POLICY ANALYSIS
FOR THE
PUBLIC AND PRIVATE
SECTORS
M.CUBED

Evaluation of East Bay Municipal Utility District's Pilot of WaterSmart Home Water Reports

Prepared by

David L. Mitchell
M.Cubed

Thomas W. Chesnutt, Ph.D., CAP™
A&N Technical Services, Inc.

Prepared for

California Water Foundation
&
East Bay Municipal Utility District

December 2013

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This report presents the results of an independent evaluation of the East Bay Municipal Utility District's (EBMUD) year-long pilot project (Pilot) of WaterSmart Software's Home Water Reports (HWRs) service. HWRs provide households with periodic information on their current water use and compare it to their past use, the average use of similar households, and the use of the most efficient similar households. This data is coupled with actionable information on ways to use water around the home more efficiently. HWRs aim to motivate households to reduce their water use through changes in behavior or adoption of more water efficient technology. The approach is based on research on social norms marketing coming out of the field of social psychology and for this reason we refer to these type of programs as social-norms-based (SNB) efficiency programs. While SNB efficiency programs have been broadly adopted by energy utilities across the United States in recent years, they are new to water utilities.

The EBMUD Pilot is the first relatively large-scale implementations of an SNB efficiency program by a large urban water utility, providing HWRs to 10,000 homes over a twelve-month period. The pilot was comprised of two experiments. The first we call the Random Group Experiment. The second we call the Castro Valley Group Experiment. In both experiments, households were selected to be in either a treatment group or a control group. Households in the treatment groups received HWRs while households in control groups did not. The Pilot ran from June 2012 through June 2013.

The Random Group Experiment consists of households representative of EBMUD's overall service area. The Castro Valley Group Experiment is comprised of a much more homogenous group of homes with characteristics thought to make them good candidates for HWRs. The goal of having two experiments was to provide insight into the effectiveness of HWRs directed at a targeted group of homes (Castro Valley Experiment) as well as into what the average effectiveness of HWRs might be if the program were expanded across EBMUDs whole service area (Random Group Experiment).

The Pilot was intended to address three primary questions:

1. First, would an SNB efficiency program like WaterSmart result in measurable reductions in household water use?
2. Second, would it increase rates of participation in other EBMUD conservation programs?

3. Third, would it increase household knowledge and awareness of water consumption and ways to use water more efficiently?

Within the context of each of the primary questions, EBMUD hoped the Pilot would yield information to address a number of additional questions of interest. These included:

1. Are households that are above (below) the norm more (less) likely to reduce their consumption of water?
2. Does whether the household receives a paper or electronic HWR affect the level of savings?
3. Is there a seasonal shape to water savings?
4. If HWRs increase participation in other conservation programs, which programs receive the greatest boost? Are households receiving HWRs that are above (below) the norm more (less) likely to participate in other conservation programs?
5. Are HWRs cost-effective? What is the expected cost of saved water from HWRs relative to other conservation program options or the cost of new water supply?

To address these questions we employ a range of statistical techniques, including robust panel data regression and dichotomous choice logit models. The following is a summary of our primary findings.

1. We find strong evidence that households in the Pilot's treatment groups reduced their water use in response to the HWRs. We estimate mean treatment effects on residential water use of 4.6% and 6.6% for the Random Group and Castro Valley Group experiments, respectively. We reject the null hypothesis of no treatment effect with better than 99% statistical confidence. Our estimates of mean treatment effect bracket the 5% mean effect estimated by WaterSmart using a less robust difference-in-differences methodology. The consistency between the WaterSmart estimates and our results is useful corroborating information.
2. We also find evidence that the magnitude of the water savings scales with level of household water use. Households in the top quartile of water use save, on average, 1% more, while households in the bottom quartile of water use save, on average, 3% less, than households in between these two categories. This suggests that if HWRs are not

going to be universally provided, utilities should consider giving households in the bottom quartile of use lower priority for receiving HWR.

3. Paper reports delivered by mail appear to be more effective in terms of water savings than electronic reports delivered by email. On average, households receiving paper reports were found to save about 1% of mean household use more than households receiving email reports. An implementing utility will still need to evaluate a host of factors, including the cost of delivering mail versus email reports, the avoided cost of saved water, and the availability of customer email addresses, to determine the preferred delivery method.
4. We estimate that the unit cost of saved water is likely to range between $250 and $590 per acre-foot for email reports and between $290 and $570 per acre-foot for paper reports.[1] The mid-point unit costs for email and paper reports are $380 and $400 per acre-foot, respectively. Even at the upper-end of the cost ranges, the unit costs are less than the cost of most other water demand management and new water supply options, indicating SNB efficiency programs could provide very cost-effective water savings.
5. We find strong evidence that households in the Pilot's treatment groups were significantly more likely to participate in audit and rebate programs offered by EBMUD than households in the control groups. Looking at both audit and rebate programs together, we estimate that households receiving HWRs were 2.3 times more likely to participate in a program than households not receiving reports. The effect appears to be strongest for audit programs, where we estimate households getting HWRs were 6.2 times more likely to participate. The effect is weaker for rebate programs (1.7 times more likely), but statistically significant. The results suggest that SNB efficiency programs can provide an effective conduit for channeling customers into other utility conservation programs.
6. Our analysis indicates that households receiving a water score of 3 on their HWR, which tells them to take action, are in fact more likely to do just that. The magnitudes of the treatment effects for both average daily use and program participation are positively correlated with water score. While our results should not be interpreted to imply that there is value to adjusting the scores to place more households in the high score category, they do suggest that targeting HWRs to homes that fall within this category is likely to

[1] Unit cost estimates have been rounded to the nearest $10 throughout this report.

yield better results in terms of average water savings and boosting program participation rates.

7. We do not find evidence that HWRs improve household knowledge of water use in the conventional sense of being able to quantitatively estimate average daily use. The proportion of homes stating they did not know their water use was essentially the same between households in the control and treatment groups. Similarly, the tendency to underestimate daily water use was also generally the same between control and treatment households. It may be that over time this will change and as households receive more HWRs they will begin to incorporate this information into their general understanding of how they use water.
8. We do, however, find strong evidence that households receiving HWRs view them as providing useful and actionable information for managing their water consumption. Households in the treatment group were 52 to 80% more likely to score EBMUD as "Excellent" in terms of explaining household water use, showing ways to save money on water bills by conserving water, and giving useful tips and tools needed to use water efficiently. Thus, HWRs appear to be effective at delivering information on ways to use water efficiently that households can, and judging by the measured effects on daily water use and program participation, do act upon.

I. INTRODUCTION

This report presents the results of our evaluation of the outcomes of the East Bay Municipal Utility District's (EBMUD) year-long pilot project (Pilot) of WaterSmart Software's Home Water Reports (HWRs) service. HWRs provide households with periodic information on their current water use and compare it to their past use, use by similar households, and efficient use. This data is coupled with actionable information on ways to use water around the home more efficiently. HWRs aim to motivate households to reduce their water use through simple to implement changes in behavior or adoption of more water efficient technology. The approach is based on research on social norms marketing coming out of the field of social psychology and for this reason we refer to these type of programs as social-norms-based (SNB) efficiency programs. While SNB efficiency programs have been broadly adopted by energy utilities across the United States in recent years, they are new to water utilities.

The EBMUD Pilot is the first relatively large-scale implementations of an SNB efficiency program by a large urban water utility. The Pilot was intended to address three primary questions:

1. First, would an SNB efficiency program result in measurable reductions in household water use?
2. Second, would it increase rates of participation in other EBMUD conservation programs?
3. Third, would it increase household knowledge and awareness of water consumption?

Within the context of each of the primary questions, it was hoped the Pilot would yield information to address a number of additional questions of interest. These included:

1. Are households that are above (below) the norm more (less) likely to reduce their consumption of water?
2. Does whether the household receives a paper or electronic HWR affect the level of savings?
3. Is there a seasonal shape to water savings?
4. If HWRs increase participation in other conservation programs, which programs receive the greatest boost? Are households receiving HWRs that are above (below) the norm more (less) likely to participate in other conservation programs?
5. Are HWRs cost-effective? What is the expected cost of saved water from HWRs relative to other conservation program options or the cost of new water supply?

The analysis that follows touches on each of these questions. To our knowledge, this report provides the first published independent evaluation of the effect of an SNB efficiency program on residential water use.[2] The remaining parts of this report are organized as follows. In Section II we provide an overview of SNB efficiency programs, including a discussion of the theoretical basis for and empirical evidence of their effectiveness. In Section III we describe the WaterSmart service that was implemented for this Pilot. In Section IV we describe the Pilot, including its goals and objectives, experimental design, and implementation. In Section V we present the results of our evaluation. This section is divided into five main parts that address Pilot outcomes in terms of household water use, participation in other conservation programs, knowledge and awareness of water use, cost effectiveness, and potential for integration with or extension of existing water use efficiency programs and strategies. In Section VI we provide a summary of Pilot outcomes and implementation lessons learned. We conclude the report in Section VII with recommendations for future research.

II. OVERVIEW OF SNB EFFICIENCY PROGRAMS

A. SOCIAL NORMS MARKETING

Social norms marketing is increasingly being used to motivate behavioral change (Andreasen, 2002). The central idea behind social norms marketing is that much of people's behavior is influenced by their perceptions of what is "normal" or "typical." According to social norms theory, if people are shown that their behavior is outside of the norm or that their perception of the norm is incorrect, they will be motivated to change the way they behave so they conform more closely to the norm. Moreover, it is believed the effect can be enhanced by coupling information on social norms with actionable information that facilitates the desired behavioral change.

Social norms marketing originated with issues related to college student drinking and substance abuse (Perkins & Berkowitz, 1986), but has evolved over the last two decades into a much more broadly applied concept. The effectiveness of social norms marketing in motivating behavioral change has been studied in a wide variety of contexts, including voting (Gerber & Rogers, 2009), environmental awareness (Goldstein, Cialdini, & Griskevicius, 2008), retirement savings (Beshears,

[2] An unpublished working paper by University of Washington researchers also examined average treatment effects of HWRs using similar panel regression techniques (Brent, et al, 2013).

Choi, Laibson, Madrian, & Milkman, 2009), charitable giving (Frey & Meier, 2004), and energy conservation (Allcott, 2011).

B. STRUCTURE OF SNB EFFICIENCY PROGRAMS

Interest in the use of social norms marketing within the energy and water utility sectors has grown significantly in the last decade. Partly this has been spurred by the transition to Advanced Metering Infrastructure (AMI) in the energy utility sector, which has substantially lowered the marginal cost of delivering detailed consumption information to customers and matching this information to the usage patterns of other similar customers. All of the major energy utilities in California are transitioning to AMI and most have coupled this technology with the provision of detailed consumption information to their customers. The largest provider of social-norms-based (SNB) efficiency program services is Opower, a private sector software-as-a-service company based in Virginia. Opower currently has contracts to run SNB efficiency programs at more than 90 energy utilities -- including 8 of the U.S.'s 10 largest -- and its programs reach more than 22 million homes worldwide.

Typical elements of SNB efficiency programs designed to promote efficient usage behavior, customer engagement, and individual consumption management include:[3]

1. Normative comparison of a customer's usage against comparable customers in the same geographical area;
2. Use of what social psychologists call "injunctive norms" which convey to the customer that efficient use of natural resources is pro-social while excessive use is anti-social;
3. Targeted conservation tips based on an analysis of a customer's past usage and individual profile;
4. Information and enticements to direct customers to other utility programs based on their previous usage patterns and individual customer profiles.

This information is delivered to customers through customized reports that they receive -- via mail or electronically -- on a monthly, bi-monthly, or quarterly basis, depending on their utility's

[3] Adapted from Sergici & Farugui (2011).

billing cycle.[4] Typically, SNB efficiency programs also provide customers access to a web portal that provides even more information on their consumption and ways in which they can improve their efficiency. Customer relationship, analytical, and reporting tools are used by the utility to respond to customer inquiries, monitor and analyze changes in usage patterns, and report on outcomes.

C. THEORETICAL BASIS FOR SNB EFFICIENCY PROGRAMS

Allcott (2011) identifies three primary mechanisms through which SNB efficiency programs may induce households to increase the efficiency of their consumption. First, providing actionable information to households on how to reduce consumption lowers the cost of implementing efficiency improvements and therefore increases household demand for them.[5] Second, given that households have incomplete knowledge about how much water is needed to achieve desired levels of water-dependent household services (e.g. a lush landscape or clean clothes), the use of social comparisons and injunctive norms may result in households adjusting their privately-optimal levels of water use when confronted with new information about what constitutes "average" and "efficient" water use for similarly situated households. Third, the use of social comparisons and injunctive norms may alter the "moral cost" of water use, thereby altering household demand for water. Households using more than the norm may be made to feel they are using more than "their fair share" and try to use less because of this. Alternatively, households using much less than the norm may be made to feel they are not getting "their just desserts" and may therefore increase their consumption.

The last case raises the possibility that use of social comparisons could induce either lower or higher consumption, depending on how households perceives their own consumption after receiving information on normative consumption for similarly situated households. If the goal of the treatment is to get households to use less of something, then inducing some households to use more of it would be an unintended and undesirable consequence of the intervention. In the social psychology literature this is termed a boomerang effect and at least one study has reported its occurrence in the context of an energy efficiency program providing normative information on household energy

[4] In the absence of AMI, the billing cycle sets the maximum frequency in which reports can be offered. However, utilities may choose to provide them less frequently than every billing cycle.

[5] Information and search costs are costs associated with finding, gathering, and processing information needed to make informed investment and consumption decisions. Consumers will balance of the cost of obtaining additional information against the benefit they expect to gain from it. Lowering information costs can therefore increase demand for goods or services where these costs had heretofore been relatively high.

consumption (Schultz, Nolan, Cialdini, Goldstein, & Griskevicius, 2007). However, Schultz et al (2007) also found that boomerang effects could be neutralized by coupling the normative information with an injunctive message conveying social approval or disapproval. Telling households with low use relative to the norm they are doing great appears to prevent them from adjusting their consumption upward, while telling households with high use relative to the norm they could do better appears to induce them to adjust their consumption downward. In his impact evaluation of Opower home energy reports, which included both information on social norms and injunctive messaging, Allcott (2011) did not find evidence of boomerang effects.[6]

Another way of thinking about the "moral" cost of using a scarce good is in terms of the value (or utility) consumers get from using less of it if they believe doing so contributes to other public goods they value, such as contributing to healthy ecosystems, protecting at-risk species, reducing greenhouse gas emissions, or benefitting public health. As pointed out by Levitt and List (2007), social norms provide a key point of reference from which consumers may judge the morality of their consumption choices.[7] Allcott (2011) posits that most consumers believe their consumption of natural resources like water and energy is closer to the social norm than it actually is. Put another way, most of us tend to believe the social norm must be close to our own level of consumption because we all want to believe we only use what we need and do so efficiently. As shown in Section V, households in the Pilot consistently and significantly underestimated their consumption of water. Similar underestimation of usage has been reported for energy consumption (The Economist, 2010). In this case, consumers may not perceive much of a gain in moral utility from using less of the resource because they already believe they are consuming near or below the socially acceptable level. When provided information reinforced with injunctive messaging that this is not the case, consumers update their beliefs about the social norm -- downward for high use consumers and upward for low use consumers. High use consumers find the moral utility they would get from using less of the good to have gone up and adjust their demands accordingly.

[6] The SNB energy efficiency programs evaluated by Allcott involved nearly 600,000 households served by 14 different energy utilities. Six of the utilities were in California and Washington, six were in the Midwest, one was in the urban Northeast, and one was in a suburban area in a Mountain state.

[7] Musings on the role of morality in economic choices is actually much older than this. The first formal treatise by an economist on the subject was Adam Smith's seminal book *The Theory of Moral Sentiments* published in 1759.

D. EMPIRICAL EVIDENCE FOR EFFECTIVENESS OF SNB UTILITY PROGRAMS

1. Evidence from the Energy Utility Sector

While SNB efficiency programs are relatively new, there have nonetheless been a number of empirical evaluations of their effectiveness. Because of its dominance in the market, most of these evaluations have focused on programs run by Opower and address impacts of SNB efficiency programs on household energy use.

Evaluations of Opower programs have typically found an average treatment effect in the range of 1.5% to 3.5% of baseline consumption. Allcott (2011), the first evaluation of a scaled SNB efficiency program to be published in a peer-reviewed journal, reported an average treatment effect in the range of 1.4% to 3.3% across seventeen separate utility experiments, with an unweighted mean treatment effect of 2%. Within California, evaluations of SNB efficiency programs run by Opower have reported average treatment effects ranging from 0.9% to 2.9% of baseline consumption (Perry & Woehleke, 2013; Wu & Osterhus, 2012; Summit Blue Consulting, LLC, 2009).[8]

SNB efficiency programs have been shown to be effective at reducing both seasonal, peak day, and peak hour energy demands (Jessoe & Rapson, 2013). Average treatment effects have also been shown to be constant or increasing over multiple years (Provencher, 2011), indicating the effectiveness of repeated treatments does not appear to diminish with time.[9] Additionally, the magnitude of the treatment effect has been shown to scale up with baseline use, meaning high use customers reduce use proportionally more than low use customers (Allcott, 2011). For example, in the SNB efficiency program experiments evaluated by Allcott (2011), the average treatment effect for households in the 30th percentile of baseline use was under 1% whereas for households in the 80th percentile it was approximately 3.5%.

[8] A report by McKinsey & Company found the long-term potential savings from SNB efficiency programs in U.S. residential energy markets to be immense -- 1.8 to 2.2 quadrillion BTUs per year, or 16% to 20% of current U.S. non-transportation residential energy use (Heck & Tai, 2013). The largest savings potential -- accounting for more than half of the total --is associated with changing temperature set points for heating and hot water systems during cold weather. Other significant potentials are associated with changing the operating parameters for air-conditioning and refrigerators. According to the report, these savings potentials are as yet largely untapped, but could be through broader use of SNB efficiency programs over a sustained period.

[9] The span of years evaluated, however, has been relatively short -- usually two to three years. The effectiveness of SNB efficiency programs over longer stretches has yet to be tested.

SNB efficiency programs also may increase customer participation in other efficiency programs, which the evaluation literature terms "uplift" or "channeling." Opower claims its home energy reports have boosted participation in other utility programs by 17% to 59%. The evaluation literature is somewhat mixed. Several studies have shown positive uplift (Provencher, Hampton, Brown, & Hummer, 2013; Opinion Dynamics Corporation, 2012) while others have shown no uplift or even negative uplift (Perry & Woehleke, 2013; Gunn, 2012).

Assessments of SNB energy efficiency program costs have found them to be cost effective relative to other energy efficiency programs (Allcott & Mullainathan, 2010; Allcott, 2011). Cost estimates for Opower-like SNB energy programs are in the neighborhood of 2.5 to 3.5 cents per kilowatt-hour saved. This is substantially below the average cost of 6.4 cents per kilowatt-hour saved estimated for conventional energy demand-side-management (DSM) programs (Arimura, Newell, & Palmer, 2009). It also is comparable to an incremental cost of 3 cents per kilowatt-hour saved estimated for DSM programs at utilities with little or no historical investment in DSM (Arimura, Newell, & Palmer, 2009).

2. Evidence from the Water Utility Sector

To our knowledge there have not been any published independent evaluations of the effectiveness of SNB efficiency programs in the water utility sector. Our evaluation of the EBMUD pilot may constitute the first such evaluation. WaterSmart has reported savings estimates in the neighborhood of 5% for its City of Cotati and EBMUD pilots. However these estimates have not been independently verified.[10] Our estimates of the average treatment effect of the EBMUD Pilot are consistent with these previous estimates. They are also consistent with average treatment effects of three WaterSmart pilots -- including the EBMUD Pilot -- reported by University of Washington researchers in an unpublished working paper (Brent, et al, 2013). If these results are replicated in

[10] WaterSmart's internal metrics rely on difference-in-differences (DID) estimators. DID estimators, however, require strong identifying assumptions -- in particular that in the absence of treatment the average outcomes of the treatment and control groups would have followed parallel paths over time (Abadie, 2005). The preferred approach for estimating the treatment effect of SNB efficiency programs is panel data regression analysis, which can more effectively control for other factors, such as weather, impacting differences in consumption between the pre and post intervention periods for the control and treatment groups (Sergici & Farugui, 2011).Typically either a fixed-effects or random-effects estimator is recommended. For information on estimation of fixed and random effects models, more generally, the reader is referred to Wooldridge (2001).

other evaluations of SNB efficiency programs for water, it would provide compelling evidence for the viability of SNB efficiency programs in the water utility space.

III. THE WATERSMART SERVICE

SNB efficiency programs can be implemented in varying ways which may yield differing results. It is therefore important to acknowledge that our evaluation results are based on a particular implementation provided by WaterSmart Software. In this section, we describe the WaterSmart SNB efficiency program that was used in the EBMUD Pilot.

A. HOME WATER REPORT DESIGN

The design of the WaterSmart HWR is very similar to the design employed by Opower for home energy reports. It is divided into two primary modules. The Social Comparison Module appears at the top of the first page of the HWR -- it is designed to be the first thing the viewer of the report sees. As illustrated in Figure 1, the Social Comparison Module of the report presents the "descriptive norm" by comparing the household to the mean and 20th percentile of its comparison group. A household's comparison group comprises geographically proximate houses of similar irrigable area and number of occupants.

The WaterSmart HWR uses injunctive norms to convey to the household how they are doing. For the EBMUD Pilot, households were told they are doing "Great" if their use was less than the 20th percentile of their usage comparison group, they are doing "Good" if their use was within the 20th and 55th percentiles, and to "Take Action" if their use was above the 55th percentile.[11] This messaging is reinforced with a large smiley face emoticon (in the shape of a water drop) whose expression -- smiley, neutral, or worried -- corresponds to where the household's water use falls within the distribution of water use for its comparison group (Figure 2).

[11] In other implementations of WaterSmart, households were placed in the "Good" group if their consumption was between the 20th and 50th percentiles. This was broadened to the 55th percentile for the EBMUD Pilot at the request of EBMUD staff.

Figure 1. WaterSmart Home Water Report Social Comparison Module

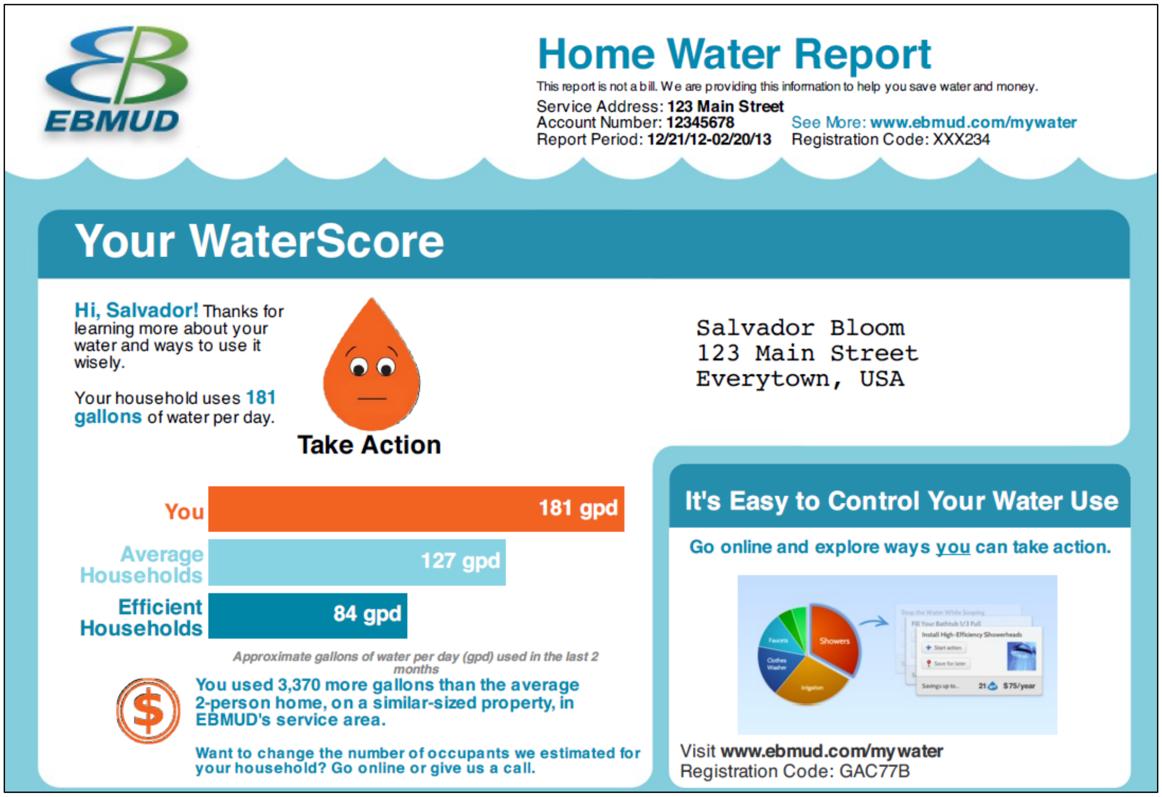
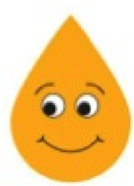


Figure 2. WaterSmart Emoticons Used With Injunctive Messages About Water Use

 Smiling Smiley Face Indicates Household Use is Considered Efficient	 Neutral Smiley Face Indicates Household Use is Okay	 Worried Smiley Face Indicates Household Use is Excessive

The second part of the report is the Suggested Actions Module. An example is shown in Figure 3. This module provides targeted recommendations of actions the household can take to use water more efficiently. The recommendations are tailored to each household based on their usage history, household characteristics, season of the year, and other factors. For example, actions related to landscape water use may be directed to households with high summer to winter use ratios or suggestions related to leaks may be directed to households showing abnormally high use compared to their prior use history.

Figure 3. WaterSmart Home Water Report Suggested Actions Module

WaterSmart HWRs are delivered to households either via mail or electronically. The initial report is delivered electronically if WaterSmart has a valid email address for the household, and via mail otherwise. Households receiving paper reports can use the web portal to opt for electronic report delivery. HWR delivery is synchronized with the customer's billing cycle. In the case of the EBMUD Pilot, HWRs were delivered bi-monthly.

B. WEB PORTAL

WaterSmart HWRs direct households to a web portal where they can get more detailed information on their water consumption and tailored recommendations for reducing their consumption. The web portal for the EBMUD Pilot is called the WaterInsight Program. Users land on a home page where they get the most current summary of their consumption relative to their comparison group as well as recommended water saving actions. From the home page they can go to pages that allow them to verify or update information about their household; track their usage in greater detail; provide additional recommendations and tips for reducing consumption; and track the actions they have taken to reduce consumption. Figure 4 shows a screenshot of the home page and some of the usage charts on the Track Usage page.

Figure 4. WaterInsight Web Portal Used in EBMUD Pilot

IV. EBMUD PILOT

In this section we provide a descriptive summary of the EBMUD Pilot.

A. PILOT GOALS AND OBJECTIVES

EBMUD hoped to address three basic questions through the Pilot. First, would an SNB efficiency program like the one implemented by WaterSmart result in measurable reductions in household water use? Second, would it increase rates of participation in other EBMUD conservation programs? And third, would it increase household knowledge and awareness of their water consumption? According to interviews with EBMUD staff, the district had been interested for some time in using billing information and other household-level data to provide information to customers on their water usage relative to other households, encourage more efficiency, and direct customers to other EBMUD conservation programs. WaterSmart's implementation of HWRs provided an attractive turnkey solution that would enable the district to test the effectiveness of doing this. At the onset of the Pilot, EBMUD staff expected that it might reduce household water use by about 2%.[12]

[12] Personal communication with EBMUD Conservation Staff, October 22, 2013.

EBMUD hoped the Pilot would address a range of additional questions stemming from the three primary questions. These included:

1. To what extent do water savings vary seasonally? Are savings primarily due to changes in outside water use, inside water use, or a combination?
2. How do water savings relate to the information households receive on their HWRs about their water consumption relative to other similarly situated households? Are households that are above (below) the norm more (less) likely to reduce their consumption of water?
3. Does the level of savings depend on whether the household receives a paper or electronic HWR?
4. If HWRs increase participation in other conservation programs, which programs receive the greatest boost? Are households receiving HWRs that are above (below) the norm more (less) likely to participate in other conservation programs?
5. Are HWRs cost-effective? What is the expected cost per gallon saved for households receiving paper versus electronic HWRs?

In addition to these objectives, both EBMUD and WaterSmart were interested in exploring whether treatment effects differed when HWRs were provided to an entire community with similar characteristics rather than to randomly selected households spread across a service area, which was how WaterSmart had implemented a previous pilot for the City of Cotati.

B. PILOT EXPERIMENTAL DESIGN

The EBMUD Pilot was comprised of two experiments. The first we call the Random Group Experiment. The second we call the Castro Valley Group Experiment. In both experiments, households were selected to be in either a treatment group or a control group. Households in the treatment groups received HWRs while households in control groups did not. The treatment period, meaning the period when homes in the treatment groups received HWRs, ran from June 2012 through June 2013. Treatment in the Random Group Experiment spans this entire period. As we explain below, for the Castro Valley Group Experiment the treatment was rolled out in phases over this period, so that the duration of treatment varied among homes in this experiment. The details of each experiment are as follows.

1. Random Group Experiment

The Random Group Experiment consisted of randomly selected households that were split evenly between the treatment and control groups. EBMUD had previously developed a stratified random sample of its single family residential customers.[13] This sample was based on three geographic zones and seven parcel size classifications, resulting in 21 strata. For the Random Group Experiment, it proportionately sampled approximately 4,000 households from these strata. About 16% of the initially sampled households were ultimately excluded from the experiment, either because of data problems identified prior to the start of the experiment or because of discontinued service or significant data anomalies during the course of the experiment. The final count of households in the Random Group Experiment is 3,286, of which 1,576 were in the control group and 1,710 were in the treatment group.

The distribution of sampled residential accounts across EBMUD's pressure zone groups is shown in Table 1. Overall the sample is representative of the geographic distribution of residential accounts within EBMUD's service territory. Group G is somewhat under-sampled while Group F is somewhat over-sampled. A table with the proportion of sampled accounts from each of EBMUD's 120 pressure zones is provided in Appendix 1. It shows the sample is generally representative of the geographic distribution of residential accounts at the pressure zone level as well.

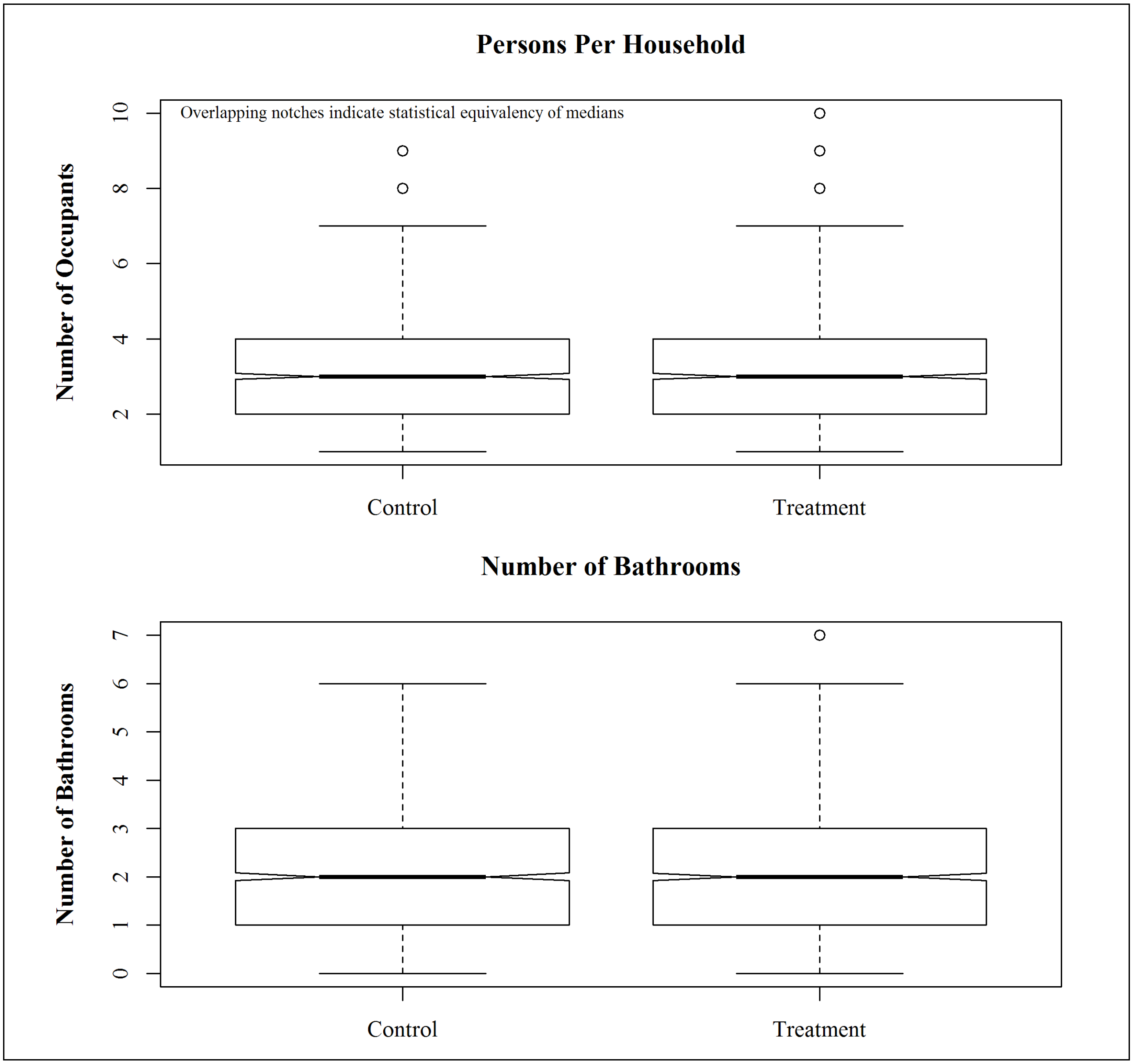
Table 1. Random Experiment Sample Distribution by Pressure Zone Group

Pressure Zone Group	No. of Zones	% Residential Accounts	% Sampled Residential Accounts		
			Total	Control	Treatment
A	32	15.6%	16.3%	8.3%	8.0%
B	23	8.7%	8.5%	4.7%	3.8%
C	15	6.6%	8.4%	1.7%	6.8%
D	20	4.4%	6.0%	2.8%	3.3%
E	8	2.2%	2.2%	1.1%	1.1%
F	17	12.5%	17.8%	9.2%	8.6%
G	4	46.9%	36.8%	18.3%	18.5%
H	1	3.1%	3.9%	1.9%	2.0%
Total	**120**	**100.0%**	**100.0%**	**47.9%**	**52.1%**

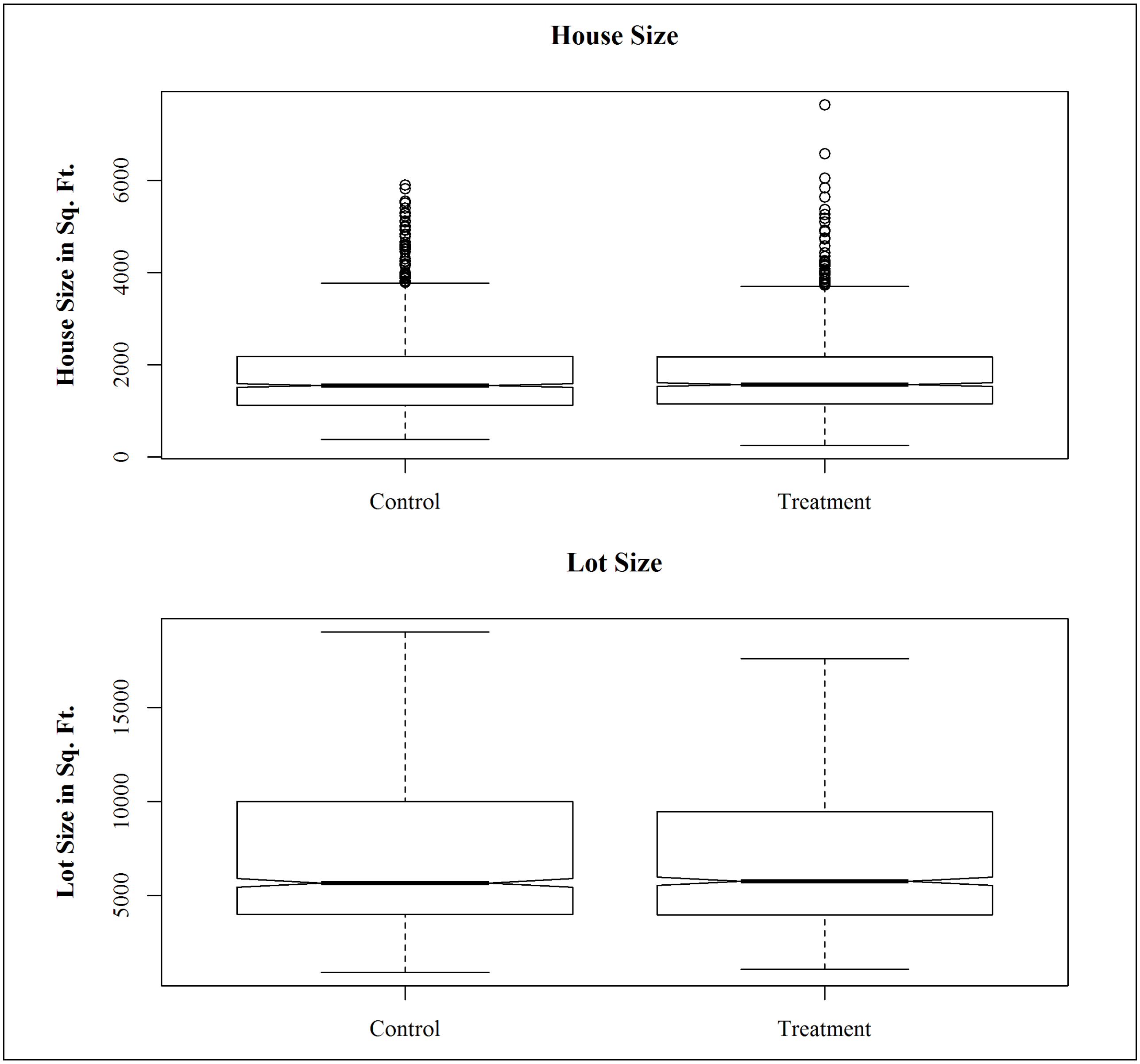
[13] EBMUD had developed the sample as part of the process it was using to calculate GPCD targets for 20x2020 (SBx7-7) compliance.

As would be expected for a random sample, the distributions of household attributes are essentially identical between the control and treatment groups, as shown in Figure 5.[14]

Figure 5. Random Experiment Sample Household Attributes

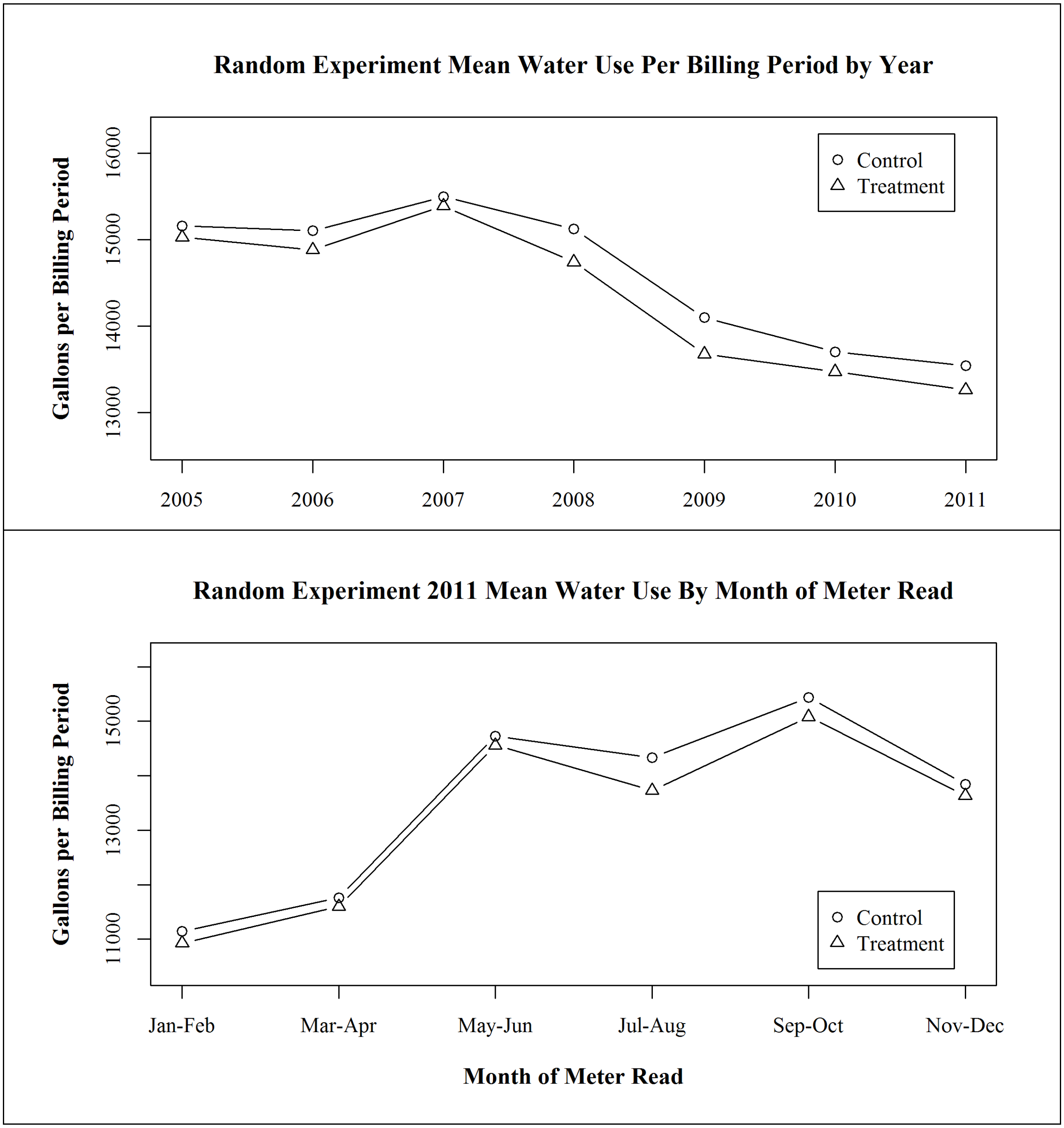


[14] The box plots in Figure 5 are interpreted as follows. The dark line segmenting the box is the median value. The top and bottom of the box are the 75th and 25th percentiles, respectively. The horizontal lines above and below the box denote the range of the distribution, excluding outliers. The circles above or below these horizontal lines represent outliers. If the notches in adjacent Control and Treatment boxes overlap, it strongly indicates the median values for the two groups are statistically the same.



The annual trend in water use prior to the start of the pilot and the seasonal pattern of water use also are similar, as shown in Figure 6. In the second panel of Figure 6, the x-axis refers to the month in which the meter was read and mean water use is for the two month period leading up to this date. For example, a meter read on 9/15 would include consumption roughly from 7/15 to 9/15. This is why Figure 6 shows a peak in mean water use for reads occurring in September and October, since reads in these months capture the bulk of summer use.

Figure 6. Random Group Temporal Patterns of Mean Water Use Per Billing Period



The first panel of Figure 6 shows a parallel trend in mean annual water use between the control and treatment groups. This is useful information for assessing the reliability of water savings estimates based on DID estimators. As previously noted, a key identifying assumption for a DID estimator is the pattern of use between the control and treatment groups would have remained the same but for the treatment.[15] The first panel of Figure 6 strongly suggests this assumption holds for the Random Group Experiment.[16]

2. Castro Valley Group Experiment

The Castro Valley Group Experiment selected more than 8,000 single-family residences in the City of Castro Valley to receive Home Water Reports. These homes comprised the treatment group. Just over 1,300 homes in the Dingee Pressure Zone, which is adjacent to Castro Valley, thought to have similar single-family residential characteristics and climate comprised the control group for this experiment.[17] The distribution of sampled accounts by pressure zone is shown in Table 2.

Table 2. Castro Valley Experiment Sample Distribution by Pressure Zone

Pressure Zone	Control	Treatment
B5A	100%	
C2A		43.5%
C4A		34.7%
C4D		0.6%
C5C		7.0%
C5D		7.9%
C5E		1.0%
C6B		3.5%
C7A		1.8%
Total	**100%**	**100.0%**

Castro Valley was selected by EBMUD for the Pilot because the community is comprised of homes thought to approximate characteristics and climate of homes that would be targeted in a future expanded program implemented throughout the EBMUD service area. On average, compared to

[15] The idea being there are no exogenous factors other than the treatment causing changes in use of the treatment group but not the control group.

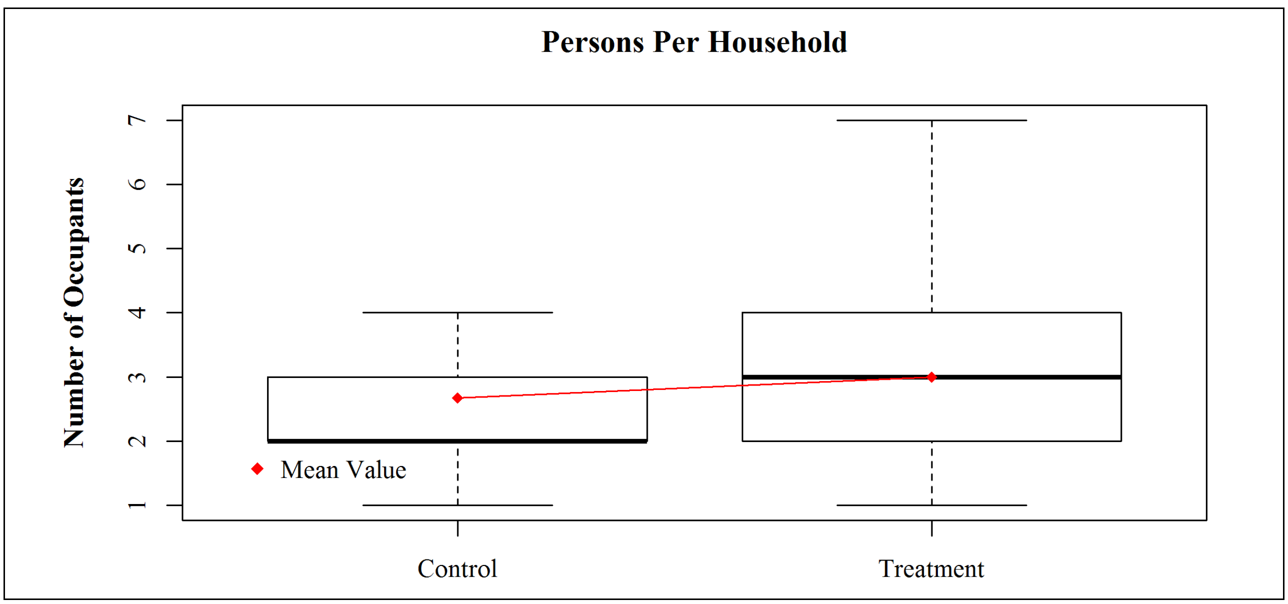
[16] For the reasons laid out in Sergici and Farugui (2011) and Chesnutt and McSpadden (1995), we employ panel data regression techniques to estimate the mean treatment effect on water use rather than a DID estimator.

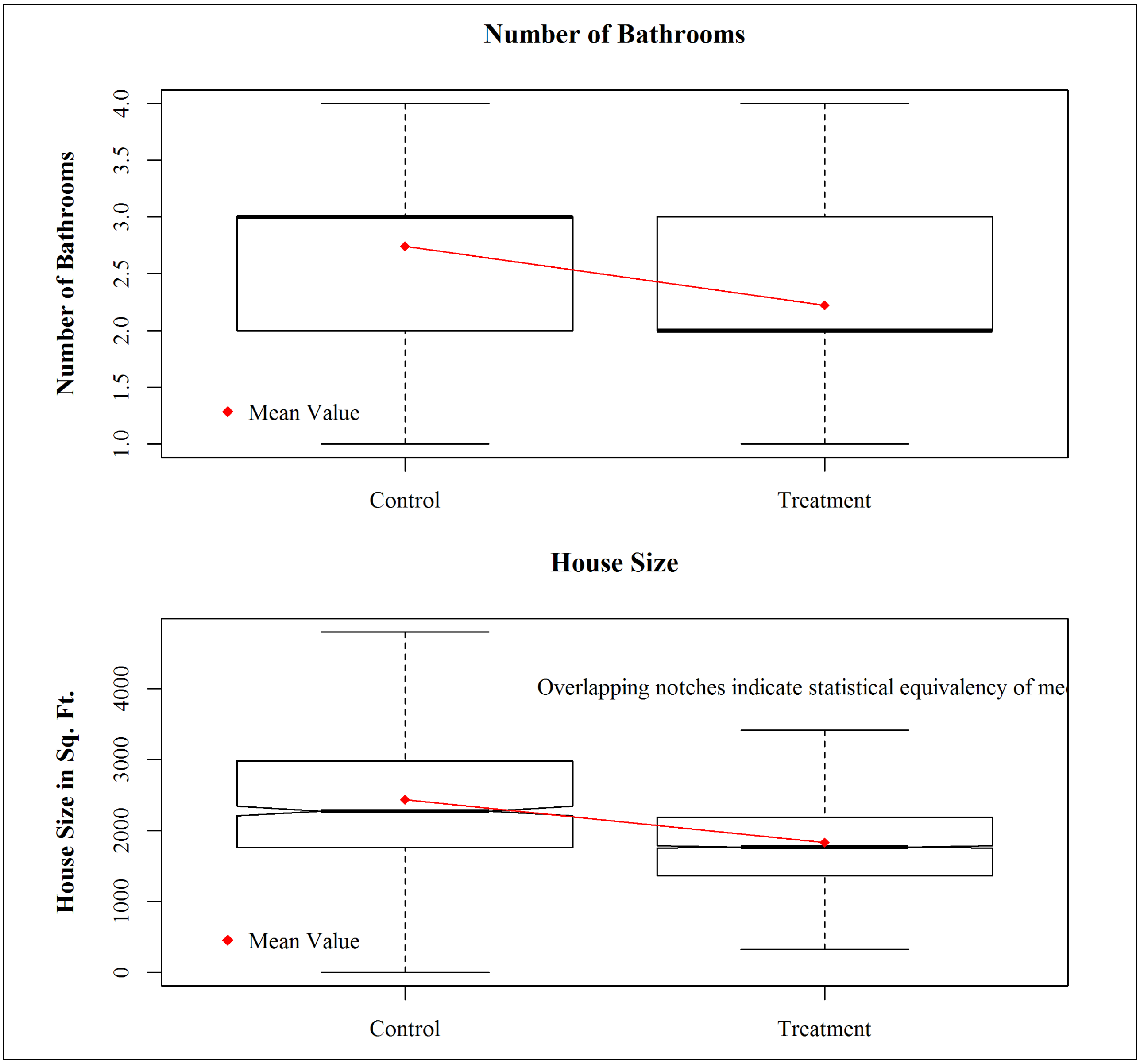
[17] These are the sample sizes developed by EBMUD and WaterSmart for the Castro Valley Group Experiment. For the analysis of treatment effect on residential water use we extended the number of households in the control group to provide better resolution on household water use. Thus, we use consumption records from 13,765 households to serve as controls in our statistical model of water use.

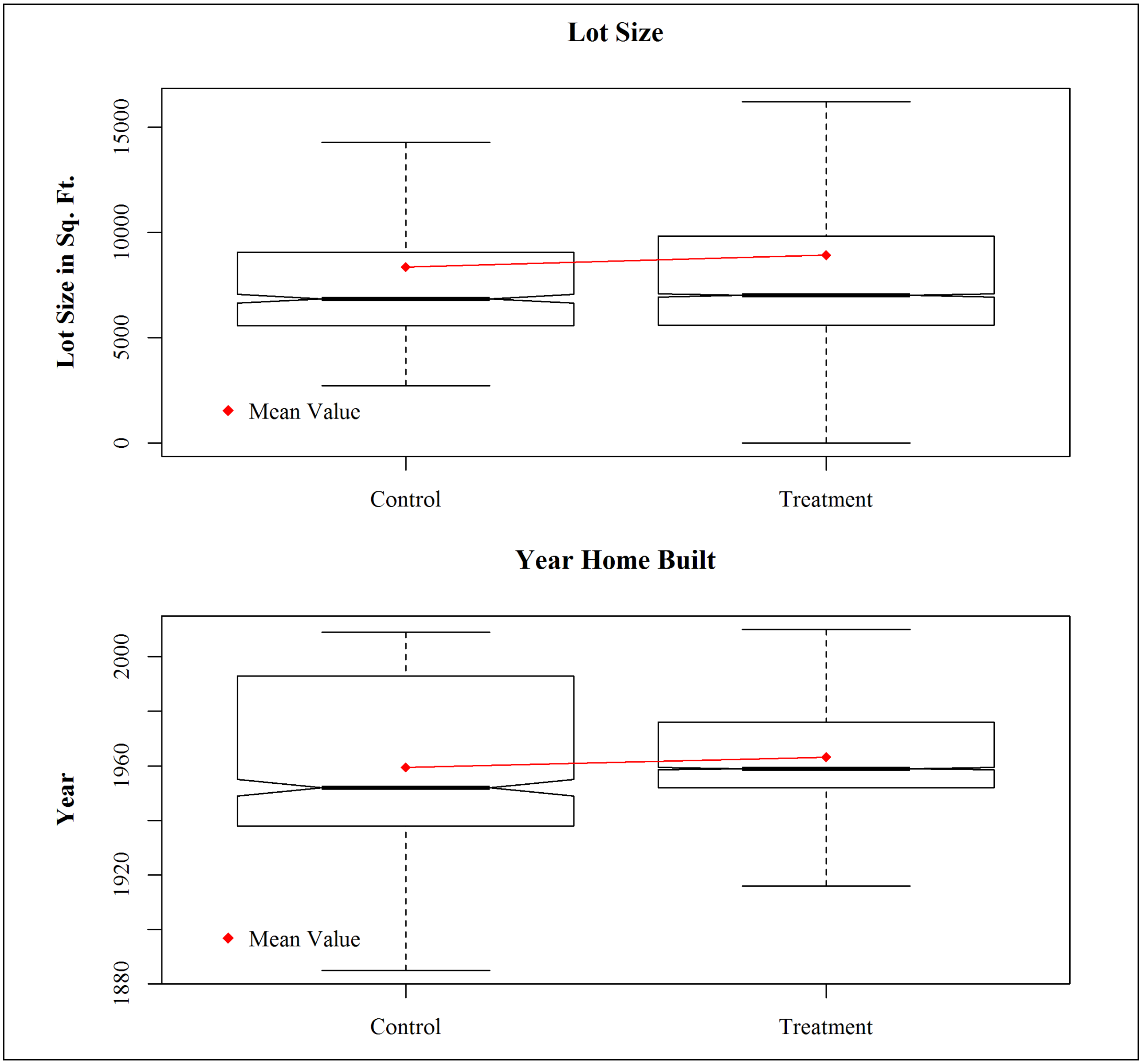
homes in the Random Group Experiment, homes in the Castro Valley Group Experiment are larger, have more bathrooms, and have larger irrigable area. They are also more homogenous, showing smaller coefficients of variation for key household characteristics. Most single family homes in Castro Valley are in the middle to upper-middle income brackets. The income distribution within the Random Group Experiment is more varied. Given the differential in home attributes between the two experiments, it was hoped the Castro Valley Group Experiment would provide insight into the effectiveness of HWRs directed at homes thought to be good targets, while the Random Group Experiment would provide insight into what the average effectiveness of HWRs might be if the program were expanded across EBMUD's whole service area.

The distributions of household characteristics for the control and treatment groups in the Castro Valley Group Experiment are summarized in Figure 7. Unlike the Random Group Experiment, the data show significant differences in household characteristics between control and treatment households. According to the data we obtained from WaterSmart, control group households tend to be larger and have more bathrooms, though fewer persons per household, on average. The age of control group homes is more varied and has a higher proportion of newer homes. Lot sizes are similar for the two groups, but water use per billing period is higher for the control group than for the treatment group.

Figure 7. Castro Valley Experiment Sample Household Attributes




Number of Bathrooms
Number of Bathrooms
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Mean Value
Control
Treatment
House Size
House Size in Sq. Ft.
0
1000
2000
3000
4000
Overlapping notches indicate statistical equivalency of me
Mean Value
Control
Treatment



As we describe in the next section, the Castrol Valley Group Experiment was rolled out in three phases. Originally the intention was to roll out the experiment in four phases, but the first two phases were ultimately combined. The rollout phases are therefore referred to as Phase 1/2, Phase 3, and Phase 4. The differences in household characteristics between the control and treatment groups by rollout phase are shown in Figure 8 and Figure 9. From these figures it is seen that Phase 1/2 and Phase 3 homes have similar attributes. Phase 4 homes are generally smaller, have fewer bathrooms, and smaller lots. These differences are only important if they have the potential to differentially affect home water use in the pre- and post-treatment periods, which could then confound estimates of the treatment effect if not controlled for in the statistical model. This is not expected to be the case for attributes like number of bathrooms or house size.[18] It could be the case for lot size since this correlates positively with irrigable area and seasonal water use. Larger lot homes may be expected to respond differently than smaller lot homes to differences in weather between the pre- and post-treatment periods. Since Phase 4 and control group homes have significantly smaller lot sizes than Phase 1/2 and Phase 3 homes, it is necessary to put appropriate weather controls into the statistical model of mean treatment effect.

The distributions of water use per billing period by rollout phase are summarized in the top panel of Figure 10. Relative to homes in the control group, median and mean water use for homes in Phase 1/2 is higher; it is about the same in Phase 3 homes; and it is lower in Phase 4 homes.

The second panel of Figure 10 shows the time trend for mean water use per billing period for each treatment phase and the control group. This panel shows that the treatment and control groups followed generally parallel trends in annual use leading up to the start of the Pilot.

The seasonal pattern of water use for 2009-2011 is shown in Figure 11. Seasonal use in 2009 and 2010 follow nearly identical patterns, suggesting a fairly stationary relationship between the different treatment phases. However, seasonal use in 2011 deviates from this pattern with a dip in water use during the summer billing period (which corresponds to the spring to early summer consumption period). The relative position of the control group shifted somewhat in 2011, perhaps in response to weather anomalies during the spring months, which again points to the need to put appropriate statistical controls on weather effects when estimating the mean treatment effect of the Pilot on water use.

[18] We employ fixed-effects regression techniques to control for the stationary differences in home attributes.

Figure 8. Castro Valley Experiment Sample Household Attributes by Rollout Phase

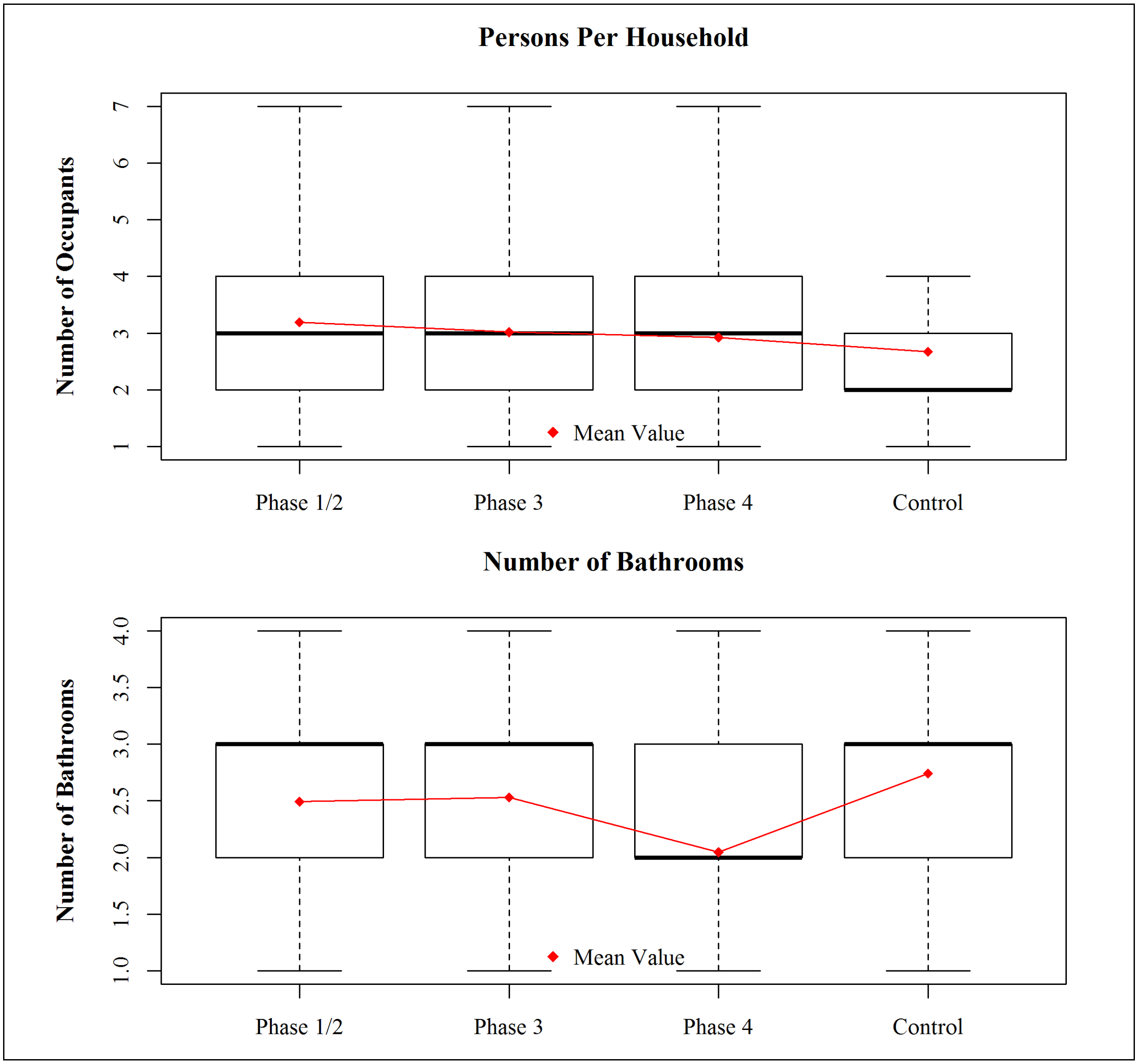


Figure 9. Castro Valley Experiment Sample Household Attributes by Rollout Phase

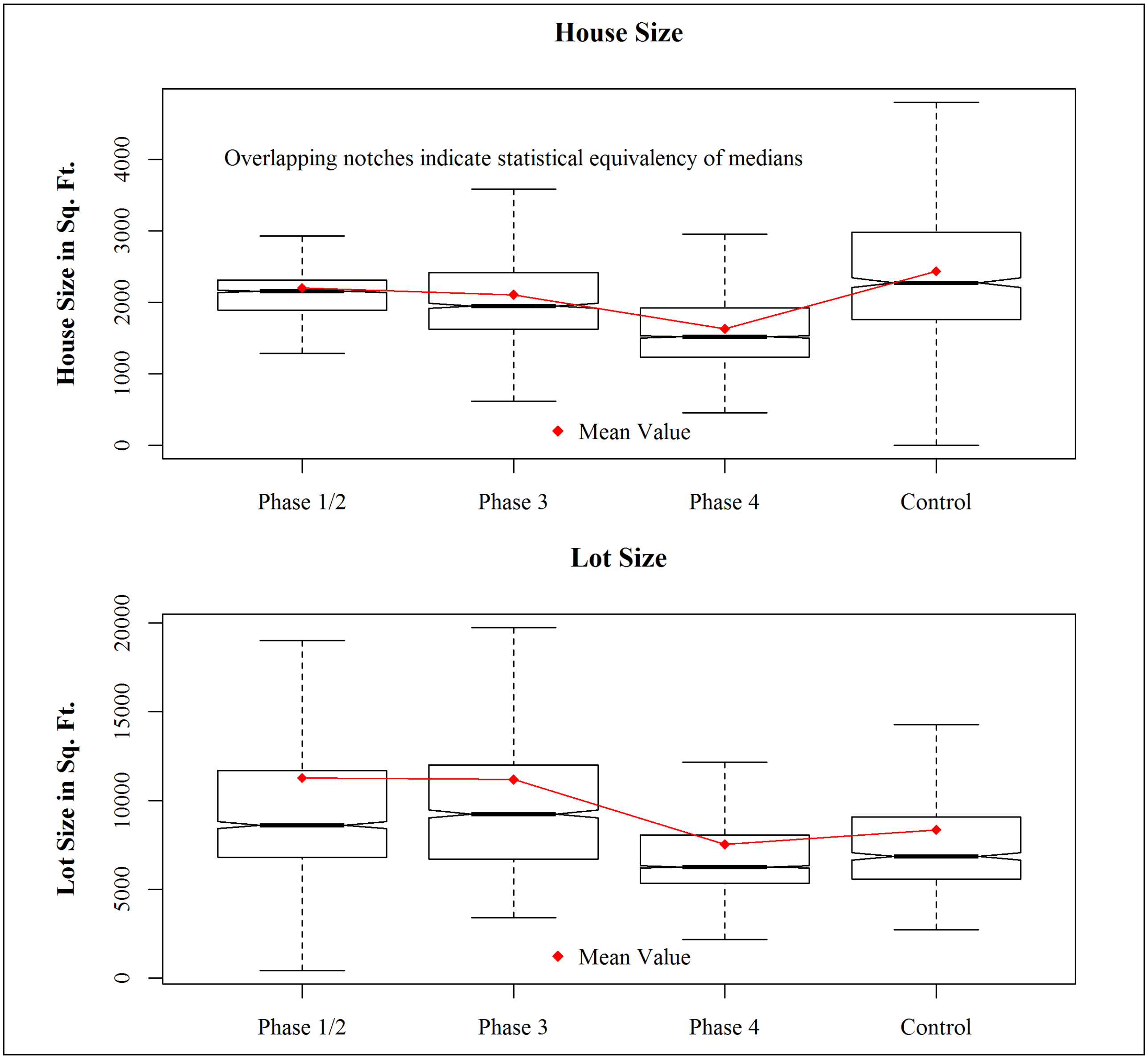


Figure 10. Castro Valley Experiment Water Use by Rollout Phase

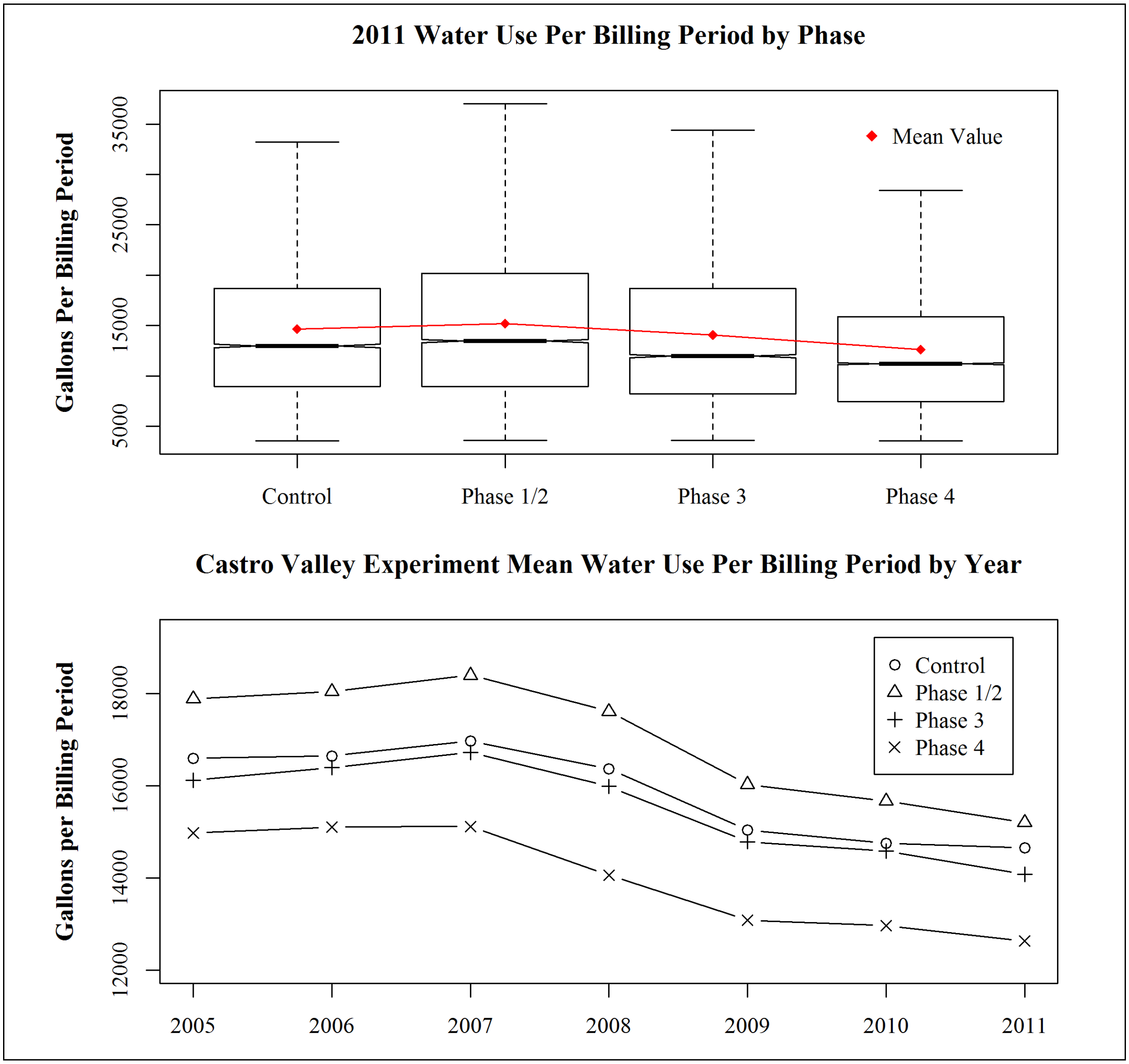
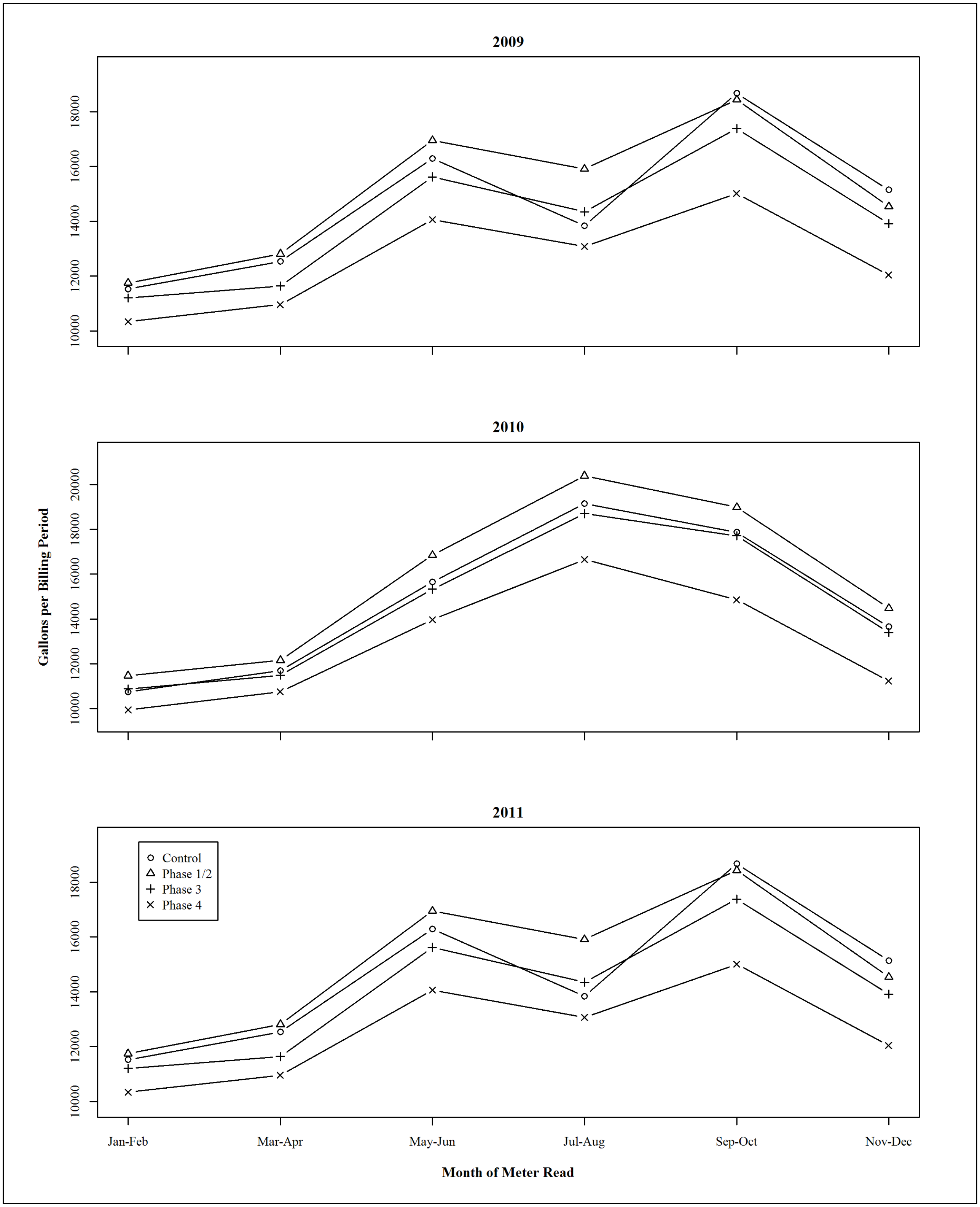


Figure 11. Castro Valley Group Seasonal Patterns of Mean Water Use Per Billing Period



C. CASTRO VALLEY HWR ROLLOUT PHASES

Primarily for administrative reasons, EBMUD chose to implement the Castro Valley Group Experiment in phases. Originally there were to be four phases. However, the first two phases were combined so ultimately the experiment was rolled out in three phases. As noted above, we refer to these phases as Phase 1/2, Phase 3, and Phase 4. Table 3 shows the number of homes in each phase along with the date they started to receive HWRs. Note there are two start dates for each phase. Which of the two dates applies for a particular home in a phase depends on its billing cycle. Homes in Phase 1/2 started receiving HWRs after June 15 or July 6 of 2012; homes in Phase 3 started receiving them after August 9 or September 7, 2012; and homes in Phase 4 started receiving them after October 7 or November 14, 2012.[19] The Pilot ran through June of 2013. Homes in Phase 1/2 received six or seven HWRs over the course of the Pilot. Homes in Phase 3 received five or six and homes in Phase 4 received four or five. Homes had continuous access to the web portal following the receipt of their first report.

Table 3. Castro Valley Group Experiment HWR Rollout

Phase	No. of Homes	HWR Start Date	No. Reports Received During Pilot
1/2	1,964	6/15/12 or 7/6/12	6 or 7
3	1,598	8/9/12 or 9/7/12	5 or 6
4	5,435	10/7/12 or 11/14/12	4 or 5

D. PRE-PILOT HOUSEHOLD SURVEY

Prior to the start of the Pilot, a household survey was administered to homes in the treatment group for the Castro Valley Group Experiment and to homes in both the control and treatment groups of the Random Group Experiment. The purpose of the survey was to collect information on household knowledge and attitudes about water use and conservation as well as information on household attributes, such as number of people in the home, number of toilets, presence and type of water using appliances, type of landscaping and irrigation, presence of pool or spa, etc. The information was used in creating the customer profiles, which were then used to tailor water saving

[19] Since the majority of the Castro Valley homes have meter read dates in the second half of the two-month billing cycle, the majority of the homes received reports in the latter months of July, September and November.

tips and other information provided on the HWRs and through the web portal. The pre-pilot survey had an approximately 20% response rate.

E. POST-PILOT HOUSEHOLD SURVEY

Following the end of the Pilot period, a second household survey was administered to homes in the treatment of control groups of both experiments. This survey was sent to an equal number of homes that did and did not respond to the pre-pilot survey. In total, surveys were sent to 4,766 households. The post-pilot survey had an approximately 31% response rate. Results from the two surveys are used in this evaluation to assess the mean treatment effect of HWRs on household knowledge and attitudes about water use and conservation.

V. EVALUATION OF PILOT OUTCOMES

In this section we present results of our evaluation of Pilot outcomes. Broadly, we address the three primary questions presented in Section IV.A:

1. Did the Pilot result in measurable reductions in household water use?
2. Did it increase rates of participation in other EBMUD conservation programs?
3. Did it increase household knowledge and awareness of their water consumption?

Within the context of each of these primary questions, we also present evaluation results for a range of secondary questions of interest. Additionally, we examine the cost-effectiveness of HWRs and their potential for integration with existing conservation programs.

A. WATER USAGE

Arguably the most important question to be addressed by the Pilot is did providing households with HWRs result in measurable reductions in household water use compared to the control group? While SNB efficiency programs are multi-dimensional in what they offer to utilities in terms of customer services and demand management, they are primarily marketed as a way to reduce customer water use.[20] As previously discussed, WaterSmart has reported savings estimates in the neighborhood of 5% for its City of Cotati and EBMUD pilots. WaterSmart's internal metrics, however, rely on less statistically robust DID methodology (Sergici & Farugui, 2011). To our

[20] Indeed, the banner across WaterSmart's homepage currently reads "Reduce Water Demand by 5% in 6 months."

knowledge, this report provides the first published independent evaluation of the effect of an SNB efficiency program on residential water use based on more robust panel data regression methodology.

1. Methodology

We use panel data regression techniques to estimate the mean effect of HWRs on household water consumption for the two experiments. Specifically, we estimated a fixed-effects model of water consumption that controls for time-variant seasonal and weather effects on consumption over the pre- and post-treatment periods as well as effects of unobserved time-invariant differences in household characteristics.

The general form of the model is given in equation (1):

$$(1) \qquad Use_{it} = \mu_i + S_t + W_t + E_{it} + \varepsilon_{it}$$

where Use_{it} is household i's average daily water use in period t, μ_i is household i's mean daily water use, S_t is the seasonal effect on average daily water use in period t, W_t is the weather effect on average daily water use in period t, E_{it} is the effect of HWRs on household i's average daily water use in period t, and ε_{it} is model error.

We model the seasonal and weather effects as continuous (as opposed to discrete bi-monthly) functions of time following the approach in Chesnutt and McSpadden (1995). We also include interactions between the seasonal and weather components to isolate season-specific weather responses.

The seasonal term, S_t, is formed by the Fourier series shown in equation (2), where d = 1, 2, 3, ... 365 is an index of the days of the year, and $\underline{d}_t$ and $\overline{d}_t$ are the first and last indexed days in period t, respectively.[21] For this analysis, we assume the number of days between billing periods, $(\overline{d}_t - \underline{d}_t + 1)$, is a constant 61 days.

[21] A Fourier series is an expansion of a periodic function f(x) in terms of a sum of sines and cosines. The use of Fourier series to represent periodic functions is called harmonic analysis, which was first employed to estimate a seasonal component in a regression context by Hannan(1960). Jorgenson (1964) extended the use of harmonics in least squares estimation to include both trend and seasonal components. Note that if t has the length of 1 day (e.g., daily observations of water use), then equation (2) simplifies to $S_t \equiv \sum_{j=1}^{6} \left\{ \beta_{1j} \cdot sin\left(\frac{2\pi \cdot j \cdot t}{365}\right) + \beta_{2j} \cdot \right.$

(2) $$S_t \equiv \sum_{j=1}^{6}\left\{\beta_{1j}\cdot\frac{1}{(\overline{d}_t-\underline{d}_t+1)}\sum_{d=\underline{d}_t}^{\overline{d}_t} sin\left(\frac{2\pi\cdot j\cdot d}{365}\right)+\beta_{2j}\cdot\frac{1}{(\overline{d}_t-\underline{d}_t+1)}\sum_{d=\underline{d}_t}^{\overline{d}_t} cos\left(\frac{2\pi\cdot j\cdot d}{365}\right)\right\}$$

The model incorporates two types of weather measures into the weather component -- rainfall and average daily evapotranspiration -- both of which are logarithmically transformed. These measures are defined in equation (3).[22]

(3) $$R_t \equiv ln\left[1+\sum_{d=\underline{d}_t}^{\overline{d}_t} Rain_d\right], ET_t \equiv ln\left[\frac{1}{(\overline{d}_t-\underline{d}_t+1)}\sum_{d=\underline{d}_t}^{\overline{d}_t} ET_d\right]$$

Because weather has a strong seasonal pattern, the weather measures are correlated with the seasonal component. To address this collinearity, the weather component is constructed as a departure from the "normal" or expected weather given the season, as shown in equation (4).

(4) $$W_t \equiv \left(R_t-\hat{R}_t\right)\cdot\beta_R+\left(ET_t-\widehat{ET}_t\right)\cdot\beta_{ET}$$

The expected values for rainfall, $\hat{R}_t$, and evapotranspiration, $\widehat{ET}_t$, are derived from regression against the seasonal harmonics. The weather measures expressed in this way are thereby separated from the seasonal effects. The seasonal component, therefore, captures all constant seasonal effects, including those caused by normal weather conditions, while the weather component captures the effect of weather departing from its normal pattern (e.g. unusually wet or dry for the given time of year).

The effect of HWRs on average daily water use is specified in equation (5), where the indicator variable, I_{it}, takes the value 1 if household i is receiving HWRs in period t and 0 otherwise.

(5) $$E_{it} \equiv I_{it}\cdot\beta_{HWR}$$

The coefficient β_{HWR} measures the mean treatment effect of HWRs on average daily water use and is expected to have a negative sign if HWRs induce lower average daily water use.[23] Note

$cos\left(\frac{2\pi\cdot j\cdot t}{365}\right)\Big\}$. The index j represents the frequency of each harmonic. Because the lower frequencies tend to explain most of the seasonal variation in average daily water use, the higher frequencies can often be omitted with little predictive loss.

[22] Total rainfall in period t is scaled by adding one in equation (3) to accommodate periods in which total rainfall is zero, in which case the logarithm of total unscaled rainfall would be undefined.

that β_{HWR} captures the effect on average daily water use of both changes to behavior and any induced participation in other conservation programs. We interact I_{it} with treatment group affiliation to separately estimate the mean treatment effect for each experiment. We also interact it with other indicators of household characteristics -- e.g., Water Score, paper vs. email report, consumption level -- to measure how the treatment effect varies with these factors.

We use Hausman's specification test to select between a fixed effects or a random effects estimator. While a random effects estimator can be more efficient, it depends on a more restricted set of assumptions about the structure of the model error, ε_{it}. Hausman's specification test indicated these assumptions were unlikely to hold and we therefore adopted a fixed effects estimation approach. The model was estimated in STATA (version 13.1) using the panel data estimator for fixed effects models with consistent standard errors for clustered data.

2. Data and Estimation

We compiled household metered consumption records from January 2006 to September 2013. We converted metered consumption to average daily use by dividing by the length of the billing period.[24] These data were then matched to the corresponding weather data for each billing period.

We collected daily weather measurements -- precipitation, maximum air temperature, and evapotranspiration -- from the CIMIS weather stations located in EBMUD's service area: Union City/Oakland Foothills (CIMIS station #149), Concord (CIMIS station #170), and Moraga (CIMIS station #178). Customer accounts were assigned to one of the three stations on the basis of zip code.

[23] Because average daily water use is logarithmically transformed prior to estimation, the coefficient β_{HWR} approximates the percentage difference in average daily water use between households that receive HWRs (i.e. receive the treatment) and households that do not.

[24] As previously noted, we treat the length of the billing period as a fixed 61 days. While this is not strictly true in all cases, doing so greatly simplifies the conversion of the billing data to average daily use with little predictive loss.

We then generated rolling bimonthly averages of rainfall, temperature, and evapotranspiration to exactly match the weather variables to the meter read dates for household water use.[25]

Data from meter reads can contain a lot of noise in the form of missing reads, duplicative reads, erroneous reads, and interpolated reads. This resulted in the elimination of some accounts from the sample for data quality reasons—too short a pre-intervention consumption history, change of residence, unconfirmed high consumption reading. Robust regression techniques were used on the remaining data to detect and address any residual data quality errors. This methodology determines the relative level of inconsistency of each observation with a given model form. A measure is constructed to depict the level of inconsistency between zero and one; this measure is then used as a weight in subsequent regressions. Less consistent observations are thereby down-weighted during model estimation.

Table 4 presents the counts on the final sample used to estimate the water use model given in equation (1).[26]

Table 4. Model Estimation Sample Sizes by Experiment

	Treatment	Control	Total
Castro Valley Group			
No. Households	10,529	13,765	24,294
No. Meter Reads	362,198	473,204	835,402
Random Group			
No. Households	1,710	1,576	3,286
No. Meter Reads	58,824	54,214	113,038

[25] A meter read represents the end of a consumption period. For example, a meter read on June 30 would represent consumption roughly from May 1 to June 30. However, meters are read on a schedule that often does not coincide with the start or end of calendar months. Thus, the need to work with daily weather data so that the weather variables can be correctly aligned with the corresponding consumption data.

[26] The sample sizes in Table 4 for the Castro Valley Group Experiment differ from the sample as originally developed by EBMUD and WaterSmart because we expanded the control group in order to get better statistical resolution on the treatment effect on water use.

3. Estimation Results

Model estimation results are presented in Table 5. The results are based on water consumption for 27,580 single family households between January 2006 and September 2013. This sample contains 1,710 households in the treatment group of the Random Group experiment, 10,529 households in the treatment group of the Castro Valley Group experiment, and over 15,000 single family control households.

The first variable in Table 5 is the overall intercept term. The estimated model also includes fixed effects intercepts for each household represented in the model, which are excluded from the table for obvious reasons of parsimony. Variables 2 thru 9 comprise the seasonal component, S_t, of the model. These correspond to the sines and cosines of the Fourier series in equation (2). These variables and their coefficients describe the shape of demand over the year given normal weather. Variables 10 thru 16 measure changes in average daily use that result from departures in weather from normal conditions. Thus, variables 10 and 11 indicate that above average rainfall pushes demand down (as one would expect), while variable 14 shows that higher than average evapotranspiration pushes demand up (again as one would expect). Interactions between season and weather are captured by variables 12 and 13 for rainfall and by variables 15 and 16 for evapotranspiration. The coefficients for these variables indicate that departures of evapotranspiration from normal produce the largest percentage effect in the spring growing season. Similarly, an inch of rainfall produces a larger effect on water use in the summer than in the winter.

The treatment effects of HWRs are captured by variables 17 thru 23. These variables represent an expanded version of equation (5) to include interactions with other household characteristics. The main effect for the Random and Castro Valley experiments are given by the coefficients for variables 17 and 18, respectively. Both are negative and statistically different from zero at better than 99% confidence, meaning the model definitively rejects the null hypothesis of no treatment effect on average daily water use.

Table 5. Household Average Daily Water Use Fixed Effects Model Estimation Results

Model Variable	Coeff.	St. Err.	t-statistic
1. Constant (Mean intercept)	5.2406	0.0003	17468.67***
2. First Sine harmonic, 12 month (annual) frequency	0.0485	0.0007	69.29***
3. First Cosine harmonic, 12 month (annual) frequency	-0.337	0.0017	-198.24***
4. Second Sine harmonic, 6 month (semi-annual) frequency	0.0000	0.0006	0.00
5. Second Cosine harmonic, 6 month (semi-annual) frequency	0.0003	0.0007	0.43
6. Third Sine harmonic, 4 month frequency	-0.0128	0.0008	-16.00***
7. Third Cosine harmonic, 4 month frequency	0.0235	0.0008	29.38***
8. Fourth Sine harmonic, 3 month (quarterly) frequency	-0.0076	0.0014	-5.43***
9. Fourth Cosine harmonic, 3 month (quarterly) frequency	-0.0092	0.0013	-7.08***
10. Deviation from logarithm of 61 day moving sum of rainfall	-0.0525	0.001	-52.50***
11. Bimonthly lag from rain deviation	-0.0051	0.0008	-6.38***
12. Interaction of contemporaneous rain with annual sine harmonic	-0.0475	0.0013	-36.54***
13. Interaction of contemporaneous rain with annual cosine harmonic	0.0126	0.0012	10.50***
14. Deviation from logarithm of 61 day moving average of CIMIS Evapotranspiration	0.2537	0.0047	53.98***
15. Interaction of CIMIS Evapotranspiration with ann. sine harmonic	0.261	0.0053	49.25***
16. Interaction of CIMIS Evapotranspiration with ann. cosine harmonic	0.1306	0.005	26.12***
17. Main Effect of HWR Intervention Random Group	-0.0564	0.0162	-3.48***
18. Main Effect of HWR Intervention in Castro Valley Group	-0.0742	0.0045	-16.49***
19. Interaction of HWR Intervention with bottom usage quartile (0-25%)	0.0292	0.0121	2.41**
20. Interaction of HWR Intervention with top usage quartile (76-100%)	-0.0116	0.0063	-1.84*
21. Interaction of HWR Intervention with Email Delivery	0.0111	0.0074	1.50
22. Interaction of HWR Intervention with Max Water Score of 2	0.0192	0.0123	1.56
23. Interaction of HWR Intervention with Max Water Score of 1	0.0546	0.0272	2.01**
Number of observations	948,440		
Number of households	27,570		
Standard Error of Individual Constant Terms (sigma_u)		0.59	
Standard Error of White Noise Error (sigma_e)		0.4172	
Time period of Consumption	Jan. 2006 - Sep. 2013		
* significant at 90% confidence level, ** significant at 95% confidence level, *** significant at 99% confidence level			

4. Mean Treatment Effect

The coefficients on variables 17 and 18 represent the treatment effect for households receiving paper reports with pre-treatment consumption that fell between the 25th and 75th percentiles. For the Random Group Experiment (variable 17), households in this category reduced consumption by approximately 5.5% (95% CI 2.4% to 8.4%).[27] Similarly situated households in the Castro Valley Group Experiment (variable 18) reduced consumption by approximately 7.1% (95% CI 6.3% to 8.0%).

To get the mean treatment effect for the full sample in each experiment we take a weighted average of the product of the intervention variables (17-23) and their estimated coefficients.[28] For the Random Group Experiment we estimate an overall mean treatment effect of 4.6%. For the Castro Valley Experiment we estimate an overall mean treatment effect of 6.6%. While the mean treatment effect for the Castro Valley Experiment is greater than for the Random Group Experiment, we cannot reject the possibility that this is due to chance, since the confidence intervals surrounding the two estimates overlap. However, a larger effect in the Castro Valley experiment is not implausible given the greater homogeneity of the homes in terms of household characteristics and water use (see Section IV.B). Indeed, a primary reason that EBMUD selected Castro Valley for the experiment was the belief that its homes would be good candidates for HWRs.

5. Impact of Household Water Use Percentile on Treatment Effect

A question relevant to the targeting of HWRs if they are not going to be provided on a universal basis is whether savings scale with level of water use. That is, is the treatment effect larger for households in the upper percentiles of consumption than for households in the lower percentiles? The model results suggest the answer is yes. We find that the treatment effect for households in the bottom quartile of use is reduced by about 2.9% while it is increased by about 1.1% for households in the upper quartile of use, relative to households with use in the inter-quartile range. The difference in treatment effect between households in the middle two quartiles and households in the upper quartile is not statistically significant. This is not the case for households in the bottom quartile, where the difference is significant. The results suggest utilities should consider giving households in the bottom quartile of use lower priority for receiving HWRs if they are not to be universally provided.

[27] While the coefficient is -0.0564, we are employing the estimator proposed by Kennedy (1981) for the expected percentage change for an indicator variable in a model with a logarithmically transformed right hand side variable, which is $exp(\hat{\beta}_{HWR} - 0.5\hat{\sigma}^2) - 1$, where $\hat{\beta}_{HWR}$ is the estimated coefficient and $\hat{\sigma}^2$ is its estimated variance. Thus, $exp(-0.0564 - 0.5(0.0162)^2) - 1 \approx -0.055$.

[28] We again employ the second order correction described in the previous footnote.

6. Impact of Water Score on Treatment Effect

We also find that treatment effect scales with HWR score. Thus, we calculate a treatment effect of 7.1% if a household in the Castro Valley treatment group received a HWR score of 3 (Take Action!); a treatment effect of 5.2% if it received a HWR score of 2; and a treatment effect of just 1.6% if it had a HWR score of 1 (Doing Great!). We view the results as indicating correlation of treatment effect with the HWR score but not necessarily causation. While it is certainly plausible that the injunctive norms implicit in the HWR scores and corresponding emoticons may motivate participation -- after all households getting a score of 3 are the only households explicitly told to take action -- it also could be the case that other underlying factors that correlate with the score are causing the response. For example, since score correlates with where households fall within the distribution of water use within their cohort, it could also be the case that households in the upper percentiles of their comparison group find more ways to reduce water use -- perhaps because they have more older water using fixtures or larger landscapes where they can make adjustments -- while households in the lower percentiles may already have efficient fixtures and perhaps minimal or already water efficient landscapes. In this case, while the injunctive norm to take action may provide some of the motivation to reduce use, other factors could also be at play. Allcott (2011) addressed this question with respect to the mean treatment effect of Opower home energy reports on energy consumption and concluded that the injunctive norms could explain no more than 15% to 30% of the differential effect in response across scores. As we discuss later in the report, we find similar effects between the HWR score and the odds of a household participating in an EBMUD audit or rebate program.

7. Impact of Paper vs Electronic Reports on Treatment Effect

Because electronic reports delivered by email offer a definite cost advantage over paper reports, there is interest in whether they generate equivalent savings. The model results suggest they do not. The coefficient on variable (21) indicates that email reports reduced the treatment effect by about 1%. We also note, however, that we cannot reject the null hypothesis of no difference at the 95% confidence level. Thus our evaluation does not provide a definitive answer to the question, other than to suggest that paper reports appear to have greater impact, on average. Even if savings are lessened by use of email reports, the cost savings may nonetheless justify their use. We take up this question later in the report.

8. Seasonal Shape of Treatment Effect

Preliminary models provide some evidence of stronger treatment effects associated with reports received in the fall and winter than in the spring and summer.[29] This finding suggests a lagged response to high water use since households using large amounts of water in the summer do not receive feedback on this until they receive reports in the late summer and fall. More research is needed to better parse the seasonality of treatment effect. In particular, a longer period of treatment spanning more than 12 months would provide better information with which to examine this question.[30]

B. CONSERVATION PROGRAM PARTICIPATION

As discussed in the previous section, embedded in the mean treatment effect on average daily water use are any changes in use resulting from increased participation in EBMUD audit and rebate programs. In this section we examine the question of whether and to what extent homes in the treatment groups of the two Pilot experiments were more likely to participate in EBMUD audit and rebate programs. Home energy reports have been reported to increase customer participation in other energy efficiency programs. This effect is sometimes referred to as "uplift" or "channeling" in the literature. Opower, the largest provider of home energy reports, claims its home energy reports have increased participation in other energy conservation programs by 17 to 59%. On its website, WaterSmart claims up to a three-fold increase in program participation for homes receiving HWRs.[31]

1. Pre-Pilot Program Participation

During the four years prior to the start of the Pilot, program participation rates were similar for the treatment and control groups of both experiments, as shown in Figure 12 and Figure 13. In particular, both groups exhibit similar trends in participation over time, with participation declining steadily from 2008 to 2011. The sample participation rates in each year are statistically equivalent between the control and treatment groups of each experiment with one exception. The one exception is 2008 rebate participation rates for the Castro Valley Experiment. In subsequent years, however,

[29] This seasonal effect was also detected by Brent, et al (2013).

[30] Recall that households in Phase 4 did not start receiving HWRs until October or November and these homes comprised the bulk of the Castro Valley treatment group.

[31] http://www.watersmartsoftware.com/our-solution.html#our-solution

the differences in the sample participation rates are not statistically significant at the 95% confidence level.[32]

From this data we conclude that rates of participation for the treatment and control groups in both experiments were very similar both in magnitude and trend leading up to the Pilot. This suggests the key identifying assumption of parallel trend needed for the difference-in-differences modeling approach is likely to hold for the two experiments with regard to program participation.

[32] If the paired confidence intervals in Figure 12 overlap, it indicates the difference between the control and treatment group participation rates are not statistically significant. Conversely, if they do not overlap, it indicates the difference is statistically significant. Note the wider confidence interval for the control group compared to the treatment group in the Castro Valley Group Experiment is due to the smaller sample size for the control group.

Figure 12. Pre-Pilot Audit Program Participation Rates

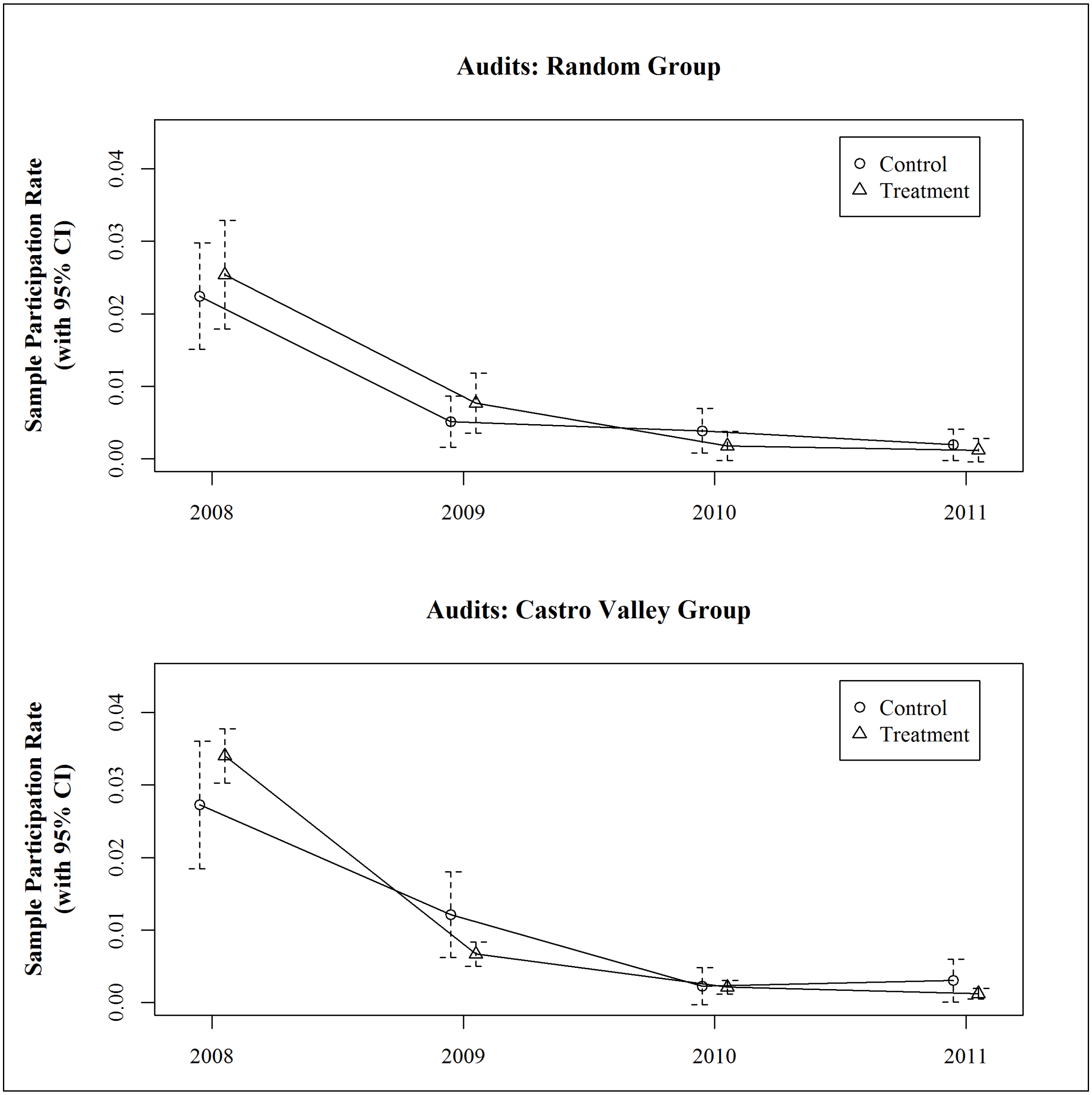
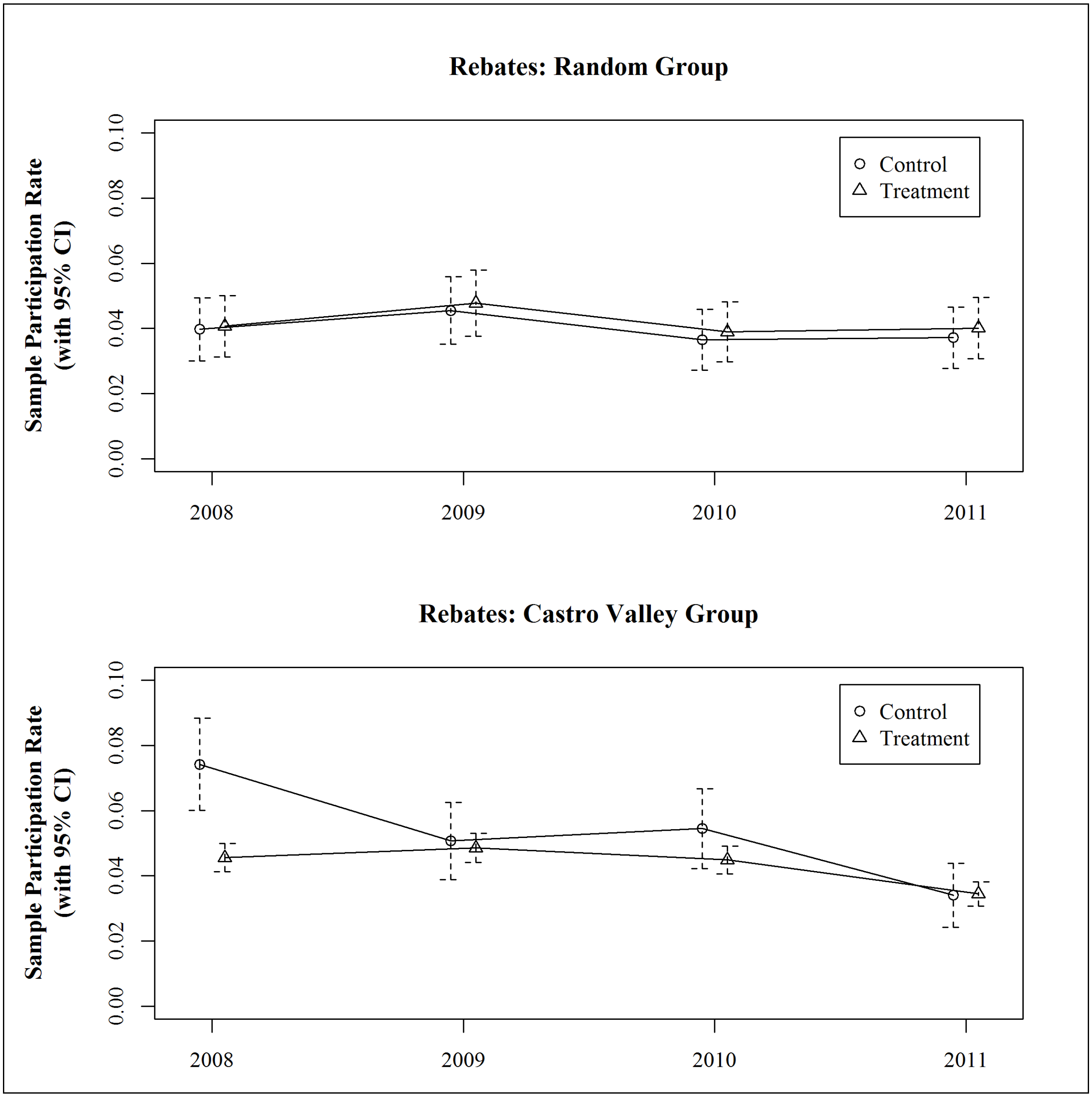


Figure 13. Pre-Pilot Rebate Program Participation Rates



2. Post-Pilot Program Participation

Participation rates in audit and rebate programs pre- and post-treatment are summarized in Figure 14 and Figure 15. These rates are for the Phase 1/2 treatment groups, which received a full year of HWRs. We define the pre-treatment period as the year prior to the start of Phase 1/2, roughly

6/15/2011 to 6/14/2012, and the post-treatment period as the year of the Pilot, roughly 6/15/2012 to 6/14/2013.

Differences in participation between the control and treatment groups in the pre-treatment period are statistically insignificant at the 95% level of statistical confidence. However, while audit participation rates for the control groups remain essentially unchanged in the post-treatment period, rates for the treatment groups increase sharply in both experiments. Given the parallel pattern in audit participation rates leading up to the Pilot (Figure 12 and Figure 13), this suggests that HWRs had a definite effect on a home's decision to request an audit. For rebates, the parallel pattern in participation between control and treatment homes is also broken -- participation rates decrease between the pre and post periods for the control groups, while they increase for the treatment groups. The effect is clearly not as large as for audits, but the trend reversal suggests there is some effect.

Figure 14. Pre- vs. Post-Treatment Audit Participation Rates

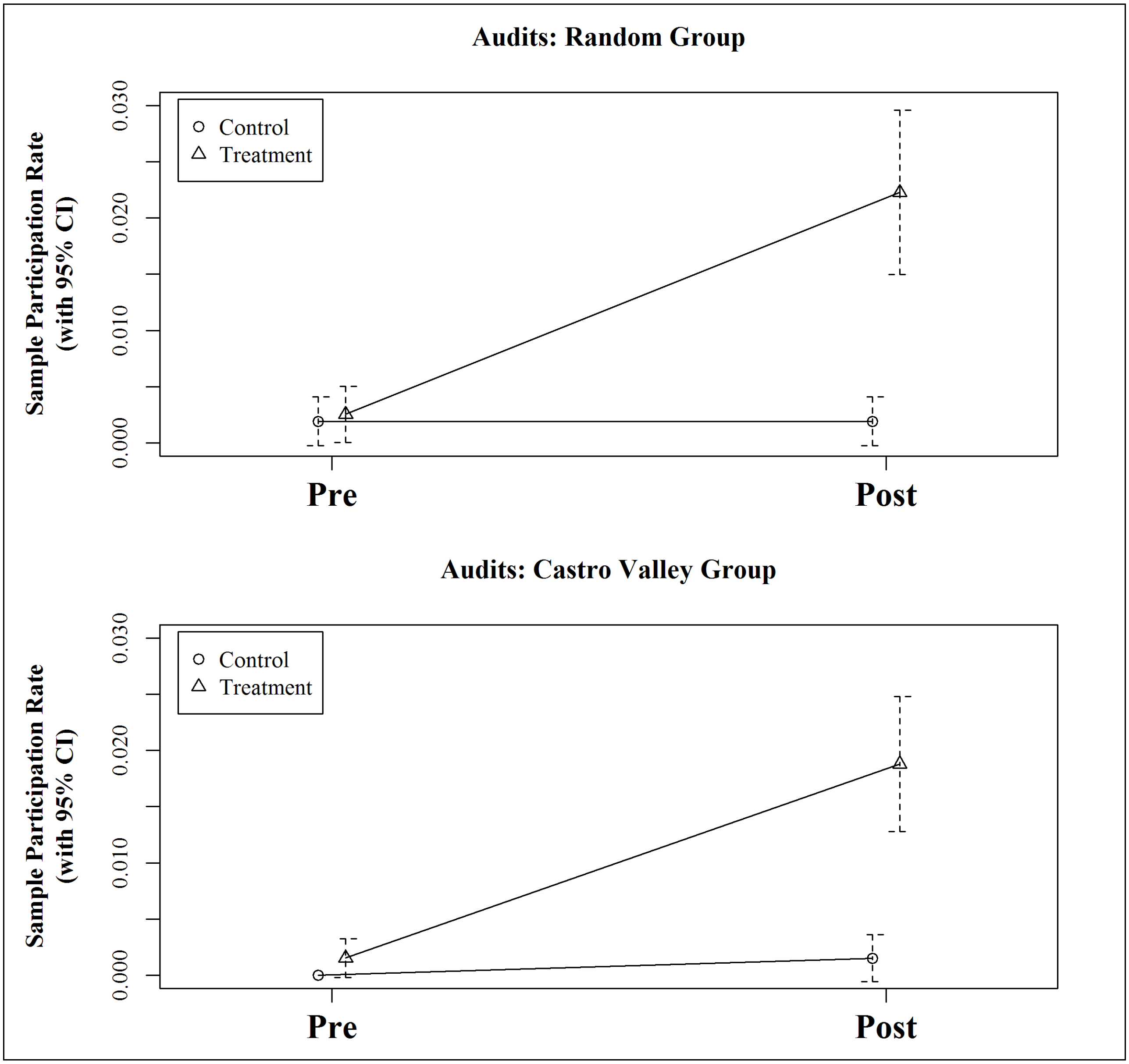
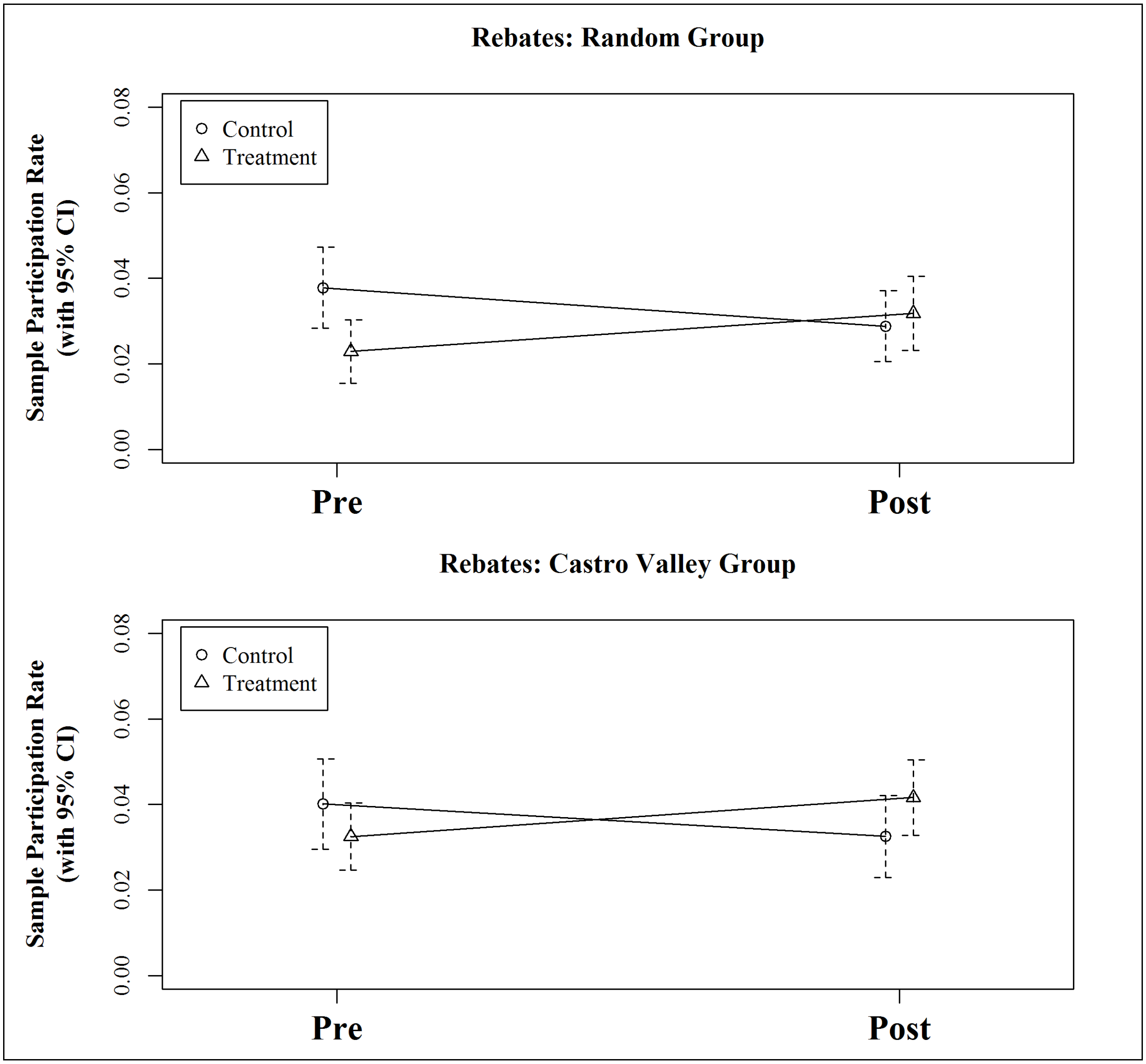


Figure 15. Pre- vs. Post-Treatment Rebate Participation Rates



3. Estimation of Mean Treatment Effect

We use logit regression techniques to estimate the strength of the effect of HWRs on the choice to participate in EBMUD rebate and water audit programs. Logit regression can be used to predict the outcome of a categorical dependent variable (in our case, the choice to participate in an

EBMUD conservation program) based on one or more predictor variables.[33] For this analysis, the key predictor variables are whether a household is in the control group or treatment group and whether the time period is pre-treatment or post-treatment. Other predictor variables are whether the household had previously participated in a rebate program, size of the household, and landscape characteristics.[34] The general specification of the model follows the DID specification in Puhani (2008) for estimating the treatment effect on a dichotomous choice variable in the context of a nonlinear regression model. The basic model is given in equation (6)

(6) $p_i = F(\beta_1 T_i + \beta_2 G_i + \beta_3 T_i \times G_i + \boldsymbol{x}_i \boldsymbol{\Theta})$, where

p_i is the probability of participation, T = 1 in the post treatment period and 0 otherwise, G = 1 if the household is in the treatment group and 0 otherwise, the $\boldsymbol{x}_i$ are additional predictor variables, and $F(\cdot)$ is the cumulative distribution function for a logistic random variable. The coefficient β_3 on the interaction term $T_i \times G_i$ measures the mean treatment effect of HWRs on participation in other EBMUD programs. The increase in the odds that a household chooses to participate in an EBMUD program given that it was in the treatment group is given by e^{β_3}.

We estimated the model given by equation (6) separately for each experiment and also for the pooled experiments. Results were similar across the individual and pooled experiments. We therefore present the results for just the pooled estimation.

Table 6 presents the results for combined participation in rebate and audit programs. That is, it shows the mean treatment effect on overall program participation, without regard to type of

[33] In the logit regression model, the probability, p, that the observed value y takes the value 1 (e.g., the household participates in a program) given the predictor variable x is $p = P[L \leq \alpha + \beta x] = F(\alpha + \beta x) = \frac{1}{1+e^{-(\alpha+\beta x)}}$, where L is a random variable that follows the logistic distribution. Given a set of observed values for y and x, maximum likelihood estimation techniques are used to estimate values for α and β. It can be shown that the log of the odds that y takes the value 1 is equal to $\alpha + \beta x$, or equivalently, the odds that y takes the value 1 is equal to $e^{\alpha+\beta x}$. If x is a binary predictor variable, such as x=1 if treatment is received and x=0 otherwise, then the change in the odds that y takes the value 1 given x=1 is equal to e^{β}. This provides a particularly convenient way to assess the strength of the effect of x on the odds that y takes the value of 1 (e.g., the odds the household participates in a program).

[34] Past participation in audit programs was not found to have a statistically significant effect and was dropped from the model.

program. The coefficient on the treatment effect variable, TG, is positive and is statistically significant at the 99% level of confidence, indicating treatment increased the likelihood of participation in other EBMUD programs.

Model results indicate that households receiving HWRs were 2.34 times more likely to participate in other EBMUD conservation programs than households not receiving HWRs.[35]

Table 6. Mean Treatment Effect on EBMUD Program Participation

Predictor Variable	Logit Model Coefficient	t-statistic
TG (treatment effect)	0.84943	4.4920 ***
T (time effect)	-0.23407	-1.6349
G (group effect)	-0.30235	-2.1869 **
PPH (persons per household)	-0.074468	-0.19239
PREV_REB (received a rebate in prior 3 years)	-0.71114	-4.1360 ***
IR_SMALL (irrigable area 4000 sqft or less)	-0.17816	-1.9393 *
CONSTANT	-2.9795	-19.494 ***
* significant at 90% confidence level, ** significant at 95% confidence level, *** significant at 99% confidence level no. obs. = 12,672		

Table 7 presents the results for participation in audit programs only. The coefficient on the treatment effect variable, TG, is positive and is statistically significant at the 95% level of confidence, indicating treatment increased the likelihood of participation in other EBMUD programs. Model results indicate that households receiving HWRs are 6.2 times more likely to participate in EBMUD audit programs than households not receiving HWRs.[36]

[35] The increase in the odds of participation is calculated by exponentiation of the coefficient for TG ($e^{0.84943} \approx 2.34$). If the probability of participation without treatment is p0 and the probability of participation with treatment is p1, then the odds ratio is defined as $\frac{p1/(1-p1)}{p0/(1-p0)}$. Given the odds ratio and knowledge of the probability of participation without treatment, one can easily calculate the probability of participation given treatment. Suppose the probability of participation without treatment is 1% and the odds ratio is 2.34. Then the probability of participation given treatment is $p1 = \frac{2.34(0.01/0.99)}{1+2.34(0.01/0.99)} \approx 0.023, or\ 2.3\%$.

[36] Again, the increase in the odds of participation is calculated by exponentiation of the treatment effect coefficient TG ($e^{1.8263} \approx 6.2$).

Table 7. Mean Treatment Effect on EBMUD Audit Program Participation

Predictor Variable	Logit Model Coefficient	t-statistic
TG (treatment effect)	1.8263	2.1975 **
T (time effect)	0.51160	0.70065
G (group effect)	0.75408	-1.0921
PPH (persons per household)	-0.20386	2.1686 **
IR_SMALL (irrigable area 4000 sqft or less)	0.57681	-2.5435 **
CONSTANT	-6.6846	-10.316 ***
* significant at 90% confidence level, ** significant at 95% confidence level, *** significant at 99% confidence level no. obs. = 12,672		

Table 8 presents the results for participation in rebate programs only. The coefficient on the treatment effect variable, TG, is positive and is statistically significant at the 95% level of confidence, indicating treatment increased the likelihood of participation in other EBMUD programs. Model results indicate that households receiving HWRs are 1.66 times more likely to participate in EBMUD rebate programs than households not receiving HWRs.

Table 8. Mean Treatment Effect on EBMUD Rebate Program Participation

Predictor Variable	Logit Model Coefficient	t-statistic
TG (treatment effect)	0.50382	2.5224 **
T (time effect)	-0.26467	-1.8107 *
G (group effect)	-0.32403	-2.3019 **
PPH (persons per household)	-0.027124	0.65012
PREV_REB (received a rebate in prior 3 years)	-0.81709	-4.1760 ***
CONSTANT	-3.2026	-20.870 ***
* significant at 90% confidence level, ** significant at 95% confidence level, *** significant at 99% confidence level no. obs. = 12,672		

Table 9 summarizes the mean increase in the odds of program participation given a household received HWRs and its corresponding 95% confidence interval. In each case, we reject the null hypothesis of no treatment effect at the 95% level of confidence.[37]

[37] The confidence intervals are calculated as $e^{\widehat{\beta}_3 \pm 1.96SE}$, where SE is the standard error on the coefficient estimate. The broad span of the confidence interval for Audits Only is driven by the very low audit counts overall for the sample of pre- and post-treatment observations. Only 86 audits were completed out of 12,672 pre- and post-treatment observations. The great majority of these audits were completed in the post-treatment period on homes in

Table 9. Mean Treatment Effect on Odds of Conservation Program Participation

EBMUD Conservation Program	Odds Ratio*	95% CI Lower	95% CI Upper
Pooled Audit and Rebate Participation	2.34	1.61	3.39
Audits Only	6.21	1.22	31.66
Rebates Only	1.66	1.12	2.45
* The odds ratio shows the increase in the odds of program participation given the household received HWRs. The null hypothesis that HWRs do not affect the odds of participation is rejected when the lower bound of the 95% CI is greater than 1.			

4. Effect of HWR Score on Likelihood of Program Participation

Recall that if household consumption is in the 20th percentile of their cohort (HWR score = 1) the report tells them they are efficient, has a large smiling emoticon, and a reinforcing message telling them they are great. If consumption is between the 20th and 55th percentiles (HWR score = 2) the report tells them they are average and gives a smiley emoticon without a reinforcing message. If consumption is above the 55th percentile (HWR score = 3) the report tells them they need to take action and reinforces the message with a worried face emoticon.

As we did for the treatment effect on average daily water use, we consider whether participation in other conservation programs is influenced by the initial score received. It seems reasonable to expect that households told they are efficient and doing great would see less reason to participate in an audit or rebate program -- why fix what's not broken -- than households told they are using too much water and need to take action. We test for this effect by extending the model in equation (6) by interacting the treatment effect term with HWR score indicator variables.

The extension of the model is given in equation (7)

$$(7)\quad p_i = F(\beta_1 T_i + \beta_2 G_i + \beta_3 T_i \times G_i + \beta_4 T_i \times G_i \times SCR1_i + \beta_5 T_i \times G_i \times SCR2_i + \boldsymbol{x}_i \boldsymbol{\Theta}),$$

where SCR1 and SCR2 are binary indicator variables that take the value 1 if the household received the indicated score and 0 otherwise. The coefficient β_4 measures the effect of an initial HWR score of 1 on participation (relative to a score of 3) while the coefficient β_5 measures the effect of an initial HWR score of 2. The change in the odds of participation given treatment is therefore given as

---

the treatment group. However, with so few observations on completed audits, the variance on the odds ratio for audits only is large.

$$Odds\ Ratio = \begin{cases} e^{\beta_3+\beta_4}\ given\ HWR\ score = 1 \\ e^{\beta_3+\beta_5} given\ HWR\ score = 2 \\ e^{\beta_3} given\ HWR\ score = 3 \end{cases}$$

If initial scores of 1 and 2 decrease the incentive to participate in programs we would expect the estimated values for β_4 and β_5 to be negative and statistically significant. Moreover we would expect the magnitude of β_4 in absolute value to exceed that of β_5. This is in general what we find, as reported in Table 10, which shows the estimated treatment effect coefficients and their t-statistics for rebates and audits combined, audits only, and rebates only. The score effects are all significant at the 99% level of confidence for the audit and/or rebate and audit only participation models. The score effects are of the expected sign and rank order for the rebate only participation model, but they do not have statistical significance at the 95% confidence level.

Table 10. Effect of Initial WaterSmart Score on Mean Treatment Effect on EBMUD Program Participation

Predictor Variable	Logit Model Coefficient	t-statistic
Model 1: Participation in audit and/or rebate programs		
TG (β_3)	1.1770	5.9543 ***
TG_SCR1 (β_4)	-0.89304	-3.9858 ***
TG_SCR2 (β_5)	-0.63286	-3.4769 ***
Model 2: Participation in audit programs only		
TG (β_3)	2.3912	2.8671 ***
TG_SCR1 (β_4)	-2.3165	-3.8936 ***
TG_SCR2 (β_5)	-1.2261	-3.6866 ***
Model 3: Participation in rebate programs only		
TG (β_3)	0.68080	3.1614 ***
TG_SCR1 (β_4)	-0.44076	-1.7778 *
TG_SCR2 (β_5)	-0.31475	-1.4671
* significant at 90% confidence level, ** significant at 95% confidence level, *** significant at 99% confidence level no. obs. = 12,672		

The increase in the odds of participation given the initial HWR score for the three participation models are shown in Table 11.[38] As with the effect of score on average daily water use,

[38] The 95% confidence intervals for the score=1 category is calculated as $e^{(\hat{\beta}_3+\hat{\beta}_4)\pm 1.96SE}$, where $SE = \sqrt{\left(SE_{\hat{\beta}_3}\right)^2 + \left(SE_{\hat{\beta}_4}\right)^2 + 2\rho SE_{\hat{\beta}_3} SE_{\hat{\beta}_4}}$ and ρ is the correlation between $\hat{\beta}_3$ and $\hat{\beta}_4$. The confidence interval for the score = 2 category is done in the same way.

we view the results as indicating correlation of treatment effect but not necessarily causation. Again, while it is plausible that the injunctive norms implicit in the HWR scores and corresponding emoticons are motivating participation it also could be the case that other underlying factors that correlate with the score are causing the response.[39] What we can say is there is clearly a differential response in participation across score categories and that some of the differentiation may result from the injunctive norm (e.g., take action) and some may be due to other underlying factors.

Table 11. Effect of Initial HWR Score on Odds of Program Participation

Initial WaterSmart Score	Odds Ratio*	95% CI Lower	95% CI Lower
Model 1: Participation in audit and/or rebate programs			
Score = 3 (e^{β_3})	3.24	2.20	4.78
Score = 2 ($e^{\beta_3+\beta_5}$)	1.72	1.09	2.72
Score = 1 ($e^{\beta_3+\beta_4}$)	1.33	0.79	2.24
Model 2: Participation in audit programs only			
Score = 3 (e^{β_3})	10.93	2.13	56.03
Score = 2 ($e^{\beta_3+\beta_5}$)	3.21	0.57	17.89
Score = 1 ($e^{\beta_3+\beta_4}$)	1.08	0.15	7.74
Model 3: Participation in rebate programs only			
Score = 3 (e^{β_3})	1.98	1.30	3.01
Score = 2 ($e^{\beta_3+\beta_5}$)	1.44	0.88	2.36
Score = 1 ($e^{\beta_3+\beta_4}$)	1.27	0.73	2.20
* The odds ratio shows the increase in the odds of program participation given the household received HWRs. The null hypothesis that HWRs do not affect the odds of participation is rejected when the lower bound of the 95% CI is greater than 1.			

The broader and arguably more important point is that households with scores of 3 are more likely to participate (for whatever reason) in other conservation programs and this is useful information for where to target HWRs if it is not possible to provide universal coverage within a service area. While our results should not be interpreted to imply that there is value to adjusting the scores to place more households in the score = 3 category, they do suggest that targeting HWRs to homes that fall within this category is likely to yield better results in terms of channeling customers to other programs than not targeting.

C. WATER USE AWARENESS

A third objective of the Pilot was to see if HWRs increase household knowledge and awareness of its water consumption. Knowledge about water use can take many forms, and thus this

[39] See Section V.A.6 for additional discussion on this topic.

is really a multi-dimensional question. For example, in the most straightforward sense, households may be considered knowledgeable about their water use if they know approximately how much they use overall or for particular purposes. However, they may also be considered knowledgeable if they have a general understanding of how to efficiently perform various water-related things (e.g. taking a shower, washing clothes, irrigating landscape) even if they are unsure of the exact quantities of water involved. In this section we consider the more straightforward sense of water use awareness as well as less tangible measures.

1. Household Estimates of Water Use

The pre- and post-Pilot customer surveys asked respondents to estimate how much water they use on average on a daily basis. Presumably, if HWRs are effective at increasing household awareness of water use, estimates from homes in the treatment group would become more accurate and less biased (i.e., not show a marked tendency to over or underestimate consumption) relative both to their pre-pilot estimates and to estimates by control group households. Ideally, we would want to compare how responses to the same question about water use differed between the pre- and post-pilot surveys. However, this is not possible because the question about water use is different in the two surveys. In the pre-pilot survey, households were asked to estimate how much water they use on average on a daily basis across the entire year. In the post-pilot survey, households were asked the same question but for the summer and for the winter. Thus the responses in the two surveys are not directly comparable. However, it is still possible to look at accuracy and bias of responses between the treatment and control groups in the post-Pilot survey to get a measure of whether HWRs improved households' quantitative estimates of their water consumption.

It is a truism among those working in the water industry that most households generally have no idea how much water they use on a daily basis. Both the pre- and post-Pilot survey responses seem to bear this out. In the pre-Pilot survey more than 40% of households either indicated they did not know their water use or left the question blank. In the post-Pilot survey, this proportion increased to over 55%. Notably, there is little difference in the response rates between the control and treatment groups, suggesting that households receiving HWRs are no more likely to think they know their water use than households not receiving HWRs.

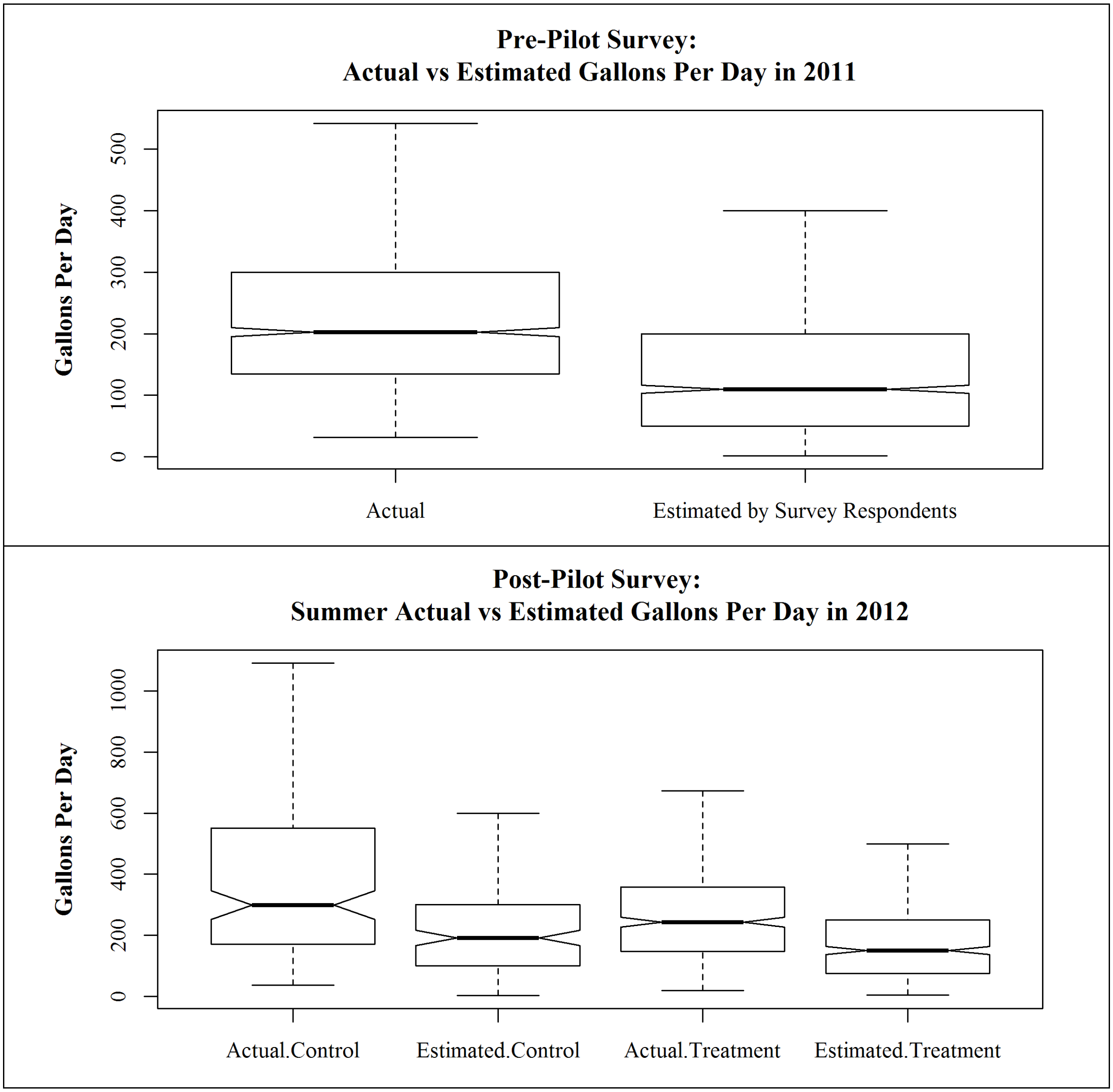
Households that did provide an estimate of average daily water use showed a significant downward bias in both the pre- and post-Pilot surveys. This is illustrated in Figure 16, which compares actual to estimated water use for those respondents providing use estimates. The tendency for households to underestimate their water consumption by a fairly wide margin is present in both periods, regardless of treatment.

In the post-Pilot survey the distribution of estimation errors is similar between the control and treatment groups, as shown in Figure 17. Both groups consistently underestimate their average use. Note that the mean error (the curved lines in Figure 17) is nearly identical for both groups and shows a tendency to increase with usage, indicating that high water use homes are more likely to underestimate their water use by a wider margin than low water use homes, which is not especially surprising.

In summary, we do not find evidence that HWRs increased the ability of households to provide an accurate quantitative estimate of their average daily water use. The proportion of homes stating they did not know or not answering the question was essentially the same between households in the control and treatment groups responding to the post-Pilot survey. Similarly, the tendency to underestimate daily water use was also generally the same between control and treatment households responding to the survey. It may be that over time this will change and as households receive more HWRs they will begin to incorporate this information into their general understanding of how they use water. Importantly, WaterSmart reported gallons per billing period rather than gallons per day on HWRs prior to April 2013, so it may be that households will be better at estimating daily water use going forward. However, it should also be noted that EBMUD has for a long while now provided average daily water use on its bills. Judging from the results of the pre-Pilot survey, however, this practice has not had much of an impact on the ability of households to estimate their water use.

It is also not obvious to us that a quantitative knowledge of daily water use is particularly relevant to most decisions about household water use. Households can make informed decisions on water use regardless of knowing precisely how much water they use. For example, to irrigate efficiently it is perhaps more important to know when and how long sprinklers should run and which parts of the yard can be effectively served with drip, than it is to know how much water the irrigation system uses. Similarly, the choice of a new washer may be made more effectively based on its potential to save on water and energy bills than on how many gallons per load it uses. Besides, our results on the effect of HWRs on household water use and decisions to participate in other conservation programs clearly show that households are responsive to the combination of social norms, injunctive messaging, and actionable information presented in HWRs regardless of their ability to accurately estimate water use. In this respect, the proof is in the pudding when it comes to whether HWRs work.

Figure 16. Pre- and Post-Pilot Estimates of Household Water Use Compared to Actual Water Use



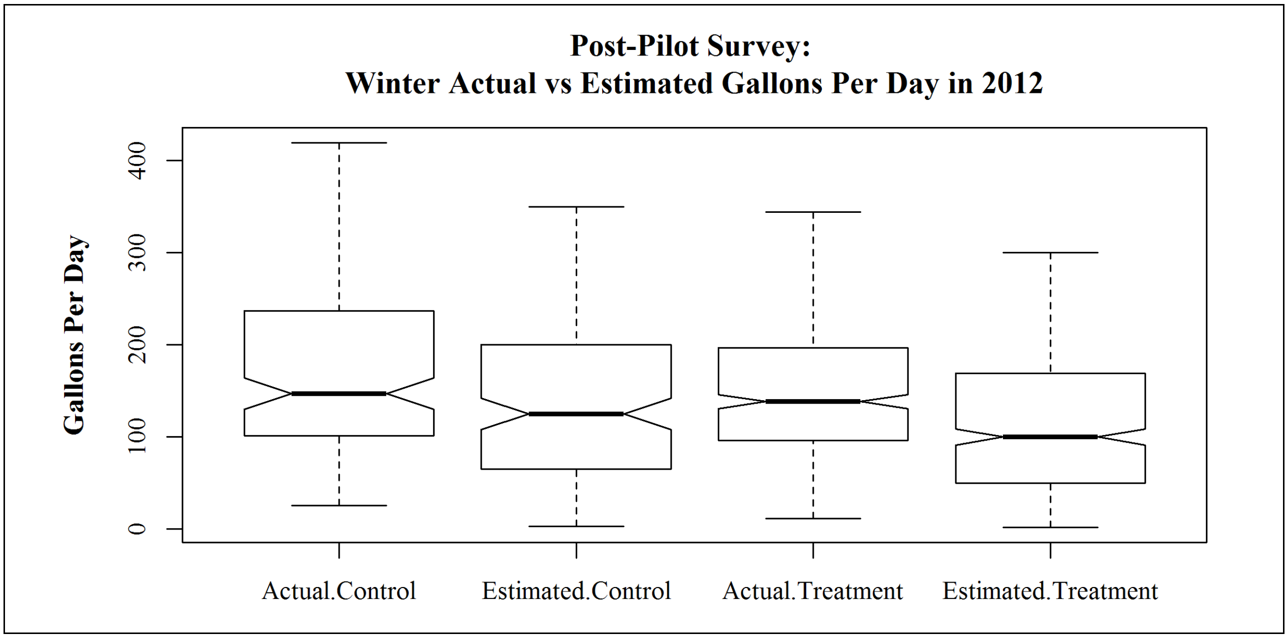
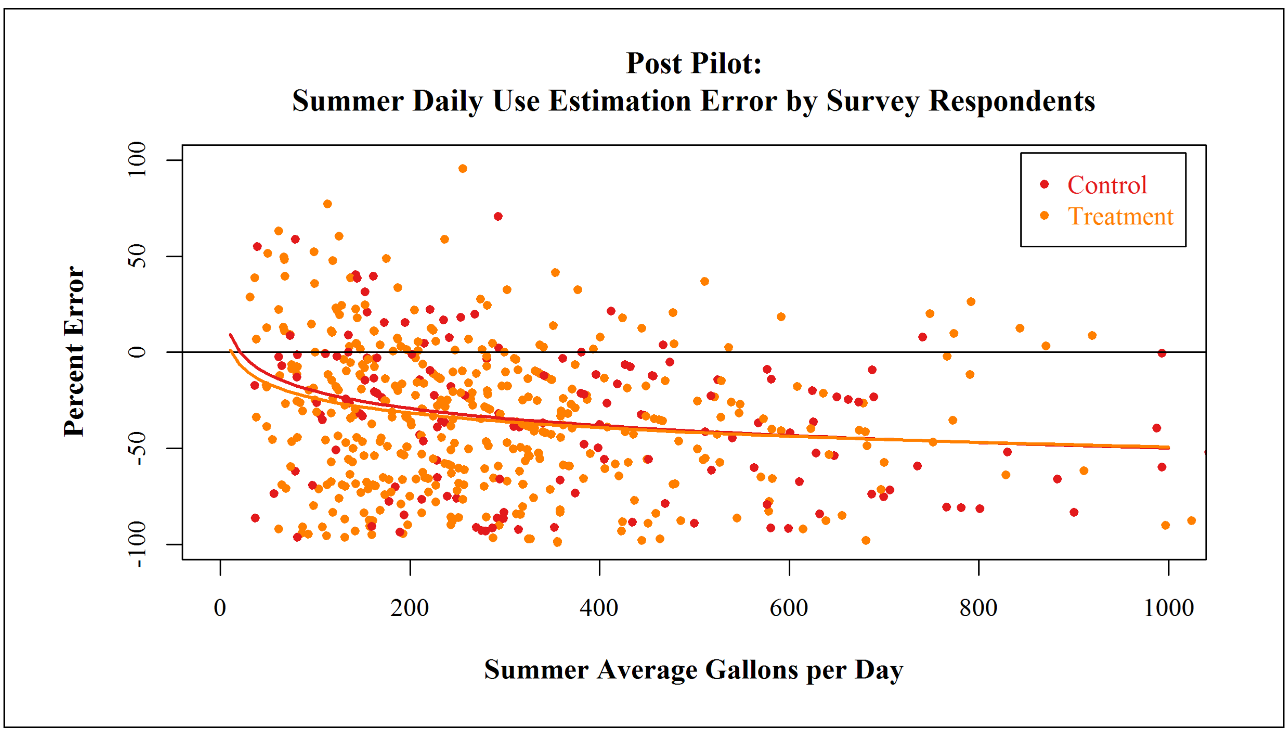
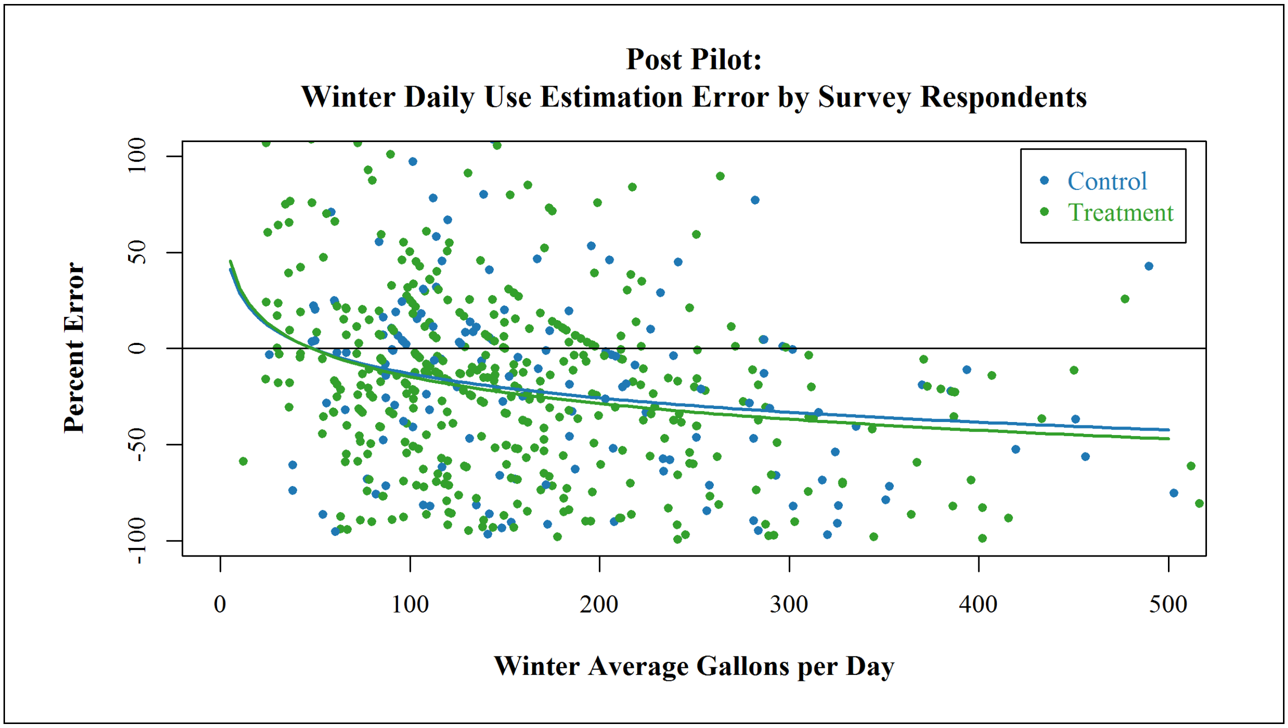


Figure 17. Post-Pilot Distribution of Water Use Estimation Error for Treatment and Control Households





2. Household Commitment Towards Water Conservation

Another dimension of a household's awareness and knowledge of its water use might be viewed in terms of its expressed beliefs in the importance of using water efficiently. The pre- and post-Pilot surveys asked respondents whether they agreed strongly with the following four statements:

1. I make an active commitment to use water efficiently.
2. It is important to me to reduce my water bills.
3. I talk with others in my household about reducing our water use.
4. I talk with friends and neighbors about ways to conserve water.

If HWRs are effective in shaping attitudes along these lines, we might expect to see more households in the treatment group indicating they agree strongly with the above statements than households in the control group. We used a difference-in-differences logit model to test for this effect. Our analysis generally does not support the hypothesis that survey respondents that were in the treatment group are more likely to agree strongly with the above statements than respondents that were in the control group.

The estimated odds ratio from the logit model and its 95% confidence interval for agreeing strongly with each of the four statements is shown in Table 12. The null hypothesis of no treatment effect is only rejected in the case of the second statement -- *It is important to me to reduce my water bills* -- and then only barely. In the other three cases, the null hypothesis of no treatment effect is not rejected, meaning that HWRs do not appear to change the odds that a household will agree strongly with the statements.

These results should be interpreted with some caution. Self-selection bias in opinion polls of this type is a common problem (Groves & Peytcheva, 2008). Typically individuals that are interested in a topic or have strongly held views on the topic are more likely to respond to voluntary polls. This can result in a biased sample that disproportionately represents individuals with particularly strong opinions or beliefs -- such as strongly supporting the need for water conservation. This type of self-selection could be present in both treatment and control group respondents.[40]

Table 12. Treatment Effect on Odds Survey Respondent Strongly Agrees with Statement

Statement	Odds Ratio*	95% CI Lower	95% CI Upper
I make an active commitment to use water efficiently	0.98	0.73	1.30
It is important to me to reduce my water bills	1.37	1.03	1.83
I talk with others in my household about reducing our water use	1.19	0.90	1.59
I talk with friends and neighbors about ways to conserve water	1.17	0.79	1.70
* The odds ratio shows the increase in the odds of strongly agreeing with the statement given the respondent received HWRs. The null hypothesis that HWRs do not affect the odds of strong agreement is rejected when the lower bound of the 95% CI is greater than 1.			

3. Getting Help on How to Save Water

The pre- and post-Pilot surveys also asked households to score EBMUD in terms of the following:

1. Explaining your water use on your bill
2. Showing you ways to save money on your water bill by conserving water
3. Giving you the tips and tools you need to use water efficiently

[40] The fact that the question about water bills is the only one showing a statistically significant treatment effect lends some support to this possibility. Self-selection is more likely to be caused by strong moral beliefs in the importance of and need for water conservation than by a strong pecuniary interest in lowering water bills.

If households view HWRs as providing useful and actionable information we would expect respondents that received HWRs to be more likely to give EBMUD a high score in these areas then other respondents. Again, we used a differences-in-differences logit model to test for this effect. In this case, we reject the null hypothesis of no treatment effect in each case, meaning we find evidence that respondents who received HWRs are more likely to score EBMUD high in terms of providing useful information for managing household water use and bills then other respondents. The estimated odds ratio from the logit model and its 95% confidence interval for scoring EBMUD as "Excellent" in each of the three above categories is shown in Table 13. The estimated mean increase in the odds of scoring EBMUD as "Excellent" ranges from 52 to 80%.

Table 13. Treatment Effect on Odds Survey Respondent Scores EBMUD as "Excellent" by Area of Assistance

Area of Assistance	Odds Ratio*	95% CI	
		Lower	Upper
Explaining your water use on your bill	1.52	1.07	2.17
Showing you ways to save money on your water bill by conserving water	1.64	1.12	2.41
Giving you the tips and tools you need to use water efficiently	1.80	1.23	2.65
* The odds ratio shows the increase in the odds of respondent score EBMUD as "Excellent" given the respondent received HWRs. The null hypothesis that HWRs do not affect the odds of an "Excellent" score is rejected when the lower bound of the 95% CI is greater than 1.			

D. COST EFFECTIVENESS

Many considerations go into a utility's decision to implement a specific demand management program. A key one is the cost of the program relative to other alternatives or doing nothing at all. A common metric for assessing the relative cost of a demand management program is to calculate the unit cost of water savings, which can then be compared to the unit cost of water savings for other demand management options as well as to the unit cost of water supply. In California, unit costs are typically expressed in dollars per acre-foot.[41]

In its most general form, the equation for calculating unit cost is given by equation (8)

$$(8) \qquad UC_i = \frac{\sum_{t=0}^{t=T_i-1} C_{it} / (1+d)^t}{\sum_{t=0}^{t=T_i-1} W_{it} / (1+d)^t}, \text{ where}$$

[41] An acre-foot of water is the volume of water that would cover an acre one foot deep in water, approximately 325,851 gallons.

C_{it} is the cost incurred by the utility in year t from implementing program i, W_{it} is the water savings expected from program i in year t, T_i is the number of years savings from program i are expected to last, and d is the discount rate. When program costs and savings last just one year, the general equation for unit cost simplifies to the ratio of annual cost to annual savings, as shown in equation (9).

$$(9) \qquad UC_i = {}^{C_i}/_{W_i} \ when\, T_i = 1$$

Equation (9) is applicable for HWRs if we make the conservative assumption that savings occur in the year in which the HWRs are received and do not persist beyond this time.[42]

1. Average Water Savings Per Household

Results from the Pilot indicate a mean treatment effect for the Random Group Experiment in the range of 4.5 to 6.5% for households receiving paper reports by mail and in the range of 3.5 to 5.5% for households receiving electronic reports by email. Because the Random Group Experiment is representative of the distribution of households for the entire EBMUD service area, these ranges provide appropriate estimates of expected water savings if the program were extended to the entire service area.

Pre-treatment mean water use for households in the Random Group Experiment was about 261 gallons per day, or about 95,265 gallons per year. Average annual household water savings would therefore be expected to range between 4,287 and 6,192 gallons for households receiving paper reports and between 3,334 and 5,240 gallons for households receiving electronic reports. Converting to acre-feet, the expected savings would be 0.0132 to 0.0190 acre-feet for paper reports and 0.0102 to 0.0161 acre-feet for email reports.

[42] While this is a common assumption made for SNB efficiency programs (Allcott, 2011), there are of course plausible scenarios where savings might persist after a household stopped receiving HWRs, such as if the household had made significant changes to its landscape or had replaced toilets or other water using appliances as a result of getting HWRs. Thus the assumption is conservative in the sense that it is likely to impart an upward bias to the unit cost estimate.

2. Average HWR Cost Per Household

EBMUD is in the process of evaluating the expansion of HWRs to more parts of its service area. As part of this evaluation it has estimated average costs per household for providing one year of HWRs based on information from potential vendors as well as its own internal costs of program implementation. For electronic HWRs delivered by email it estimates an annual cost of $4.50 to $5.00 per household. For paper HWRs delivered by mail it estimates an annual cost of $6.40 to $6.60 per household. These are average costs over three years assuming the program is scaled up from its present level of less than 15,000 households to 100,000 households by the end of the third year.

Because some costs are fixed and others are variable, the average cost depends on the number of households in the program. Thus unit costs for other service areas may be more or less than what EBMUD has estimated depending on the scale of the program. For assessing program cost-effectiveness, we assume a cost range for paper and email reports of $4.00 to $6.00 and $5.50 to $7.50 per household, respectively.

3. Unit Costs of Water Savings

Expected program unit costs, in dollars per acre-foot, are summarized in Table 14.[43] The unit cost range for conserved water from email reports is $250 to $590 per acre-foot; for paper reports, it is $290 to $570 per acre-foot. The mid-point unit costs for email and paper reports are $380 and $400 per acre-foot, respectively.[44]

[43] Unit cost estimates have been rounded to the nearest $10.

[44] While email reports have slightly lower unit costs, paper reports offer somewhat more water savings. The economic advantage of one over the other depends on the avoided cost of saved water. For example, if 10,000 reports can be provided by email at a cost of $5/report and by mail at a cost of $6.50/report, and the expected savings for email and mail reports are 0.0161 and 0.01315 acre-feet, respectively, which is the midpoint of the expected savings range, then the paper reports would cost $15,000 more than the email reports but also save 29.5 acre-feet more. An avoided cost of saved water of $508 or more would give the economic advantage to paper over email reports, since the value of the incremental water savings would exceed their incremental cost. If the avoided cost of saved water was less than $508 (but still above the unit cost for email reports), the economic advantage would be with the email reports. In reality, most utilities are likely to start with either all paper or a mix of paper and email reports, since delivery of email reports requires a utility to have working email addresses for its targeted

Even at the upper-end of the cost range, the unit costs are competitive with most other options for water demand management. A review of unit costs from conservation master plans conducted for the California Water Foundation in 2012 found unit costs to typically range from $450 to $950 per acre-foot, with a central tendency in the neighborhood of $700 per acre-foot (M.Cubed, 2012). The unit costs for HWRs are mostly below this range, especially in the case of email HWRs.

Table 14. Unit Cost of Saved Water from HWRs Program in $/AF

Email Reports		Average Annual Water Savings				
		3.5%	4.0%	4.5%	5.0%	5.5%
Annual Cost/ Household	$4.00	$390	$340	$300	$270	$250
	$4.50	$440	$390	$340	$310	$280
	$5.00	$490	$430	$380	$340	$310
	$5.50	$540	$470	$420	$380	$340
	$6.00	$590	$510	$460	$410	$370

Paper Reports		Average Annual Water Savings				
		4.5%	5.0%	5.5%	6.0%	6.5%
Annual Cost/ Household	$5.50	$420	$380	$340	$310	$290
	$6.00	$460	$410	$370	$340	$320
	$6.50	$490	$450	$400	$370	$340
	$7.00	$530	$480	$440	$400	$370
	$7.50	$570	$510	$470	$430	$400

The unit costs in Table 14 are also competitive with most other options for new water supply. Costs vary significantly by type of supply. For recycled water, costs can range from the low hundreds to over $2,000 per acre-foot. A review of 26 Bay Area recycled water projects found an average cost of about $1,100 per acre-foot (M.Cubed, 2007). Costs for desalination range even higher. Recent cost estimates for five proposed desalination projects in Southern California range from $1,191 to $2,340 per acre-foot (California Natural Resources Agency, 2013). The California Department of Water Resources (DWR) estimates the implicit cost of water supply from proposed new conveyance for the Delta at between $302 to $408 per acre-foot at the Delta. Additional costs would accrue for transmission, treatment (for urban users), and distribution, which could add up to several hundred dollars to the price paid by urban water users, putting the cost of the water at the

residential customers, which is often not the case. As the program is implemented, customers can be directed to the web portal to select the type of delivery -- email or paper -- according to preference.

point of use (which is the appropriate cost when comparing to demand management costs) in the $500 to $700 range.[45] The unit cost of HWRs, even at the upper end of the range, are competitive or more than competitive with each of these supply options.

Targeting reports to households in the upper percentiles of consumption would lower the unit costs even further. Recall that we estimate households in the upper quartile of water use save, on average, about 1% more than households in the inter-quartile range. For households in this category, the mid-point unit cost of HWRs would be $310 and $340 per acre-foot for paper and email reports, respectively.

E. PROGRAM INTEGRATION

Another important consideration that goes into a utility's decision to implement a specific demand management program is how well it integrates with or enhances the existing programs it offers. From the perspective of EBMUD conservation staff, the Pilot highlighted several key advantages of an SNB efficiency program in terms of overall customer service and extension of its existing programs.

In terms of customer service, EBMUD staff reported finding the customer analytics accessed through what WaterSmart calls the Utility Dashboard to be extremely useful. This gave them detailed information on home water use and household attributes which they could access when interacting with customers over the phone to address questions about water use, bills, or other issues. With this information at their fingertips they could better determine how to help the customer and better direct customers to other programs that could help them reduce their water use. In this regard, EBMUD staff believe an SNB efficiency program like the one tested in the Pilot will help them achieve a long-term goal of having more specific and relevant dialog about water use with their customers.

The Utility Dashboard also gave EBMUD staff the ability to update or extend the customer profile data in real time. This proved to be important during the Pilot when customers called with questions about the water score they received on the HWR. With the Utility Dashboard, staff could

[45] These estimates are based on the assumption of no cost overruns for the conveyance facilities relative to current cost estimates, which would be unusual for a large public infrastructure project of this sort. For less sanguine estimates of implicit supply cost of new Delta conveyance, see http://hydrowonk.com/blog/2013/09/16/what-would-be-californias-water-supply-situation-without-the-bdcp-and-what-it-means-for-tunnels/.

quickly determine whether the score that was generated was based on accurate information on household attributes. If the information was inaccurate, EBMUD staff could update the information while on the phone with the customer.

EBMUD staff also reported the ability to customize the HWRs gave them options for crafting content and messaging that could evolve with their overall program direction and objectives. In particular, as EBMUD shifts the focus of its conservation programs to landscape water savings, staff anticipate using HWRs to emphasize outdoor water use efficiency and to channel more customers into landscape audit and rebate programs.

The potential scalability of an SNB efficiency program is also viewed as important to EBMUD staff. The ability to ramp up or down the program at little cost gives them options for either targeting specific customer groups with HWRs or rolling them out on a much broader scale, potentially as part of a drought response. With respect to drought management, EBMUD staff expect HWRs will play an increasingly important role by giving them the ability to customize drought response messages, more effectively communicate the importance of curbing water use, and even possibly developing customer-specific water shortage allocations.

Overall, EBMUD staff reported that an SNB efficiency program like the one tested in the Pilot will give them new and better ways to provide customer service related to water use efficiency and to more effectively market and channel customers into complementary conservation programs.

VI. SUMMARY OF OUTCOMES AND IMPLEMENTATION LESSONS

A. PILOT OUTCOMES

To summarize the results of the evaluation, our principal findings on Pilot outcomes are as follows:

1. We find strong evidence that households in the Pilot's treatment groups reduced their water use in response to the HWRs. We estimate mean treatment effects of 4.6% and 6.6% for the Random Group and Castro Valley Group experiments, respectively. Our estimates of mean treatment effect bracket the 5% mean effect estimated by WaterSmart using a less robust DID methodology. The consistency of results between the DID estimates and our results is useful corroborating information.

2. We also find evidence that the magnitude of the effect scales with level of household water use. Households in the top quartile of water use save, on average, 1% more, while households in the bottom quartile of water use save, on average, 3% less, than households in between these two categories. This suggest utilities should consider giving households in the bottom quartile of use lower priority for receiving HWRs if they are not going to be universally provided.
3. Paper reports delivered by mail appear to be more effective in terms of water savings than electronic reports delivered by email. On average, households receiving paper reports were found to save about 1% of mean household use more than households receiving email reports. Whether this translates into an economic advantage for an implementing utility, however, will depend on the cost of delivering mail versus email reports as well as the avoided cost of water saved.
4. We estimate that the unit cost of saved water is likely to range between $250 and $590 per acre-foot for email reports and between $290 and $570 per acre-foot for paper reports.[46] The mid-point unit costs for email and paper reports are $380 and $400 per acre-foot, respectively. Even at the upper-end of the cost ranges, the unit costs are less than the cost of most other options for water demand management and new water supply, indicating SNB efficiency programs could provide very cost-effective water savings.
5. We find strong evidence that households in the Pilot's treatment groups were significantly more likely to participate in audit and rebate programs offered by EBMUD than households in the control groups. Looking at both audit and rebate programs together, we estimate that households receiving HWRs were 2.3 times more likely to participate in a program than households not receiving reports. The effect appears to be strongest for audit programs, where we estimate households getting HWRs were 6.2 times more likely to participate. The effect is less strong for rebate programs (1.7 times more likely), but statistically significant. The results suggest that SNB efficiency programs can provide an effective conduit for channeling customers into other conservation programs the utility is promoting.
6. Our analysis indicates that households receiving a water score of 3 (Take Action!) are in fact more likely to do just that. The magnitudes of the treatment effects for both average daily use and program participation are positively correlated with water score. While our

[46] Unit cost estimates have been rounded to the nearest $10 throughout this report.

results should not be interpreted to imply that there is value to adjusting the scores to place more households in the score = 3 category, they do suggest that targeting HWRs to homes that fall within this category is likely to yield better results in terms of average water savings and boosting program participation rates.

7. We do not find evidence that HWRs improve household knowledge of water use in the conventional sense of being able to quantitatively estimate average daily use. The proportion of homes stating they did not know their water use was essentially the same between households in the control and treatment groups. Similarly, the tendency to underestimate daily water use was also generally the same between control and treatment households. It may be that over time this will change and as households receive more HWRs they will begin to incorporate this information into their general understanding of how they use water.
8. We do find evidence that households receiving HWRs view them as providing useful and actionable information for managing their water consumption. Households in the treatment group were 52 to 80% more likely to score EBMUD as "Excellent" in terms of explaining household water use, showing ways to save money on water bills by conserving water, and giving useful tips and tools needed to use water efficiently. Thus, HWRs appear to be effective at delivering information on ways to use water efficiently that households can, and judging by the measured effects on daily water use and program participation, do act upon.

B. IMPLEMENTATION LESSONS

Implementation lessons from the Pilot are still emerging and we expect them to evolve as more experience is gained with SNB efficiency programs. Some preliminary lessons from the Pilot include:

1. Good data management provides one of the most important keys to successful implementation of SNB efficiency programs. If the program is outsourced to a third-party company, this requires the establishment of robust protocols for data handling, quality control, and security. Privacy issues are of paramount concern since HWRs rely on customer-specific information that needs to be safeguarded from improper use.
2. Regular communication between utility staff and the SNB efficiency program service provider is essential. Throughout the Pilot, staffs of EBMUD and WaterSmart met on a

routine basis to review progress and interim results, discuss challenges, and plan next steps. These meetings allowed them to bring to the table emerging issues and to address them before they became significant problems.

3. Surveying households prior to implementation to gather additional information on household characteristics and water use attitudes provides essential information for binning customers into cohorts and tailoring the messaging of the initial HWRs. It is important to make sure sufficient resources have been set aside for this task. EBMUD staff reported being caught somewhat off guard by the high rate of responses and resources required to to process the survey data, but also noted its importance to successfully launching the Pilot.
4. Prior to implementation it is also important to educate customer service representatives about the new program and train them on how to respond to or direct customers with inquiries or complaints about the information in their HWR. In the Pilot, the most common complaint was from customers receiving a water score of 3 who felt they had been scored incorrectly. EBMUD worked with its customer service representatives to turn these calls from complaints to opportunities for customer outreach by first verifying with the customer the information upon which the score was based and second by providing them information on ways to more effectively use water around the home and to alert them to audit and rebate programs that may directly benefit them.
5. It is important to experiment with how information is presented in the HWR. In the Pilot, EBMUD quickly discovered that customers responded negatively when told their use was being compared to their neighbor's but seemed to be okay if told their use was being compared to similarly situated homes.
6. Phasing implementation afforded EBMUD and WaterSmart the opportunity to fine-tune the process as they gained feedback and experience with producing and delivering the reports and responding to customer inquiries. Phasing enabled them to implement an adaptive management approach to implementing the program.

VII. RECOMMENDATIONS FOR FUTURE RESEARCH

These are early days for the application of SNB efficiency programs to residential water use and there is still much to be learned in terms of efficacy, cost, and implementation. Some questions that future research can help address include:

1. Are treatment effects persistent or do they fade with time? Literally, only time will tell. As noted in Section II.D, empirical evaluations of home energy reports have found the treatment effect to persist and even strengthen over three years. These studies, however, have only considered a relatively short amount of time. Whether something similar will be the case for HWRs is an important topic of inquiry. There are several related questions: (1) Do effects persist even if HWRs are discontinued, perhaps because of induced changes in water using appliances and fixtures? (2) Do HWR savings grow, stay constant, or decline with time? (3) Does so-called demand hardening impose a limit on the changes in household water use that can reasonably be expected from HWRs?
2. To what extent are water savings driven by changes in outdoor water use? Our preliminary models provided some evidence of stronger treatment effects associated with reports received in the fall and winter than in the spring and summer. This may indicate a lagged response to receiving information about high summer water use, but the phasing of the Pilot, which resulted in the largest block of Castro Valley homes not getting their first HWR until October or November, could also be involved.
3. Does the frequency in which HWRs are provided matter? Our results found that providing HWRs on a bi-monthly basis yielded significant reductions in water use. Would providing reports more (less) frequently result in a larger (smaller) effect on water use? In the case of home energy reports, Allcott (2011) concluded more frequent reports did yield larger savings, but not by enough to justify the added expense. But he did not consider the potential cost advantage of providing reports less frequently just to households in the lower percentiles of use.
4. Are differences in water savings and program participation associated with the HWR water score due primarily to the injunctive messaging or other underlying factors?
5. What is a reliable range of water savings to expect from SNB efficiency programs if implemented broadly across the state? At present we have very few data points from which to gauge this. Only a handful of pilot implementations have been completed and while they seem to suggest initial water savings in the range of 4 to 6% more evaluations of outcomes under varied conditions are needed to know if this range is stable.

BIBLIOGRAPHY

Abadie, A. (2005). Semiparametric Difference-in-Differences Estimators. *The Review of Economic Studies, 72*(1), pp. 1-19.

Allcott, H. (2011, October). Social Norms and Energy Conservation. *Journal of Public Economics, Vol. 95*(Issues 9-10), pp. 1082-1095.

Allcott, H., & Mullainathan, S. (2010, March 5). Behavior and Energy Policy. *Science, 327*, 1204-1205.

Andreasen, A. R. (2002). Marketing Social Marketing in the Social Change Marketplace. *Journal of Public Policy & Marketing, 21*(1), pp. 3-13.

Arimura, T. H., Newell, R. G., & Palmer, K. (2009). *Cost-Effectiveness of Electricity Energy Efficiency Programs.* Resources for the Future, Discussion Paper 09-48.

Bergstrom, T. C., & Goodman, R. P. (1973). Private Demands for Public Goods. *American Economic Review, 63,* 280-296.

Beshears, J., Choi, J., Laibson, D., Madrian, B., & Milkman, K. (2009). *The Effect of Providing Peer Information on Retirement Savings Decisions.* Bosta, MA: Working Paper, Harvard Univerity.

Brent, D. A., Cook, J. H., & Olsen, S. (October, 2013). *Heterogeneity in Response to Social Norms for Water Conservation.* Working Paper, Department of Economics, University of Washington.

California Natural Resources Agency. (2013, July 17). *Bay Delta Conservation Plan.* Retrieved from Bay Delta Conservation Plan Public Meeting Presentation, July 17, 2013: http://baydeltaconservationplan.com/Libraries/Dynamic_Document_Library/July_Public_Meeting_Presentation_Final.sflb.ashx

Chesnutt, T. W., & McSpadden, C. N. (1995). *Determinants of Phoenix Water Demand.* Prepared for City of Phoenix.

Frey, B., & Meier, S. (2004). Social Comparisons and Pro-Social Behavior: Testing 'Conditional Cooperation' in a Field Experiment. *American Economic Review, 95*(5), 1717-1722.

Gerber, A., & Rogers, T. (2009). Descriptive Social Norms and Motivation to Vote: Everybody's Voting and So Should You. *Journal of Politics, Vol. 71*, pp. 1-14.

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels. *Journal of Consumer Research, 35*(3), pp. 472-482.

Groves, R. M., & Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. *Public Opinion Quarterly, 72*(2), 167-189.

Gunn, R. (2012). *Energy Efficiency/Demand Response Plan: Year 4 Evaluation Report: Home Energy Reports (Draft).* Prepared for Commonwealth Edison Company.

Hannan, E. J. (1960). The Estimation of Seasonal Variation. *Australian Journal of Statistics, 2*(1), 1-15.

Heck, S., & Tai, H. (2013). *Sizing the Potential of Behavioral Energy-Efficiency Initiatives in the US Residenital Market.* Prepared by McKinsey&Company in collaboration with Opower.

Jessoe, K., & Rapson, D. (2013). *Knowledge in (Less) Power: Experimental Evidence from Residential Energy Use.* Berkeley, CA: Energy Institute at Haas.

Jorgenson, D. W. (1964). Minimum Variance Linear Unbiased Seasonal Adjustment of Economic Time Series. *Journal of the American Statistical Association, 59*(307), 681-724.

Kennedy, P. E. (1981). Estimation with correctly interpreted dummy variables in semilogarithmic equations. *Journal of Economic Review, 71*(4), 801.

Levitt, S. D., & List, J. A. (2007). What Do Labratory Experiments Measuring Social Preferences Reveal about the Real World? *The Journal of Economic Perspectives, 21*(2), pp.153-174.

M.Cubed. (2007). *Importance of Recycled Water to the San Francisco Bay Area.* Prepared for Bay Area Clean Water Agencies.

M.Cubed. (2012). *Review of Unit Cost Ranges for CWF Water Efficiency Strategies.* Prepared for California Water Foundation.

Opinion Dynamics Corporation. (2012). *Impact and Process Evaluation of 2011 (PY4) Ameren Illinois Company Behavioral Modification Program (Draft).* Prepared for Ameren Illinois Company.

Perkins, H. W., & Berkowitz, A. D. (1986). Perceiving the Community Norms of Alcohol Use Among Students: Some Research Implications for Campus Alcohol Education Programming. *International Journal of the Addictions, 21*(9-10), pp. 961-976.

Perry, M., & Woehleke, S. (2013). *Evaluation of Pacific Gas and Electric Company's Home Energy Report Initiative for the 2010-2012 Program.* Prepared by Freeman, Sullivan & Co. for Pacific Gas and Electric.

Provencher, B. (2011). *Evaluation Report: OPOWER SMUD Pilot Year 2.* Boulder, CO: Prepared by Navigant Consulting for Opower.

Provencher, B., Hampton, J., Brown, G., & Hummer, J. (2013). *Home Energy Reports Program: Program Year 2012 Evaluation Report, Appendix H.* Prepared by Navigant Consulting, Inc. for AEP Ohio.

Puhani, P. A. (2008). *The Treatment Effect, the Cross Difference, and the Interaction Term in Nonlinear "Difference-in-Differences" Models.* Bonn, Germany: Discussion Paper No. 3478, IZA.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, H. J., & Griskevicius, V. (2007). The Constructive, Destructive, and Reconstructive Power of Social Norms. *Psychological Science, 18*(5), pp. 429-434.

Sergici, S., & Farugui, A. (2011). *Measurement and Verification Principles for Behavior-Based Efficiency Programs.* San Francisco, CA: Brattle Group, Inc.

Shaikh, S. L., & Larson, D. M. (2003). A Two-Constraint Almost Ideal Demand Model of Recreation and Donations. *The Review of Economics and Statistics, 85*(4), 953-961.

Summit Blue Consulting, LLC. (2009). *Impact Evaluation of OPOWER SMUD Pilot Study.* Prepared for Opower.

The Economist. (2010, August 19). Watts Up? People Habitually Underestimate their Energy Consumption. *The Economist*.

Wechsler, H., Nelson, T. E., Lee, J. E., Seibring, M., Lewis, C., & Keeling, R. P. (2003). Perception and Reality: A National Evaluation of Social Norms Marketing Interventions to Reduce College Students' Heavy Alcohol Use. *Journal of Student Alcohol, 64*(4), pp. 484-94.

Wooldridge, J. M. (2001). *Econometric Analysis of Cross-Section and Panel Data.* Boston, MA: Massachusetts Institute of Technology.

Wu, M., & Osterhus, T. (2012). *Impact & Persistence Evaluation Report: Sacramento Municipal Utility District Home Energy Report Program.* Prepared by Integral Analytics, Inc.

ACKNOWLEDGEMENTS

"Data! Data! Data! ... I can't make bricks without clay." Sherlock Holmes exclaimed in *The Adventure of the Copper Beeches*. This cry is familiar to every statistician the world over. Fortunately in our case we had the staffs from EBMUD and WaterSmart to ensure we had the data we needed for this evaluation. We owe particular thanks to Ora Chaiken and Chad Haynes of WaterSmart and Mike Hazinski, Dave Wallenstein, and Richard Harris of EBMUD. Whenever we cried "Data! Data! Data!" they responded and got us what we needed. We would also like to thank Bill Jacoby and Mike Myatt of the California Water Foundation who shepherded this project in its beginning phases, and Ronnie Cohen, also with California Water Foundation, who kept it on track once we were underway. We also owe thanks to our funders, California Water Foundation and EBMUD, who gave us the necessary resources to do this work. All of these individuals and organizations have contributed to this project in essential ways, but only the authors are responsible for the content of this report.

APPENDIX 1: Random Group Experiment Sample Distribution by Pressure Zone

Group	Pressure Zone	% Residential Accts	% of Randomly Sampled Accts		
			Total	Control	Treatment
A	A0A	4.29%	5.59%	2.76%	2.82%
A	A11A	0.04%	0.00%	0.00%	0.00%
A	A11C	0.18%	0.03%	0.00%	0.03%
A	A1A	1.56%	1.29%	0.74%	0.55%
A	A1B	0.07%	0.09%	0.06%	0.03%
A	A2A	2.69%	1.96%	1.07%	0.89%
A	A2AA	0.11%	0.03%	0.03%	0.00%
A	A2B	1.08%	2.06%	1.01%	1.04%
A	A2C	0.16%	0.25%	0.09%	0.15%
A	A2D	0.14%	0.21%	0.09%	0.12%
A	A2E	1.93%	2.12%	1.07%	1.04%
A	A3B	0.03%	0.03%	0.03%	0.00%
A	A4A	0.02%	0.00%	0.00%	0.00%
A	A4AA	0.03%	0.09%	0.03%	0.06%
A	A4B	0.33%	0.25%	0.12%	0.12%
A	A4BA	0.09%	0.15%	0.09%	0.06%
A	A4BB	0.08%	0.06%	0.03%	0.03%
A	A4C	0.24%	0.12%	0.06%	0.06%
A	A4D	0.05%	0.12%	0.09%	0.03%
A	A4G	0.04%	0.09%	0.06%	0.03%
A	A4K	0.01%	0.00%	0.00%	0.00%
A	A4L	0.00%	0.00%	0.00%	0.00%
A	A4M	0.20%	0.31%	0.09%	0.21%
A	A4N	0.03%	0.03%	0.03%	0.00%
A	A5A	0.76%	0.52%	0.28%	0.25%
A	A5B	0.06%	0.06%	0.03%	0.03%
A	A5C	0.02%	0.00%	0.00%	0.00%
A	A7B	0.11%	0.00%	0.00%	0.00%
A	A7C	0.03%	0.03%	0.00%	0.03%
A	A7D	0.67%	0.49%	0.21%	0.28%
A	A9B	0.16%	0.12%	0.09%	0.03%
A	A9D	0.38%	0.21%	0.12%	0.09%
B	B11A	0.06%	0.12%	0.03%	0.09%
B	B11B	0.49%	0.31%	0.09%	0.21%
B	B11C	0.18%	0.15%	0.12%	0.03%
B	B11D	0.00%	0.00%	0.00%	0.00%
B	B13A	0.27%	0.21%	0.12%	0.09%
B	B2A	2.38%	1.81%	1.01%	0.80%
B	B2AA	0.19%	0.34%	0.18%	0.15%
B	B3A	2.01%	2.27%	1.20%	1.07%
B	B4AA	0.17%	0.37%	0.15%	0.21%
B	B4B	0.08%	0.15%	0.15%	0.00%
B	B4C	0.04%	0.06%	0.06%	0.00%
B	B5A	1.00%	0.46%	0.46%	0.00%
B	B5B	0.05%	0.09%	0.06%	0.03%
B	B5C	0.01%	0.03%	0.00%	0.03%
B	B5D	0.25%	0.37%	0.25%	0.12%

Group	Pressure Zone	% Residential Accts	% of Randomly Sampled Accts		
			Total	Control	Treatment
B	B7A	0.85%	0.77%	0.25%	0.52%
B	B7B	0.14%	0.37%	0.18%	0.18%
B	B7C	0.01%	0.00%	0.00%	0.00%
B	B9A	0.17%	0.09%	0.09%	0.00%
B	B9AA	0.01%	0.00%	0.00%	0.00%
B	B9B	0.17%	0.34%	0.18%	0.15%
B	B9CA	0.03%	0.06%	0.00%	0.06%
B	B9D	0.18%	0.09%	0.06%	0.03%
C	C1A	0.98%	1.04%	0.34%	0.71%
C	C2A	1.81%	2.33%	0.00%	2.33%
C	C2B	0.46%	0.68%	0.34%	0.34%
C	C2C	0.02%	0.00%	0.00%	0.00%
C	C4A	1.47%	1.57%	0.00%	1.57%
C	C4B	0.28%	0.49%	0.31%	0.18%
C	C4C	0.52%	0.83%	0.43%	0.40%
C	C4D	0.02%	0.03%	0.00%	0.03%
C	C5C	0.20%	0.43%	0.00%	0.43%
C	C5D	0.22%	0.21%	0.00%	0.21%
C	C5E	0.03%	0.03%	0.00%	0.03%
C	C6A	0.15%	0.28%	0.12%	0.15%
C	C6B	0.10%	0.15%	0.00%	0.15%
C	C7A	0.05%	0.06%	0.00%	0.06%
C	C8A	0.27%	0.31%	0.12%	0.18%
D	D11B	0.03%	0.03%	0.00%	0.03%
D	D5A	2.26%	3.62%	1.78%	1.84%
D	D5AA	0.01%	0.00%	0.00%	0.00%
D	D5AB	0.06%	0.03%	0.00%	0.03%
D	D5AC	0.04%	0.12%	0.03%	0.09%
D	D5AD	0.01%	0.03%	0.00%	0.03%
D	D5AE	0.01%	0.00%	0.00%	0.00%
D	D7A	0.89%	1.14%	0.49%	0.64%
D	D7AB	0.02%	0.00%	0.00%	0.00%
D	D7B	0.42%	0.46%	0.21%	0.25%
D	D7BA	0.00%	0.00%	0.00%	0.00%
D	D7C	0.05%	0.03%	0.03%	0.00%
D	D7F	0.18%	0.12%	0.00%	0.12%
D	D7J	0.10%	0.15%	0.12%	0.03%
D	D7K	0.03%	0.03%	0.00%	0.03%
D	D7KA	0.01%	0.00%	0.00%	0.00%
D	D9A	0.05%	0.03%	0.00%	0.03%
D	D9C	0.16%	0.25%	0.09%	0.15%
D	D9E	0.02%	0.00%	0.00%	0.00%
D	D9J	0.03%	0.00%	0.00%	0.00%
E	E3A	1.81%	1.87%	1.01%	0.86%
E	E3AA	0.00%	0.00%	0.00%	0.00%
E	E3AE	0.01%	0.03%	0.00%	0.03%
E	E5B	0.27%	0.34%	0.12%	0.21%
E	E5C	0.02%	0.00%	0.00%	0.00%
E	E7AA	0.01%	0.00%	0.00%	0.00%

Group	Pressure Zone	% Residential Accts	% of Randomly Sampled Accts		
			Total	Control	Treatment
E	E7B	0.02%	0.00%	0.00%	0.00%
E	E9A	0.02%	0.00%	0.00%	0.00%
F	F10A	0.04%	0.03%	0.03%	0.00%
F	F11A	0.03%	0.00%	0.00%	0.00%
F	F13A	0.00%	0.00%	0.00%	0.00%
F	F3A	3.46%	5.03%	2.76%	2.27%
F	F4A	2.24%	3.41%	1.75%	1.66%
F	F5B	1.04%	1.32%	0.71%	0.61%
F	F5BA	0.10%	0.09%	0.03%	0.06%
F	F5BF	3.59%	5.28%	2.61%	2.67%
F	F5BG	0.00%	0.00%	0.00%	0.00%
F	F7B	0.03%	0.00%	0.00%	0.00%
F	F7D	0.05%	0.12%	0.06%	0.06%
F	F7E	1.31%	1.50%	0.74%	0.77%
F	F7G	0.00%	0.00%	0.00%	0.00%
F	F8A	0.51%	0.74%	0.28%	0.46%
F	F9A	0.01%	0.03%	0.03%	0.00%
G	G0A	33.54%	24.59%	11.91%	12.68%
G	G1AA	10.60%	9.91%	5.34%	4.57%
G	G1AB	1.61%	1.10%	0.40%	0.71%
G	G1BA	1.16%	1.17%	0.61%	0.55%
H	H1A	3.09%	3.87%	1.87%	2.00%
F	F7F	0.09%	0.18%	0.12%	0.06%
F	F9B	0.03%	0.09%	0.09%	0.00%
	Grand Total	100.00%	100.00%	47.88%	52.12%