

TECHNICAL REPORT

Community Advantage Panel Survey
Technical Sampling Report:
Owners, 2003-2014

Sarah F. Riley

May 2015

Funding provided by the Ford Foundation



The UNC Center for Community Capital at the University of North Carolina at Chapel Hill is the leading center for research and policy analysis on the transformative power of capital on households and communities in the United States.

The center's in-depth analyses help policymakers, advocates and the private sector find sustainable ways to expand economic opportunity to more people, more effectively.

Roberto G. Quercia, Director
Lucy Gorham, Executive Director

UNC Center for Community Capital
The University of North Carolina at Chapel Hill
1700 Martin Luther King Blvd | Campus Box 3452, Suite 129
Chapel Hill NC 27599-3452 | (877) 783-2359 | (919) 843-2140
communitycapital@unc.edu | www.ccc.unc.edu

Contents

1	Introduction	3
2	Self-Help Lending in Context	4
3	Survey Design and Sampling Methods	8
3.1	Overview	8
3.2	The Baseline Interview	8
3.3	Follow-Up Interviews	9
3.4	Survey Attrition	13
4	Weight Construction	18
4.1	Overview of Weight Construction	18
4.2	Base Weights	18
4.3	Baseline Non-Response Weights	19
4.4	Baseline Post-Stratification Weights	21
4.5	Baseline Final Weights	24
4.6	Weights for Years 1-11	25
5	CAPS vs. CPS and AHS	27
5.1	Overview	27
5.2	Data Preparation	27
5.3	Differences and Similarities	28
5.3.1	Current Population Survey	28
5.3.2	American Housing Survey	29

List of Tables

1	Unweighted Owner Baseline Demographics	15
2	Owners Gross Response Rates by Survey Year and Cohort	16
3	Attriters vs. Non-Attriters	17
4	Fitted Logistic Regression Predicting Baseline Non-Response	20
5	Summary Statistics for Baseline Non-Response Weights	21
6	Baseline Post-stratification Weighting Cells	23
7	Summary Statistics for Baseline Final Weights	24
8	Effect of Non-Response and Final Weights on Key Baseline Survey Demographics	26
9	Comparison of Weighted CAPS and CPS Owners Demographics	31
10	Comparison of Weighted CAPS and AHS Owners Demographics	32

1 Introduction

In 1998, the Ford Foundation, Self-Help, Fannie Mae, and the UNC Center for Community Capital (the center) initiated the Community Advantage Secondary Mortgage Market Demonstration 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 subset of CAP mortgage recipients (the owners sample) and a group of matched renters (the renters sample)¹ in order to evaluate this program. The product of this data collection is the Community Advantage Panel Survey (CAPS).

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 CAPS as it pertains to the owners sample.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 comparable subsets of the Current Population Survey (CPS) and the American Housing Survey (AHS). 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 have the following characteristics:

- 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 Minority.

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 govern-

ment 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 were 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 2015, twelve years' worth of owners survey data have been collected in total. Key demographic information was collected each year via the Universal Core module, but otherwise the survey modules have varied from year to year. See the survey specifications (i.e., questionnaires) for details about which modules were administered in each survey 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 was 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 had been purchased by Self-Help at the time of the baseline survey. These loans have the following characteristics:

- 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 acquired loans on a rolling 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 1 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, about 73% of respondents were between the ages of 18 and 40. Nearly all of the respondents (92%) were employed, and the majority (62%) were living in the South. Thirty percent of respondents had received at most a high school diploma or GED, while 25% 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 7% 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 location and an income ceiling 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 panel⁶. 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

⁶Each year, a small number of cases also are excluded due to incapacity or death. The spouse is substituted for the original respondent in those cases where one is available and willing to complete the survey.

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. Survey participants were also excluded if they had not completed any interviews in the prior three years. 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%.

As in prior years, the phone survey conducted in Year 8 collected information about household financial conditions. Moreover, a series of new questions was added concerning attitudes toward debt repayment and strategic default. Eligibles were again selected as those individuals who had been eligible for the prior year's interview and who did not provide a hostile refusal. Of the 2,488 individuals who were eligible to be surveyed in Year 8, a total of 2,018 completed the survey. Thus, the response rate in Year 8 was 81%.

In Year 9, a total of 2,325 owners were determined to be eligible for the survey; the eligibility criteria were the same as those used in prior years, except that survey participants were also excluded if they had not completed any interviews in the prior three years and had not been located in the past two years. The Year 9 survey collected information about wealth and assets, mortgages, home improvements, and savings behavior that was similar to the data that had been collected in Years 2 and 5. Those owners who had previously completed the Year 2 or Year 5 in-home survey were interviewed in person, while all other Year-9 eligibles were given the opportunity to participate by phone. A total of 2,055 owners completed the Year 9 survey, yielding a gross response rate of 88%. Of these respondents, 1,141 completed the interview by phone, and 914 completed the interview in person. The

response rates by mode were 86% and 92%, respectively.

In Year 10, a total of 2,286 owners were determined to be eligible for the survey. To be considered eligible, these owners had to have been eligible for the Year 9 survey and to have never refused participation in a hostile manner or asked to be permanently removed from the survey panel. The Year 10 survey collected information about economic challenges, stress, financial services, moral hazards, and transportation habits. A total of 1,924 owners completed the Year 10 survey, yielding a gross response rate of 84%.

In Year 11, a total of 2,149 owners were determined to be eligible for the survey. To be considered eligible, these owners had to have been eligible for the Year 10 survey and to have never refused participation in a hostile manner or asked to be permanently removed from the survey panel. As in the previous year, the Year 10 survey collected information about economic challenges, stress, financial services, and moral hazards. It also collected social capital information similar to that collected in Year 4. A total of 1,832 owners completed the Year 11 survey, yielding a gross response rate of 85%.

Table 2 presents gross response rates for each of these groups, measured as the number of completed interviews relative to the total number of eligible 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 twelve 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. They were also slightly more likely to be located in the South and less likely to be located in the West. Finally, consistent respondents were 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 income and employment status, as indicated in Table 3.

Table 1: Unweighted Owner Baseline Demographics

Variable Name	<i>N</i>	Owners	%
Gender			
Male	2,019		53.9
Female	1,724		46.1
Age			
18-25 years old	709		19.0
26-30 years old	875		23.4
31-35 years old	657		17.6
36-40 years old	490		13.1
41-45 years old	359		9.6
46-50 years old	288		7.7
51-60 years old	248		6.6
61 years old or more	116		3.1
Race			
White	2,292		61.2
Black	730		19.5
Hispanic	595		15.9
Other	126		3.4
Education attainment			
11th grade or less	328		8.8
High school graduate/GED	776		20.8
Some 2-year college	628		16.8
2-year degree	540		14.4
Some 4-year college	492		13.2
Bachelors degree	567		15.2
Some graduate school	139		3.7
Graduate degree	239		6.4
Vocational or other license	32		0.9
Marital status			
Partner or Companion	416		11.1
Married	1,714		45.8
Widowed	67		1.8
Divorced	594		15.9
Separated	73		2.0
Never Married	875		23.4
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	443		11.8
\$20,000-25,000	623		16.6
\$25,000-30,000	676		18.1
\$30,000-40,000	1,313		35.1
\$40,000-50,000	446		11.9
\$50,000 or more	242		6.5

Note: The income buckets are upward inclusive.

Table 2: Owners Gross Response Rates by Survey Year and Cohort

Survey Year	Cohort	Fielded By	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	1,284	77%
3	2006 Phone	SRU	2,498	2,118	85%
	2006 Soft Refusals Phone	RTI	1,234	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%
8	2011 Phone	RTI	2,488	2,018	81%
9	2012 Phone	RTI	1,328	1,141	86%
	2012 In-home	RTI	997	914	92%
10	2013 Phone	RTI	2,286	1,924	84%
11	2014 Phone	RTI	2,149	1,832	85%

Table 3: Attriters vs. Non-Attriters

Variable Name	Attriters		Non-Attriters	
	N	%	N	%
Gender**				
Male	1,515	56.2	504	48.1
Female	1,181	43.8	543	51.8
Age**				
18-25 years old	546	20.3	163	15.6
26-30 years old	658	24.4	217	20.7
31-35 years old	477	17.7	180	17.2
36-40 years old	350	13.0	140	13.4
41-45 years old	239	8.9	120	11.5
46-50 years old	179	6.6	109	10.4
51-60 years old	163	6.1	85	8.1
61 years old or more	83	3.1	33	3.2
Race**				
White	1,603	59.5	689	65.8
Black	509	18.9	221	21.1
Hispanic	488	18.1	107	10.2
Other	96	3.6	30	2.9
Education attainment**				
11th grade or less	277	10.3	51	4.9
High school graduate/GED	599	22.2	177	16.9
Some 2-year college	437	16.2	191	18.2
2-year degree	382	14.2	158	15.1
Some 4-year college	359	13.3	133	12.7
Bachelors degree	375	13.9	192	18.3
Some graduate school	93	3.5	46	4.4
Graduate degree	150	5.6	89	8.5
Vocational or other license	22	0.8	10	1.0
Marital status*				
Partner or Companion	306	11.4	110	10.5
Married	1,271	47.2	443	42.3
Widowed	47	1.8	20	1.9
Divorced	388	14.4	206	19.7
Separated	60	2.2	13	1.2
Never Married	620	23.0	255	24.4
Household size**				
1	562	20.9	271	25.9
2	782	29.0	301	28.8
3	534	19.8	227	21.7
4	439	16.3	142	13.6
5+	379	14.1	106	10.1
Employment status				
Working	2,485	92.2	971	92.7
Looking for Work (Unemployed)	93	3.5	27	2.6
Retired	45	1.7	18	1.7
Out of Labor Force	73	2.7	31	3.0
Geographic coverage*				
Midwest	693	25.7	263	25.1
Northeast	67	2.5	30	2.9
South	1,632	60.6	669	63.9
West	301	11.2	85	8.1
Income				
Less than \$20,000	326	12.1	117	11.2
\$20,000-25,000	444	16.5	179	17.1
\$25,000-30,000	465	17.3	211	20.2
\$30,000-40,000	936	34.7	377	36.0
\$40,000-50,000	340	12.6	106	10.1
\$50,000 or more	185	6.9	57	5.4

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.⁷ 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

⁷For 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%)⁸ 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 4.⁹ 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

⁸The 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.

⁹We 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 4: Fitted Logistic Regression Predicting Baseline Non-Response

Variable (N=7,134)	Coefficient (Std. Error)	Odds Ratio
Intercept	0.39 (0.05)	.
Male	-0.19 (0.05)	0.83
Race/Ethnicity		
(White or Black)	.	.
Hispanic or Other Minority	-0.24 (0.06)	0.78
Geography		
(Other States)	.	.
North Carolina	0.43 (0.06)	1.53
California	-0.74 (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 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 5.

Table 5: Summary Statistics for Baseline Non-Response Weights

Statistic	Value
N	3,743
Sum	7,220
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

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 ratio 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 6](#).

Table 6: Baseline Post-stratification Weighting Cells

No Credit Score		Weighted Panel	SH Generalization Sample	<i>wPS</i>
OLTV	Debt-to-Income Ratio			
≤ 95%	≤ 38%	71.25	633	8.88
> 95%	> 38%	12.36	142	11.49
	≤ 38%	317.34	1,328	4.18
	> 38%	34.32	266	7.75
Credit Score ≤ 720				
OLTV	Debt-to-Income Ratio			
≤ 95%	≤ 38%	766.22	3,729	4.87
> 95%	> 38%	125.72	687	5.46
	≤ 38%	3,959.51	13,116	3.31
	> 38%	217.37	989	4.55
Credit Score > 720				
OLTV	Debt-to-Income Ratio			
≤ 95%	≤ 38%	252.50	2,078	8.23
> 95%	> 38%	72.49	444	6.12
	≤ 38%	1,205.79	4,648	3.85
	> 38%	96.41	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.

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 Table 7 below.

Table 7: Summary Statistics for Baseline Final Weights

Statistic	Value
N	3,743
Sum	28,843
Minimum	4.8
Maximum	47.2
Mean	7.7
Median	6.7
Standard Deviation	3.4
Variance Inflation Factor ($F = 1 + CV^2$)	1.2

Table 8 presents a comparison of how the baseline non-response and final weights affect the distribution of key demographic characteristics of the baseline survey sample. Overall, the weights tend to reduce the fractions of White and Black respondents while increasing the fraction of Hispanics and Other Minorities. They also increase the fraction of men in the sample and generate small increases in average respondent age, income, and household size. The distributions of educational attainment, marital status, and employment status remain largely unchanged.

4.6 Weights for Years 1-11

The final weight construction process for Years 1-11 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 also 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 was 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

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

Table 8: Effect of Non-Response and Final 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		7,220		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	7,220	0.57	28,843	0.56
Survey Age	3,685	35.19	7,217	35.21	28,830	35.65
18-25		0.19		0.19		0.18
26-30		0.23		0.23		0.23
31-35		0.18		0.18		0.18
36-40		0.13		0.14		0.14
41-45		0.10		0.10		0.10
46-50		0.08		0.08		0.08
51-60		0.07		0.07		0.07
> 60		0.03		0.03		0.03
Survey Income	3,743	\$32,057	7,220	\$32,515	28,843	\$32,171
<= \$20K		0.12		0.12		0.13
\$21-30K		0.36		0.35		0.35
\$31-40K		0.34		0.33		0.32
\$41-50K		0.12		0.12		0.12
> \$50K		0.06		0.07		0.07
Survey Relative Income (% AMI)	3,740	0.59	7,213	0.60	28,813	0.59
<= 50		0.34		0.34		0.36
51-80		0.57		0.56		0.55
81-100		0.06		0.06		0.06
> 100		0.03		0.03		0.03
Survey Education	3,741		7,217		28,831	
11th grade or less		0.09		0.10		0.10
HS/GED		0.21		0.21		0.21
Some 2-yr college/trade-school		0.17		0.17		0.17
Associate/trade-school degree		0.14		0.14		0.14
Some 4-yr college		0.13		0.13		0.12
Bachelor's degree		0.15		0.15		0.15
Some grad school		0.04		0.04		0.04
Graduate or professional degree		0.06		0.06		0.06
Survey Marital Status	3,739		7,213		28,806	
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		7,220		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		7,220		28,843	
1		0.22		0.21		0.21
2		0.29		0.28		0.28
3		0.20		0.20		0.20
4		0.15		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 CAPS vs. CPS and AHS

5.1 Overview

In this section, we compare the weighted sample of owners who completed baseline to the weighted May 2003 Current Population Survey (CPS) and the weighted 2003 American Housing Survey (AHS)¹⁰. Because CAPS was intended to be representative of households meeting Self-Help lending criteria and certain age requirements, we subset the CPS and the AHS using similar criteria so that these national surveys can be more readily compared to CAPS with respect to key demographic attributes. Statistically significant differences exist across these surveys with respect to all household characteristics considered. The largest discrepancy involves geographic coverage.

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.

The American Housing Survey (national sample) is a nationally representative survey of approximately 60,000 housing units that is administered by the Census Bureau every two years. The survey is designed to capture information about the characteristics of the US housing stock, including housing costs and physical attributes. The survey also captures the demographic characteristics of the people who inhabit each of the targeted housing units.

Because the target populations of the CPS and the AHS are considerably more broad than that of CAPS, and because the head-of-household is typically the survey respondent

¹⁰These administrations of the CPS and AHS fall roughly at the median of the baseline owner interview dates.

for CAPS, we restrict both the CPS and the AHS 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 and AHS to include only those households that meet a simplified version of the Self-Help charity criteria. In particular, the household income must be no greater than 80% of the area median income (AMI) 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 these restricted samples of CPS and AHS participants.

5.3 Differences and Similarities

5.3.1 Current Population Survey

Table 9 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 more likely to be male (56% vs. 50%) and less likely to be White (58% vs. 61%). In particular, CAPS respondents are considerably more likely to be Hispanic (21% vs. 16%).

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 41% of CAPS owners were 30 years of age or younger, compared to only 11% for the CPS. In addition, only 28% of CAPS owners were 41 years of age or older, while 66% 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 64% as likely as CAPS owners to have completed a bachelor's degree. Over 90% of CAPS owners were employed as of the baseline interview, compared with 69% of CPS owners.

CAPS owners and CPS owners look relatively similar with respect to their overall household size distribution, except that CAPS contains a lower proportion of one-person house-

holds (21% vs. 28%). Both samples have about a 50% likelihood of being married, while CAPS respondents are more likely to have never been married (26% vs. 17%) and less likely to be divorced (17% vs. 21%). The income distribution of CAPS owners is also similar to that of CPS owners, with the exception that only 13% of CAPS owners are in the lowest income bracket (< \$20,000), compared with 27% for the CPS, and 34% of CAPS owners are in the \$30-40K income bracket, compared with only 29% of CPS owners.

The greatest difference between CAPS owners and the CPS sample concerns geographic coverage. CAPS has very little coverage in the Northeast (3% vs. 16%), and it over-represents the South (56% vs. 40%).

Some of the above differences can be explained in part by noting that the CAPS respondents only recently became 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.

5.3.2 American Housing Survey

Table 10 provides demographic proportions for both the CAPS owners who completed the baseline interview and owners in the AHS. Relative to the AHS owners, CAPS owners are slightly more likely to be male (56% vs. 53%), more likely to be Hispanic (21% vs. 10%), and less likely to be White (58% vs. 69%). In addition, CAPS owners tend to be younger and more educated on average than AHS owners, and they are also more likely to be employed. About 40% of CAPS owners are 30 years of age or younger, compared to 9% for the AHS. In addition, only 28% of CAPS owners are 41 years of age or older, while 70% of AHS owners fall into this category. With regard to educational attainment, more than 68% of CAPS owners have completed at least some college, compared with 50% of AHS owners. Over

90% of CAPS owners are employed, compared with 63% of AHS owners.

CAPS owners and AHS owners look roughly similar with respect to their overall household size distributions, except that AHS households are less likely to have only one member. The income distribution of CAPS owners is also similar to that of AHS owners, with the exceptions that only 13% of CAPS owners are in the lowest income bracket (< \$20,000), compared with 25% for the AHS, and that 34% of CAPS owners are in the \$30-40K income bracket, compared with only 27% of AHS owners.

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

Table 9: Comparison of Weighted CAPS and CPS Owners Demographics

Variable Name	CAPS (<i>N</i> = 29K) %	CPS (<i>N</i> = 16M) %
Gender		
Male	56.2	50.1
Female	43.8	49.2
Age		
18-25 years old	18.0	4.1
26-30 years old	22.6	6.6
31-35 years old	17.8	10.4
36-40 years old	13.9	12.6
41-45 years old	9.8	14.1
46-50 years old	7.9	13.2
50-60 years old	6.9	25.2
61 years old or more	3.5	13.7
Race		
White	57.6	60.5
Black	17.6	16.2
Hispanic	21.1	16.5
Other	3.7	6.8
Education attainment		
11th grade or less	10.3	18.7
High school graduate/GED	21.4	37.3
Some college or associate degree	43.2	27.9
Bachelor's degree or higher	25.1	16.03
Marital status		
Married	53.4	49.9
Widowed	2.1	6.5
Divorced	16.8	21.2
Separated	2.1	5.5
Never Married	25.6	16.9
Household size		
1	21.2	27.7
2	27.7	26.9
3	19.8	17.1
4	16.5	15.8
5+	14.9	12.5
Employment status		
Working	91.4	68.8
Looking for Work (Unemployed)	3.4	4.6
Retired	2.0	9.2
Out of Labor Force	3.2	17.4
Geographic coverage		
Midwest	24.3	21.3
Northeast	2.8	15.7
South	56.3	40.2
West	16.7	22.8
Income		
Less than \$20,000	13.0	26.9
\$20,000-25,000	16.5	13.7
\$25,000-30,000	17.3	14.9
\$30,000-40,000	33.7	29.4
\$40,000-50,000	12.3	8.5
\$50,000 or more	7.2	6.6

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 weights.

Table 10: Comparison of Weighted CAPS and AHS Owners Demographics

Variable Name	CAPS ($N = 29K$) %	AHS ($N = 17M$) %
Gender		
Male	56.2	53.4
Female	43.8	46.6
Age		
18-25 years old	17.6	3.1
26-30 years old	22.8	6.2
31-35 years old	17.8	9.4
36-40 years old	13.9	11.3
41-45 years old	9.8	13.8
46-50 years old	7.9	13.8
50-60 years old	6.9	27.7
61 years old or more	3.5	14.7
Race		
White	57.6	69.3
Black	17.6	14.6
Hispanic	21.1	10.1
Other	3.7	5.9
Education attainment		
11th grade or less	10.3	16.7
High school graduate/GED	21.4	33.1
Some college or associate degree	43.2	29.8
Bachelor's degree or higher	25.1	20.4
Marital status		
Married	53.4	47.2
Widowed	2.1	7.6
Divorced	16.8	24.3
Separated	2.1	4.8
Never Married	25.6	16.1
Household size		
1	21.2	28.3
2	27.7	30.5
3	19.8	17.1
4	16.5	13.8
5+	14.9	10.3
Employment status		
Working	91.4	62.8
Looking for Work (Unemployed)	8.6	37.2
Geographic coverage		
Midwest	24.3	25.8
Northeast	2.8	17.1
South	56.3	38.3
West	16.6	18.8
Income		
Less than \$20,000	13.0	25.4
\$20,000-25,000	16.5	10.3
\$25,000-30,000	17.3	12.7
\$30,000-40,000	33.7	27.0
\$40,000-50,000	12.3	15.1
\$50,000 or more	7.2	9.6

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 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).

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).