

Supporting Statement for Paperwork Reduction ACT Submission
FAMILY SELF-SUFFICIENCY EVALUATION
OMB# 2528-0296

B. COLLECTION OF INFORMATION USING STATISTICAL METHODS

B1. Respondent Universe, Sampling Selection, and Expected Response Rates

The Family Self-Sufficiency (FSS) evaluation, funded by the US Department of Housing and Urban Development (HUD), has enrolled 2,551 households across 18 public housing authorities (PHAs). The implementation research will cover all the sites in the evaluation. On-site data collection will be conducted at nine PHAs, and data will be collected by phone for the other nine PHAs. With respect to staff interviews, MDRC expects to be able to interview the supervisor(s) and 2-3 case managers at each of the nine sites we will visit in person and the supervisor at each of the other nine sites where interviews are conducted by phone.

Additionally, for the sites selected for on-site data collection, the implementation research team will conduct interviews with up to 10 participants at each PHA, either individually or in a focus group. MDRC will select a random sample for the participant interviews, and the sample will be stratified based on work status at study enrollment and program engagement. Letters will be mailed out to selected sample members encouraging them to volunteer to be interviewed. MDRC will review the interview logistics with the nine sites selected for participant interviews.

For data on the sample that will be derived from administrative records, we expect high data coverage for both the program and control groups. Similarly, based on prior experience, we don't anticipate response issues with the staff interviews. With respect to the participant interviews, again based on prior experience conducting interviews with voucher holders, we expect that interviews will be completed with at least 7 or more participants at each of the nine sites (thus allowing for some non-response – and this potential response rate is also informed by the evaluation's recently completed 36-month survey, which achieved a high response rate – 77 percent).

B2. Procedures for Data Collection and Statistical Analysis

The study involves randomly assigning a total 2,656 households to one of two groups:

- **FSS group.** These individuals have access to the core elements of the FSS program – case management as well as rent escrow provisions.
- **Control group.** These individuals will not be enrolled in FSS and will not have access to FSS case management or escrow.

Random assignment occurred after participants signed the informed consent and completed a baseline survey. The impact sample consists of 2,556 households from the total of 2,656 that enrolled.

Statistical Impact Analysis

The core impact analysis for the evaluation will assess the overall and independent effects of the FSS program by comparing the key outcomes of this treatment group to the outcomes of the control group. The study will track both the program and the control groups for a number of years using administrative and survey data to measure outcomes.

The impact analysis will examine the program's effects on a wide range of outcomes, including education, employment, income, financial security, and material well-being.

Statistical models

As noted in Supporting Statement A and detailed in the original OMB submission for the FSS evaluation (OMB control number 2528-0296, Expiration Date 07/31/2016), this evaluation is a randomized controlled trial. The power of this experimental design comes from the fact that random assignment ensures that the treatment and control groups are alike in all aspects of the distribution of observed and unobserved baseline and pre-baseline characteristics. As a result, any post-baseline differences between the two groups can be interpreted as effects of the intervention.

The estimation strategy for survey-based outcomes is the same as that used for those collected from administrative records. We will use regression adjustment to increase the power of statistical tests that are performed, in which the outcome, such as "employment during Year 1" is regressed on an indicator for program group status and a range of other background characteristics.

The general form of the regression models which will be used to estimate program impacts is as follows:

$$Y_i = \alpha + \beta P_i + \delta X_i + \epsilon_i$$

where

Y_i is the outcome measure for sample member i ;

P_i equals one for program group members and zero for control group members;

X_i is a set of background characteristics for sample member i ; and

ϵ_i is a random error term for sample member i .

The coefficient β is interpreted as the impact of the program on the outcome. The regression coefficients, δ , reflect the influence of background characteristics.

We may vary the functional form and estimation method depending on the scale of measurement of the outcome for which impacts are estimated; for example, continuous outcomes will be estimated using Ordinary Least Squares (OLS) regression. We can use a more complex set of methods depending on the nature of the dependent variable and the type of issues being addressed, such as: logistic regressions for binary outcomes (e.g., employed or not); Poisson regressions for outcomes that take on only a few values (e.g., months of employment); and quantile regressions to examine the distribution of outcomes for continuous outcomes.

The evaluation will examine many outcomes across a number of domains. When multiple outcomes are examined, the probability of finding statistically significant effects increases, even when the intervention has no effect. For example, if 10 outcomes are examined in a study of an ineffective treatment, it is likely that one of them will be statistically significant at the ten percent level by chance.

While the statistical community has not reached consensus on the appropriate method of correcting for this problem, we propose to address it by being parsimonious in our selection of outcome variables. In particular, we plan to identify a set of “primary” outcomes and subgroups before beginning the impact analysis. All other outcomes and subgroups will be considered “secondary” and will be used to provide context for the primary impact findings or to generate hypotheses about impacts. Schochet (2008) suggests that this strategy is flexible enough to credibly test the key hypotheses about the program, while at the same time allowing the analyst to examine a range of outcomes in a more exploratory manner in order to uncover policy-relevant information.

The main analysis will be conducted over the pooled sample across all PHAs. While sample sizes will not permit PHA-specific impacts, MDRC will examine impacts for certain clusters of programs, such as smaller versus larger programs, programs that have a strong focus on employment, or programs that are operating in strong versus weak local labor markets. The ability to conduct this type of analysis depends on the variation in program features and contexts across participating PHAs, which will be critical to capture in the implementation analysis.

Minimum Detectable Effect Size

MDRC has enrolled 2,656 households, across the 18 PHAs participating in the evaluation, with half assigned to the program group and half assigned to a control group. As noted above, we derive a research sample size of 2,551 households from the full sample. MDRC will work to achieve a survey response rate of at least 80 percent, creating an effective survey sample of approximately 2,000 households. A sample size of 1,000 per research group is large enough to detect policy relevant impacts on outcomes measured through the survey both for the survey sample as well as key subgroups.

It is useful to consider the concept of Minimum Detectable Effects (MDEs) to explore the size of program impacts that are likely to be observed or detected for a set of outcomes and a given sample size. Since these are estimates, the actual MDEs may be smaller or larger than what is shown here. The estimates shown are likely to be conservative, since they assume that baseline

variables are not used in the impact model to improve precision. Pre-random assignment values of key outcomes, such as employment and earnings, are likely to be highly predictive of post-random assignment values of the same outcome. In this case, the increased precision brought about by including these variables in the impact model can reduce the MDEs considerably.

Table 1 presents MDEs for the proposed sample size. The first column presents data for the survey sample and the second column presents data for a subgroup within the survey sample, assuming that subgroup makes up half of the larger sample. The first several rows present MDEs for work, education and earnings. For the survey sample, the evaluation could detect effects (increases or decreases) as small as 5.5 percentage points on employment rate at the time of survey. Because sample sizes are more likely to be smaller for subgroups, the MDE for a subgroup is somewhat larger, at 8.0 percentage points. MDEs for earnings are shown in the table but are harder to predict, given the difficulty of predicting the variance of earnings. The final row presents MDEs in terms of effect sizes (or the impact on a given outcome divided by the standard deviation of that outcome). Effect sizes are a useful way to present and compare impacts on outcomes that are measured in different units, such as family well-being scales. For each of the proposed sample sizes, the effects sizes are typically considered small to moderate in the evaluation literature.

In sum, the proposed sample size is adequate for detecting effects on a range of outcomes that are relatively modest but still but meaningful from a policy standpoint. This pattern holds for the survey sample and for key subgroups.

The Family Self-Sufficiency Evaluation
Table 1
Minimum Detectable Effects
Research Group i versus Research Group j

	Survey Sample (1000 per group)	Subgroup (500 per group)
<u>Percentage point effects</u>		
Employed at time of survey	5.5	8.0
Has any degree or diploma	5.3	7.7
<u>Dollar effects</u>		
Earnings	1,002	1,417
Effect size	0.11	0.16

Notes: MDEs are calculated based on a two-tailed significance test and assuming an R-squared in the impact model of 0. Average values for employment, education and earnings outcomes are taken from the Opportunity NYC Work Rewards sample. Average values are 44% for employment at time of survey; 63 percent for degree or diploma, and \$6,874 (standard deviation of \$9,500) for earnings. Effect sizes are measured as the impact on a given outcome divided by its standard deviation

B3. Maximizing Response Rates and Issues of Nonresponse

Administrative data sources do not rely on participant response.

For the staff and participant interviews proposed under this submission, MDRC will work with each housing agency to select a diverse mix of study participant to serve as key informants. Using a purposive sampling approach, selection of study participants will focus on identifying a sample with varying program experiences and outcomes and those who reflected different work statuses (employed/unemployed) at the time of study enrollment. Given the small sample size, this approach will allow the research to learn and understand experiences from a diversity of perspectives and circumstances.

During the interview, participants will be encouraged to share their experiences; informed that there are no program consequences (i.e., loss of benefits) for not answering any particular question; and told that their name will not be associated with any information that they provide. Based on MDRC's experience conducting qualitative research, individuals who participate in these types of interviews are interested in sharing their experiences, making non-response less of an issue

B4. Pre-Testing

The interview protocols are designed to serve as semi-structured interview guides for discussions with program staff and participants. The development of these guides has been informed by the evaluation team's understanding of the program literature, existing evidence, and deep familiarity with the FSS programs in the study. The development of the protocols is also informed by the team experience drafting protocols for field research. Once in the field, we expect to try the protocols at one site, take stock, and then see if any adjustments are necessary before the interview guides are used in other sites.

B5. Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data

Overall, the information for the FSS evaluation is being collected by MDRC and its subcontractors. With HUD oversight, MDRC and its subcontractors are responsible for developing the study documents included in this submission. The statistical aspects of the overall evaluation design were developed in consultation with MDRC senior economist and impact analyst, Cynthia Miller.