CONSUMER FINANCIAL PROTECTION BUREAU
INFORMATION COLLECTION REQUEST – SUPPORTING STATEMENT

Supporting Statement for Randomized Evaluation of the Credit Matters Loan at St. Louis Community Credit Union and Credit Matters Counseling offered by BALANCE Financial Fitness Program

(OMB CONTROL NUMBER: 3170-XXXX)

Table of Contents

B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Respondent Universe and Sampling Methods

We propose using a randomized controlled trial to evaluate the effectiveness of three financial capability interventions on economically vulnerable consumers: (1) the Credit Matters Loan, a bundled credit-building loan product available at St. Louis Community Credit Union (SLCCU), (2) Credit Matters Counseling offered over the telephone by BALANCE Financial Fitness Program, and (3) nudges to continue making loan payments into savings after loan expiration. The Credit Matters Loan is the intervention of primary interest. To date there are few rigorous randomized evaluations of financial capability programs in the United States. These evaluations have important implications for policymakers and practitioners alike.

Our sample will consist of members of SLCCU that consent to participate in this study, with treatment randomized at the individual level. SLCCU branches are located in the St. Louis metropolitan area and the credit union caters primarily to economically vulnerable consumers.

Participants will self-select into the study in response to recruitment efforts to identify individuals seeking to build or repair their credit. Recruiters will be set up at a table in a prominent place in the branch with a banner that will seek to attract people who are interested in building their credit. The banner will indicate that respondents will be given gift cards for participation in the survey. The recruiter will invite interested credit union members to participate in study.

We will place two surveyors in each of five branches listed below, for three and a half weeks each. Surveyor capacity will likely be more of a constraint than branch volume. We expect to conduct 160 surveys per branch per week, for a total of 2,500 surveys over the recruitment period.

Table 1. Survey Branch and Monthly Volume

Branch	Unique Visitors
Forest Park	9,293
Graham Rd.	2,907
Dellwood	1,598
Chippewa	2,172
University City	3,800

Note: Volume size is the number of members based on monthly snapshot of over the counter transactions by unique member ID

We will obtain informed consent from each participant that lays out the voluntary nature of their participation and explains how their participation will aid in our research efforts. We will use administrative data for our analysis for most variables. However, surveying will be required to gather information on other outcomes of interest that are not available through administrative data, such as informal savings, debt at the household level, and economic preferences. Surveys will be administered in-person for the baseline survey and by phone for the endline survey 12 months later. Non-monetary incentives will be offered in the form of a $5 gift card to encourage participation in the study and a $10 gift card to incentivize endline survey completion. Participants will have the choice between a gift card to

a local grocery store or gas station (with attached convenience store). Further efforts to increase response rates are detailed below.

2. Information Collection Procedures

Statistical methods for stratification and sample selection

Our sample will be drawn from SLCCU's existing client base. Participants will self-select into the study in response to recruitment efforts to identify individuals seeking to build or repair their credit. The population is well-defined prior to treatment since it is composed of all members of SLCCU.

Randomization will occur at the individual level. In all regressions, we will control for survey location, survey date, and surveyor, and report heteroskedasticity-robust (Huber-White) standard errors.[1] We will conduct orthogonality checks to test whether the treatment and control samples are balanced on observed covariates at baseline in order to support using the experimental results to compare outcome measures and interpret differences causally as impacts of the intervention.[2]

Estimation procedure

We will investigate the research questions of interest using randomized encouragement designs. Encouragement designs are used to evaluate existing programs where withholding access to the programs is not feasible for ethical or practical reasons. Encouragement design evaluations have been used to measure the impact of studying on test scores, the impact of receiving information about tax deferred accounts on financial behavior, and the use of a commitment savings device on savings.[3]

Subjects will be recruited into the Credit Matters study and will provide oral consent to participate. A surveyor will then administer a baseline survey to the participant. At the end of the survey, the surveyor will give the participant a brief description of the Credit Matters Loan. If the participant is interested in the loan product, they will be randomly placed into two Credit Matters Loan treatment arms: (1) an "easy on-ramp" group that is encouraged to open the loan on the spot, or (2) a "high-angle on-ramp" group that is required to complete an additional step before opening the loan. Participants in the "high on-ramp"

---

1 Huber, P. J. (1967), "The Behavior of Maximum Likelihood Estimates Under Non-standard Conditions," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley: University of California Press; White, H. (1980), "A Heteroskedastic-Consistent Covariance Matrix Estimator and a Direct Test of Heteroskedasticity," *Econometrica*, 48; Eicker, F. (1963), "Asymptotic Normality and Consistency of the Least Squares Estimator for Families of Linear Regressions," *Annals of Mathematical Statistics*, 34; and Eicker, F. (1967), "Limit Theorems for Regressions With Unequal and Dependent Errors," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA: University of California Press. Heteroskedasticity, or non-constant variance in unobserved variables, can bias the estimated standard errors of regression coefficients. Using robust standard errors corrects for this potential bias, and is the more conservative approach for hypothesis testing.
2 We will compare means of treated and non-treated individuals in the sample across several characteristics. Specifically, we will perform an Analysis of Variance, which tests multiple joint hypotheses of differences in means. Doing so will validate the random sampling procedure, should we (as anticipated) fail to reject the joint null of no differences in the means between the treated and non-treated.
3 Duflo, E., Saez, E. (2003). "The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment". *Quarterly Journal of Economics*, 118 (3), 815–842; Holland, P.W. (1988). "Causal inference, path analysis, and recursive structural equations models". *Sociological Methodology* 18, 449–484; Karlan, D., Nelson, S., Shafir, E. and Zinman, J (2012). Super Savers? A Randomized Evaluation of Commitment Savings and Financial Counseling in New York City. *Working Paper*

group will be required either to complete an online financial assessment prior to opening a Credit Matters Loan or simply return for loan opening at a later date.

In addition, all study participants will be randomly assigned to one of two BALANCE treatment arms: (1) encouraged to call BALANCE for credit counseling, or (2) not encouraged to call BALANCE for credit counseling. Participants assigned to the encouraged group will receive a description of the counseling services offered by BALANCE and a nudge to call. The assignment to the BALANCE treatment arm will be independent of the respondent's interest in the Credit Matters Loan, and independent of their assignment to the Credit Matters Loan treatment arm. This design is known as "cross-cutting" randomization, and has been used in a number of prior studies to examine the impact of multiple interventions individually, and in combination with one another.[4]

Finally, at the six month mark of the loan term, a random subset of Credit Matters Loan borrowers in the study will be offered a "Pay Yourself Back," program, where they will be encouraged to automate payments into their savings accounts after successful repayment of the loan. The suggested amount of the automated payment will be the same size as their monthly Credit Matters Loan payments.

We will measure the **effectiveness of the product** by comparing outcomes for those who receive the easy on-ramp to those in the high-angle on-ramp group (adjusted for fraction of consumers who took up the loan in the two groups). A comparison of outcomes for those participants who choose not to take up the loan with those in the high-angle on-ramp group will allow us to examine whether **selection into the product** is indicative of overall creditworthiness and future credit behavior. We will also examine whether selection can be predicted using differences in responses to survey questions (unobservable to lender, but observable to researcher) and SLCCU data (observable to lender).

We will measure the **effectiveness of credit education** by comparing outcomes for participants who were encouraged to call BALANCE with those who did not receive an encouragement. We will also examine the **interaction effect of credit education combined with a Credit Matters loan** by comparing outcomes for individuals who receive both encouragements (product and phone counseling) to the groups that receive only one. Finally, we will measure the effectiveness of the "**Pay Yourself Back**" program by comparing Credit Matters Loan borrowers who were encouraged to take up Pay Yourself Back and those who did not receive an encouragement for Pay Yourself Back.

To estimate the effect of the intervention on outcomes we will use appropriate econometric regression techniques that account for the distribution of the error terms and the outcome of interest, such as difference-in-difference estimators using Ordinary Least Squares, quantile or median regressions, or fixed effects. Difference-in-difference estimators estimate the impact of an intervention by comparing pre- and post-treatment trends for those treated (the *treatment* group) and those not treated (the *control* group), with and without controls for covariates.[5] Ordinary Least Squares approximates the conditional mean of the dependent (outcome) variable, whereas median or quantile regression estimates the conditional median (or other quantiles) of the outcome variable.[6] Fixed effects models control for individual level

4 Kremer, M. (2003). "Randomized evaluations of educational programs in developing countries: Some lessons". American Economic Review 93 (2), 102–106; Dupas P. (2006). "Relative risks and the market for sex: Teenagers, sugar daddies, and HIV in Kenya". Mimeo, Dartmouth College

5 Donald, S. and K. Lang (2007). "Inference with Difference-in-Differences and Other Panel Data." *The Review of Economics and Statistics*, 89(2).

6 Ordinary Least Squares and the various methods that employ it (such as Difference-in-Difference and Fixed Effects estimators) estimate the conditional mean of the outcome given treatment and control variables. Quantile regression allows for analysis of the expected effects of the treatment and control variables on the outcome for observations away from the mean (non-average); for example, we might be interested in the impact of treatment on

heterogeneity, reducing the threat of omitted variables bias.[7] Our analysis will account for circumstances that contaminate the initial randomization, such as attrition, spillovers and partial compliance, through the use of appropriate econometric methods.

<u>Degree of accuracy needed for the purpose described in the justification</u>

A sufficient sample size is needed in order to precisely estimate whether intervention had an impact on the outcomes of interest. We have estimated the number of individuals needed in our sample by conducting power calculations. The power of a statistical test is the probability that the null hypothesis will be rejected conditional on it being incorrect. Because the selection of a test statistic affects both the power of the test (the likelihood of rejecting the null hypothesis when it is true) and the required sample size, the test statistic must be selected in a way to minimize sample size while maximizing power. While larger sample sizes improve the ability to detect effects and choose between hypotheses, it is also more costly to collect. To that end, power calculations help guide researchers to appropriately select sample sizes that will have sufficient power to detect some minimum detectable treatment effect (explained more below).[8] Specifically,

$$n \geq \left(\frac{z_{\alpha/2} - \Phi^{-1}(1-\tau)}{MDE/\sigma_\beta}\right)^2$$

$z_{\alpha/2}$ is the test statistic associated with the selected size of the test. The research team has selected 5% as the statistical significance level for the evaluation, which is the probability of detecting a positive or negative effect (using a two-tailed test) when in fact the intervention has no effect. This is the probability of a type-1 (false positive) error generally accepted as significant in peer-reviewed journal articles. The power of the research design is the probability that, for a given effect size and a given statistical significance level, we will be able to reject the hypothesis of zero effect if the product or service provision does indeed have a real effect. The power is represented by τ. Φ^{-1} is the inverse cumulative density function for the standard normal distribution.

The research team has determined the minimum detectable treatment effect (MDE), the smallest effect size that we would possibly want to detect, for our intervention of primary interest, the Credit Matters Loan.[9] It is important to estimate this conservatively so that the analysis can detect relevant changes in outcomes of interest, even if lower than expected. Programs with a small effect size require a larger sample and if the sample size is underestimated there may be too much noise in the data to detect a non-zero effect (even if there is one).

Determining the minimum detectable effect is an important process that incorporated input from both SLCCU and the research team. For the outcome measure of change in FICO (Fair, Isaac and Company) credit score, the research team has set the MDE at a 35 point increase (which represents an approximately 6% increase) based on de-identified credit score data provided by Cities for Financial Empowerment (see

an individual in the 20th percentile, or quantile, of the outcome variable. Quantile regression allows us to do that. See Buchinsky, M. (1998). "Recent Advances in Quantile Regression: A Practical Guideline for Empirical Research." *Journal of Human Resources*, 33(1).

7 Hausman, J. (1978). "Specification Tests in Econometrics." *Econometrica*, 46(6).

8 Bloom, Howard S. (2006). The Core Analytics of Randomized Experiments for Social Research. MDRC Working Paper. Bloom presents the results in terms of the Minimum Detectable Effect. Given our exposition purposes, it is here expressed in terms of the sample size

9 We expect interaction effects across treatments to be negligible.

below).

Since this is the first rigorous study of its kind, it is difficult to predict precisely what the observed effect size might be. Anecdotal evidence, however, suggests it is reasonable to expect to see effect sizes in excess of 35 points. Innovative Changes, a non-profit organization that helps economically vulnerable individuals manage financial needs, advertises that consumers with low credit scores can expect a credit builder loan to increase their credit score by 37 points on average.[10] Mission Asset Fund reports that participants in their credit building lending circles saw their credit scores increase by 49 points on average.[11] [12]

To estimate the necessary sample size, we also need to estimate the standard error of the estimated program effect. To do so, the research team established a mean and variance of the outcome in the absence of the evaluation. For the outcome of change in FICO score, the research team has established that the mean FICO score in the absence of the evaluation for representative SLCCU members would be approximately 604, with standard deviation of 91.9 points. This data is based on de-identified data from Cities for Financial Empowerment (CFE) financial counseling clients, which was used as representative data given that SLCCU does not consistently collect credit scores on its members, and that the population of CFE clients and SLCCU members are similar on observable characteristics, such as income level. For the outcome of change in savings, the research team has established, based on de-identified SLCCU administrative data, that mean savings levels (including both checking and savings balances) in the absence of the evaluation for representative SLCCU members likely to participate in our study is $514, with a standard deviation of $1,035.

When randomization is over encouragement to participate and not treatment itself (as in this case), we also must estimate the correlation between treatment and incentives, using guidance from the partner institution on how responsive individuals will be to encouragement to participate. We will first ask respondents about interest in the Credit Matters loan. Based on discussions with SLCCU, we have estimated that 66% of SLCCU members who self-selected into the survey in response to recruitment materials focused on credit building will express interest in the loan. This group will then be randomized into two groups, one that receives additional encouragement to take up the loan, and one group that must complete an additional step before opening the product. We have estimated that 40% of the group that is encouraged to take up the Credit Matters loan will do so, and that a negligible portion of the group that must complete an additional step will ultimately take up the loan despite their treatment assignment.

Using these inputs, the research team determined the level of power for which we are aiming ex-ante and determined the necessary sample size. In order to detect the MDE outlined above in credit score with 95% confidence, the sample must be approximately 2,400 total respondents. (Note that respondents here means individuals that complete the survey, and that the assumptions in the paragraph above indicate how many of those individuals we expect to take up the product). Additionally, with a sample of 2,400 participants, we will also be able to observe increases in savings (including both checking and savings balances) of approximately $400, roughly two-thirds of the secured loan amount.

Unusual problems requiring specialized sampling procedures

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

---

10 http://www.innovativechanges.org/about/what-we-do/responsible-consumer-loans/credit-builder-loan
11 http://www.consumer-action.org/news/articles/2012_p2p_lending_survey/
12 Rigorous evaluation of credit-building products is still necessary to determine the impact of these products on credit scores with precision.

<u>Any use of periodic (less frequent than annual) data collection cycles to reduce burden</u>

The evaluation design relies in part on administrative data, in order to reduce data collection burdens. Soft credit report pulls will be conducted with the consent of consumers in order to learn about changes in credit scores and account information will be reported from SLCCU. Additionally, surveys will be administered (in order to collect information on outstanding debt, self-reported savings, and information on financial attitudes, financial knowledge, and behavioral indicators). We will examine changes in outcomes for both treatment and control groups based on baseline and endline surveys, but will make data collection as efficient and un-burdensome as possible, based in part on previous experience of both the RAND Corporation and IPA, and in part through the piloting process.

3. Methods to Maximize Response Rates and Address Issues of Non-Response

Given that the mandate of this project is to evaluate existing products and services, we will be implementing an "encouragement design," wherein consumers are encouraged to take up the Credit Matters Loan product that is already available to them. In order to experimentally estimate the impact of treatment offer, we will measure the average effect of "intent to treat" (ITT). This effect differs from "treatment on the treated" (TOT). Those who are non-compliant with the treatment of encouragement, i.e. do not take up the product, are accounted for under this estimate, as are "crossovers," members of the control group (not encouraged) who take up the product or service anyways.[13] In this case of randomizing over encouragement, we will estimate the effects of the program using instrumental variable regression to estimate Local Area Treatment Effect (LATE). This is a relevant consideration for policymaking decisions as the products and services we evaluate are voluntary, and not mandated; furthermore, even mandated programs do not have complete compliance.

Given the evaluation design, we will be able to take advantage of in-person interviews at SLCCU for the baseline survey. We will deal with systematic non-response to the endline survey with state-of the art statistical techniques that account for non-random sample attrition. Again, in order to maximize survey response rates, we will offer non-monetary incentives in the form of gift cards to encourage participation Participants will have the choice between a gift card to a grocery store, or gas station (with attached convenience store). More importantly, our primary outcome of interest (change in credit scores) will be measured using administrative data that is not subject to typical evaluation attrition.

4. Testing of Procedures or Methods

The baseline survey has been piloted to refine and streamline the data collection process, making sure that the most relevant information is collected from participants in the most efficient way possible. The baseline survey instrument is attached to this posting as an appendix. It contains questions on personal data (for the purpose of verifying respondent identity and matching respondents to administrative data), demographic data, and financial preference data. The pilot instrument is also attached as an appendix to this posting; it contains the same information as the baseline survey, but without the section on personal data since no personally identifiable information (PII) was collected at the pilot stage.

Because the internal validity of the results of any evaluation is dependent on the quality of the data collected, we have piloted the survey instrument at an SLCCU branch that will not participate in the full study. OMB approved the request for the pilot under the generic fast track approval process (OMB#

---

[13] Bloom, Howard S. (2006). The Core Analytics of Randomized Experiments for Social Research. MDRC Working Paper.

3174-0024).   During the pilot we conducted 59 surveys.

We learned a number of lessons from the pilot. There were a few survey questions that elicited confusion and we have re-worded the questions to make them clearer to respondents. We have also re-ordered some survey questions to improve the time it takes to complete the survey.

During and after baseline survey data collection, the quality of the data will be systematically verified through "back-checking" (i.e., performing an audit by resurveying of a random subset of individuals who were already surveyed to ensure that responses are consistent). This will be conducted by a second and independent team of surveyors.

Data entry will be undertaken using software such as EPI-DATA, SQL or MS-ACCESS which have the capability of conducting range and consistency checks. The data will be double-entered, where each element of data (with personally identifiable data and de-identified data entered separately) is entered twice by separate data entry operators. Both rounds of entry are then compared, and all deviations between the two rounds must be re-entered and scrutinized. Data entry will be conducted following the field work. Any errors will be resolved by referring to the original questionnaire or re-surveying, depending on the nature of the error. The entered data is then checked for completeness and correctness by conducting a third entry on a sub-sample (10% of the total sample). This process will allow us to calculate the error rate of the final dataset. The data will be re-entered if the error rate is higher than 0.5%.

5 Contact Information for Statistical Aspects of the Design

This study is being conducted by two non-profit contractors, The RAND Corporation and Innovations for Poverty Action, under contract to the Consumer Financial Protection Bureau.

Contact information:

CFPB:
Sarah Bainton Kahn
Policy Analyst, Financial Empowerment
Consumer Education & Engagement
Consumer Financial Protection Bureau
Office: 202-435-7030
sarah.bainton@cfpb.gov

The RAND Corporation:
Principal Investigator: Kata Mihaly, PhD
Economist
RAND Corporation
1200 South Hayes Street
Arlington, VA 22202
Office: 703-413-1100 x5393
Email: kmihaly@rand.org

Co-Principal Investigator: Jeremy Burke, PhD
Economist
RAND Corporation
1200 South Hayes Street

Arlington, VA 22202
Office: 703-413-1100 x5328
Email: jeremyb@rand.org

Innovations for Poverty Action:
Principal Investigator: Dean Karlan, PhD
Professor of Economics
Yale University
P.O. Box 208269
New Haven, CT 06520
Office: (203) 432-4479
Email: dean.karlan@yale.edu