# Supporting Statement for Paperwork Reduction Act Submission
# Pre-Purchase Homeownership Education and Counseling Demonstration and Impact Evaluation
# OMB #2528-0293

**Part B. Statistical Methods**

> 1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.

The HUD Pre-Purchase Homeownership Counseling Demonstration and Impact Evaluation enrolled 5,854 low-, moderate-, or middle-income (LMMI), first-time homebuyers from three national lenders across 28 jurisdictions nationally.[1] Each of the data collections covered by this submission relate to the universe of study participants, which is defined as the set of all individuals who enrolled in the study (excluding those who withdrew from the study).

The study seeks to collect data through the Long-Term Follow-Up Survey and tracking letter from all study participants who enrolled in the study. The study team also seeks to continue to collect the following administrative data:

- The lenders' origination and servicing data for any study participant who originates a home purchase loan with one of the study's participating lenders;

- Origination and servicing data from the Federal Housing Administration (FHA) for any study participant who received an FHA loan; and,

- Credit bureau data from one of the three major credit bureaus.

Additionally, the study team seeks to continue to collect consent from each co-borrower listed on a home purchase mortgage loan of a study participant. The universe of respondents for the co-borrower consent collection therefore includes all co-borrowers of all enrolled study participants.

The study administered a Short-Term Follow-Up Survey to study participants 12 months after study enrollment. The response rate for the Short-Term Follow-Up Survey (administered by telephone and in-person field data collection) was 78.4 percent. The expected response rate of the Long-Term Follow-Up Survey is approximately 70 percent. The study design includes a schedule of strategic contacts with the study participant during the enrollment period and then throughout the study follow-up period to keep them involved and engaged in the study. The study team has also conducted intensive tracking with study participants since study enrollment. Additional information about our

---

[1]  In the fall of 2013, the study piloted recruitment and random assignment in three sites. Full study enrollment across all 28 sites began in January 2014 and was completed in February 2016.

tracking strategy is provided in our response to B.3. With this design, ongoing tracking, and incentives to thank them for their participation, the study team believes we are well-positioned to achieve the target of a 70 percent response rate for the Long-Term Follow-up Survey.

**Sampling Plan**

*Jurisdictions*

Participant enrollment was implemented in 28 jurisdictions. The primary criterion for selecting jurisdictions was the expected volume of eligible study participants. However, the process of selecting the sample of jurisdictions also considered the ability of local in-person housing counseling agencies to provide services to enrolled participants throughout the jurisdiction. Exhibit B.1 lists the study's 28 jurisdictions.

**Exhibit B.1: The Study's 28 Jurisdictions**

| List of 28 Study Sites |
| --- |
| Atlanta-Sandy Springs-Marietta, GA |
| Boston-Cambridge-Quincy, MA-NH |
| Chicago-Joliet-Naperville, IL-IN-WI |
| Dallas-Fort Worth-Arlington, TX |
| Detroit-Warren-Livonia, MI |
| Houston-Sugar Land-Baytown, TX |
| Las Vegas-Paradise, NV |
| Los Angeles-Long Beach-Santa Ana, CA |
| Miami-Fort Lauderdale-Pompano Beach, FL |
| Minneapolis-St. Paul-Bloomington, MN-WI |
| New York-Northern New Jersey-Long Island, NY-NJ-PA |
| Orlando-Kissimmee-Sanford, FL |
| Philadelphia-Camden-Wilmington, PA-NJ-DE-MD |
| Phoenix-Mesa-Glendale, AZ |
| Portland-Vancouver-Hillsboro, OR-WA |
| Raleigh-Cary, NC |
| Riverside-San Bernardino-Ontario, CA |
| Sacramento-Arden-Arcade-Roseville, CA |
| San Antonio-New Braunfels, TX |
| San Diego-Carlsbad-San Marcos, CA |
| San Francisco-Oakland-Fremont, CA |
| San Jose-Sunnyvale-Santa Clara, CA |
| Seattle-Tacoma-Bellevue, WA |
| St. Louis, MO-IL |
| Stockton, CA |
| Tampa-St. Petersburg-Clearwater, FL |
| Virginia Beach-Norfolk-Newport News, VA-NC |
| Washington-Arlington-Alexandria, DC-VA-MD-WV |

Note: The 28 jurisdictions listed above are ordered alphabetically.

*Lenders*

The study team has partnered with three national lenders for the implementation of this study. The participating lenders were chosen purposively, based on their volume of loan originations and willingness to partner for this study.

### *Study Participants*

The universe of potential study participants was defined in the previous submission (Control #2528-0293) and includes all eligible customers who communicated with a study participating lender during the enrollment period, excluding customers who could not be contacted, refused the initial offer of study participation, did not meet the study's eligibility requirements, or did not provide consent. The sample of enrolled study participants includes all individuals who enrolled in the study, excluding those individuals who withdrew from the study.

No statistical methods are involved in identifying study participants for completion of the Long-Term Follow-Up survey. The study will administer the Long-Term Follow-Up survey to the universe of enrolled study participants.

*Co-Borrowers*

No statistical methods are involved in identifying the co-borrowers for consent collection. The study will seek to collect consent from each co-borrower in the study participant universe. Among the 59.3 percent of study participants who purchased a home according to responses to the Short-Term Follow-Up Survey, about one-fourth (25.6 percent) planned to have a co-borrower at baseline, and a similar share of purchasers actually ended up purchasing a home with a co-borrower at follow-up (22.5 percent). During the administration of the Short-Term Survey the study team collected consents from 257 co-borrowers.

2. Describe the procedures for the collection of information including: statistical methodology for stratification and sample selection, estimation procedure, degree of accuracy needed for the purpose described in the justification, unusual problems requiring specialized sampling procedures, and any use of periodic (less frequent than annual) data collection cycles to reduce burden.

**Analysis Plan**

The previous submission explains the implementation analyses, analysis and documentation of random assignment, and the baseline analyses in detail. This submission discusses the impact analyses which will be conducted using administrative data and the Long-Term Follow-Up Survey data.

*Impact Analyses*

The overarching question guiding this evaluation is, *What are the impacts of homebuyer education and counseling on low-, moderate-, and middle-income prospective first-time homebuyers?* We consider the intervention's impact on study participants in three broad domains of outcomes:

- *Preparedness and search*—These outcomes are related to the decision of whether to purchase a home or not, the search for affordable homes, and selection of appropriate mortgages.

- **_Financial capability_**—These outcomes are related to participants' general financial knowledge, behavior and traditional financial markers such as debts and savings, access to affordable credit, and credit profile.

- **_Sustainable homeownership_**—These outcomes are related to homebuyers' mortgage payment behaviors, including those behaviors that can play a role in avoiding foreclosure and accruing and protecting home equity.

The experimental evaluation design balances the pre-existing characteristics of the study participants assigned to each of the treatment groups and to the control group. Therefore, the impact of being offered a treatment can be estimated by simply comparing mean outcomes of those offered treatment, relative to those in the control group. For example, if we find that study participants who were offered homebuyer education and counseling have higher homeownership rates than control group members, we can interpret the difference in homeownership rates as the causal impact of being offered homebuyer education and counseling services. We will estimate this difference in means with multivariate regression, as controlling for baseline characteristics of study participants will increase precision. These analyses will rely on the data from the baseline data described by the study's prior submission for baseline measures of covariates. The primary impact analyses for the final evaluation will be conducted using outcomes collected from administrative data and the Long-Term Follow-Up Survey.

The randomization of participants to treatment and control groups produces an estimate that reflects _the overall impact of being offered homebuyer education and counseling_— referred to as the Intent to Treat (ITT) estimate. Commonly, adjustment of the ITT allows estimation of the impact of homebuyer education and counseling on those who actually complete the prescribed homebuyer education and counseling services—referred to as the Treatment on the Treated (TOT) estimate. The two will differ because of incomplete take-up (i.e., some people who are offered a treatment will choose not to take it up). We will compute TOT estimates by two-stage least squares regression.

As permitted by randomization into multiple treatment groups, we will also estimate the impact of the intervention _by service delivery mode,_ as follows:

- The _impact of offering in-person homebuyer education and counseling services_, computed as the difference in mean outcomes between treatment group members offered in-person services and their control group counterparts.

- The _impact of offering remote homebuyer education and counseling services,_ computed as the difference in mean outcomes between treatment group members offered remote services and their control group counterparts.

- The differential effect of in-person services _compared to_ remote services.

The impact analyses will also estimate the intervention's impacts for subgroups defined by baseline demographic and socioeconomic characteristics of the individual study participants, as well as housing characteristics of the area in which they lived.  For example, subgroups may be defined with respect to borrower personal characteristics (from the baseline survey); financial characteristics (from the baseline credit bureau data and baseline survey); and neighborhood characteristics (which requires knowledge of the borrower's address from the baseline survey). This subgroup analysis is motivated by interest in providing information to policymakers and practitioners about targeting and providing services. Similar to the method used to produce the overall impact, we analyze the subgroup impacts

by pooling all of the sample assigned to any treatment group and comparing their mean outcomes to those of the control group.

As an extension to the analyses described above, we will identify an experimental subgroup comprising treatment group members who are "most likely" to participate in services and their control group counterparts who would have been "most likely" to participate in services had they been offered services. Doing so enables us to compute an experimental estimate of the impact on the subgroup of study participants most likely to participate in services.

**Justification of Level of Accuracy**

This section presents minimum detectable effects for the impact analyses discussed in the previous section. Minimum detectable effects (MDEs) are the smallest true effects of an intervention that researchers can expect to detect as statistically significant when analyzing samples of a given size. Exhibit B.2 reports MDE values for four different estimates of interest: (1) *the overall impact of offering homebuyer education and counseling services*; (2) *the impact of offering in-person services*; (3) *the impact of offering remote services*; and (4) *subgroup impacts*.  For each estimate of interest, we report the sample size and MDEs for outcomes constructed using both administrative data and Long-Term Follow-Up Survey data (for which we anticipate approximately full coverage) and outcomes constructed using only Long-Term Follow-Up Survey data (for which we expect 30 percent missing outcome data).

The MDEs are shown for dichotomous outcomes—variables like foreclosure or home purchase — that take on a value of 0 or 1. We present MDEs for outcomes with a control group mean of 1 percent (which is equal to the MDE for outcomes with a control group mean of 99 percent); outcomes with a control group mean of 10 percent (or 90 percent); and outcomes with a control group mean of 50 percent.  The MDEs are calculated using standard assumptions regarding power, significance level, and R-squared values. These assumptions are: 80 percent power; a two-tailed test at the 0.10 significance level; and an R-squared value of 0.15.

Panel A of Exhibit B.2 presents the sample size and MDE values for *the overall impact of offering homebuyer education and counseling services*, which is computed by comparing outcomes for the pooled sample of all treatment group members to the control group. For a dichotomous outcome variable with a mean control group value of 50.0 percent, the MDE estimate for outcomes constructed using both administrative data and survey data is 3.1 percentage points. This value implies that the average value for the treatment group would have to increase to 53.1 percent for the impact of counseling to be significant. For outcomes constructed using only survey data, the required average treatment group value is 53.7 percent.

Panel B of Exhibit B.2 presents the sample size and MDE values for *the impact of offering in-person services*. For a dichotomous outcome variable with a mean control group value of 50.0 percent, the MDE estimate for outcomes constructed using both administrative data and survey data is 7.4 percentage points. This value implies that the average value for the treatment group would have to increase to 57.4 percent for the impact of counseling to be significant. For outcomes constructed using only survey data, the required average treatment group value is 58.9 percent.

Panel C of Exhibit B.2 presents the sample size and MDE values for *the impact of offering remote services*. For a dichotomous outcome variable with a mean control group value of 50.0 percent, the MDE estimate for outcomes constructed using both administrative data and survey data is 3.3

percentage points. This value implies that the average value for the treatment group would have to increase to 53.3 percent for the impact of counseling to be significant. For outcomes constructed using only survey data, the required average treatment group value is 53.9 percent.

Panel D of Exhibit B.2 presents the sample size and MDE values for *subgroup impacts,* where we assume that each subgroup of interest comprises 50.0 percent of the full sample. For a dichotomous outcome variable with a mean control group value of 50.0 percent, the MDE estimate for outcomes constructed using both administrative data and survey data is 4.7 percentage points. This value implies that the average value for the treatment group would have to increase to 54.7 percent for the impact of counseling to be significant. For outcomes constructed using only survey data, the required average treatment group value is 55.6 percent.

The analysis will also estimate impacts on subgroups defined by characteristics observed after random assignment and exogenous subgroups with small sample sizes. Such analyses will be exploratory in nature, and the study is not designed to confidently detect impacts for these subgroups.

**Exhibit B.2: Sample Sizes and Minimum Detectable Effects For Planned Impact Analyses**

| Outcome | Control Group Sample Size | Treatment Group Sample Size | Total Sample Size | MDE for Outcomes with Control Group Mean of… | | |
|---|---|---|---|---|---|---|
| | | | | 1 or 99 Percent | 10 or 90 Percent | 50 Percent |
| **Panel A: Overall Impact of Homebuyer Education and Counseling** | | | | | | |
| Outcomes Based on Administrative Data and Survey Data | 2,448 | 3,322 | 5,770 | 0.6 | 1.8 | 3.1 |
| Outcomes Based on Survey Data | 1,714 | 2,325 | 4,089 | 0.7 | 2.2 | 3.7 |
| **Panel B: Impact of Offering In-Person Services**<br>**Outcomes Based on Administrative Data and Survey Data**<br>**Outcomes Based on Survey Data** | | | | | | |
| Outcomes Based on Administrative Data and Survey Data | 1,184 | 806 | 1,990 | 1.5 | 4.5 | 7.4 |
| Outcomes Based on Survey Data | 829 | 564 | 1,393 | 1.8 | 5.3 | 8.9 |
| **Panel C: Impact of Offering Remote Services** | | | | | | |
| Outcomes Based on Administrative Data and Survey Data | 2,448 | 2,516 | 4,964 | 0.7 | 2.0 | 3.3 |
| Outcomes Based on Survey Data | 1,714 | 1,761 | 3,475 | 0.8 | 2.4 | 3.9 |
| **Panel D: Subgroup Impacts [c]** | | | | | | |
| Outcomes Based on Administrative Data and Survey Data | 1,224 | 1,661 | 2,885 | 0.9 | 2.8 | 4.7 |
| Outcomes Based on Survey Data | 857 | 1,163 | 2,020 | 1.1 | 3.4 | 5.6 |

[c] Assumes subgroup comprises 50.0 percent of the full sample.
*Notes*: Sample sizes exclude withdrawals. Analysis assumes α = 0.10, two-tailed t-test, and R-squared of 0.15. For outcomes constructed using both administrative data and Long-Term Follow-Up Survey data we assume 0 percent of the sample is missing outcome data. Short

Term Impact Report analyses indicated that outcomes constructed using short-term follow-up survey data, credit bureau data, lenders' loan origination and servicing data, and Federal Housing Administration data cover 99+ percent of the sample. We assume a 70 percent response rate to the Long-Term Follow-Up Survey and therefore assume outcomes constructed using only Long-Term Follow-Up Survey Data will be missing for 30 percent of the sample.

**Unusual Problems Requiring Specialized Sampling Procedures**

There are no unusual problems associated with this sample.

**Any Use of Periodic (less frequent than annual) Data Collection Cycles to Reduce Burden**

Not applicable to this study.

3. Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.

The study team has undertaken extensive efforts to maximize the response rates for the collected data. These efforts are described below.

The loan origination and servicing data from the FHA and participating lenders and credit bureau data from one of the three major credit bureaus will continue to be collected with the processes begun during the initial phase of the study.

The collection of the Long-Term Follow-Up Survey and co-borrower consent each include layered data collection strategies in order to maximize the response rate. The Long-Term Follow-Up survey will ask the study participant to confirm whether they purchased a home with one or more co-borrowers. If any co-borrower has not yet provided consent to the study team during previous data collection efforts, the interviewer will ask the study participant whether the co-borrower is available, and attempt to collect co-borrower consent verbally or arrange a time to callback to collect the consent.

The expected response rate of the Long-Term Follow-Up Survey is approximately 70 percent. The Long-Term Follow-Up Survey will use a telephone plus field approach to collect the survey responses. The CATI center will conduct interviews with all study participants that can be reached by telephone. For those study participants who are not reached by telephone, the data collection team will conduct in-person interviews using a field data collection approach. The study team administered a Short-Term Follow-Up Survey to study participants 12 months after study enrollment. The response rate for the Short-Term Follow-Up Survey (using the same approach – telephone and in-person field data collection) was 78.4 percent.

During the baseline and interim data collection period, the Abt team developed a robust tracking system that utilizes both passive and active measures and involves reaching out to study participants regularly. This design includes a schedule of strategic contacts with the study participant during the enrollment period and then throughout the study follow-up period to keep them involved and engaged in the study. The tracking plan builds upon the study's existing data system and supplements it with data obtained through passive searches every six months of proprietary databases—including the

National Change of Address database. These sources are supplemented by mailed biannual requests for the study participants to review and return a form that includes contact information for the participant, as well as up to three people who always know how to reach the participant. This robust tracking plan helps the study team keep the participant contact information accurate. These efforts began during the study enrollment period and will continue until the study participant completes the Long-Term Follow-Up Survey.

The tracking strategy is essential to achieving the highest possible response rate for the Long-Term Follow-up Survey. The high response rates achieved for the Short-Term Follow-up survey reflects the success of these approaches. In addition, as noted in Part A.9, we will use incentives to thank participants for responding to the Long-Term Follow-up Survey. Study participants will receive $35 for completing the survey. With this design, ongoing tracking, and incentives to thank study participants for their participation, the study team believes we are well-positioned to achieve the target of a 70 percent response rate for the Long-Term Follow-up Survey.

4. Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of test may be submitted for approval separately or in combination with the main collection of information.

The study team relies on review of each instrument by Abt Associates and other study team staff, HUD personnel, and the study's advisory panel to ensure that the instruments are clear, flow well and are as concise as possible, and collect the data necessary for analysis. Additionally, before the Long-Term Follow-Up Survey is implemented in the field, the study team will conduct up to nine pre-tests of the data collection instrument with a small sample of study participants.  These pre-tests will provide information on the average length of the survey, and identify any final modifications to improve the clarity and flow of the instrument.  The study team will provide to OMB, the results from these pre-tests and any final modifications to the data collection instrument prior to their review.

5. Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.

HUD's Office of Policy Development and Research will work with the contractor, Abt Associates, to conduct the proposed data collection. Marina L. Myhre, Ph.D., a Social Science Analyst in HUD's Office of Policy Development and Research, Program Evaluation Division, serves as Contracting Officer's Technical Representative (COTR). Her supervisor is Ms. Carol Star. Dr. Myhre and Ms. Star can be contacted at (202) 402-5705 and (202) 402-6139, respectively. The study's Principal Investigators are Dr. Laura Peck and Debbie Gruenstein Bocian from Abt Associates. Dr. Peck can be reached at (301) 347-5537. Debbie Gruenstein Bocian can be reached at (301) 968-4424. Donna DeMarco serves as the study's Project Director and can be contacted at (617) 349-2322. Dr. Shawn Moulton is the study's Director of Analysis and can be reached at (617) 520-2459.