistical Analysis, Estimation, and Power

The statistical analysis of the sample is straightforward given the design of the study.  The following exemplify the core estimation and tests.  The calculations illustrate the type of statistical estimates we expect to perform for designated credit score ranges, based on a national study having randomly selected consumers with credit sores that match the national distribution.  We also plan to illustrate the sample results with a variety of graphs and descriptive statistics, and we will conduct additional statistical inferences for the sample statistics that are likely to generate the most interest.

First, we will estimate the fraction of consumers in the population having credit report information with at least one (or more) confirmed material error(s).  Based on the second pilot study, we expect overall that $p = .10$; that is, we expect to observe 10% of the 1000 observations to have at least one confirmed material error in their credit reports.[7]  The associated 95% confidence interval with n =1000 is (.082, .120).  Thus, we would have a margin of error in this study of about $\pm$ 2 percentage points.  The power of this test against the alternative hypothesis that $p = .07$ is 92%; the power against $p = .14$ is 97%.

We are especially interested in individuals with mid to low FICO scores, along with further refinements.  Currently, the mean for the national distribution of FICO scores is approximately 690 and the median is approximately 710.  The second pilot study indicates that the observations with FICO scores below 700 would have $p = .21$ (the probability that a randomly selected consumer with a credit history in that score range has at least one confirmed material error).  The associated 95% confidence interval with n = 390 is (.171, .254).[8]  Thus, for this subsample of particular interest, we would have a margin of error of about $\pm$ 4 percentage points.  The power of this test against the alternative hypothesis that $p = .13$ is 99%; the power against $p = .29$ is 96%.

---

[6]    We presently expect that this aggregate list will remain under 10,000 individuals.

[7]    The estimate $p = .1$ uses the best available information, which comes from the second pilot study.  This estimate is tentative; just 8 of the 128 individuals (about 6%) in the second pilot study have one or more confirmed material errors on their credit reports.  Our estimate of 10% adjusts for the fact that the second pilot study and the national study have different numbers of individuals in each FICO group (i.e., the national study is expected to have relatively more people with low FICO scores and relatively more confirmed material errors).

[8]    At the time of the second pilot study (2007-2008), the national proportion of people with FICO scores below 700 was approximately 39%.  For the national study, we will use the most current figures.

We are also interested in the impact of the correction of confirmed material errors on FICO scores. Using the second pilot study, we expect 90% of the sample to have no confirmed material errors, so the impact of correcting confirmed material errors on the overall sample mean is likely to be small. Considering the subsample with mid to low FICO scores that have at least one confirmed material error, the results of the second pilot study suggests that the average FICO score for this subsample may rise from 603 to 626. With n = 210, the associated 95% confidence interval for the new mean is (614, 638).

### 3. Methods to Maximize Response Rates/Reliability of Sample Plan

In order to maximize the response rate, we plan to financially compensate all individuals who choose to participate and who complete the requirements of the study. The data from the second pilot study indicate that, without financial incentives, the participation rate will be very low overall and especially low for individuals with low FICO scores. In an effort to overcome these problems, the FTC will use two different solicitation letters, depending on the individual's FICO score. The solicitation letters sent to individuals with FICO scores equal to or above 700 will offer the participants $25 upon completion of their work. Solicitation letters sent to people with FICO scores below 700 will offer the participants $25 upon agreeing to participate and a further $50 upon completion of their work (thus a total payment of $75). This latter group is especially reluctant to participate and needs a significantly stronger financial incentive to do so.

### Information About Non-Respondents

As explained above, the study contractor obtains from the CRA the FICO scores and redacted credit reports for all non-respondents relative to the final SC list. (Again, the credit reports of non-respondents will be redacted by the CRA to remove any and all personally identifying information.) Once the study contractor has all the redacted credit reports, it will assist BE staff in determining whether non-respondents have significantly different FICO scores or credit histories from study participants.

Ideally, we would want to observe whether credit reports of non-respondents are significantly different in regards to inaccurate information from the reports of participants, but this is not possible. What is possible is to compare participants and non-respondents on a wide range of key observable credit and non-credit related factors, both within FICO groups and overall. Credit related factors would include such items as reportable bankruptcies, the number of credit cards, and the number of missed credit payments. Non-credit related factors would include the number of reported household moves and the number of consumer identities. If credit reports of participants and non-respondents are similar on the observable factors, then differences in unobservable factors would also be presumptively small.

While we cannot observe whether the credit reports of non-respondents are significantly different in regards to inaccurate information from those of the participants, we can observe whether the credit reports of "high effort" participants are different in terms of negative and inaccurate information from those of "willing" participants. The study contractor will classify respondents into "high effort" versus "willing," based on the mailing to which they respond (there are two mailings to everyone on the SC list) and how quickly they respond. "High effort"

participants are presumptively similar to non-respondents in many characteristics, so comparing the "high effort" to the "willing" participants provides insight into differences between non-respondents and participants. The contractor will further assist BE staff in determining whether "high effort" participants have significantly different FICO scores, credit histories, alleged material error rates, and confirmed material error rates from "willing" participants.

### 4. Testing of Procedures or Methods Undertaken

The proposed design of the national study is largely based on the results of the second pilot study. The FTC's 2008 Report to Congress gives a review of the second pilot study and its related numerical information pertinent to credit report accuracy. As explained in that report, the credit scores of the 128 randomly selected participants do not adequately reflect the national distribution of credit scores (below-average scores were under represented in the sample and above-average scores were over represented). All of the numerical estimates in Specifications 1 and 2 are obtained from the second pilot study; they are presented as the best estimates available at this time.[9]

### 5. Individuals Consulted on Statistical Aspect of the Study

The study design has been reviewed internally by Dr. Jesse B. Leary, Deputy Assistant Director of the Bureau of Economics (202-326-3480)[10] and by FTC economists Dr. Peter Vander Nat (202-326-3518) and Dr. Paul Rothstein (202-326-2213).

---

[9] A national study may well find a different rate of confirmed material errors than the second pilot study. Below-average credit scores were under represented in that study; further, consumers with such scores tend to have relatively more confirmed material errors than those with above-average scores (a matter highlighted in the 2008 Report to Congress (at 10-11)).

[10] Dr. Leary is the main author of a recent statistical report submitted to Congress by the FTC under the FACT Act (see, *Credit-Based Insurance Scores: Impact on Consumers of Automobile Insurance*, Federal Trade Commission, July 2007).