

COMMUNITY ADVANTAGE PANEL SURVEY
TECHNICAL SAMPLING REPORT:
Owners, 2003-2010

March 2011

By
Sarah F. Riley
HongYu Ru

Center for Community Capital
The University of North Carolina at Chapel Hill

Prepared with financial support from the Ford Foundation

Center for Community Capital
Research and analysis on the transformative power of capital



UNC
COLLEGE OF
ARTS & SCIENCES

Contents

1	Introduction	3
2	Self-Help Lending in Context	4
3	Survey Design and Sampling Methods	7
3.1	Overview	7
3.2	The Baseline Interview	8
3.3	Follow-Up Interviews	9
3.4	Survey Attrition	12
4	Weight Construction	16
4.1	Overview of Weight Construction	16
4.2	Base Weights	16
4.3	Baseline Non-Response Weights	17
4.4	Baseline Post-Stratification Weights	20
4.5	Baseline Final Weights	20
4.6	Weights for Years 1-7	21
4.7	Using the Weights	22
5	Baseline CAPS Owners vs. Owners in the Current Population Survey	27
5.1	Overview	27
5.2	Data Preparation	27
5.3	Differences and Similarities	28

List of Tables

1	Owners Survey Modules and Completes by Year	8
2	Unweighted Owner Baseline Demographics	13
3	Owners Gross Response Rates by Survey Year and Cohort	14
4	Attriters vs. Non-Attriters	15
5	Fitted Logistic Regression Predicting Baseline Non-Response	18
6	Summary Statistics for Baseline Non-Response Weights	19
7	Summary Statistics for Baseline Final Weights	21
8	Effect of Baseline Non-Response Weights on Key Auxiliary Measures	23
9	Baseline Post-stratification Weighting Cells	24
10	Effect of Post-Stratification Weights on Key Self-Help Variables	25
11	Effect of Non-Response Weights and Post-Stratification Weights on Key Base- line Survey Demographics	26
12	Comparison of Weighted CAPS and CPS Owners Demographics	30

1 Introduction

In 1998, the Ford Foundation, Self-Help, Fannie Mae, and the UNC Center for Community Capital (the Center) formed a partnership, the Community Advantage Program (CAP), to demonstrate the feasibility of extending mortgage credit to low-income and minority households. Since 2003, and with the support of the Ford Foundation, the Center has gathered panel survey data from a sample of CAP mortgage recipients and a group of matched renters¹ in order to evaluate this program. The product of this data collection is the Community Advantage Panel Survey.

The information collected by the survey has varied by survey year, with the exception of the Universal Core module, which has been administered each year to collect key demographic information. Aside from the Universal Core, the primary survey modules concern the social and financial impacts of home ownership, including mortgage details, neighborhood quality, social capital, financial literacy, wealth and asset accumulation, and changes in household spending patterns as a result of economic conditions.

This document² provides a technical overview of the design of the Community Advantage Panel Survey (CAPS), as it pertains to the CAPS home owners.³ This overview discusses sampling methods, selection criteria, and the construction of weights. In addition, it discusses CAP in the context of other programs designed to serve the needs of prospective low-to-moderate-income homeowners and then compares the weighted summary statistics of key CAPS owner demographics with those from a comparable subset of the Current Population Survey (CPS). These comparisons put the experiences of CAPS owners in context with respect to the national low-to-moderate income and minority population of owners in the US.

¹Renters were matched to owners based on baseline geographic location and an income ceiling.

²We wish to thank Andy Peytchev at RTI International for helpful comments on a prior version. Any errors remain our own.

³A separate document discusses the matched renters.

2 Self-Help Lending in Context

Established in 1980, Self-Help is a nonprofit organization that advocates for and facilitates lending to individuals and communities that are unlikely to be served by mainstream financial institutions. Among other lending activities, Self-Help operates a secondary market program for private lenders who make home mortgage loans to individuals in low-income or minority communities.

The institutions with which Self-Help has partnered as a part of this secondary market program originate and typically service loans that are then purchased by Self-Help. Self-Help eventually sells many of these loans to Fannie Mae while retaining recourse. Although the specific lending rules that are applied at loan origination vary by originating institution, most of these loans are made to borrowers whom Self-Help targets as a part of its mission. In particular, Self-Help's charitability criteria stipulate that these borrowers must meet the following criteria:

- The borrower's household income is at or below 80% of the US Department of Housing and Urban Development's (HUD's) household area median income (AMI) of the metropolitan statistical area (MSA) or, for non-MSAs, the borrower's household income is at or below 80% of the HUD's household AMI for that county. AMI's have to be based on the loan date year.

OR

- The borrower's household income is greater than 80% but less than or equal to 115% of the HUD's household AMI for that MSA *AND* fulfills one or more of the following conditions:
 - The borrower lives in a low-income census tract: low-income census tracts located in MSA's are tracts where household median income is less than 80% of the MSA's

household median income. For non-MSA census tracts, a low-income census tract or county is one where household median income is less than 80% of the greater of the state or national non-MSA median household income. AMI's have to be based on the loan date year.

- The borrower lives in a minority census tract: a minority census tract is one where minority composition exceeds 30%. Census tracts have to be based on the loan date year.
- The borrower is a minority. That is, the borrower is Native American, Asian, Native Hawaiian or Other Pacific Islander, African American, Hispanic, or Other.

In addition, although Self-Help loans do require full documentation, they are available to borrowers who may have lower credit ratings or higher debt-to-income ratios than conventional mortgage loan products require, and in some cases they require little or no down payment. Furthermore, Self-Help provides its own mortgage insurance on loans that have very low down payments.

Thus, Self-Help's goals toward improving access to capital are similar in spirit to the historical goals of the Government Sponsored Entities (such as Fannie Mae and Freddie Mac), the Community Reinvestment Act, and the Federal Housing Authority. While each of these organizations or legislations has operated in a slightly different way, each has sought to make it easier for people who would not otherwise be served by the capital markets to get loans.

More specifically, the charters of Fannie Mae and Freddie Mac have historically specified three primary goals for the mortgage purchases that these organizations make. These goals are as follows:

- Low-to-moderate income goal: At least X percent of the dwelling units financed should be for families with income no greater than the area median income, defined as the median income for the metropolitan area or non-metropolitan county. For 2009, the

target fraction was $X = 43\%$.

- Special affordable goal: At least X percent of the dwelling units financed should be for very-low-income families (those with incomes not greater than 60 percent of area median income) or low-income families (those with income not greater than 80 percent of area median income) in low-income areas. For 2009, the target fraction was $X = 18\%$.
- Under-served areas goal: At least X percent of the dwelling units financed should be located in low-income Census tracts to low-to-middle-income Census tracts with high minority populations. For 2009, the target fraction was $X = 32\%$.

Similarly, the Community Reinvestment Act (CRA) of 1977, which was subsequently revised in 1989 and 1995, encourages lending institutions to meet the credit needs, as well as the depository needs, of the communities in which they operate. Banks undergo an evaluation assessing the extent of their compliance with the CRA, and for large banks this evaluation consists of three tests: a lending test, an investment test, and a service test. Non-commercial, residential loans made to families must be made to borrowers who either have household incomes no greater than 80% of area median income or live in census tracts where the median income is no greater than 80% of area median income, in order to qualify under the CRA lending test.

Finally, the Federal Housing Authority (FHA) underwrites home mortgage loans made to low-to-moderate-income households who might not otherwise qualify for financing. As with Self-Help loans, FHA loans typically have lower down payment and credit score requirements, as well as more flexible debt-to-income ratio requirements, than conventional mortgage products.

Because Self-Help's lending activities do dovetail so closely with these federal programs, the CAP survey is positioned to provide researchers with an opportunity to inform government policy by studying survey participants.

3 Survey Design and Sampling Methods

3.1 Overview

The target population of the survey is a set of nearly 29,000 borrowers who received Self-Help loans (the Self-Help Generalization Sample) that originated between 1998 and 2004 and met the Self-Help charitability criteria discussed in the previous section. A subset of 7,223 of these borrowers were put into calling at the beginning of the survey, as the rest of the loans in the Self-Help Generalization Sample had not yet been purchased by Self-Help and were, therefore, not available for sampling at that time. From among these 7,223 cases, 3,743 owners completed the baseline interview.

As of the beginning of 2011, eight years' worth of survey data had been collected. Key demographic information was collected each year via the Universal Core module, but otherwise the survey modules have varied from year to year. In particular, the table below presents an overview of the survey modules of questions that have been fielded each year and the mode of survey administration, as well as the total number of respondents who completed the survey in each year.

Table 1: Owners Survey Modules and Completes by Year

Year	Survey Year	Modules	Mode	Completes
2003	0	Home purchasing info Counseling	Phone	3,743
2004	1	Social Capital, Parenting	Phone	2,614
2005	2	Savings Wealth and Assets	Phone In-home	2,701 1,284
2006	3	Mortgages Rising Energy Costs Financial Literacy Sense of Community	Phone	2,380
2007	4	Social Capital 2, Parenting 2 Medical Costs, Credit Scores Collective Efficacy	Phone	2,079
2008	5	Wealth and Assets 2 Mortgages 2, Savings 2 Housing Experiences Home Improvements Foreclosures, Stress	Phone/In-home	2,376
2009	6	Economic Challenges	Phone	2,229
2010	7	Economic Challenges 2	Phone	2,088

Note: The Universal Core was administered each year.

3.2 The Baseline Interview

The baseline (Year 0)⁴ interviews were conducted during the period from March 2001 to January 2004. A set of 7,223 owner cases were provided to the Survey Research Unit at UNC-Chapel Hill (SRU), the organization that the Center contracted to carry out the actual data collection activities. With the exception of a few loans that were excluded from the sample for data abnormalities⁵, these 7,223 owners comprise all Self-Help borrowers whose loans met the following criteria:

- Mortgage originated between September 3, 1999, and May 21, 2003.

⁴We label the survey years beginning with Year 0, rather than Year 1, so that the survey years for owners and their matched renters are comparable.

⁵Efforts were made to exclude college students and loans that were made for the purpose of a refinance rather than for a purchase. In addition, a handful of originators, including the NC State Employees Credit Union, were excluded.

- First payment date between November 1, 1999, and July 1, 2003.
- Purchased by Self-Help between February 10, 2000, and May 31, 2003.
- Income and racial characteristics conform to the Self-Help charity criteria discussed in the previous section.

Because Self-Help acquires loans on an on-going basis, the baseline sample of 7,223 loans was released to SRU for calling in four sequential groups, or draws. The first draw consisted of the 988 eligible loans that were available in the fall of 2000. The second, third, and fourth draws each consisted of 2,438 loans, 2,625 loans, and 1,172, respectively.

Of these 7,223 cases that were put into calling, a total of 3,743, or 52%, completed the baseline survey. Table 2 below presents unweighted frequencies and proportions for the group of borrowers who completed the baseline survey. Slightly less than half of the respondents are female, and approximately 40% are minorities. At the time of the baseline interview, over 73% of respondents were between the ages of 18 and 40. Nearly all of the respondents were employed, and the majority (62%) were living in the South. Thirty-four percent of respondents had received at most a high school diploma or GED, while 24% had obtained a bachelor's degree or graduate degree. About half of the sample was living with a spouse or partner, and about half reported a household size of two members or fewer. The majority of respondents received annual incomes between \$20,000 and \$50,000, with only 13% making more than \$50,000.

3.3 Follow-Up Interviews

For the first follow-up interview (Year 1), only those owners who had not moved since loan origination were asked to complete the survey. The original intent was to track owners who remained owners and who retained their original Self-Help loans. This requirement was relaxed in later survey years in order to allow information about mobility to be collected. The Year 1 survey was conducted by phone and collected information about social capital and

parenting behaviors, as well as the demographic and financial information that is collected every year as part of the Universal Core module. A total of 2,614 owners completed the Year 1 survey, for a response rate of 70%, relative to the pool of eligible owners who had completed baseline.

The second follow-up survey (Year 2) collected information about savings behavior, wealth and asset accumulation, and mortgages. Unlike the previous two surveys, the Year 2 survey was divided into two parts, namely a telephone interview that collected universal core and savings information, and an in-person interview that collected some demographics information and data for the other two survey modules. In particular, it was thought that an in-home collection of wealth and assets data might both make respondents more comfortable in discussing their finances and also result in more accurate data because the respondents could check their financial records during the interviews, if necessary. Individuals who had moved since the prior interview were also administered an additional Movers module to collect information about mobility.

While SRU administered the telephone survey, the Center also contracted with RTI International (RTI) to conduct the in-person interviews. Aside from the specific survey questions that were asked and the mode of administration, these two surveys differ with respect to the group of cases that were considered eligible. Primarily those owners who had been previously matched with a comparison group of renters based on income and location and had not previously been coded as a hostile refusal were eligible for the in-home survey. The decision to restrict the in-home survey to this subgroup of owners and to a small number of unmatched rural owners was made in order to contain survey costs. As a result, 2,701 owners completed the Year 2 phone survey, while 1,284 owners completed the Year 2 in-home survey. A subgroup of 1,086 owners completed both of these surveys.

The third follow-up survey (Year 3) was conducted by phone and was administered by SRU. In Year 3, all owners who completed the baseline survey were given the opportunity to participate, regardless of whether or not they had moved since baseline. The Year 3

survey collected information about family experiences with energy costs, financial literacy, and sense of community. A total of 2,380 owners completed the Year 3 survey.

In conjunction with the Year 3 survey, a supplemental version of the Year 3 survey was administered by RTI via telephone to those owners who had refused to complete either or both of the Year 2 surveys but had not refused in a hostile manner or asked to be permanently removed from the survey panel. This supplemental, or 'soft-refusal,' survey included the survey questions from Year 2 concerning savings, mortgages, and wealth and assets, in addition to the Year 3 survey questions, as a means of recovering these data for these non-respondents. Overall, 262 soft-refusal cases completed this supplemental survey.

The fourth follow-up survey (Year 4) was likewise administered by RTI over the phone and collected data about medical costs, credit scores, and collective efficacy, in addition to a second round of questions concerning social capital and parenting. These latter questions paralleled those that were asked in Year 1 and provide a means of assessing change over time. Of the 3,183 owners who were eligible for the Year 4 survey, a total of 2,079 owners completed the Year 4 interview.

In Year 5, a total of 2,796 owners were determined to be eligible for the survey. To be considered eligible, these owners had to have participated in more than two of the prior survey years and to have never refused participation in a hostile manner or asked to be permanently removed from the survey panel. The Year 5 survey collected information about wealth and assets, mortgages, and savings behavior that was similar to the data that had been collected in the Year 2 and soft-refusal surveys. Those owners who had previously completed the Year 2 in-home survey were interviewed in person, while all other Year-5 eligibles were given the opportunity to participate by phone. A total of 2,376 owners completed the Year 5 survey, yielding a gross response rate of 85%.

The Year 6 survey was administered by phone and collected information about changes in household financial conditions and expectations for the future. A total of 2,639 owners were eligible for the Year 6 survey. To be considered eligible, these owners had to have been

eligible for the Year 5 survey and to have never refused participation in a hostile manner or asked to be permanently removed from the survey panel. A total of 2,229 owners completed the Year 6 survey, for a gross response rate of 84%.

The Year 7 survey likewise was administered by phone and collected information about changes in household financial conditions and expectations for the future. A total of 2,592 owners were eligible for the Year 7 survey. These owners had been eligible for the Year 6 survey and had never refused participation in a hostile manner or asked to be permanently removed from the survey panel. A total of 2,088 owners completed the Year 7 survey, for a gross response rate of 81%.

Table 3 presents gross response rates for each of these groups, measured as the number of completes relative to the total number of cases initially put into calling in each year. The table also summarizes which institution fielded each of the data collections.

3.4 Survey Attrition

Those survey respondents who completed an interview in all seven survey years (i.e., non-attriters) differ in several respects from all other survey participants who completed the baseline survey (i.e., attriters). In particular, those respondents who consistently participated were more likely to be female, White, and slightly older on average. In addition, consistent respondents were more likely to have completed at least a college degree and less likely to have not completed high school. Finally, consistent respondents were slightly more likely to have fewer than four household members and were less likely to be married or partnered. Attriters and non-attriters look similar with respect to the other key demographic distributions, including income, geography, and employment status, as indicated in Table 4.

Table 2: Unweighted Owner Baseline Demographics

Variable Name	<i>N</i>	Owners	%
Sex			
Male	2,019		53.9
Female	1,724		46.1
Age			
18-25 years old	718		19.3
26-30 years old	864		23.3
31-35 years old	645		17.4
36-40 years old	489		13.2
41-45 years old	351		9.5
46-50 years old	287		7.7
51-60 years old	248		6.7
61 years old or more	108		2.9
Race			
White	2,292		61.2
Black	730		19.5
Hispanic	595		15.9
Other	126		3.37
Education attainment			
11th grade or less	370		9.9
High school graduate/GED	903		24.1
Some 2-year college	660		17.6
2-year degree	510		13.6
Some 4-year college	393		10.5
Bachelors degree	555		14.8
Some graduate school	115		3.1
Graduate degree	222		6.0
Vocational or other license	13		0.4
Marital status			
Partner or Companion	416		11.1
Married	1,714		45.9
Widowed	67		1.8
Divorced	596		16.0
Separated	74		2.0
Never Married	870		23.3
Household size			
1	833		22.3
2	1,083		28.9
3	761		20.3
4	581		15.5
5+	485		13.0
Employment status			
Working	3,456		92.3
Looking for Work (Unemployed)	120		3.2
Retired	63		1.7
Out of Labor Force	104		2.8
Geographic coverage			
Midwest	956		25.6
Northeast	97		2.6
South	2,301		61.5
West	386		10.3
Income			
Less than \$20,000	382		10.2
\$20,000-25,000	531		14.2
\$25,000-30,000	582		15.6
\$30,000-40,000	906		24.2
\$40,000-50,000	845		22.6
\$50,000 or more	497		13.3

Note: The income buckets are upward inclusive.

Table 3: Owners Gross Response Rates by Survey Year and Cohort

Survey Year	Cohort	Fielded By	Initial Eligibles	Completes	Response Rate
0	2003 Phone	SRU	7,223	3,743	52%
1	2004 Phone	SRU	3,743	2,614	70%
2	2005 Phone	SRU	3,718	2,701	73%
	2005 In-home	RTI	1,671	1,190	71%
	2005 Movers In-home	RTI	244	93	38%
3	2006 Phone	SRU	2,596	2,118	82%
	2006 Soft Refusals Phone	RTI	1,237	262	21%
4	2007 Phone	RTI	3,183	2,079	65%
5	2008 Phone	RTI	1,617	1,296	80%
	2008 In-home	RTI	1,179	1,080	92%
6	2009 Phone	RTI	2,639	2,229	84%
7	2010 Phone	RTI	2,592	2,088	81%

Table 4: Attriters vs. Non-Attriters

Variable Name	Attriters		Non-Attriters	
	N	%	N	%
Sex**				
Male	1,395	56.5	621	48.7
Female	1,072	43.5	655	51.3
Age**				
18-25 years old	500	20.5	218	17.1
26-30 years old	600	24.6	264	20.8
31-35 years old	423	17.4	222	17.4
36-40 years old	323	13.3	166	13.0
41-45 years old	213	8.7	138	10.8
46-50 years old	161	6.6	126	9.9
51-60 years old	148	6.1	100	7.9
61 years old or more	69	2.8	39	3.1
Race**				
White	1,432	58.0	860	67.4
Black	478	19.4	252	19.8
Hispanic	466	18.9	129	10.1
Other	91	3.7	35	2.7
Education attainment**				
11th grade or less	286	11.6	84	6.6
High school graduate/GED	603	24.5	300	23.5
Some 2-year college	419	17.0	241	18.9
2-year degree	345	14.0	165	12.9
Some 4-year college	279	11.3	114	8.9
Bachelors degree	320	13.0	235	18.4
Some graduate school	71	2.9	44	3.5
Graduate degree	129	5.2	93	7.3
Vocational or other license	13	0.5	0	0.0
Marital status*				
Partner or Companion	284	11.5	132	10.3
Married	1,163	47.3	551	43.2
Widowed	42	1.7	25	2.0
Divorced	352	14.3	244	19.1
Separated	59	2.4	15	1.2
Never Married	561	22.8	309	24.2
Household size**				
1	511	20.7	322	25.2
2	698	28.3	385	30.2
3	493	20.0	268	21.0
4	411	16.7	170	13.3
5+	354	14.3	131	10.3
Employment status				
Working	2,275	92.2	1,181	92.6
Looking for Work (Unemployed)	85	3.4	35	2.7
Retired	41	1.7	22	1.7
Out of Labor Force	66	2.7	38	3.0
Geographic coverage				
Midwest	629	25.5	327	25.6
Northeast	57	2.3	40	3.1
South	1,492	60.6	809	63.4
West	286	11.6	100	7.9
Income				
Less than \$20,000	257	10.4	125	9.8
\$20,000-25,000	352	14.3	179	14.0
\$25,000-30,000	396	16.1	186	14.6
\$30,000-40,000	563	22.8	343	26.9
\$40,000-50,000	565	22.9	280	21.9
\$50,000 or more	334	13.5	163	12.8

Note: * p-value is less than 0.05, ** p-value is less than 0.01.

The income buckets are upward inclusive.

4 Weight Construction

4.1 Overview of Weight Construction

To construct sampling weights for the owners who completed the CAP survey, we follow standard survey weighting methodology. The weighting process involves four major steps:

- Construct the base weights. These weights take into account the probability that an individual was asked to participate in the survey.
- Construct the non-response weights. These weights correct for some of the bias that may be present in substantive survey variables as a result of differences in the attributes of respondents and non-respondents.
- Construct the post-stratification weights. These weights correct for some of the bias that may result from under-coverage in the surveyed sample relative to a larger target group of individuals to whom the results should ideally be generalizable.
- Construct the final weights as the product of the base weights, the non-response weights, and the post-stratification weights.

In constructing the non-response weights, we use a logistic regression to estimate response propensities. In constructing the post-stratification weights, we use a weighting class approach.⁶ We construct the latter weighting classes based on a handful of auxiliary variables that we believe to be of key importance for analysis of the survey.

4.2 Base Weights

Because the baseline owners survey pool was deliberately selected to meet specific criteria and is not a random sample, the probability of selection for each of these individuals is

⁶For discussions of common approaches to weighting, see [Biemer and Christ \(2008\)](#), [Kalton \(1986\)](#), and [Kalton and Flores-Cervantes \(2003\)](#).

identical and equal to 1. In other words, all members of the frame population were selected for inclusion in the survey. Therefore, the baseline weight for individual i in the sampling frame is given by

$$w_{Bi} = \frac{1}{\text{probability of selection}} = 1$$

4.3 Baseline Non-Response Weights

To construct baseline non-response weights, we consider auxiliary variables that are available from the Self-Help database, as well as survey paradata, such as the subgroup, or draw (1, 2, 3 or 4), of Self-Help loans in which the case was originally released to SRU for calling. The Self-Help variables include some demographics, loan characteristics, and mortgage payment behavior and are available for the majority (99%)⁷ of both respondents and non-respondents. The paradata are available for all respondents and non-respondents.

To select auxiliary variables that are related to the baseline response propensity, we perform a logistic regression with a response indicator as the dependent variable. The primary predictors of baseline response, or those that are consistently statistically significant at the 5% level or better, along with their corresponding odds ratios, are presented in Table 5.⁸ Location was the most important predictor of baseline response. Individuals located in North Carolina were much more likely to complete the survey than individuals in other

⁷The small number of cases for which no auxiliary data are available comprise those cases that disappeared from the Self-Help database following a database upgrade at the end of 2006.

⁸We considered the following additional variables while refining the regression model but found these to be insignificant predictors: annual income, relative income at the MSA level, credit scores, age, origination year, whether or not there was a coborrower, the original loan-to-value ratio, worst-ever delinquency, the date on which Self-Help acquired the loan, the mortgage note date, and urban/rural status. The states North Carolina and California were selected for inclusion because these are the top two states with regard to loan concentration in the Generalization Sample and because there appear to be strong predictive effects associated with these states. We also considered the next two largest states, Oklahoma and Ohio, but these were not significant predictors of non-response.

Table 5: Fitted Logistic Regression Predicting Baseline Non-Response

Variable (N=7,134)	Coefficient (Std. Error)	Odds Ratio
Intercept	0.39 (0.05)	.
Male	-0.18 (0.05)	0.83
Race/Ethnicity		
(White or Black)	.	.
Hispanic or Other Minority	-0.25 (0.06)	0.78
Geography		
(Other States)	.	.
North Carolina	0.43 (0.06)	1.53
California	-0.73 (0.10)	0.48
Self-Help Loan Draw		
(Draw 1 or Draw 4)	.	.
Draw 2 or Draw 3	-0.23 (0.05)	0.79
Debt-to-income ratio		
(Debt-to-income ratio \leq 38%)	.	.
Debt-to-income ratio $>$ 38%	-0.31 (0.10)	0.73

NOTE: All coefficients are significant at the 1% level. This analysis excludes observations for which Self-Help data are not available.

states, possibly because the panel survey is orchestrated by the University of North Carolina. Similarly, eligibles in California were among the least likely to complete the survey. In addition, males and non-Black minority groups, as well as borrowers with very high debt-to-income ratios were also somewhat less likely than females, Whites and Blacks, and those with debt-to-income ratios below 38% to complete the baseline interview. Finally, those cases that were initially put into calling during the middle of the baseline data collection period (Draws 2 and 3) were less likely to complete the survey than those cases in Draws 1 and 4.

Using the propensity scores generated from this logistic regression, we created weights to correct for baseline non-response. Specifically, the weight for each baseline case is the inverse of its propensity score, for those cases for which the Self-Help data are available. For those 35 respondent cases for which Self-Help data are not available, we assign a weight of $\frac{89}{35}$, as there are 89 cases in the larger baseline sample of 7,223 eligibles for which Self-Help data

are also missing.

More formally, the baseline non-response weight for each respondent i is the inverse of the probability that he responded to the baseline survey ($y_i = 1$), given the covariates x_i , as follows:

$$w_{NRi} = \frac{1}{\hat{p}_i}$$

where

$$\hat{p}_i = Prob(y_i = 1|x_i) = \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)}$$

and where α and β are the set of estimated coefficients of the logistic regression. Summary statistics for the baseline nonresponse weights are provided in Table 6.

Table 6: Summary Statistics for Baseline Non-Response Weights

Statistic	Value
N	3,743
Sum	7,221
Minimum	1.4
Maximum	4.7
Mean	1.9
Median	1.9
Standard Deviation	0.46
Variance Inflation Factor ($F = 1 + CV^2$)	1.06

The effect of the baseline weights on the key auxiliary variables can be seen from Table 8. The first column in the table presents means and proportions for the unweighted set of baseline survey completers, while the second column presents comparable summary statistics for the full set of cases that were originally put into calling for the baseline interview. The third column presents a weighted version for the cases in the first column.

As indicated in Table 8, the weighting process brings many of the sample means and proportions of the baseline survey respondents more in line with those of the baseline group of eligibles who were put into calling. In particular, we observe increases in mean credit

score, income, and original property value and in the proportion of males in the sample, while we observe decreases in the mean original loan-to-value ratio and in the proportion of Whites in the sample. The mean age, original note rate, debt-to-income ratio, and relative income remain largely unchanged.

4.4 Baseline Post-Stratification Weights

We create the baseline post-stratification weights with the aim of bringing the baseline-non-response-weighted CAP survey data in line with the Self-Help Generalization Sample from the perspective of key lending criteria. In particular, we create these weights based on credit score, debt-to-income ratio, and original loan-to-value ratio, for those cases for which Self-Help data are available. The post-stratification weight w_{PSi} for individual i is given by the inverse of the frequency of weighted CAPS respondents to the frequency of individuals in the Self-Help Generalization Sample meeting the criteria for a given cell. Those 89 cases for which Self-Help data are not available are assigned the average post-stratification weight of all other cases. The frequencies for each of the post-stratification weighting cells are presented in Table 9.

4.5 Baseline Final Weights

The baseline final weights are the product of the base weights, the baseline non-response weights, and the baseline post-stratification weights. More formally, the final weight for case i is given by

$$w_i = w_{Bi} * w_{NRi} * w_{PSi}$$

Summary statistics for the baseline final weights for all owners retained in the survey are presented in the table below.

The marginal effects of the post-stratification weights relative to that of the baseline

Table 7: Summary Statistics for Baseline Final Weights

Statistic	Value
N	3,743
Sum	28,843
Minimum	4.8
Maximum	47.3
Mean	7.7
Median	6.7
Standard Deviation	3.4
Variance Inflation Factor ($F = 1 + CV^2$)	1.2

non-response weights on the key Self-Help variables can be seen in Table 10, which compares the panel with final weights to both the Self-Help Generalization Sample and the panel with baseline non-response weights. Similarly, Table 11 presents a comparison of how each set of weights affects key demographic baseline survey responses.

4.6 Weights for Years 1-7

The final weight construction process for Years 1-7 mimics that of the baseline weights. However, the set of variables used as predictors of non-response differs to take account of additional variables and different non-response patterns, relative to baseline. In particular, for each subsequent year of data collection, we derive response propensity scores based on respondent race, gender, education and age at baseline, the number of calls placed to the respondent at baseline, and whether the respondent’s loan had ever been delinquent as of the last interview of the respondent for the survey year(s) in question. For example, a weight for respondents who completed Year 1 and Year 5, given that they had completed baseline, would take into consideration whether they had ever been delinquent as of the Year 5 interview date. The demographic variables were selected because they are systematically related to attrition in the survey, as discussed in section 3.4. The initial call attempts and the mortgage delinquency information were added because these non-survey variables are

consistently significant predictors of attrition in this sample.

After correcting for non-response, we include an additional, intermediate post-stratification correction based on weight classes, so that the race and relative income distribution of a given set of survey respondents matches the corresponding population totals for the 7,223 individuals who were put into calling at baseline. As done for the baseline final weights, the final post-stratification step then weights the respondents for credit score, debt-to-income ratio, and original loan-to-value ratio, up to the Self-Help Generalization Sample.

More formally, let the w_{Bi} and w_{NRi} respectively denote the base weight and the baseline non-response weight, as before. In addition, let w_{NRij} denote the non-response weight for a respondent who completed the survey in selected subsequent years j . Finally, let w_{PS1ij} and w_{PS2ij} respectively denote the two post-stratification adjustments mentioned above. Then, the final weight for a survey respondent i in year(s) j following baseline is given by

$$w_{ij} = w_{Bi} * w_{NRi} * w_{NRij} * w_{PS1ij} * w_{PS2ij}$$

4.7 Using the Weights

The final survey weights are designed to reduce any bias in the survey data that may be related to demographic and lending measures and to unit non-response. These survey weights do not address any bias that may result from item missing data in specific survey questions. Therefore, we recommend that researchers multiply-impute any item-missing data in their subsample of interest prior to applying the weights. For discussions of multiple imputation and other methods of dealing with item non-response, see [Rassler and Riphahn \(2006\)](#), [Little \(2003\)](#), and [Graham, Olchowski, and Gilreath \(2007\)](#).

Table 8: Effect of Baseline Non-Response Weights on Key Auxiliary Measures

Variable	Unweighted Panel		Baseline Eligibles		Panel Weighted for Baseline Non-Response	
	N	Mean	N	Mean	Sum of Weights	Mean
SH Race	3,629		7,134		6,962	
White		0.65		0.61		0.62
Black		0.18		0.16		0.16
Hispanic		0.14		0.19		0.18
Other Minority		0.03		0.04		0.04
SH Male	3,453	0.55	6,797	0.58	6,646	0.58
SH Age	3,612	33.37	7,094	33.66	6,930	33.28
18-25		0.22		0.21		0.22
26-30		0.18		0.19		0.18
31-35		0.30		0.28		0.29
36-40		0.10		0.10		0.10
41-45		0.08		0.09		0.08
46-50		0.05		0.05		0.05
51-60		0.05		0.05		0.05
> 60		0.02		0.02		0.02
SH Income	3,702	\$31,853	7,121	\$32,436	7,122	\$32,247
<= \$20K		0.12		0.11		0.12
\$21-30K		0.36		0.36		0.35
\$31-40K		0.34		0.33		0.33
\$41-50K		0.12		0.13		0.12
> \$50K		0.06		0.07		0.07
SH Relative Income (% AMI)	3,686	59.10	7,077	59.26	7,089	59.32
<= 50		0.34		0.33		0.34
51-80		0.57		0.58		0.57
81-100		0.06		0.06		0.06
> 100		0.03		0.03		0.03
SH Credit Score	3,500	674.50	6,680	674.76	6,696	675.65
No Score		0.06		0.06		0.06
<= 580		0.05		0.05		0.05
581-620		0.12		0.11		0.11
621-660		0.23		0.23		0.23
661-720		0.32		0.32		0.32
> 720		0.23		0.22		0.23
SH Debt-to-Income Ratio (%)	3,688	0.27	7,077	0.27	7,085	0.27
<= 38		0.94		0.92		0.92
39-50		0.05		0.07		0.07
> 50		0.01		0.01		0.01
SH Original LTV (%)	3,708	96.09	7,134	95.70	7,132	95.89
<= 90		0.10		0.12		0.11
91-95		0.07		0.08		0.07
96-97		0.44		0.45		0.46
> 97		0.39		0.35		0.36
SH Original Value	3,708	\$84,011	7,134	\$88,863	7,132	\$87,637
<= \$50K		0.15		0.14		0.14
\$51-75K		0.32		0.31		0.31
\$76-100K		0.30		0.28		0.28
\$101-200K		0.22		0.24		0.25
> \$200K		0.01		0.03		0.02
SH Original Note Rate (%)	3,708	7.6	7,134	7.6	7,132	7.6
< 6		0.01		0.01		0.01
6-7		0.35		0.35		0.35
8-9		0.52		0.54		0.53
> 9		0.11		0.11		0.10
SH Geographic Coverage	3,708		7,133		7,132	
Midwest		0.26		0.26		0.26
Northeast		0.03		0.03		0.03
South		0.61		0.57		0.57
West		0.10		0.15		0.15

Note: The letters 'SH' indicate that the variable comes from the Self-Help data, rather than from survey responses. Note: The buckets for each attribute are contiguous. For example, a bucket labeled '39-50' actually includes the set (38,50]. Note: Mean origination credit scores are calculated for the subset of observations that are non-missing and, therefore, excludes individuals who had no origination credit score.

Table 9: Baseline Post-stratification Weighting Cells

No Credit Score	Debt-to-Income Ratio	Weighted Panel	SH Generalization Sample	<i>wPS</i>
OLTV				
≤ 95%	≤ 38%	71.35	633	8.87
> 95%	> 38%	12.38	142	11.47
	≤ 38%	317.86	1,328	4.18
	> 38%	343.40	266	7.73
Credit Score ≤ 720				
OLTV	Debt-to-Income Ratio			
≤ 95%	≤ 38%	766.28	3,730	4.87
> 95%	> 38%	125.72	687	5.46
	≤ 38%	3,959.21	13,115	3.31
	> 38%	217.61	989	4.54
Credit Score > 720				
OLTV	Debt-to-Income Ratio			
≤ 95%	≤ 38%	252.42	2,078	8.23
> 95%	> 38%	72.51	444	6.12
	≤ 38%	1,205.52	4,648	3.86
	> 38%	96.38	431	4.47
Missing SH Data		89	352	3.95
Total Cases		7,223	28,843	3,743
Mean Weight				3.95
Variance Inflation Factor ($F = 1 + CV^2$)				1.09

Note: The frequency of missing data in the Self-Help Generalization Sample is an estimate based on the average post-stratification weight of the other cases.

Table 10: Effect of Post-Stratification Weights on Key Self-Help Variables

Variable	Panel Weighted for Baseline Non-Response		SH Generalization		Panel with Final Weights	
	Sum of Weights	Mean	N	Mean	Sum of Weights	Mean
SH Race	6,962		28,014		27,650	
White		0.62		0.57		0.61
Black		0.16		0.17		0.16
Hispanic		0.18		0.18		0.19
Other Minority		0.04		0.08		0.04
SH Male	6,646	0.58	26,996	0.56	26,424	0.58
SH Age	6,930	33.39	27,856	34.96	27,535	33.82
18-25		0.22		0.21		0.21
26-30		0.18		0.20		0.18
31-35		0.29		0.20		0.29
36-40		0.10		0.13		0.11
41-45		0.08		0.07		0.08
46-50		0.05		0.07		0.06
51-60		0.05		0.03		0.05
> 60		0.02		0.02		0.02
SH Income	7,122	\$32,247	27,640	\$32,541	28,441	\$31,904
<= \$20K		0.12		0.12		0.13
\$21-30K		0.35		0.35		0.35
\$31-40K		0.33		0.32		0.33
\$41-50K		0.12		0.14		0.12
> \$50K		0.07		0.07		0.07
SH Relative Income (% AMI)	7,089	59.32	27,303	60.32	28,277	58.47
<= 50		0.34		0.31		0.36
51-80		0.57		0.59		0.55
81-100		0.06		0.07		0.06
> 100		0.03		0.03		0.03
SH Credit Score	6,696	675.65	26,122	682.35	26,122	680.60
No Score		0.06		0.08		0.08
<= 580		0.05		0.04		0.04
581-620		0.11		0.10		0.10
621-660		0.23		0.21		0.21
661-720		0.32		0.31		0.29
> 720		0.23		0.27		0.27
SH Debt-to-Income Ratio (%)	7,085	0.27	27,245	0.28	28,224	0.28
<= 38		0.92		0.89		0.90
39-50		0.07		0.09		0.09
> 50		0.01		0.02		0.02
SH Original LTV (%)	7,132	95.89	28,488	94.07	28,491	94.53
<= 90		0.11		0.18		0.16
91-95		0.07		0.09		0.11
96-97		0.46		0.42		0.41
> 97		0.36		0.31		0.32
SH Original Value	7,132	\$87,637	28,488	96,687	28,491	90,400
<= \$50K		0.14		0.11		0.14
\$51-75K		0.31		0.27		0.29
\$76-100K		0.28		0.28		0.28
\$100-200K		0.25		0.31		0.26
> \$200K		0.02		0.04		0.03
SH Original Note Rate (%)	7,132	7.6	28,491	7.3	28,491	7.6
< 6		0.01		0.07		0.02
6-7		0.35		0.39		0.36
8-9		0.52		0.50		0.52
> 9		0.10		0.05		0.10
SH Geographic coverage	7,132		28,490		28,491	
Midwest		0.26		0.15		0.24
Northeast		0.03		0.04		0.03
South		0.57		0.62		0.56
West		0.15		0.19		0.17

Note: The letters 'SH' indicate that the variable comes from the Self-Help data, rather than from survey responses. Note: The buckets for each attribute are contiguous. For example, a bucket labeled '39-50' actually includes the set (38,50]. Note: Mean origination credit scores are calculated for the subset of observations that are non-missing and, therefore, excludes individuals who had no origination credit score.

Table 11: Effect of Non-Response Weights and Post-Stratification Weights on Key Baseline Survey Demographics

Variable	Unweighted Panel		Panel weighted for Baseline Non-Response		Panel with Final Weights	
	N	Mean	Sum of Weights	Mean	Sum of Weights	Mean
Survey Race	3,743		7,221		28,843	
White		0.61		0.58		0.58
Black		0.20		0.18		0.18
Hispanic		0.16		0.20		0.21
Other Minority		0.03		0.04		0.04
Survey Male	3,743	0.54	7,223	0.57	28,843	0.56
Survey Age	3,700	35.09	7,149.84	35.12	28,528	35.56
18-25		0.19		0.19		0.18
26-30		0.23		0.23		0.23
31-35		0.17		0.17		0.18
36-40		0.13		0.14		0.14
41-45		0.09		0.09		0.10
46-50		0.08		0.08		0.08
51-60		0.06		0.07		0.07
> 60		0.03		0.03		0.03
Survey Income	3,743	\$36,078	7,221	\$36,186	28,843	\$35,862
<= \$20K		0.10		0.10		0.11
\$21-30K		0.30		0.29		0.30
\$31-40K		0.24		0.25		0.25
\$41-50K		0.23		0.22		0.21
> \$50K		0.13		0.14		0.13
Survey Relative Income (% AMI)	3,740	0.65	7,213	0.64	28,813	0.64
<= 50		0.31		0.31		0.33
51-80		0.45		0.45		0.45
81-100		0.14		0.14		0.13
> 100		0.09		0.09		0.09
Survey Education	3,741		7,217		28,831	
11th grade or less		0.10		0.11		0.12
HS/GED		0.24		0.24		0.24
Some 2-yr college/trade-school		0.18		0.17		0.17
Associate/trade-school degree		0.14		0.13		0.13
Some 4-yr college		0.11		0.10		0.10
Bachelor's degree		0.15		0.15		0.15
Some grad school		0.03		0.03		0.03
Graduate or professional degree		0.06		0.06		0.06
Survey Marital Status	3,737		7,209		28,792	
Partner or Companion		0.11		0.11		0.11
Married		0.46		0.48		0.48
Widowed		0.02		0.02		0.02
Divorced		0.16		0.15		0.15
Separated		0.02		0.02		0.02
Never Married		0.23		0.22		0.23
Survey Employment Status	3,743		7,221		28,843	
Working		0.92		0.92		0.91
Looking for work (unemployed)		0.03		0.03		0.03
Retired		0.02		0.02		0.02
Out of labor force		0.03		0.03		0.03
Survey household size	3,743		7,221		28,843	
1		0.22		0.21		0.21
2		0.29		0.28		0.28
3		0.20		0.20		0.20
4		0.16		0.17		0.16
5+		0.13		0.15		0.15

Note: The variables in this table indicate survey responses or variables derived from survey responses. They reflect SH data only when survey responses were missing due to item non-response and SH data were available to fill in these missing values. Note: The buckets for each attribute are contiguous. For example, a bucket labeled '39-50' actually includes the set (38,50].

5 Baseline CAPS Owners vs. Owners in the Current Population Survey

5.1 Overview

In this section, we compare the weighted sample of owners who completed baseline to the weighted Current Population Survey (CPS) for May 2003, as this administration of the CPS falls roughly at the median of the baseline owner interview dates.

Because CAPS was intended to be representative of households meeting Self-Help lending criteria and certain age requirements, we subset the CPS using these criteria so that the two samples can be more readily compared with respect to key demographic attributes. The largest differences between the CAPS owners and the CPS owners involve geographic coverage, educational attainment, and age.

5.2 Data Preparation

The Current Population Survey is a survey of 50,000 households that is conducted monthly by the Census Bureau and the Bureau of Labor Statistics. The survey is designed to be representative of the non-institutionalized civilian population in the US, and it collects information on demographics and other household characteristics in order to provide an integrated picture of the US labor force and its experiences.

Because the target population of the CPS is considerably more broad than that of CAPS, and because the head-of-household is the survey respondent for CAPS, we restrict the CPS to the reference person of each household where the reference person is an owner. The reference person is identified as the person whose name is on the property deed for that household. In addition, we exclude full-time students and respondents over the age of 65, as efforts were made to exclude such individuals from the CAPS pool. Finally, we further subset the CPS to include only those households that meet a simplified version of the Self-Help charity

criteria. In particular, the CPS household income must but be no greater than 80% of the area median income (AMI) at the MSA level or be no greater than 115% of the AMI if the respondent is a minority. Thus, the analysis below considers how similar the CAPS owners are to this refined group of CPS participants.

5.3 Differences and Similarities

Table 12 provides demographic proportions for both the CAPS owners who completed the baseline interview and owners in the CPS. Relative to the CPS owners, CAPS owners are slightly more likely to be male (56% vs. 50%) and considerably more likely to be Hispanic (21% vs. 16%). However, the proportions of Whites and Blacks in these two samples are roughly comparable.

In addition, CAPS owners tend to be younger and more educated on average than CPS owners and are also considerably more likely to be employed. About 40% of CAPS owners were 30 years of age or younger, compared to only 12% for the CPS. In addition, only 28% of CAPS owners were 41 years of age or older, while 65% of CPS owners fall into this category. With regard to educational attainment, CPS owners are nearly twice as likely as CAPS owners not to have finished high school. Moreover, CPS owners are only 70% as likely as CAPS owners to have completed at least some college. Over 90% of CAPS owners were employed as of the baseline interview, compared with 70% of CPS owners.

However, CAPS owners and CPS owners look very similar with respect to their overall household size distribution, and both have about a 50% likelihood of being married. The income distribution of CAPS owners is also similar to that of CPS owners, with the exception that only 11% of CAPS owners are in the lowest income bracket (< \$20,000), compared with 24% for the CPS.

The greatest difference between CAPS owners and the CPS sample concerns geographic coverage. CAPS has very little coverage in the Northeast, and it slightly over-represents the South.

Some of the above differences can be explained in part by noting that the CAPS respondents have only recently become home owners, while the CPS does not provide a means of excluding those respondents who may have owned their homes for a longer period of time. In particular, the fact that CAPS owners are somewhat younger on average than their CPS counterparts is likely related to this tenure limitation. Similarly, that CAPS owners are more likely never to have been married and are more likely to be participating in the labor force may easily be related to this average difference in age.

Overall, we can say that CAPS owners are largely representative of the US non-student, non-institutionalized, low-to-moderate income population of recent home owners below the age of 65 who also would have met Self-Help's charitability criteria as of May 2003, if one considers income, race, and gender. However, the CAPS sample emphasizes individuals who are slightly younger and better educated and much less likely to be from the Northeast than is the case for this general population. These differences should be kept in mind by researchers who make inferences from the CAPS data.⁹

⁹These qualitative conclusions do not change if retired and/or rural (i.e., non-MSA) respondents are excluded from the CPS.

Table 12: Comparison of Weighted CAPS and CPS Owners Demographics

Variable Name	CAPS (<i>N</i> = 28K) %	CPS (<i>N</i> = 19M) %
Sex		
Male	56.2	49.6
Female	43.8	50.4
Age		
18-25 years old	18.0	4.7
26-30 years old	22.5	6.9
31-35 years old	17.6	11.1
36-40 years old	14.1	12.7
41-45 years old	9.5	14.3
46-50 years old	8.0	13.6
50-60 years old	7.0	24.3
61 years old or more	3.3	12.4
Race		
White	57.6	58.7
Black	17.6	16.8
Hispanic	21.1	16.3
Other	3.7	8.2
Education attainment		
11th grade or less	11.8	17.5
High school graduate/GED	24.4	35.4
Some college or associate degree	40.2	27.9
Bachelor's degree or higher	23.6	19.1
Marital status		
Married	53.4	51.3
Widowed	2.1	5.8
Divorced	16.9	20.1
Separated	2.2	5.4
Never Married	25.5	17.2
Household size		
1	21.2	25.1
2	27.7	25.9
3	19.8	17.4
4	16.5	17.0
5+	14.9	14.7
Employment status		
Working	91.4	70.2
Looking for Work (Unemployed)	3.4	4.6
Retired	2.0	8.3
Out of Labor Force	3.2	16.8
Geographic coverage		
Midwest	24.3	21.3
Northeast	2.8	15.7
South	56.3	40.2
West	16.6	22.8
Income		
Less than \$20,000	11.1	23.8
\$20,000-25,000	13.8	11.9
\$25,000-30,000	15.8	12.8
\$30,000-40,000	24.8	26.5
\$40,000-50,000	21.0	10.9
\$50,000 or more	13.4	14.1

Note: The income buckets are upward inclusive.

Note: All differences are statistically significant with $p < 0.0001$.

Note: Sample sizes presented for each sample are the sum of 30 weights.

References

- Biemer, Paul P., and Sharon L. Christ, 2008, Weighting survey data, in Edith de Leeuw, Joop J. Hox, and Don A. Dillman, ed.: *International Handbook of Survey Methodology* (L. Erlbaum Associates, New York).
- Graham, John W., Allison E. Olchowski, and Tamika D. Gilreath, 2007, How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory, *Prevention Science* 8.
- Kalton, Graham, 1986, Handling Wave Nonresponse in Panel Surveys, *Journal of Official Statistics* 2(3).
- , and Ismael Flores-Cervantes, 2003, Weighting Methods, *Journal of Official Statistics* 19(2).
- Little, Roderick J., 2003, Bayesian methods for unit and item nonresponse, in R. L. Chambers, and C. J. Skinner, ed.: *Analysis of Survey Data* (John Wiley and Sons, Ltd.).
- Rassler, Susanne, and Regina T. Riphahn, 2006, Survey Item Nonresponse and Its Treatment, *Allgemeines Statistisches Archiv* 90.