Do Water Audits Promote Economic Welfare? Evidence from a Natural Field Experiment*

Jesper Akesson
The Behavioralist

Robert W. Hahn
University of Oxford & Technology Policy Institute

Rajat Kochhar
University of Southern California

Robert D. Metcalfe
University of Southern California & NBER

This Version: November 26, 2022

Abstract

Water suppliers in water-scarce regions are showing greater interest in using non-price mechanisms that can help encourage conservation. One such mechanism is a water audit, which involves assessing water use in the home and providing tailored suggestions regarding how to conserve water. Yet, very little is known about the efficacy, efficiency, and cost-effectiveness of water audits. This paper helps fill this research gap by implementing a natural field experiment in the United Kingdom. We implement a natural field experiment by randomly allocating 45,000 water customers to a control group or to groups that are provided with different encouragements to take-up an online water audit. Our analysis yields three main findings. First, providing participants with financial incentives to participate in the audit significantly increases audit take-up, with an elasticity of around 0.5. Using the positive encouragements, we find that the water audit reduces household water consumption by 17 percent for about two months. Second, notwithstanding these large improvements in water conservation, incentivizing uptake of the audit does not appear to pass a benefit-cost test. We also implement a marginal value of public funds approach to considering benefit and costs, and reach a similar conclusion. Third, we find that targeting of high users could double the effectiveness of the financial incentive interventions.

Keywords: water conservation; natural field experiment; benefit-cost analysis; marginal value public funds

*Acknowledgements: We thank Tom Andrewartha, David Black, Keith Colquhoun, Jonathan Dennis, Dustin Garrick, Robin Gilthorpe, Karen Guz, Nathan Hendren, Alex Plant, Louise Richardson, Cathryn Ross, Ruth Scott, Dale Whittington, Crawford Winton, and Benjamin Wisniewski for helpful comments and guidance. Akesson received support for helping to implement this experiment. The views in this paper represent those of the authors and are not necessarily those of the institutions with which they are affiliated. AEA Registry: AEARCTR-0009486

1 Introduction

Water scarcity has become an issue across the world, and there are increasing concerns that climate change could exacerbate this situation (Pörtner et al. 2022, Vorosmarty et al. 2000). In response to concerns about scarcity, water suppliers and regulators are showing greater interest in identifying non-price mechanisms that could help encourage conservation. One increasingly common mechanism is a water audit, which help identify behavioral and technological inefficiencies in the home, and provide tailored recommendations for promoting conservation. Environmental Protection Agency, USA (2013) considers water audits to be the critical first step towards identifying and quantifying the water uses and losses. Based on data from the Water Research Foundation, a total of 4,575 audits were conducted by water utilities in just five US states[1] between 2011-2014 (Sturm et al. 2015). However, very little is known about the efficacy, efficiency, and cost-effectiveness of audits.

This paper helps fill this research gap by implementing a natural field experiment that encourages residential customers to take a home water audit. We partnered with a water utility to examine the effectiveness of these audits. We randomly encourage some customers to take up the audit, and we randomize the type of encouragement that customers receive—from various financial incentives to environmental appeals and moral suasion. Our experimental design allows us both to study how different encouragements influence take up, and how water audits influence consumption.

Our paper contributes to recent research on water and energy conservation in three ways. First, to the best of our knowledge, this is the first paper to use a natural field experiment to estimate the causal impact of water audits on consumption. Second, most previous research on water conservation studies the effect of non-pecuniary interventions, such as moral suasion and social comparisons (see Nauges & Whittington (2019) for a review of studies). We introduce several treatments, two of which use financial incentives, and compare them to non-financial incentives. Third, while researchers have noted the need for rigorous benefit-cost analysis of water policies based on causal estimates, we develop and operationalize two frameworks for implementing such analysis: one is a standard benefit-cost framework; and a second is a less traditional approach based on the marginal value of public funds (MVPF).

The experimental design involves sending letters to residential customers, and in some cases, follow-up emails, that encourage them to take a water audit. This audit consists of logging into the company's online water audit, answering questions on water use habits and home features, and receiving recommendations for reducing consumption. The online tool provided information on free water-saving devices offered by the utility, and

[1] California, Delaware, Georgia, Tennessee, and Texas

helped customers book an in-home audit if appropriate. We measure water consumption after the interventions and compare it with the water consumption of a control group. We randomly allocate 45,000 customers to a control group and one of 6 treatment groups.

The control group received no communication. Treatment group 1 received an encouragement letter that was in use by NWL prior to the trial, while the remaining five groups received newly designed letters, each catering to a different motivation for water conservation. Treatment group 2 received a simplified version of the letter sent to treatment group 1, which made the call to action more salient. The third treatment group received letters reminding them of the scarcity of water, while treatment group 4 was sent a letter comparing the household's consumption to that of their neighbors (*i.e.*, moral suasion). Treatment groups 5 and 6 received letters that provided different levels of monetary incentives (£10 and £15) to encourage completion of the home water audit.

The causal impact of audits on water consumption is estimated using an encouragement design with two-stage least squares. For the first stage, we estimate the impact of the randomized encouragement on the take-up of the audit. Using the results from the first-stage, we then examine the impact of the audit on consumption, which yields a local average treatment effect (LATE).

We have four main findings. First, relative to the *Vanilla* letter, all letters led to a significant increase in take-up of the diagnostic, with the *Incentives* treatment having the maximum impact. Specifically, households exposed to the *Incentives 10* treatment had a 4.5 percent higher rate of take-up relative to the *Vanilla* group, while the increase was 5.7 percent for households in the *Incentives 15* group. Importantly, the impact of the two *Incentives* treatment is statistically different from each other. This amounts to a price elasticity of 0.53, implying that a 10 percent increase in incentives leads to a 5.3 percent increase in take-up. Thus, increasing the amount of financial incentive could be a fruitful strategy to increase participation, though it may have repercussions on the cost benefit analysis. In addition, we find that reminder emails increase adoption of the audit.

Second, encouraging metered subjects to participate in an online water audit with financial incentives reduces household consumption by about 43 liters per day for an average of 65 days, or about 17 percent.[2] Our best estimate of the £15 treatment is that it reduces consumption for metered households by 44 liters per day, while the £10 treatments reduces consumption for the same subgroup by 43 liters per day. This suggests that the size of the subsidy for completing the audit may not be that important for water conservation, unlike take-up. We employ a LATE methodology to estimate these effects, and the results are robust to using different combinations of the letters as instruments. We also undertake an external validity exercise, where we explore how our results would gener-

[2]We can only measure the short-term effects of the audit because the consumption data covers 65 days following the delivery of the interventions.

alise to customers who currently do not have meters. Weighting each household by the inverse probability of being metered revises our estimates on water conservation downwards to 34 liters per day, though the results still remain significant with our preferred specification.

Third, notwithstanding the substantial improvements in water conservation, the interventions that use financial incentives do not appear to pass a benefit-cost test that (only) includes the greenhouse gas benefits associated with these reductions. Because our analysis does not quantify other potentially important benefits, such as ecosystem improvements and reductions in infrastructure costs, we define a lower bound on other benefits needed to pass a benefit-cost test. We find that the social cost of carbon in the base case would need to be 22 times higher for benefits to just equal costs. Using our base case estimate of $51 per ton (Interagency Working Group, USG 2021), this means the social cost of carbon would need to be about $1200 per ton for benefits to just equal costs if there were no other benefits.[3] Another interpretation of this finding is that a cubic meter of water conservation would need to yield other benefits or reduced investment costs of about $6.0 for this intervention to be worthwhile.[4]

Our benefit-cost framework builds on pioneering work by Hendren and others who use a marginal value of public funds approach (*e.g.*, Finkelstein & Hendren (2020) and Hendren & Sprung-Keyser (2020)). A key advantage of this approach is that it separates the problem of estimating the welfare impact of the subsidy from the problem of estimating the welfare impact of the intervention that could pay for the subsidy, such as a tax. We believe this approach has the potential to add further insight by considering the cost of raising revenue separately. While it is convenient to assume a lump sum tax will be used for analytical simplicity, it is not necessary. Recent work on audits in the energy area, discussed below, uses the assumption of lump sum transfers. We calculate *MVPF* under two scenarios of marginal costs for the utility: short- and long-run. Our results indicate that the *MVPF* in the base case is -0.074 under the short-run marginal cost scenario, and increases to 0.0048 under the long-run marginal cost scenario. The value under the former scenario implies that the government would be spending $1 to generate negative benefits. This arises because, first, short-run marginal cost is sufficiently lower than price, and water savings would lead to a significant fall in revenues for water utilities. Second, monetary value of greenhouse gas savings from water conservation are not high enough to compensate for the loss in revenues. Under the long run scenario, the number is slightly positive, mainly as a result of lower revenue losses because long run marginal cost is closer to the

[3]See Section 4.2 for details. The word "ton" refers to a metric ton.

[4]In what follows, we will round all estimates in the text to two significant digits, while all estimates in regression tables will be rounded to three decimal places. All estimates are either in 2020 pounds or dollars. Estimates of dollars from earlier studies have also been converted from original year dollars to 2020 dollars using the Consumer Price Index (CPI) for Urban Consumers (US Bureau of Labor Statistics (2021)).

price charged by water utilities.

Fourth, we consider the issue of targeting and explore whether such targeting allows the benefits of an intervention to exceed the cost (Allcott 2011, Ayres et al. 2013, Ferraro & Miranda 2013, Brent et al. 2015, Wichman et al. 2016, Knittel & Stolper 2019, Brent et al. 2020, Gerarden & Yang 2021). We examine the targeting of high users, who are defined as users with pre-treatment consumption higher than the median consumption. Our results complement the literature, in that targeting of high users doubles the reduction in consumption compared to the reduction estimated without heterogeneous treatment effects (83 versus 43 liters per day). This suggests that audits can be targeted to improve their efficiency. We take this a step further, and ask whether targeting could increase the net benefits such that our economic welfare estimates are more attractive. Nonetheless, we find that targeting is not sufficient for benefits to exceed costs in our base case. Though cost-effectiveness increases by 38 percent, the social cost of carbon in the base case would need to still be 11 times higher (as compared to 22 times before) for benefits to just equal costs. The *MVPF* under the short-run marginal cost scenario becomes worse (-0.16 versus -0.074 earlier), primarily due to higher water savings leading to larger producer surplus losses, while the value under the long-run marginal cost increases marginally from 0.0048 to 0.011. These results imply that though targeting increases water conservation in our experiment, the improvement in net benefits are not sufficient to substantially enhance welfare.

The basic intuition behind our results can be explained simply. The short-term reductions in greenhouse gas emissions from water conservation are comparatively small, on the order of 1.6 tons for 65 days for our base case. And while the experimental cost per person is also relatively small, on the order of $1.2 per consumer (not including the producer surplus loss), this leads to a cost effectiveness (CE) of $1000 per ton, which is much higher than most estimates for the SCC (Interagency Working Group, USG (2021)). If we assume our results persist over a long period of time, the cost-effectiveness calculus looks more attractive because the benefits from conservation increase. See Section 4 for a more extended discussion of welfare.

Our results build on a growing literature that examines the impacts of different kinds of interventions on resource use, such as water and energy. There are, however, relatively few rigorous estimates of the impacts of audits on water or energy use. A useful summary of studies analyzing water audits is provided by Ansink et al. (2021). They review a growing literature that evaluates the impact of information provision and technology adoption through water audits on potential water savings. However, they do not identify any natural field experiments that address online audits.

Apart from audits, various other interventions related to rates, rebates, and enforcement regulations have been used to induce behavioral change so as to promote water conservation, with effect sizes comparable to ours. Browne, Gazze & Greenstone (2019) disentangle the effect of different residential water conservation policies adopted by a utility during the 2011-2017 California drought. They find large effect of rate changes (elasticity between 20-44 percent) and outdoor water schedule regulations (water use decreased by 21-24 percent), which are similar in magnitude to our result that participation in audits leads to a substantial 17 percent decline in consumption relative to pre-treatment consumption. West et al. (2021) examine the effects of automating the enforcement of water conservation regulations, and find similar large effects, with treated households curtailing their water consumption by 31 percent. Therefore, price change and enforcement policies also lead to effect sizes in the same range as our results. The one exception is Browne, Gazze, Greenstone & Rostapshova (2019), who implement a field experiment in California, randomly assigning visual or automated enforcement methods to detect water-use violations. Their effect sizes are small, with automated enforcement decreasing water consumption by about 3 percent.

Another important intervention is behavioral nudges that use different kinds of messaging. There have been several experiments and quasi-experiments examining the implementation of social norm messaging (*e.g.*, Ferraro & Price (2013), Brent et al. (2015), Jaime Torres & Carlsson (2016), Datta et al. (2015)). Nauges & Whittington (2019) provide a review of the literature on the impact of information treatment on water and energy use. Most studies, whether in the energy or the water sector, find that social norm information treatments reduce consumption by about 2 to 5 percent for a period of time, with greater reductions typically observed when the intervention includes social norm comparisons as opposed to interventions providing technical advice or raising awareness. Our paper integrates social norms messaging with online audits (*Moral Cost* letter), allowing us to study the effect of the former on diagnostic completion and their combined effect on water consumption. The *Moral Cost* letter has a high impact on diagnostic completion, second only to the *Incentives* treatment, with the former increasing take-up by 1.6 percent compared to the *Vanilla* letter. However, the combined effect on post-treatment consumption is more subdued, with consumption falling by only 2.9 liters per day (1.1 percent) relative to the *Vanilla* treatment group.

Several authors have used their estimates of the expected change in water or energy use to estimate the cost effectiveness and economic welfare implications of water conservation strategies. These include studies on the impacts of metering, social norm messaging, and the nature of the regulatory intervention. Ferraro & Price (2013) find that social norm messaging augmented by technical advice reduces consumption by 4.8 per cent, which implies a cost of \$0.17 per cubic meter reduced for the utility. Bernedo et al. (2014) demonstrate that persistent long-term impacts of the policy studied by Ferraro & Price (2013) imply that the cost per gallon saved is 60 percent lower (\$0.07 per cubic meter) than the figure derived using only contemporaneous treatment effects. This is many orders of magnitude lower than the estimates for our experiment[5], which are in the range of \$1.3 - \$3.7. This highlights the importance of long term impacts, which we are unfortunately unable to derive due to paucity of data. Jessoe et al. (2021) experimentally test the effect of social norms messaging about residential water use on electricity consumption. Taking into account the electricity conservation spillover increases the net benefits of their intervention from \$2.9 per household to \$4.0, an increase of 39 percent. If we could capture similar spillover benefits in our data, which we currently assume to be negligible for our analysis, the cost-effectiveness would be more attractive, especially if the effects last longer than 65 days. Ornaghi & Tonin (2021) study the effects of metering on economic welfare, and suggest that the benefits of reducing over-consumption outweigh the costs of installing and operating a meter. This has important implications for our setting as approximately 70 percent of the households in our sample are unmetered. Our reweighted LATE estimates in Section 3.4 indicate that the potential gains from providing them with meters and then conducting an audit could be large (35 liters per day reduction), which would offset the costs of metering.[6] [7]

The evidence on the cost effectiveness of water audits is limited. To the best of our knowledge, Ansink et al. (2021) provide the only cost effectiveness assessment of water audits, and find that technology is more cost-effective than information provision by a factor of two for a water audit program in England. However, the selection into their audit program was not random, and the focus was exclusively on households with above-average water use.

The experimental literature on the impact of conservation measures in the energy sector is also well developed. However, of particular interest are Allcott & Greenstone (2017) and Fowlie et al. (2018), who study the welfare impact of audits, with their results in close harmony to ours. The former study models home energy efficiency investment decisions to evaluate two large residential energy efficiency programs in Wisconsin. These programs involved a home energy audit followed by decisions on which recommended investments to undertake. They implement a large field experiment in Wisconsin, and find that the programs reduced economic welfare. A comparison of the observed investment

[5] In Section 4.1, we calculate the cost effectiveness of our experiment, and compare it to the estimates in the literature.

[6] Nauges & Whittington (2019) argue using illustrative calculations that it is far from obvious that social norm messaging instruments will pass a benefit-cost test, especially in low- and middle-income countries. Our results indicate that in the absence of persistent impacts and lack of spillover benefits, the same would hold true for even high-income countries like the UK.

[7] Also see Mansur & Olmstead (2012), who assess the potential welfare gains of switching from non-market to market-based regulation of water supply during periods of drought. Using data on residential water usage from urban areas in the US and Canada, the authors show that there are substantial welfare gains from a price-based approach. Though we do not study the effect of market-based regulations, it would be beneficial to compare their cost-effectiveness with experiments involving behavioral interventions.

costs with the present discounted value of energy savings indicates the programs has an internal rate of return (IRR) of -4.1 percent, while a revealed preference model finds that the programs reduce welfare by $0.18 per dollar of subsidy. Though we do not calculate an IRR, our negative *MVPF* of -0.074 has the same implication: the costs to the government of the intervention are higher than the benefits. In Fowlie et al. (2018), the authors measure the welfare gains from the Weatherization Assistance Program, a residential energy efficiency program in Michigan. The program involves conducting an energy audit of the home before implementing a weatherization retrofit, with the purpose of recommending specific efficiency improvements. The paper uses experimental and quasi-experimental variation in participation to identify the returns to investments. Their results suggest that the upfront investment costs are about twice the actual energy savings, and the projected savings are more than three times the actual savings. This again implies that the costs outweigh the benefits, making it undesirable for utilities or governments to undertake certain investments aimed at energy conservation.

The paper proceeds as follows. Section 2 provides the details on the audit program and randomized trial. In Section 3, we describe our empirical strategy and present the results from the experiment. Section 4 presents a welfare analysis, including information on cost effectiveness. Conclusions and areas for future research are discussed in Section 5.

2 Background and Experimental Design

In 2018, Northumbrian Water commissioned Save Water Save Money (SWSM) to provide its online water audit tool for Northumbrian's customers.[8] The tool, hosted on the company website, asked customers questions about their water use habits and homes. The main purpose of the tool was to help customers understand their water consumption, and identify ways in which they can save water and money. The tool also informed customers about free water-saving devices that NWL offers, and helped them book an in-home water audit if appropriate. The questionnaire on the platform took approximately ten minutes to complete.

NWL was interested in getting its customers to take their online water audit, and understanding the impact of the audits on consumption. We were interested in helping NWL with these objectives, and, in addition, understanding the impact of different behavioral interventions on economic welfare. In order to encourage the use of the SWSM platform, we designed a set of customer communications using theories from behavioral science. We used one of NWL's existing direct mailers as a template, and designed 5 new direct mailers (see the discussion in Appendix D). The only difference between the five commu-

[8] The water audit can be accessed at this url: https://www.getwaterfit.co.uk/questions/ (last accessed: June 07, 2022)

nications was the application of different behavioral science ideas.

We implemented an RCT to test the effectiveness of the redesigned letters, and to understand how the SWSM platform influences water consumption. This RCT included 45,000 NWL customers, spread across three post code areas. The customers that participated in the trial were randomly allocated to one of six treatment groups that received letters or a control group that received no letter. Subsequently, customers for whom NWL had email contact details were also randomly allocated to groups that either received or did not receive an email reminder about the online audit tool. The reminder emails followed the same theme as the initial letters that customers received. This design allows us to estimate the effects of particular letters and reminders on take-up of the audit.[9]

There were six letter treatments. Treatment 1 (*Vanilla*) informed customers that they can save water and money by using the free online platform. It also noted that many other customers had saved money with the platform, and told them how to access it. Treatment 2 (*Simplified*) was similar to the *Vanilla* communication but it simplified the content, making the main message and the call to action more salient. Treatment 3 (*Altruism*) added to the message of the *Simplified* mailer by reminding the consumers that water is a scarce resource, and asked them to help conserve it in their local area. Treatment group 4 (*Moral Cost*) received a letter that complemented the *Simplified* Mailer by telling customers that people in their region were making a change in an effort to save water, and invited them to join their neighbors. Furthermore, for consumers with relatively high water consumption, it informed them that they were in the top 50 per cent of consumption, whereas for the bottom 50 per cent, it congratulated them on being efficient. The final two treatment groups, Treatment 5 and Treatment 6, were offered pecuniary incentives (*£10 Incentive* and *£15 Incentive*) for completing the water audits. The former supplemented the *Simplified* mailer by emphasizing monetary savings, and offered a £10 incentive for using the platform, while the latter communication changed the incentive from £10 to £15.

The data used to randomize the trial participants and to measure outcomes came from three anonymized sources: NWL's administrative data on meter readings; the SWSM platform, which was used to code responses to the diagnostic questionnaire; and Customer Relationship Management (CRM) data identifying whether reminder emails were opened.[10]

The experiment took place over four months between December 2018 and March 2019. We collected baseline data for purposes of randomization and analysis of pre-treatment consumption from January 2017. All direct mailers were posted on 8$^{\text{th}}$ December 2018, and email reminders were sent on 6$^{\text{th}}$ February 2019.

Table A.1 presents summary statistics on the observable characteristics of the households across different treatment groups. Using an F-test of joint significance, we find that

[9]We are not aware of other studies of water audits that have estimated the effect of email reminders.
[10]CRM is a tool to help manage and analyze customer interactions and data on websites.

the differences are not statistically significant at conventional levels, which suggests that that the various treatments are balanced on pre-treatment observable variables.

3 Results

We begin by reporting the effect of the letters on the take-up of the audit program, and then analyze the impact of the interventions on water consumption. We then use a LATE analysis to measure the effect of completing the diagnostic on water conservation. Our analysis of consumption is limited to households with meters. To measure the likely impact of the interventions if scaled up to include non-metered households, we reweight our estimates to reflect the broader population of consumers. We also use the data on reminders to study the effect of reminders on completing the audit. However, because the reminder emails were sent near the end of the study period, we cannot analyze their impact on consumption.

3.1 Likelihood of Engagement

To examine the effects of the behavioral interventions on the share of households that complete the diagnostic, we run the following regression:[11]

$$y_i = \alpha + \sum_j \beta_j T_{ij} + \gamma \mathbf{X}_i + \epsilon_i \tag{1}$$

where, y_i is a dummy variable that equals 1 if household i completed the water audit, and 0 otherwise. α represents the average take-up of the audit for the excluded treatment group. T_{ij} is a dummy that equals 1 if household i received treatment j, and 0 otherwise, where j refers to the different treatment groups. The coefficient of interest, β_j, is the average treatment effect (ATE) of the different letters on the likelihood of completing the audit. $\mathbf{X}_i$ represents a vector of dummy controls, γ is a vector of estimates of their impact, and ϵ_i is an error term.

Table 1 presents the estimates. The excluded group in models (1) and (2) is the *Vanilla* letter, and the excluded group in models (3) and (4) is the *Simplified* letter. The control vector here includes $Rural_i$, which is a dummy that equals 1 if household i lived in a rural area, and $Meter_i$, which is also a binary variable that equals 1 if household i has a water meter.[12] We present the results both with and without controls included in the regressions. Our results indicate that relative to the *Vanilla* treatment arm, all the letters

[11] The raw data from the field experiment on the number of households that completed the diagnostic, and how that differs across metered and unmetered households, is presented in Table A.2. We do not have data on the water-saving devices ordered by different households, and if they booked an in-home audit.

[12] We do not have data on household covariates like income, number of members, etc.

increased the take-up of the diagnostic significantly, with the *Incentive* treatment arm performing the best. Within the *Incentives* treatment arm, the higher financial incentive of £15 had a marginally greater impact (5.7 per cent versus 4.5 per cent).[13] The effect of the *Altruism* letter becomes insignificant when the reference treatment arm is the *Simplified* letter (column 3), but the impact of the *Incentives* and *Moral Cost* letters continues to be positive and significant. The results do not differ when we control for whether a household is situated in a rural area and has a water meter. We conclude that behavioral interventions can help to promote the use of audit tools, with financial incentives being more effective than others.

Table 1: ATE Estimates of Letters on Diagnostic Completion

	Completed Diagnostic			
	Vanilla		Simplified	
	(1)	**(2)**	**(3)**	**(4)**
Simplified	0.007^{***}	0.007^{***}		
	(0.002)	(0.002)		
Altruism	0.005^{**}	0.005^{**}	-0.002	-0.002
	(0.002)	(0.002)	(0.003)	(0.003)
Incentives 10	0.045^{***}	0.045^{***}	0.039^{***}	0.038^{***}
	(0.004)	(0.004)	(0.004)	(0.004)
Incentives 15	0.057^{***}	0.057^{***}	0.050^{***}	0.051^{***}
	(0.005)	(0.005)	(0.005)	(0.005)
Moral Cost	0.016^{***}	0.016^{***}	0.009^{***}	0.009^{***}
	(0.003)	(0.003)	(0.003)	(0.003)
Intercept	0.019^{***}	0.008^{***}	0.025^{***}	0.013^{***}
	(0.002)	(0.002)	(0.002)	(0.003)
Controls	✗	✓	✗	✓
Observations	37,298	37,298	29,838	29,838

Notes: Robust standard errors are in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.
All regressions report the average treatment effect estimates of different behavioral interventions on diagnostic completion (Equation (1)). The dependent variable for all models is *Completed Diagnostic*, a dummy variable that equals 1 if the household completed the water diagnostic, and 0 otherwise. Models (1) and (2) exclude the observations in the Control group, with the Vanilla letter comprising the reference treatment arm. Models (3) and (4) exclude the observations in the Control and Vanilla groups, with the Simplified letter serving as the reference group. Models (2) and (4) include the dummy variables *Meter* and *Rural* as controls. The former equals 1 if the household has a water meter attached to it, and the latter equals 1 if the household is located in a rural area.

[13] These impacts are statistically different from each other at a 1 per cent level of significance

3.2 Effect of the Behavioural Interventions on Water Consumption

Although the letters were successful in promoting the take-up of the water audit tool, the main objective was to encourage water conservation. In this section, we estimate the effects of the different communications on household water consumption. The letters can work in one of two ways: first, by directly encouraging an individual to conserve water after being influenced by the content of the letter; and second, through take-up of the audit. Unfortunately, we cannot estimate the direct effect of the letters because the time period between receiving the letter and completing the audit is too small. This section, thus, focuses on the overall impact on consumption of the direct encouragement and take-up of the audit.

To estimate the effect of the treatment on consumption, we run the following regression of post-treatment water consumption (y_i) on an indicator for whether the household was treated (T_i), and a vector of controls:

$$y_i = \alpha + \beta T_i + \gamma \mathbf{X}_i + \epsilon_i \tag{2}$$

where T_i equals 1 if the household received any treatment letter, and 0 if it was in the control group. With respect to water consumption, we have two data points for each household: one pre-treatment and one post-treatment. Moreover, consumption information was not available at a uniform frequency for all participants, which makes calculation of monthly consumption data difficult. Though pre-consumption data are available since 2017, post-consumption data were available only up until February 2019 for a majority of the sample. Therefore, we are only able to estimate the short-term impact of the audit. Furthermore, the data on water consumption is available only for the 30 percent of households who had a meter installed.[14] However, this is not a concern for our econometric identification strategy because, as shown in Table A.1, all treatment groups were balanced on the proportion of households with a water meter, pre-diagnostic consumption, and the number of consumers in each treatment group in the top 50$^{\text{th}}$ percentile of consumption (*high-use* households).[15] Furthermore, we run balance tests on observable characteristics just for metered households in Table A.1 (columns (6) and (7)) and find no significant differences between different treatment groups. Additionally, to address concerns about possible bias in sample selection due to the focus on metered households, , we reweight our LATE estimates inSection 3.4 so that the metered sample is similar to the general population.

To analyze the heterogeneity among different treatments, we ran a regression similar to Equation (1), with y_i now denoting post-treatment water consumption for household

[14]Further details on the computation of pre- and post-treatment water consumption are provided in Appendix C.

[15]Calculations on pre-diagnostic consumption and number of *high-use* customers are only feasible for households that have meters.

i. The vector of covariates, $\mathbf{X}_i$, now consists of $Rural_i$ and an additional covariate, *Pre-Treatment Water Consumption*$_i$, which measures the daily water usage of a household before the letters were sent.[16] For all regressions, variables related to water consumption are measured in liters per day. The effect of receiving a letter on consumption (Equation (2)) is presented in column (1) of Table 2, while the heterogeneity results are reported in columns (2)-(4).

We find evidence that all behavioral interventions reduced water consumption, measured in liters per day. Column (1) provides the average treatment effect of receiving any letter on post-treatment consumption. Though the estimate is negative, it is insignificant. Columns (2) through (4) estimate the effect for each behavioral intervention, with the reference group as the control, *Vanilla*, and *Simplified* letter, respectively. With reference to the control group, all treatment arms except *Vanilla* experienced a fall in consumption after letters were sent out; however, only the monetary incentives led to a statistically significant decrease. Though point estimates suggest that *Incentives* 15 had a larger impact than *Incentives* 10 (3.5 versus 4.7 liters per day), the two are not significantly different from each other. When we exclude the control group, and the *Vanilla* letter becomes the omitted category (column (3)), the drop in consumption is significant across all remaining categories, with the decrease in consumption ranging from 2.9 liters per day under *Moral Cost* to 6.3 liters per day under *Incentives 15*. In percentage terms, this decrease amounts to between 1.1 percent and 2.5 percent of the pre-treatment water consumption, a small but economically meaningful impact.[17] The effect of the *Incentives 15* treatment is more than twice the effect of the *Moral Cost* one, and the effect sizes are statistically different from each other. In general, pecuniary incentives lead to a significantly larger decrease in consumption when compared with other behavioral interventions.

Finally, dropping both the control group and the *Vanilla* group with the *Simplified* letter as the reference category (column 4) leads to only the £15 financial incentive remaining significant. To summarize, the *Incentives* group significantly reduced their consumption regardless of the reference group, while the other treatments had a significant negative impact only when we compare them to the *Vanilla* arm.

[16] We do not control for $Meter_i$ as the sample only includes households for which we had water consumption data, and this limits our sample to households with meters.

[17] The small decrease in consumption due to the *Moral Cost* letter (which also combined a social comparison message) is in contrast to the literature (Ferraro & Price 2013), which finds that social comparison messages have a greater impact on water conservation than prosocial messages or technical information alone.

Table 2: ATE Estimates of Letters on Post-Treatment Consumption

	Post-Treatment Water Consumption			
	Control (1)	Control (2)	Vanilla (3)	Simplified (4)
Treated	-1.392			
	(1.264)			
Vanilla		1.627		
		(1.600)		
Simplified		-1.595	-3.217^{**}	
		(1.635)	(1.592)	
Altruism		-1.593	-3.216^{**}	-0.031
		(1.625)	(1.581)	(1.613)
Incentives 10		-3.543^{*}	-5.177^{***}	-2.077
		(1.923)	(1.886)	(1.916)
Incentives 15		-4.687^{**}	-6.328^{***}	-3.259^{*}
		(1.919)	(1.882)	(1.910)
Moral Cost		-1.302	-2.925^{*}	0.191
		(1.592)	(1.546)	(1.581)
Intercept	117.635^{***}	116.782^{***}	119.045^{***}	44.042^{***}
	(43.032)	(42.439)	(42.634)	(6.524)
Controls	✓	✓	✓	✓
Observations	11,700	11,700	9,770	7,795

Notes: Robust standard errors are in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.
All regressions report the average treatment effect estimates of different behavioral interventions on post-treatment water consumption (Equation (2)), measured in liters per day. The dependent variable for all models is *Post-Treatment Water Consumption*, a continuous variable that measures the water consumption of a household after the treatment date December 8, 2018. Pre-treatment consumption and post-treatment consumption were available for only a subset (30 percent) of the households. Households with unreasonably large differences between pre- and post-treatment consumption (absolute value greater than 50 percent) were dropped from the sample. The data were trimmed at 1 and 99 percentile of pre-treatment consumption. The model names reflect the reference group for each regression. The regressor of interest in Model (1), *Treated*, is a dummy variable that equals 1 for all households that received any letter. Models (1) and (2) include all observations, with the control treatment arm constituting the reference group. Model (3) excludes the observations in the control group, with the *Vanilla* letter comprising the reference group. Model (4) excludes the observations in the control and *Vanilla* group, with the *Simplified* letter acting as the reference group. All models include *Rural* and *Pre-Treatment Consumption* as controls. *Rural* is a dummy variable that equals 1 if the household is located in a rural area. *Pre-Treatment Consumption* is a continuous variable that measures the water consumption of a household before the treatment date of December 8, 2018.

3.3 Effect of Diagnostic Completion on Water Consumption

We now turn our attention to estimating the impact of completing the audit on water consumption. To do this, we employ an Instrumental Variable (IV) strategy and two stage least squares (2SLS).

Table 3: LATE Estimates of Diagnostic Completion on Post-Treatment Consumption

	Post-Treatment Water Consumption			
	(1)	**(2)**	**(3)**	**(4)**
Complete Diagnostic	-45.446^{***}	-43.442^{***}	-58.922^{*}	-49.147
	(15.335)	(16.691)	(34.400)	(93.367)
Intercept	10.402^{***}	10.276^{***}	13.769^{***}	15.040
	(1.525)	(2.122)	(4.041)	(12.469)
Instruments	All Treatment	Incentives+Simplified	Incentives	Incentives 15
F-stat in First Stage	39	66	16	3
Controls	✓	✓	✓	✓
Observations	11,700	5,830	3,900	1,974

Notes: Robust standard errors are in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.
All regressions report the local average treatment effect estimates of diagnostic completion on post-treatment water consumption. The dependent variable for all models is *Post-Treatment Water Consumption*, a continuous variable that measures the water consumption of a household, in liters per day, post the treatment date of 08-Dec-2018. Pre-treatment consumption and post-treatment consumption were available for only a subset (30 per cent) of the households. Households with unreasonably large differences between pre and post-treatment consumption (absolute value greater than 50 per cent) were dropped from the sample. The data was then trimmed at 1 and 99 percentile of pre-treatment consumption. The regressor of interest is *Complete Diagnostic*, which is a dummy variable that equals 1 for all households who completed the water diagnostic. The IV in Model (1) is a vector of dummies for all the different treatment arms. The IV in Model (2) is a vector that includes dummies for Incentives 10, Incentives 15, and Simplified treatment arms. The IV in Model (3) is a vector of dummies for Incentives 10 and Incentives 15 groups, while the IV in Model (4) is only the Incentives 15 group. The sample in model (2) consists of the *Incentives*, *Simplified* and control group, while the sample in model (3) includes only *Incentives* and the *Simplified* group. Model (4) only includes the *Incentives* group. All models include *Rural* and *Pre-Treatment Consumption* as controls. *Rural* is a dummy variable that equals 1 if the household is located in a rural area. *Pre-Treatment Consumption* is a continuous variable that measures the water consumption of a household, in liters per day, before the treatment date of 08-Dec-2018.

The first stage involves running the following regression:

$$\textit{Diagnostic Completion}_i = \alpha + \sum_j \beta_j Z_{ij} + \gamma \mathbf{X}_i + \epsilon_i \tag{3}$$

where *Diagnostic Completion*$_i$ is a dummy that equals 1 if household i completed the online diagnostic, and Z_{ij} is the instrument used. The number of instruments vary depending on the specification, and j, therefore, refers to the different instruments. β_j is the estimate of the j^{th} instrument. $\mathbf{X}_i$ is a vector of household covariates as before, and consists of $Rural_i$ and *Pre-Treatment Water Consumption*$_i$. γ_i is a vector of estimates of the impact of the household covariates, and ϵ_i is the error term. The second stage involves using the residuals from Equation (3), $\hat{Z}_i$, to run the following regression:

$$y_i = \alpha + \beta \hat{Z}_i + \gamma \mathbf{X}_i + \epsilon_i \tag{4}$$

where y_i represents post-treatment water consumption, and $\mathbf{X}_i$ is the same vector of household covariates used in the first stage.

We use different combinations of instruments for our LATE estimates, all of which give us similar results. The results are presented in Table 3. The model in column (1) uses all the letters as instruments. Therefore, Z_i is a vector of length $j = 6$, with each element of the vector a dummy variable for the different letters. This model satisfies the *relevance* condition as letters do tend to increase adoption of the water audit tool (see Table 1). The estimates in column (1) suggest that completing the diagnostic led to a significant fall in consumption of 45 liters per day (17 percent). However, a potential problem with the instrument in column (1) is that the *exclusion restriction* may not strictly be satisfied, as certain letters could directly impact water consumption through their message of altruism or moral suasion (the direct impact). Therefore, in column (2), we reduce the sample to the following four groups: *Incentive* 10, *Incentive* 15, *Simplified* and the control group. Z_i now represents a vector of 3 instruments, namely *Incentives 10*, *Incentives 15*, and *Simplified* groups.[18] We are reasonably confident of satisfying the *exclusion restriction* here because there were few differences between the *Incentives* and the *Simplified* letter, with the exception that the former used a monetary incentive. These letters simply asked the customers to download the water audit application, without any inducement to an environmental or altruistic cause, and therefore our assumption is that they should not impact water consumption directly. Our results under this specification indicate that completing the diagnostic reduces water consumption by 43 liters per day (17 percent).

Next, in column (3) we run a regression that is even more likely to be consistent with the *exclusion restriction* by restricting our sample to *Incentives 10*, *Incentives 15*, and *Simplified* groups. Our set of instruments now include the *Incentives 10* and *Incentives 15* groups. This specification is more robust because if the very act of receiving a letter influenced water consumption, then the presence of the control group in the regression would violate the *exclusion restriction*. Model (3) avoids this problem by only including customers who received the *Incentives* or *Simplified* letters. The effect size increases to 59 liters per day (23 percent), but it is significant only at the 10 percent level. This is most likely due to issues of statistical power as our sample size has decreased considerably. Finally, there is a possibility that the *Simplified* letter affected the customers attitude towards water conservation directly, and therefore, including it would be a violation of the *exclusion restriction*, making our LATE estimates biased. Hence, in column (4), our sample only includes the *Incentives 10* and *Incentives 15* groups. Z_i is now a single instrument represented by a dummy for the *Incentives 15* group. Though we get a negative coefficient of 49 liters per day, which is similar to our earlier specifications, it is insignificant because of low statistical power. Given the small sample size in the last two models, we use model (2) as our preferred specification.

[18] This formulation appears to satisfy the relevance condition *i.e.*, the correlation between the endogenous regressor and the IV is significantly different from 0. As column (2) in Table 2 shows, the letters do tend to increase the likelihood of completing the audit.

Our results suggest that there is a meaningful effect of completing the water audit tool on water consumption, ranging from 17-23 percent of pre-treatment consumption. However, the duration of this effect is not known; nor do we know whether the audit may have stimulated the adoption of new technologies over time.

In Appendix A.2, we examine whether there were any heterogeneous treatment effects of the behavioral interventions. Specifically, we test whether households with consumption greater than the median (*high-use households*) conserved more after receiving the letter. This is relevant because if audits differ in terms of their impact across groups, it may be more effective to target a behavioral intervention based on a customer's attributes. We find that the effect of the *Incentive* 10 and *Incentive* 15 letters on consumption is 7.5 and 8.4 liters per day, respectively, for the high users, and both effects are statistically significant. On the other hand, the effect is indistinguishable from 0 for the low users, irrespective of the intervention. A related analysis for the LATE, presented in Appendix A.2 (Table A.4), reveals a similar pattern. Completion of the audit had a large impact on the high users (between 78 liters to 89 liters per day), but it was not statistically different from 0 for the low users.[19] Thus, average treatment effects mask crucial heterogeneity.[20]

3.4 External Validity

The main results of our experiment are for metered houses in NWL's service region. In this section we explore how the results could generalize to other NWL consumers who do not currently have meters. Our reweighting exercise reduces the estimates of water savings from 43 liters per day (17 percent) to 34 liters per day (14 percent), but they still remain statistically significant.

The reweighting exercise is important because the sample used for estimating the effect sizes of the interventions consists solely of metered households. This may lead to concerns about the extent to which these findings generalize to other populations. We cannot say whether our numerical estimates generalize to populations outside the region that NWL serves; however, within our sample we can explore the extent to which the sample might be affected by including all customers as opposed to just those customers that have meters. Though we show that all observable covariates for the metered households are balanced across all treatment groups (Appendix A.1), we can test the sensitivity of the results by reweighting the study sample to match the demographic composition of the general population of NWL customers. We reweight the metered sample so that it looks like the general

[19] The difference between the ATE and LATE estimate (7.5-8.4 liters per day versus 78-89 liters per day) is large because the former measures the effect of the letter on water consumption for all treated households, while the latter looks at the impact on households who completed the audit.

[20] We also checked whether different interventions encouraged different categories of households to take-up the water-audit tool, but did not find any significant differences in take-up between high and low users.

population that was sampled, and that yields a reweighted LATE. One important caveat is that the reweighted LATE is conditional on unmetered households getting a meter. If this is not the case, the impact of an intervention on a metered household is likely to be very different from the same intervention for an unmetered one because information on water use via meters could significantly alter water consumption.

To implement the reweighting, we conduct the following four steps (similar to Stuart et al. (2011)). First, we determine the household demographics (X_i) we use to reweight. We choose all of the observable variables that were provided to us by the utility: rural-urban classification, availability of email address, and residential post-code.[21] Second, we use a logistic regression to model the probability ($\hat{p}_i$) of being metered with the covariates as predictors. $\hat{p}_i$ thus denotes the estimated probability of sample selection for household i. Third, we follow inverse probability of treatment weighting (IPTW) to weight each household (see Hahn & Metcalfe (2021) for weighting using sub-classification, which is a coarser method than IPTW). IPTW methods give each individual their own weight, which is calculated as the inverse propensity scores, *i.e.*, in our setting, the inverse probability of being metered: $w_i(X_i) = 1/\hat{p}_i(X_i)$. Lastly, we estimate the LATE using the weights w_i we generated as a population weight.

Table 4: Reweighted LATE Estimates

	Post-Treatment Water Consumption			
	(1)	**(2)**	**(3)**	**(4)**
Complete Diagnostic	-41.634^{**}	-33.854^{**}	6.403	-220.795
	(20.607)	(16.184)	(40.670)	(358.316)
Intercept	7.819^{***}	7.125^{***}	4.991	33.334
	(1.790)	(1.845)	(4.356)	(43.110)
Instruments	All Treatment	Incentives+Simplified	Incentives	Incentives 15
F-stat in First Stage	39	66	21	1
Controls	✓	✓	✓	✓
Observations	11,700	5,830	3,900	1,974

Notes: Robust standard errors are in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.
All regressions report the reweighted local average treatment effect estimates of diagnostic completion on post-treatment water consumption. The dependent variable for all models is *Post-Treatment Water Consumption*, a continuous variable that measures the water consumption of a household, in liters per day, post the treatment date of 08-Dec-2018. Pre-treatment consumption and post-treatment consumption were available for only a subset (30 per cent) of the households. Households with unreasonably large differences between pre and post-treatment consumption (absolute value greater than 50 per cent) were dropped from the sample. The data was then trimmed at 1 and 99 percentile of pre-treatment consumption. The regressor of interest is *Complete Diagnostic*, which is a dummy variable that equals 1 for all households who completed the water diagnostic. The IV in Model (1) is a vector of dummies for all the different treatment arms. The IV in Model (2) is a vector that includes dummies for Incentives 10, Incentives 15, and Simplified treatment arms. The IV in Model (3) is a vector of dummies for Incentives 10 and Incentives 15 groups, while the IV in Model (4) is only the Incentives 15 group. The sample in model (2) consists of the *Incentives*, *Simplified* and control group, while the sample in model (3) includes only *Incentives* and the *Simplified* group. Model (4) only includes the *Incentives* group. All models include *Rural* and *Pre-Treatment Consumption* as controls. *Rural* is a dummy variable that equals 1 if the household is located in a rural area. *Pre-Treatment Consumption* is a continuous variable that measures the water consumption of a household, in liters per day, before the treatment date of 08-Dec-2018.

[21] We do not have data on income or the number of household members, so use the data on post codes as a proxy.

To estimate the re-weighted LATE, we run the same regression as in Section 3.3, but include the weights in the estimation. The reweighted LATE estimates are presented in Table 4. For our preferred specification in column (2) (*Incentives 10*, *Incentives 15* and *Simplified* group as IV), re-weighting slightly reduces the point estimates from 43 liters per day to 34 liters per day (13 per cent), and the coefficients still remain statistically significant. The reweighted LATE in column (1), where we use all the letters as instruments, is also very similar to the unweighted LATE in Table 3, with the effect size a significant 42 liters per day. For models (3) and (4), where the instruments are both *Incentives 10* and *Incentives 15*, and only *Incentives 15*, respectively, the effects sizes are insignificant. We do not read too much into these coefficients because of low statistical power, but they have been presented for comparison. In conclusion, we are reasonably confident that our results are not driven by sample selection, and can be scaled up with similar effects to unmetered households, provided they are metered before the intervention.[22]

3.5 Effect of Reminders on Diagnostic Completion

Customers that provided NWL with email contact details (13,989 households) were also randomly allocated to groups that either received or did not receive an email reminder. The randomization was limited to households who had not completed the water audit by February 2019. The reminder emails followed the same themes as the initial direct mailers that the customers received. This allows us to estimate the impact of receiving a reminder email. We run the following regression to estimate the effect of reminders:

$$y_i = \alpha + \phi R_i + \sum_j \beta_j T_{ij} + \sum_j \pi_j R_i \times T_{ij} + \gamma \mathbf{X}_i + \epsilon_i \tag{5}$$

where, y_i is a dummy for diagnostic completion, R_i is a dummy that equals 1 if the household i received a reminder email, and T_{ij} is a dummy that turns on if household i initially received treatment j. $\mathbf{X}_i$ is a vector of household covariates, specifically dummies for whether the household was located in a rural area, and whether it had a water meter attached to it. The constant α represents the average diagnostic completion rate for households that were not sent a reminder and that belonged to the excluded group in the regression analysis, ϕ is the estimate for the average effect of any reminder on diagnostic completion, and β_j represents the effect of the initial treatment allocation on diagnostic completion. Our main coefficient of interest is π_j which is the estimate on the interaction term. It represents the effect of reminders belonging to the j^{th} treatment group on diagnostic completion. ϵ_i signifies the error term. The sample only includes households who had not completed the audit on the date the reminder emails were sent out. Table 5 presents the results.

[22]Metered and unmetered households may differ on unobserved variables, in which case the results may not generalize.

Table 5: ATE Estimates of Reminders on Diagnostic Completion

		Vanilla	Simplified
	(1)	**(2)**	**(3)**
Reminder	0.026^{***}	0.015^{***}	0.026^{***}
	(0.002)	(0.004)	(0.005)
Reminder × Simplified		0.011^{*}	
		(0.006)	
Reminder × Altruism		0.003	−0.007
		(0.006)	(0.006)
Reminder × Incentives 10		0.015^{*}	0.004
		(0.008)	(0.009)
Reminder × Incentives 15		0.011	0.001
		(0.008)	(0.009)
Reminder × Moral Cost		0.030^{***}	0.019^{**}
		(0.007)	(0.008)
Intercept	-0.007^{***}	-0.008^{***}	-0.009^{***}
	(0.002)	(0.002)	(0.002)
Controls	✓	✓	✓
Observations	11,031	11,031	8,752

Note: Robust standard errors are in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.
All regressions report the average treatment effect estimates of reminders on diagnostic completion (Equation (5)). The dependent variable for all models is *Completed Diagnostic*, a dummy variable that equals 1 if the household completed the water diagnostic, and 0 otherwise. Models (1) and (2) exclude the observations in the control group, with the *Vanilla* letter comprising the reference treatment arm in model (2). Model (3) excludes the observations in both the control and *Vanilla* group, with the *Simplified* letter constituting the reference group. The estimates on the various treatment arms (*Simplified, Altruism, Incentives 10, Incentives 15*, and *Moral Cost*) are omitted from the table in the interest of space, but are all statistically insignificant. Observations only include households for whom NWL had email contact details, provided they did not complete the diagnostic before the reminder emails were sent. Therefore, 631 households for whom NWL had email details, but who had completed the water diagnostic before the reminders were sent, were excluded from the analysis. All regressions include *Meter* and *Rural* as controls. The former equals 1 if the household has a water meter attached to it, and the latter equals 1 if the household is located in a rural area.

In column (1) in Table Table 5, we estimate the direct impact of any reminder on the likelihood of completing the diagnostic. To do so, we modify Equation (5) and run the model without the effect of initial treatment groups (T_{ij}), and the interaction terms between the treatment groups and reminders ($R_i \times T_{ij}$). Our findings suggest that reminders increased the likelihood of completing the diagnostic by 2.6 percent as compared to the group that did not receive the reminders. Next, in columns (2) and (3), we estimate the impact of each specific reminder. The omitted category in column (2) is the *Vanilla* letter, with control group excluded from the sample.[23] With reference to the *Vanilla* group, reminders to the *Moral Cost* group have the highest additional impact of 3 percentage points, while the magnitude of impact for *Incentives 10* and *Simplified* groups is also significant. Notably, the impact of the *Moral Cost* reminder is significantly different from the impact of these other two treatments. In the final specification in column (3), the omitted group is *Simplified*, with both control and *Vanilla* groups excluded from the sample.

4 Welfare Analysis

In this section, we examine whether promoting online water audits improves economic welfare. We consider the impacts of different interventions from our experiment on various measures of economic welfare. We have three main findings. First, there is not sufficient information on water conservation benefits to say whether specific interventions would pass a benefit-cost test. Second, the cost effectiveness of these interventions does not appear to be attractive relative to interventions studied by other researchers for promoting water conservation. Third, a benefit-cost analysis based on greenhouse gas benefits alone, and a comparable MVPF analysis, do not appear to suggest that the investment is worthwhile. However, without information on the full range of quantitatively significant benefits from water conservation, and the impact of our interventions over time, it is difficult to make informed statements about whether audits are likely to pass different benefit-cost tests. Nonetheless, we do a bounding exercise that allows us to estimate what other benefits would need to be for some of our interventions to just pass a benefit-cost test.

4.1 Cost Effectiveness

We consider the cost effectiveness of this intervention in detail and then compare it with other interventions in the literature.

[23]We omit estimates of effect of the initial treatment assignment (β_j) from Table 5 for brevity. However, all of the estimates of β_j are insignificant for all specifications.

Table 6: Different Measures of Cost Effectiveness

Case Parameter	Base Case (1)	No Producer Surplus Loss (2)	Vanilla Letter (3)	Targeting High Users (4)	Two Year Duration (5)
Cost of Mailing	420	420	-	200	420
Producer Surplus Loss	200	-	290	200	2200
Time Cost	68	68	68	30	68
[A]: Total Cost (in £)	690	490	360	430	2,700
[B]: Effectiveness (in m^3)	240	240	340	240	2,600
Cost Effectiveness (£/m^3)	2.9	2.1	1.0	1.8	1.0

Notes: This table shows how the cost effectiveness changes using different assumptions. Cost effectiveness is measured in terms of pounds per cubic meter of water conserved in 2020 £. It is computed as the total cost divided by the effectiveness (A/B). See text for details on the various cases.

4.1.1 Cost effectiveness of the natural field experiment

We measure three categories of costs: the cost of sending letters, the lost producer surplus associated with the decline in production, and the value of time in filling out the survey. Our base case is the *Incentive 10* treatment, and we measure its effectiveness relative to not sending out a letter, and to sending out the *Vanilla* letter. Results are presented in Table 6.

We describe the parameters[24] for the base case below. The cost of mailing represents the postal cost of sending letters to 1,040 participants (the sample size in the *Incentive 10* group) at a cost of 41 pence per letter, which was the Royal Mail's standard tariff in 2020-21 for bulk orders containing less than 2,500 items. There would also be costs associated with paper, ink and time, but we assume they are negligible. The *Producer Surplus Loss* is defined as the total loss in net revenue (*i.e.*, revenue minus cost) caused by water savings. Since the length of our post treatment consumption data differs across households, we assume that water savings last for 65 days, which is the average number of days post-treatment for which we have consumption data. Given a consumer price of £1.3 per cubic meter, a short-run marginal cost of 44 pence, and average savings of 3.5 liters per day per household, the producer surplus loss over the 65-day period is £200. The *Time Cost* is defined as the monetary value of time associated with filling out the survey and is computed as the product of the average time taken by a household to complete the survey (7 minutes) and 50 percent of the median UK hourly wage rate of £14 per hour, (available from the Annual Survey for Hours and Earnings, Office for National Statistics, UK ASHE

[24] For a list of all the parameters and their sources, see Table B.1

(2021)).[25] The sum of these items gives a total cost of £700.[26] Effectiveness is measured by the per capita reduction in water consumption relative to the case of no letter (3.5 liters per day for 65 days) multiplied by the number of people in the £10 incentive group, which gives 240 cubic meters. The cost effectiveness is given by total cost divided by effectiveness, or £2.9 per cubic meter.

The other four cases are variations on the base case. The first variation labeled *No Producer Surplus Loss* sets producer surplus losses to zero, yielding a cost effectiveness of £2.1 per cubic meter, which is a 29 percent decline relative to the base case. [27] The reason we present the case of *No Producer Surplus* is because many studies do not consider changes in producer surplus in computing cost effectiveness. In our view, this may be particularly important in cases involving utilities, where prices may differ substantially from marginal private costs (Reguant (2019), Hahn & Metcalfe (2021)). Many of the nudges that are carried out for water involve utility customers, and thus, this change should be included where possible.

The second variation changes the benchmark for comparison from the control group to the *Vanilla* letter. We do this analysis because the utility planned to send out this letter to their customers without our intervention. The cost effectiveness declines from the base case by 64 percent. The decline results from the reduction in mailing costs to zero, and the increase in water savings per household.

The third variation targets only high users, who are defined as users above the median consumption threshold of 220 liters per day. This leads to an increase in the average reduction in consumption from 3.5 to 7.5 liters per household per day. The cost effectiveness is reduced by 38 percent as a result, from £2.9 per cubic meter in the base case to £1.8 per cubic meter.[28] This suggests that targeting could be an important strategy for improving cost effectiveness and increasing net benefits, which is consistent with other studies (*e.g.*, see Ferraro & Price (2013), Ferraro & Miranda (2013), and Brent et al. (2020)).

The fourth variation considers the impact of a change in duration of the persistence of the effects due to the intervention.[29] Though we make the conservative assumption that our effects last for only 65 days on average, this is because of limitations of our data. In fact, various studies (Ansink et al. (2021), Bernedo et al. (2014)) find a long-term persistent impact of their interventions, going up to 6 years in some instances. If we assume the

[25] According to White (2016), for local personal travel, value of travel time savings (VTTS) is estimated at 50 percent of hourly median household income. We follow the VTTS convention for our calculations

[26] We do not include the cost of financial incentives in this calculation because they are treated as a transfer. We consider how this affects the calculation below.

[27] Note that a decline in the measure of cost effectiveness is arguably an *improvement*. Either costs go down or effectiveness, as measured by conservation, increases (or both).

[28] A similar calculation for the £15 intervention reveals that cost effectiveness is reduced by 30 percent.

[29] We ignore discounting here because it is not central.

benefits last for a full two years, this directly impacts the quantity of water conserved. Conservation increases by a factor of 11 (2,600 v/s 240 cubic meters), and cost effectiveness decreases from £2.9 in the base case to £1.0 per cubic meter, or by 65 percent.

We also considered the impact of our base case of £10 versus the £15 intervention. In the £15 incentive case, both costs and effectiveness increase, but effectiveness increases by more than the costs. The result (not shown in the table) is that the effectiveness of the £15 intervention is £2.5 per cubic meter, 15 percent lower than the £10 intervention.

Finally, we considered the impact of underestimating costs in two ways. First, it is possible that the cost of sending letters if provided by a private firm would be roughly twice the cost of postage.[30] This would increase the cost-effectiveness by a factor of two. We also considered the impact of including the costs of the incentive treatment in the the definition of costs. This has the effect of increasing cost effectiveness by about 120 percent from £2.9 to £6.5 in the base case.

4.1.2 Comparison with other studies

There are a small number of other studies that compute the cost effectiveness of water conservation measures using modern causal identification strategies. These studies are summarized in Table 7.

The table provides an estimate of the cost effectiveness of different water conservation studies in dollars per cubic meter of water conserved. The table illustrates five key points. First, cost effectiveness estimates vary over a large range, from \$0.06 per cubic meter in Ferraro & Miranda (2013) to \$8.1 per cubic meter in Ansink et al. (2021). Second, our study appears to fall in the mid-range of existing estimates. Third, most existing cost effectiveness estimates using causal studies do not include changes in producer surplus as an indirect cost. Fourth, only a small number of studies report estimates of both cost effectiveness and the quantity of water reductions associated with that activity. Finally, the persistence of the treatment effect is important for cost effectiveness, as can be seen from the difference between the cost effectiveness numbers in Ferraro & Price (2013) and Bernedo et al. (2014) (\$0.17 versus \$0.07 per cubic meter). Both studies analyze the same field experiment, but the former assumes the effect lasts for four months, while the latter estimates that effects are statistically detectable six years later. Finally, it is not clear in most cases whether these applications scale, and over what domain. This is a problem with many studies of this type (List (2022)).

[30] This is based on private conversations with two experts.

Table 7: Comparison of Cost-Effectiveness with Other Studies

Variables Studies	CE Estimate ($/m^3)	Intervention	CE Provided in Initial Study	Quantity Estimated ('000 m^3)	Direct Costs Estimated	Other Costs Estimated
Ferraro & Price (2013)	0.12 - 0.17	Social norm letter	Yes	700	$1.2 per hh	Forgone Revenues: $1.5mn to $1.6mn
Ferraro & Miranda (2013)	0.06 - 0.11	Social norm letter	Yes	No	$1.2 per hh	No
Bennear et al. (2013)	2.2 - 7.6	Rebate on high efficiency toilets	Yes	3.0	$170 per hh	Forgone Revenues: $9,800
Bernedo et al. (2014)	0.07	Social norm letter	Yes	1,700	$1.2 per hh	No
Brent et al. (2015)	0.50 - 0.75	Social comparison letter and home water report	Yes	No	$11 per hh per year	No
Datta et al. (2015)	-	Social comparison letter	No	6.7	$440	-
Brent et al. (2020)	-	Social comparison messages	No	27 - 37	No	-
Ansink et al. (2021)	3.8 - 8.1	Water audit with information and technological component	No	No	*Information arm*: $38 per hh; *Technology arm*: $29 per hh	No
NWL Study	1.3 - 3.7	Online audit with £10 incentive	Yes	0.24	$0.5 per hh	$0.3 per hh

Notes: Notes: *hh* refers to household. Details of the calculation for this table are provided in Table B.2 and Table B.3. Cost effectiveness is measured in terms of dollars per cubic meter of water conserved in 2020 dollars. Quantity is measured in terms of thousands of cubic meters. Ferraro & Price (2013), Ferraro & Miranda (2013), and Bernedo et al. (2014) analyze the same field experiment and therefore, the direct cost estimates are the same.

4.2 Benefit-Cost Analysis

The previous section considered the cost effectiveness of our intervention. In principle, one could do a full-blown benefit-cost analysis (BCA). We start with a simplified BCA, and then consider a Marginal Value of Public Funds (MVPF) approach in the next section that is more detailed. Our purpose in this section is to present a framework for a BCA that allows us to ask one simple question: how large do other benefits (*i.e.*, those not quantified in our analysis) need to be to just offset costs that we estimate? Other benefits could include ecosystem benefits as well as reductions in investment costs (see discussion below).

The benefits in our analysis result from greenhouse gas emission reductions associated with a reduction in water consumption. Non-carbon greenhouse gas emissions have been converted to CO_2-equivalents for use in our analysis.[31] The carbon footprint numbers for the water supply, use, and disposal system have been sourced from the Environmental Agency, a leading public body for the environment in England and Wales (Emma et al. 2008).

To define benefits formally, we introduce some notation. Let Δg be the total change in water consumption due to the intervention over the time period of our analysis. Let V be the incremental greenhouse gas benefits that result from one cubic meter reduction in water consumption[32]. The benefits from the intervention, B, are then $-V\Delta g$. The costs, C, are given by the direct incremental costs (cost of mailing the letters and financial incentives provided to the households who complete the audit) of the experiment, E, and any other losses in producer surplus that may result.[33] These losses can be represented by the difference in the price of water, p, and short-run marginal cost of water, c (presumed to be constant for simplicity), multiplied by the change in water consumption, Δg. That is, the producer surplus losses are $(p - c)\Delta g$. We can now estimate net benefits as follows:

$$\begin{aligned} \text{Net benefits} &= B - C \\ &= -V\Delta g + (p - c)\Delta g - E \end{aligned} \tag{6}$$

Note that we have not included any measure of consumer surplus. This is because we invoke the envelope theorem, and assume that consumers who switch are just as well off

[31] The contribution of different greenhouse gases to total water industry emissions are: carbon dioxide (74 per cent), nitrous oxide (14 per cent), and methane (12 per cent) (Emma et al. 2008).

[32] See Table B.1 for a full breakdown of the greenhouse gas benefits based on different stages of water supply and use.

[33] These costs may be better approximated by the prices charged by a private sector for do these tasks. We consider a sensitivity on costs below to address this issue.

as they were before. [34] If they are better off, then the measure of $B-C$ we estimate is an underestimate.[35] We also do not include other benefits from water conservation, which may be substantial, but for which we do not have an estimate. These include possible savings from reduced capital and operating costs associated with expanded supply (Maddaus (2011)). In addition, ecosystem services, such as habitat, biodiversity, fishing, recreation, erosion protection, aesthetic value, and non-use values that can result from conservation are not included (See Bishop & Weber (1996) for a more extended discussion on the impact of demand reduction on water utilities and the environment).

Estimates for the various parameters in Equation (6) are shown in Table 8 along with the detailed results on net benefits. We perform the analysis using two different assumptions about cost: a short-run marginal cost (SRMC) of £0.44 per cubic meter, and a long-run marginal cost (LRMC) of £0.98 per cubic meter. The cost numbers were estimated based on sources from NWL (NWL Financial Statements (2009, 2021)). The SRMC, in our case, is equivalent to the base operating expenditure per cubic meter of water, or the marginal operating cost. It takes capacity as given, and includes costs associated with electricity for water transport, storage and treatment, and abstraction charges by environmental agencies.[36] LRMC, on the other hand, is the sum of marginal operating and marginal capacity costs.[37] The LRMC was calculated based on the annualized cost of the last major water resource investment undertaken by NWL – expanding Abberton reservoir in 2009 – and equals £0.54 per cubic meter.[38]

Before explaining the results, it is useful to highlight one key point. The price for water in this application appears to exceed the estimated marginal social cost (MSC) based on quantified benefits. The current price is £1.3 and the estimated MSC is the sum of marginal private costs (either £0.44 if we use the SRMC or £0.98 if we use the LRMC) and marginal external costs (£0.27). This gives an estimated MSC of either £0.71 if we use the SRMC (£0.44 + £0.27) or £1.3 if we use the LRMC (£0.98 + £0.26). This observation implies that any conservation measure, even if it had no costs attached, would not pass a narrowly prescribed benefit-cost test because price already exceeds the estimated marginal social

[34] For a model that motivates our welfare equation based on nudge theory, see Allcott and Kessler (2019). These authors assume a lump sum tax finances the nudge and quasi-linear utility for consumers. For an application that includes externalities, see Akesson, Hahn and Metcalfe (draft).

[35] We explore this issue in a sensitivity analysis below.

[36] Marsden Jacob (2004) state that for all practical purposes in the water industry, estimating SRMC by reference to operating costs is reasonable. Moreover, conversations with NWL representatives suggested that setting SRMC equal to the short-run average costs was a reasonable assumption.

[37] Costs associated with investments as a result of an incremental increase in demand.

[38] This is similar to the concept of long-run incremental cost in Mann et al. (1980), and includes both the capital costs associated with a change in capacity and volume sensitive costs. However, in this case, it may be an underestimate because it does not appear to include investments in raw water and wastewater treatment facilities, and water and sewer networks. Such costs could increase the LRMC substantially, but NWL did not have an estimate.

cost. Stated another way, because price is greater than the estimated marginal social cost, consumers may be consuming too little water relative to what might be viewed as economically efficient. This, of course, assumes that the estimated marginal social cost is a valid measure. Below, we argue that is unlikely to be the case in many instances because other benefits associated with water conservation have not been quantified. This is the reason we perform the bounding exercise contained in Table 3 to estimate what those other benefits would need to be to just offset costs.

We consider three different cases for estimating net benefits associated with the SRMC and the LRMC. The first uses the base case with the *£10 Incentive*, and it is compared to the case of no letter. The second uses *£15 Incentive* intervention with the same comparison group. The third uses the *Vanilla* letter as the benchmark with the *£10 Incentive*. We use the two incentive treatments because those interventions are the ones that resulted in an economically significant reduction in water consumption. The rationale for considering a different benchmark is that NWL was going to send out the *Vanilla* letter anyway. V is computed based on the Social Cost of Carbon (SCC), which is the monetary value of the net harm to society associated with adding a small amount of carbon to the atmosphere in a given year. We use an estimate of the SCC of \$51 per metric ton of CO_2 (in 2020 dollars), which assumes a discount rate of 3 percent (Interagency Working Group, USG (2021)). Below, we also consider how using different values for the SCC would affect the benefit-cost analysis.

The table shows that the measured benefits fall short of the measured costs in all three scenarios under both the cost structures. This may not be particularly surprising in light of the fact that we are not quantifying many benefits. The measured benefits are slightly higher, albeit still negative with the LRMC because the producer surplus loss due to reduced consumption is lower as a result of assuming the higher cost. The second to last row of each panel in the table shows that other benefits would need to be between £2.5 to £7.2 per cubic meter reduced for the total benefits to just offset the total costs. Though this gives us the required monetary value of other benefits needed to break even, we can also calculate how much the other benefits have to be in relation to the greenhouse gas benefits. The final row of both the panels show that other benefits would need to be 9 times (in the case of LRMC; 11 times in the case of SRMC) as great as carbon emission benefits for benefits to justify costs.

One could also ask how much the SCC would have to increase for benefits to just equal costs when other benefits are excluded (or assumed to be zero). The answer is that in the base case with LRMC, the SCC would need to increase by about 2,000 percent to \$1,100 per ton, and to \$1,200 per ton using the SRMC. These numbers are much higher than most estimates for the SCC.

We also examine how increasing the persistence of our treatments affects the results.

Table 8: Simple Benefit-Cost Analysis

Case Parameter	Units	Base Case (£10 Incentive) (1)	£15 Incentive (2)	Vanilla Letter as Benchmark (3)
V	$£/m^3$	0.27	0.27	0.27
p	$£/m^3$	1.3	1.3	1.3
Δg	m^3	-240	-290	-340
$-V\Delta g$	£	64	80	94
E	£	1,300	1,900	850
***Panel A* (*SRMC*)**				
c	$£/m^3$	0.44	0.44	0.44
$(p-c)\Delta g$	£	-200	-250	-290
$B - C$ (Equation (6) above)	£	-1,400	-2,100	-1,000
Breakeven Other Benefits $= -(B - C)$	£	1,400	2,100	1,000
Breakeven Other Benefits / Δg	$£/m^3$	6.0	7.2	3.0
Breakeven Other Benefits / GHG benefits	*times*	22	26	11
***Panel B* (*LRMC*)**				
c	$£/m^3$	0.98	0.98	0.98
$(p-c)\Delta g$	£	-64	-80	-94
$B - C$ (Equation (6) above)	£	-1,300	-1,900	-860
Breakeven Other Benefits $= -(B - C)$	£	1,300	1,900	860
Breakeven Other Benefits / Δg	$£/m^3$	5.4	6.7	2.5
Breakeven Other Benefits / GHG Benefits	*times*	20	24	9

Notes: We implement the equation for net benefits, Equation (6). *Panel A* shows the results for short-run marginal costs (c=£0.44 per cubic meter). *Panel B* shows the results for long-run marginal cost (c=£0.98 per cubic meter). See appendix *** for details.

The answer can be found analytically by differentiating the net benefits equation with respect to Δg, which gives $-V + (p - c)$. This expression is £0.03 using the LRMC and £0.57 per cubic meter using the SRMC. Because both these numbers are positive, this says that net benefits actually increase with greater consumption, if we assume other benefits are zero. The reason is that the producer surplus losses per unit of consumption exceed the greenhouse gas benefits. If we ignore the producer surplus losses then the greenhouse gas benefits that result in net benefits of zero within the 65-day period corresponds to an SCC of $1,000 per ton.

The relationship between price and costs will also affect the benefit-cost analysis. If utility regulators set price equal to cost, then there would be no producer surplus. If we assume that the price were set equal to the SRMC, then estimated net benefits in the base case would increase by £200, from -£1,400 to -£1,200. Alternatively, if we assume that prices were set equal to the LRMC, then estimated net benefits in the base case would increase by £72, from -£1,300 to -£1,200.

We also considered the impact of substantially underestimating the LRMC, which may be the case for the reasons outlined above. As a bounding exercise, we consider an LRMC of £5 (Whittington et al. 2009), an order of magnitude higher than our point estimate. In this case, if these costs are avoided with conservation, the net benefits go from being negative to positive. For example in the £10 incentive case net benefits switch from -£1,300 to £520.

Finally, in some situations, it might be argued that people who made changes to their behavior as a result of taking the audit may actually benefit relative to the status quo. This could arise because of benefits from information that changes behavior or from benefits from the act of conserving (i.e., "warm glow"). Unfortunately, we do not have information on the extent to which such people benefited. However, we did perform a sensitivity analysis that assumes that people who take the audit benefit by the amount of the transfer they receive. The result was to increase net benefits in the base case from -£1,300 to -£430. Future work could address whether people who responded to a financial nudge were actually better off if they changed their behavior (Bernheim & Taubinsky (2018)).

4.3 MVPF framework

In this section we apply an MVPF approach to assessing benefits and costs. The core of the MVPF approach is to consider the after-tax benefits to all groups in society from a small change in expenditure on a particular intervention and compare that with the net cost to the government (Hendren & Sprung-Keyser (2020), Finkelstein & Hendren (2020)). In general, the higher the benefits and the lower the net cost to the government, the more attractive the intervention is, other things equal.

We introduce some notation to clarify our estimation procedure. Define after-tax benefits as WTP or willingness to pay, and define G as the net cost to the government. Our measure of MVPF is WTP/G.

First, we consider WTP. Define dg/dn as the change in water consumption for a small change in expenditure on the intervention (say £1), and define t_c as the tax rate on profits of the firm (which in this case is a utility). Then

$$\begin{aligned} WTP &= (1-t_c)(p-c)(dg/dn) - V(dg/dn) \\ &= ((1-t_c)(p-c) - V)(dg/dn) \end{aligned} \tag{7}$$

Equation (7) says conservation is worth *considering* if the loss in after-tax unit profits is more than compensated for by the environmental gain (assuming net costs, G, are positive).

Now consider the net cost to the government. This is given by

$$G = 1 - t_c(p - c)(dg/dn) \tag{8}$$

This says that the net *cost* to the government is the direct cost of the intervention (£1) plus the loss in firm revenues from a £1 increase in expenditures.

The formula for MVPF is thus:

$$\begin{aligned} MVPF &= WTP/G \\ &= \frac{((1 - t_c)(p - c) - V)(dg/dn)}{(1 - t_c(p - c)(dg/dn))} \end{aligned} \tag{9}$$

Table 9: MVPF Calculations

Case Parameter	Base Case (£10 Incentive) (1)	£15 Incentive (2)	Vanilla Letter as Benchmark (3)
Panel A (SRMC)			
c	0.44	0.44	0.44
WTP	-0.076	-0.062	-0.17
G	1.0	1.0	1.1
$MVPF = \frac{WTP}{G}$	-0.074	-0.060	-0.16
Panel B (LRMC)			
c	0.98	0.98	0.98
WTP	0.0049	0.0040	0.011
G	1.01	1.01	1.02
$MVPF = \frac{WTP}{G}$	0.0048	0.0039	0.010

Notes: This table computes the MVPF for the three scenarios described in Table Table 8 using Equation (9). *Panel A* shows the results for short-run marginal costs. *Panel B* shows the results for long-run marginal cost. The values for t_c and dg/dn are the same for *Panel B* as *Panel A*. The values for V and p are the same as those in Table 8. See appendix ** for details.

Table 9 summarizes three MVPF calculations. It mirrors the net benefit calculations. For the short-run marginal cost scenario, MVPF ranges from -0.16 to -0.062. The negative sign here arises because WTP is negative and the net cost to the government is positive. This analysis is similar to our simple BCA in that it suggests the investment is not worth making unless other benefits not included here are significant. Using LRMC instead of SRMC increases the after-tax benefits due to a fall in producer surplus loss. The MVPF is positive in this case, but still remains small in absolute terms.

As can be seen from Equation (7), increasing the social cost of carbon, which is proportional to V, would increase the MVPF. For example, increasing V to 0.68 [39] in the case of the SRMC would make mean that WTP, and hence $MVPF$ were zero using the other base case assumptions.

There are many uncertainties in the preceding analysis. The largest uncertainties may relate to categories that we have not quantified, including benefits not quantified and possible cost savings from deferring capital investment. In addition, there are uncertainties in many of the key parameters such as costs. In some cases, we think these uncertainties could change the direction of the benefit-cost analysis. That is why we did the bounding analysis.

5 Conclusion

Water suppliers and regulators are showing greater interest in assessing non-price mechanisms to encourage conservation as scarcity becomes more of an issue. One approach that is being used is water audits, which offer customers recommendations on how they could reduce their water consumption.

This paper explores how online water audits affect cost effectiveness and economic efficiency using a natural field experiment. We have three main findings. First, encouraging subjects to participate in an online water audit with financial incentives reduces household consumption by 43 liters per day, or about 17 percent. However, we also find that the size of the financial incentive used to encourage conservation only matters marginally, with £10 and £15 incentives having roughly the same effect. This finding suggests that it may be worth testing lower levels of financial incentives in future experiments to see how such incentives affect both water consumption and the willingness to take the audit. Second, notwithstanding these improvements in water conservation, the intervention does not appear to pass a benefit-cost test that only includes the benefits of reducing greenhouse gas emissions. Because our analysis does not quantify other potentially important benefits and cost savings of conservation, such as ecosystem benefits and reductions in infrastructure costs, we define a lower bound on other benefits needed for benefits to just offset costs. We find that other benefits would need to be about 22 times as high as greenhouse gas benefits for benefits to just offset costs using a standard analysis.[40] Using a marginal value of public funds approach for measuring benefits and costs yields similar conclusions. Third, we find that targeting of high users could roughly double the effectiveness of interventions with financial incentives.

[39] This amounts to an SCC of $130 per ton of CO_2e

[40] If the social cost of carbon increases over time, as much research suggests, such interventions could become more attractive.

There are several areas for future research that we think could be fruitful. First, we think it would be useful to develop better measures of the cost effectiveness and net benefits associated with different kinds of interventions aimed at promoting water conservation. Table 7, which reviews behavioral economics research in this area, reveals how little we know about the cost effectiveness of different interventions. It would be useful for decision makers in charge of water conservation to know something about the likely costs and effectiveness of the group of interventions they are considering. The same is true of net benefits. Very few studies using causal methods for estimating water conservation have tried to address the net benefit question. We think using both a standard net benefit framework as well as the MVPF framework could provide useful inputs to decision making. Just as Hendren & Sprung-Keyser (2020) developed and compared several estimates of MVPFs in the education and health areas, it could be useful to undertake a similar exercise for water and energy interventions. Second, it would be very useful to try to quantify some of the other benefits associated with water conservation in monetary terms. Related to that, it would be useful to get better measures of the full marginal external cost of water consumption, and how this varies over time and space (Hanemann et al. (2006), Garrick et al. (2017)). Third, better information is needed on private costs, in particular the short-run and long-run marginal costs associated with water supply in different regions, as these will also be critical in assessing the net benefits of conservation. Armed with more accurate information on the marginal social cost and its relationship to price, policy makers will be in a better position to design more equitable and efficient policies that promote conservation when it is needed.

References

Allcott, H. (2011), 'Social norms and energy conservation', *Journal of Public Economics* **95**(9-10), 1082–1095.

Allcott, H. & Greenstone, M. (2017), Measuring the welfare effects of residential energy efficiency programs, Technical report, National Bureau of Economic Research.

Ansink, E., Ornaghi, C. & Tonin, M. (2021), 'Technology vs information to promote conservation: Evidence from water audits'.

ASHE, U. (2021), 'Annual survey of hours and earnings'. data retrieved from Office of National Statistics, https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/ashe1997to2015selectedestimates.

Ayres, I., Raseman, S. & Shih, A. (2013), 'Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage', *The Journal of Law, Economics, and Organization* **29**(5), 992–1022.

Bennear, L. S., Lee, J. M. & Taylor, L. O. (2013), 'Municipal rebate programs for environmental retrofits: An evaluation of additionality and cost-effectiveness', *Journal of Policy Analysis and Management* **32**(2), 350–372.

Bernedo, M., Ferraro, P. J. & Price, M. (2014), 'The persistent impacts of norm-based messaging and their implications for water conservation', *Journal of Consumer Policy* **37**(3), 437–452.

Bernheim, B. D. & Taubinsky, D. (2018), 'Behavioral public economics', *Handbook of Behavioral Economics: Applications and Foundations 1* **1**, 381–516.

Bishop, D. B. & Weber, J. A. (1996), *Impacts of demand reduction on water utilities*, American Water Works Association.

Brent, D. A., Cook, J. H. & Olsen, S. (2015), 'Social comparisons, household water use, and participation in utility conservation programs: Evidence from three randomized trials', *Journal of the Association of Environmental and Resource Economists* **2**(4), 597–627.

Brent, D. A., Lott, C., Taylor, M., Cook, J., Rollins, K. & Stoddard, S. (2020), 'What causes heterogeneous responses to social comparison messages for water conservation?', *Environmental and Resource Economics* **77**(3), 503–537.

Browne, O., Gazze, L., Greenstone, M. & Rostapshova, O. (2019), Enforcement and deterrence with certain detection: An experiment in water conservation policy, Technical report, Working Paper.

Browne, O. R., Gazze, L. & Greenstone, M. (2019), Determinants of water conservation: Evidence from a California drought, Technical report, Working Paper.

Datta, S., Datta, S., Josï, J., Zoratto, L., Calvo-Gonzï, O., Darling, M. & Lorenzana, K. (2015), *A behavioral approach to water conservation: evidence from Costa Rica*, World Bank Washington, DC, USA.

Emma, R., Feiferi, L., Fida, C. & Paul, R. (2008), 'Greenhouse gas emissions of water supply and demand management options', *UK Environmental Agency Science Report, SC070010* .

Environmental Protection Agency, USA (2013), 'Water audits and water loss control for public water systems', *Environmental Protection Agency, Washington DC, USA* .

Exchange Rates UK (2020), https://www.exchangerates.org.uk/. Accessed: 2022-04-30.

Ferraro, P. J. & Miranda, J. J. (2013), 'Heterogeneous treatment effects and mechanisms in information-based environmental policies: Evidence from a large-scale field experiment', *Resource and Energy Economics* **35**(3), 356–379.

Ferraro, P. J. & Price, M. K. (2013), 'Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment', *Review of Economics and Statistics* **95**(1), 64–73.

Finkelstein, A. & Hendren, N. (2020), 'Welfare analysis meets causal inference', *Journal of Economic Perspectives* **34**(4), 146–67.

Fowlie, M., Greenstone, M. & Wolfram, C. (2018), 'Do energy efficiency investments deliver? evidence from the weatherization assistance program', *The Quarterly Journal of Economics* **133**(3), 1597–1644.

Garrick, D. E., Hall, J. W., Dobson, A., Damania, R., Grafton, R. Q., Hope, R., Hepburn, C., Bark, R., Boltz, F., De Stefano, L. et al. (2017), 'Valuing water for sustainable development', *Science* **358**(6366), 1003–1005.

Gerarden, T. D. & Yang, M. (2021), 'Using targeting to optimize program design: evidence from an energy conservation experiment', *Energy, COVID, and Climate Change, Online-June* pp. 7–9.

Hahn, R. W. & Metcalfe, R. D. (2021), 'Efficiency and equity impacts of energy subsidies', *American Economic Review* **111**(5), 1658–88.

Hanemann, W. M. et al. (2006), 'The economic conception of water', *Water Crisis: myth or reality* **61**, 74–76.

Hendren, N. & Sprung-Keyser, B. (2020), 'A unified welfare analysis of government policies', *The Quarterly Journal of Economics* **135**(3), 1209–1318.

HM Government (2019), 'Environmental reporting guidelines: including streamlined energy and carbon reporting guidance'.

Interagency Working Group, USG (2021), Technical support document: social cost of carbon, methane, and nitrous oxide interim estimates under executive order 13990, Technical report, Tech. rep., White House. URL https://www. whitehouse. gov/wp-content/uploads ....

Jaime Torres, M. M. & Carlsson, F. (2016), 'Social norms and information diffusion in water-saving programs: Evidence from a randomized field experiment in Colombia'.

Jessoe, K., Lade, G. E., Loge, F. & Spang, E. (2021), 'Spillovers from behavioral interventions: Experimental evidence from water and energy use', *Journal of the Association of Environmental and Resource Economists* **8**(2), 315–346.

Knittel, C. R. & Stolper, S. (2019), Using machine learning to target treatment: The case of household energy use, Technical report, National Bureau of Economic Research.

List, J. A. (2022), *The voltage effect: How to make good ideas great and great ideas scale*, Currency.

Maddaus, W. O. (2011), 'Realizing the benefits from water conservation', *Journal of contemporary water research and education* **114**(1), 2.

Mann, P. C., Saunders, R. J. & Warford, J. J. (1980), 'A note on capital indivisibility and the definition of marginal cost', *Water Resources Research* **16**(3), 602–604.

Mansur, E. T. & Olmstead, S. M. (2012), 'The value of scarce water: Measuring the inefficiency of municipal regulations', *Journal of Urban Economics* **71**(3), 332–346.

Marsden Jacob, A. (2004), 'Estimation of long run marginal cost (LRMC)', *Queensland Competition Authority* .

Nauges, C. & Whittington, D. (2019), 'Social norms information treatments in the municipal water supply sector: some new insights on benefits and costs', *Water Economics and Policy* **5**(03), 1850026.

NWL Charges Scheme (2020), 'Northumbrian water limited household charges scheme'.

NWL Financial Statements (2009), 'Northumbrian water limited annual report and financial statements'.

NWL Financial Statements (2021), 'Northumbrian water limited annual report and financial statements'.

Ornaghi, C. & Tonin, M. (2021), 'The effects of the universal metering programme on water consumption, welfare and equity', *Oxford Economic Papers* **73**(1), 399–422.

Pörtner, H., Roberts, D., Poloczanska, E., Mintenbeck, K., Tignor, M., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., Möller, V. et al. (2022), 'Ipcc, 2022: Summary for policymakers', *Climate Change* .

Reguant, M. (2019), 'The efficiency and sectoral distributional impacts of large-scale renewable energy policies', *Journal of the Association of Environmental and Resource Economists* **6**(S1), S129–S168.

Royal Mail (2021), 'Price guide: Advertising mail (including sustainable), business mail, publishing mail, subscription mail, door to door, response services and inbound mail services'. data retrieved from Royal Mail, https://www.royalmail.com/sites/royalmail.com/files/2020-12/Marketing_services_and_general_correspondence_price_guide_1_January_2021_v2.pdf.

Stuart, E. A., Cole, S. R., Bradshaw, C. P. & Leaf, P. J. (2011), 'The use of propensity scores to assess the generalizability of results from randomized trials', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **174**(2), 369–386.

Sturm, R., Gasner, K. & Andrews, L. (2015), *Water audits in the United States: A review of water losses and data validity*, Water Research Foundation.

US Bureau of Labor Statistics (2021), 'All urban consumers (cpi-u)', *Washington, DC: US Department of Labor, Bureau of Labor Statistics* .

Vorosmarty, C. J., Green, P., Salisbury, J. & Lammers, R. B. (2000), 'Global water resources: vulnerability from climate change and population growth', *science* **289**(5477), 284–288.

West, J. D., Fairlie, R. W., Pratt, B. E. & Rose, L. (2021), Automated enforcement of irrigation regulations and social pressure for water conservation, Technical report, National Bureau of Economic Research.

White, V. (2016), 'Revised departmental guidance on valuation of travel time in economic analysis', *Office of the Secretary of Transportation, US Department of Transportation, Available at: transportation. gov/sites/dot. gov/files/docs/2016% 20Revised% 20Value% 20of% 20Travel% 20Time% 20Guidance. pdf* .

Whittington, D., Hanemann, W. M., Sadoff, C. & Jeuland, M. (2009), *The challenge of improving water and sanitation services in less developed countries*, Now Publishers Inc.

Wichman, C. J., Taylor, L. O. & Von Haefen, R. H. (2016), 'Conservation policies: who responds to price and who responds to prescription?', *Journal of Environmental Economics and Management* **79**, 114–134.

Appendices

The Appendix is divided into four sections. Appendix A provides balance tables, statistics on diagnostic completion, and additional results on heterogeneity of treatment effects based on pre-treatment water consumption. We also compare the characteristics of households who complete the diagnostic versus those who did not. The section concludes by analyzing the interaction of households with the reminder emails and how it differed across treatment arms. Section 2 provides details on the welfare section. We first report all the parameters and their sources, and subsequently present our calculations of the cost effectiveness of other studies in the literature. Section 3 sheds light on the measurement of pre- and post-treatment water consumption data using an illustrative example. Finally, in Section 4, we provide samples of the different letters and reminders that were sent to households.

A Baseline Balance and Additional Results

A.1 Balance Table

Our various treatments are balanced on pre-treatment covariates. Table A.1 provides a measure of the balance on observed covariates across different treatment groups. Column (1) reports the number of people in each treatment group. Columns (2) to (4) provide the percentage of population with a water meter, living in a rural area, and for whom the utility had an email id, respectively. Column (5) reports balance on the number of consumers for whom we had water data available. We also check for balance within the sub sample of customers with meters as our LATE estimates only use metered households. In this regard, columns (6) and (7) report the number of metered households living in rural areas, and who provided NWL with an email id. The only significant differences (at 10 percent) are: a) *Vanilla* households have a lower probability (67 percent versus 70 percent in the control group) of living in rural areas, and b) fewer customers (41 percent versus 44 percent in the control group) in the *Incentives 10* group had registered their email ids with NWL. We, therefore, control for these covariates in our regressions. Finally, columns (8) and (9) provide balance on pre-treatment water consumption, and how many consumers within each treatment group fell in the top 50$^{\text{th}}$ percentile of water consumption for the entire sample.

We calculated the p-value on t-test of equality of means with control group, and the same is reported in brackets. We find that the covariates for the treatment arms are not significantly different from the covariates in the control group.[41] Column (8) reports the p-values from F-tests of joint significance of all the regressors from an OLS regression where the dependent variable is a dummy variable taking a value of 0 if the customer is assigned to the Control group, and it takes a value of 1 for customers assigned to the treatment group in each respective row. A significant F-test would represent that covariates can predict participation in a particular group, but all of them are insignificant. Finally, the p-values reported in the last row are from the F-test of joint significance of the treatment dummies from an OLS regression where the dependent variable is the observable covariate and the independent variables are dummies for different treatment groups. A significant F-test would indicate that in at least one treatment group the mean of the covariate is different than the others. Again, we fail to reject the null hypothesis that all coefficients are 0.

Table A.2 provides the raw data from the RCT on the number of households that completed the diagnostic. These figures are further broken down based on the number of metered and unmetered households. As reported in column (4), the majority of households that completed the audit were metered, and this is consistent across all treatment groups.

A.2 Heterogeneity Based on Pre-Treatment Water Consumption

Targeting households based on some pre-treatment covariates may be a more cost-effective intervention for utilities if there is heterogeneity in response. We find that our letters have a greater impact on high water users, and the result holds with the LATE estimate of impact of the online audit.

To show that high-use households are more likely to be influenced by these interventions, we run the following econometric model:

$$y_i = \alpha + \sum_j \beta_j T_{ij} + \phi \ \textit{High-Use}_i + \sum_j \eta_j \ \textit{High-Use}_i \times T_{ij} + \gamma \mathbf{X}_i + \epsilon_i \tag{10}$$

where, y_i denotes post-treatment water consumption for household i, T_{ij} is a dummy that equals 1 if household i received treatment j, where j refers to the different treatment groups. $\textit{High-Use}_{\text{i}}$ is also a dummy and equals 1 if household i had a pre-treatment water consumption greater than the median of the sample. γ is a vector of estimates for the different dummy controls, represented by $\mathbf{X}_i$, for household i. These controls include $Rural_i$ and $\textit{Pre-Treatment Water Consumption}_{\text{i}}$ in liters per day. Finally, ϵ_i is the error term. If households with higher pre-treatment usage were incentivized more to conserve water, we would expect η_j to be negative and significant.

[41] It is important to note that water consumption data was only available for metered customers and, therefore, columns (6) to (9) pertain to the sub sample with meters.

Table A.1: Baseline Balance Across Treatment Groups

	Number of Customers	Has a Water Meter	Lives in Rural Area	Provided an Email	Water Consumption Data Available	Lives in Rural Area (Metered h/h)	Provided an Email (Metered h/h)	Pre-Diagnostic Consumption (m^3/day)	Top 50% Consumers	F-test of Joint Significance
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
All Customers	44,757	.429 (.002)	.701 (.002)	.313 (.002)	.324 (.002)	0.685 (.003)	0.434 (0.004)	.258 (.003)	.499 (.004)	
Control	7,459	.427 (.006)	.706 (.005)	.312 (.005)	.320 (.005)	0.695 (0.008)	0.441 (0.010)	.257 (.007)	.499 (.010)	
Vanilla	7,460	.428 (.006) [.887]	.696 (.005) [.206]	.316 (.005) [.638]	.327 (.005) [.339]	0.674 (0.008) [0.066]	0.433 (0.010) [.581]	.266 (.007) [.360]	.498 (.010) [.921]	{.296}
Simplified	7,460	.428 (.006) [.848]	.703 (.005) [.697]	.313 (.005) [.920]	.321 (.005) [.852]	0.692 (0.008) [.779]	0.435 (0.010) [0.709]	.256 (.007) [.855]	.511 (.010) [.427]	{.501}
Altruism	7,460	.429 (.006) [.797]	.699 (.005) [.355]	.314 (.005) [.809]	.324 (.005) [.591]	0.685 (0.008) [.386]	0.442 (0.010) 0.900	.253 (.007) [.661]	.500 (.010) [.966]	{.486}
Incentives 10	3,789	.436 (.008) [.356]	.699 (.007) [.488]	.320 (.008) [.372]	.332 (.008) [.189]	0.685 (0.011) [.445]	0.409 (0.014) [0.062]	.248 (.010) [.451]	.482 (.014) [.300]	{.252}
Incentives 15	3,670	.423 (.008) [.670]	.698 (.008) [.409]	.303 (.008) [.313]	.324 (.008) [.633]	0.678 (0.012) [.219]	0.440 (0.014) [0.985]	.254 (.010) [.752]	.472 (.015) [.121]	{.499}
Moral Cost	7,459	.430 (.006) [.741]	.701 (.005) [.519]	.309 (.005) [.684]	.323 (.005) [.700]	0.683 (0.008) [.270]	0.432 (0.010) [.529]	.263 (.007) [.557]	.513 (.010) [.373]	{.656}
F-test of Joint Significance		{.958}	{.916}	{.733}	{.883}	{0.595}	{0.581}	{.698}	{.229}	

Notes: Robust standard errors from OLS regressions are in parenthesis.
p-value on t-test of equality of means with control group is in brackets; P-value on F-tests is in braces.
All data was provided by Northumbrian Water Limited. Column (1) reports the number of customers assigned to each treatment. Columns (2) to (5) report the mean value of each customer characteristic, derived from an OLS regression of the characteristic of interest on a series of dummy variables for each treatment group. The excluded (comparison) group in these regressions is the control group. Robust standard errors are reported in parenthesis throughout. Columns (6) and (7) report balance on customer characteristics for a specific sub sample: *metered* households. Columns (8) and (9) provide the balance on pre-treatment water consumption and number of *high users*, defined as households with pre-treatment water consumption greater than the median of the metered sample. Column (8) reports the p-values from F-Tests of joint significance of all the regressors from an OLS regression where the dependent variable is a dummy variable taking value 0 if the customer is assigned to the control group, and it takes value 1 for customers assigned to treatment group j, and the independent variables are the variables in columns (2) to (4) and (8) to (9). The sample only includes observations for which we have consumption data available, and therefore, *Water Consumption Data Available* (column (5)) is excluded as the same will always be 1 for every observation. The p-values reported in the last row are from the F-test of joint significance of the treatment dummies in each column regression where the sample includes all customers, except for columns (6) to (9), in which case the sample only includes observations for which we have consumption data available.

Table A.2: Statistics on Diagnostic Completion and Metered Households

	Number of Customers	Completed Audit	Metered Customers	Metered Customers who Completed Audit
		(% of Customers)	(% of Customers)	(% of Completed Audit)
	(1)	(2)	(3)	(4)
All Customers	44,757	1,287 (2.9)	19,180 (42.9)	860 (66.8)
Control	7,459	3 (0.0)	3,184 (42.7)	3 (100.0)
Vanilla	7,460	140 (1.9)	3,193 (42.8)	102 (72.9)
Simplified	7,460	189 (2.5)	3,196 (42.8)	133 (75.6)
Altruism	7,460	176 (2.4)	3,200 (42.9)	119 (67.6)
Incentives 10	3,789	242 (6.4)	1,652 (43.6)	136 (56.2)
Incentives 15	3,670	278 (7.6)	1,551 (42.3)	161 (57.9)
Moral Cost	7,459	259 (3.5)	3,204 (43.0)	206 (79.5)

Notes: All data was provided by Northumbrian Water Limited. Column (1) reports the number of customers assigned to each treatment group. Column (2) reports the number of customers who completed the online diagnostic. Percentage of households which completed the audit relative to total number of households in the treatment group are reported in parenthesis. Column (3) reports the number of customers who had a water meter installed in their homes. Percentage of metered households relative to total number of households in the treatment group are reported in parenthesis. Column (4) reports the number of metered households who completed the audit. Percentage of metered households which completed the audit relative to total number of households which completed the audit in the treatment group are reported in parenthesis.

Table A.3: Heterogeneous Treatment Effects Based on Pre-Treatment Usage

	Post-Treatment Water Consumption			
	Control	**Control**	**Vanilla**	**Simplified**
	(1)	**(2)**	**(3)**	**(4)**
High-Use	9.716^{***}	9.729^{***}	7.328^{***}	5.361^{*}
	(2.702)	(2.701)	(2.822)	(2.918)
High-Use × Treated	-4.891^{*}			
	(2.503)			
High-Use × Vanilla		-2.199		
		(3.196)		
High-Use × Simplified		-5.821^{*}	-3.614	
		(3.220)	(3.171)	
High-Use × Altruism		-5.223	-3.025	0.679
		(3.223)	(3.170)	(3.194)
High-Use × Incentives 10		-7.863^{**}	-5.655	-2.053
		(3.920)	(3.882)	(3.899)
High-Use × Incentives 15		-7.398^{*}	-5.195	-1.554
		(3.977)	(3.934)	(3.958)
High-Use × Moral Cost		-3.960	-1.760	1.913
		(3.148)	(3.092)	(3.123)
Intercept	7.771^{***}	7.770^{***}	10.546^{***}	10.879^{***}
	(1.664)	(1.663)	(1.739)	(1.995)
Controls	✓	✓	✓	✓
Observations	11,700	11,700	9,770	7,795

Notes: Robust standard errors are in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.
All regressions report the average treatment effect estimates of different behavioral interventions on post-treatment water consumption (Equation (10)). The dependent variable for all models is *Post-Treatment Water Consumption*, a continuous variable that measures the water consumption of a household, in liters per day, post the treatment date of 08-Dec-2018. Pre-treatment consumption and post-treatment consumption were available for only a subset (30 per cent) of the households. Households with unreasonably large differences between pre- and post-treatment consumption (absolute value greater than 50 per cent) were dropped from the sample. The data was trimmed at 1 and 99 percentile of pre-treatment consumption. The model names reflect the reference group for each regression. The regressor of interest, *High-Use*, is a dummy that equals 1 if the household had a pre-treatment water consumption greater than the median of the sample. *Treated* is a dummy variable that equals 1 for all households who received any letter. The estimates on *Treated* and the various treatment arms (*Vanilla*, *Simplified*, *Altruism*, *Incentives 10*, *Incentives 15*, and *Moral Cost*) are omitted from the table in the interest of space. Models (1) and (2) include all observations, with the control treatment arm constituting the reference group. Model (3) excludes the observations in the control group, with the *Vanilla* letter comprising the reference group. Model (4) excludes the observations in the control and *Vanilla* group, with the *Simplified* letter acting as the reference group. All models include *Rural* and *Pre-Treatment Consumption* as controls. *Rural* is a dummy that equals 1 if the household is located in a rural area. *Pre-Treatment Consumption* is a continuous variable that measures the water consumption of a household, in liters per day, before the treatment date of 08-Dec-2018.

Table A.3 presents the results. With reference to the control group, the interventions had an additional significant negative impact of 4.9 liters per day on treated consumers in the high usage category (column (1)). When we include individual dummies for different behavioral communications (column (2)), our findings suggest that the *Simplified* and *Incentive* letters have a significantly higher impact on high-use households. This heterogeneity, however, does not persist when our reference group changes to *Vanilla* or *Simplified* in columns (3)-(4), but the effect sizes are still negative and large. Thus, we do find evidence of heterogeneous treatment effects based on water usage prior to treatment, especially when we compare the interventions to the control group.

We also test for heterogeneity in our LATE estimates by running the following regression:

$$y_i = \alpha + \beta T_i + \phi \, \textit{High-Use}_i + \eta \, \textit{High-Use}_i \times T_i + \gamma \mathbf{X}_i + \epsilon_i \tag{11}$$

where T_i represents *Completed Diagnostic*, which is an indicator for whether the household completed the audit or not. The coefficient of interest is η, which represents the additional impact of completing the diagnostic on high users compared with low users. As discussed in Section 3.3, we need to use IV's for *Completed Diagnostic*, with the IV for the interaction term, $\textit{High-Use}_i \times T_i$, just the IV for $\textit{Completed Diagnostic}_i$ interacted with $\textit{High-Use}_i$. The results are presented in Table A.4.

For all specifications, the coefficient on *Completed Diagnostic* is negative but insignificant. Notably, the coefficient on the interaction term is negative and significant for our first two specifications, and also much higher than the coefficients in Table 3 where we do not distinguish between high- and low-use households. This implies that audits had a far greater impact on high-use households than low-use households. The estimate in column (3), when our sample includes *Simplified* and *Incentives* group, is negative but insignificant. However, as discussed earlier, this is most likely due to low statistical power because we lose a major portion of our sample. For our preferred specification in column (2), the fall in consumption is 83 liters per day for the high-use consumer. This represents a percentage reduction of 24 percent relative to pre-treatment consumption for high-use households. Thus, the online audit incentivized high-use households to conserve more water.

Both the results above lend credence to the theory that behavioral interventions can have heterogeneous impacts on consumers depending on their pre-treatment usage. Therefore, utilities can target the households with high consumption as they seem more likely to be incentivized.

Table A.4: LATE Estimates of Heterogeneous Treatment Effects

	Post-Treatment Water Consumption			
	(1)	**(2)**	**(3)**	**(4)**
Complete Diagnostic	-12.637	-5.097	-11.215	-47.769
	(10.745)	(11.459)	(39.373)	(63.047)
High-Use	9.851^{***}	13.475^{***}	13.818^{**}	8.364
	(2.651)	(3.420)	(6.973)	(28.261)
High-Use ×	-65.567^{**}	-82.943^{**}	-129.464	4.134
Complete Diagnostic	(32.642)	(36.576)	(111.668)	(241.173)
Intercept	10.292^{***}	10.460^{***}	10.944^{***}	16.654^{*}
	(1.603)	(2.311)	(2.077)	(9.449)
Instruments	All Treatment	Incentives+Simplified	Incentives	Incentives 15
F-stat in First Stage	20, 17	34, 30	40, 16	2, 1
Controls	✓	✓	✓	✓
Observations	11,700	5,830	11,700	1,974

Notes: Robust standard errors are in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.
All regressions report the local average treatment effect estimates of diagnostic completion on post-treatment water consumption. The dependent variable for all models is *Post-Treatment Water Consumption*, a continuous variable that measures the water consumption of a household, in liters per day, post the treatment date of December 08, 2018. Pre-treatment consumption and post-treatment consumption were available for only a subset (30 per cent) of the households. Households with unreasonably large differences between pre and post-treatment consumption (absolute value greater than 50 per cent) were dropped from the sample. The data was then trimmed at 1 and 99 percentile of pre-treatment consumption. The regressor of interest is *High-Use* × *Complete Diagnostic*. *Complete Diagnostic* is a dummy variable that equals 1 for all households who completed the water diagnostic. *High-Use* is a dummy that equals 1 if the household had a pre-treatment water consumption greater than the median of the sample. For all the models, the instrument for the interaction term is the IV for the endogenous variable, *Complete Diagnostic*, interacted with *High-Use*. The IV in Model (1) is a vector of dummies for all the different treatment arms. The IV in Model (2) is a vector that includes dummies for Incentives 10, Incentives 15, and Simplified treatment arms. The IV in Model (3) is a vector of dummies for Incentives 10 and Incentives 15 groups, while the IV in Model (4) is only the Incentives 15 group. The sample in model (2) consists of the Incentives, Simplified and control group, while the sample in model (3) includes only Incentives and the Simplified group. Model (4) only includes the Incentives group. All models include Rural and Pre-Treatment Consumption as controls. Rural is a dummy variable that equals 1 if the household is located in a rural area. *Pre-Treatment Consumption* is a continuous variable that measures the water consumption of a household, in liters per day, before the treatment date of December 08, 2018.

A.3 Characteristics of Households that Complete the Diagnostic

Continuing with our theme of targeting, we find that different behavioral interventions influenced different set of households to take up the water audit tool. This is relevant because if the different letters differ in terms of which households they influence, it may be easier to target the right behavioral intervention based on the customer attributes.

Table A.5 provides the average value of the household characteristics across different treatment arms for the subset of households who completed the diagnostic. The columns represent different interventions and each row represents a household characteristic, ranging from the type of residence and the number of different water-consumption devices installed, to its water and energy usage. The last column reports the p-value from a joint F-test of whether the household characteristic varies across the different groups. The results indicate that the financial incentives treatment influenced a relatively larger number of unmetered households to commit to the audit. Therefore, households who were un-

Table A.5: Characteristics of Households which Complete the Diagnostic

	Vanilla (1)	**Altruism** (2)	**Simplified** (3)	**Incentives 10** (4)	**Incentives 15** (5)	**Moral Cost** (6)	***F-Test*** (7)
Rural	0.61	0.66	0.65	0.70	0.68	0.65	0.71
Metered	0.73	0.68	0.70	0.56	0.58	0.80	9.18^{***}
Number of:							
Showers	1.29	1.31	1.31	1.19	1.21	1.27	1.76
Toilets	1.99	1.97	1.93	1.73	1.73	1.97	4.29^{***}
Basins	1.95	1.81	1.83	1.64	1.63	1.81	4.05^{***}
Bathtubs	0.90	0.92	0.93	0.88	0.89	0.90	0.37
Kitchen Utility Taps	1.33	1.25	1.44	1.26	1.37	1.28	3.03^{***}
People at Home	2.25	2.11	2.10	2.22	2.23	2.17	0.69
Cost of Water (*£/year*)	386.72	365.96	402.65	387.85	353.27	383.91	0.70
Frequency (per week):							
Showers	10.36	10.22	9.71	9.95	10.83	10.34	0.81
Baths	2.85	2.97	2.89	3.19	2.86	2.81	0.29
Boiling Water	27.39	24.51	25.16	23.79	24.10	26.12	1.78
Wash Up by Hand	12.96	13.26	15.41	14.16	12.87	13.13	1.73
Dishwasher	2.22	2.28	2.15	1.62	1.73	2.17	2.08^{*}
Washing Machine	5.22	4.85	4.51	4.98	4.51	4.36	1.16
Watering Garden	2.11	2.26	1.89	1.97	1.80	1.92	0.77
Shower Duration (*mins*)	6.49	6.83	7.05	7.68	7.04	6.73	2.10^{*}
Water Use ('000 litres/yr):							
Bathroom	85.11	81.47	83.52	90.22	90.76	86.68	0.87
Kitchen	32.56	30.15	30.72	31.66	30.53	30.46	0.49
Outdoor	1.08	1.49	1.40	1.03	1.03	1.09	1.53
Household	118.74	113.11	115.65	122.92	122.32	118.22	0.63
Per Person	54.12	55.52	55.58	56.22	56.19	55.02	0.27
Energy Use ('000 kWh/yr):							
Bathroom	1.22	1.21	1.23	1.34	1.34	1.27	0.65
Kitchen	0.67	0.63	0.64	0.65	0.63	0.64	0.35
Household	1.90	1.85	1.87	1.99	1.97	1.91	0.39
Type of Residence:							
Cottage/Bungalow	0.09	0.06	0.10	0.07	0.08	0.12	1.25
Detached	0.31	0.35	0.28	0.21	0.23	0.34	3.83^{***}
Flat	0.04	0.03	0.07	0.02	0.04	0.06	1.33
Semi-Detached	0.41	0.38	0.41	0.48	0.44	0.33	2.65^{**}
Terrace	0.15	0.18	0.15	0.21	0.21	0.15	1.44
Observations	140	176	189	242	278	259	

Notes: $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.
All data are from the water diagnostic survey, and the number of observations, therefore, include only the homes which completed the diagnostic. Columns 1 to 6 report the mean value of each household characteristic for the respective treatment groups. *Rural*, *Metered*, and all five variables related to *Type of Residence* are binary. *Cost of Water (£/year)* is self-reported and only includes homes which pay for their own water. The variables related to *Water Use (litres/year)* and *Energy Use (kWh/year)* are calculated by NWL based on the answers provided by the households in the diagnostic. *Energy Use (kWh/year)* is the total amount of energy used by a household in a year to heat water. The final column, *F-Test*, reports the p-value from a joint orthogonality test of equality of means between the six treatment groups.

able to monitor their daily consumption were more likely to complete the diagnostic if offered monetary rewards. Furthermore, the average number of basins and toilets were lower in households that completed the diagnostic owing to the *Incentives* treatment, suggesting that financial inducement is a strong motivator for smaller or poorer households. In other words, financial incentives also influenced households that would not reasonably be expected to use the water audit tool.

A.4 Interaction of Households with Reminders

Sending reminders to consumers may be an important method to reinforce the impact of behavioral interventions. Therefore, it is important to know how customers interact with reminders and their impact on take-up of the audit. We find that customers interaction with the reminder email depends on the content of the reminder, with Moral Cost reminder doing well in terms of positive engagement.

Email reminders were randomly sent to the subset of customers that had not completed the diagnostic by February 2019. Using CRM data, we can count the number of people who opened the reminder emails, or opened the reminder email and clicked on the link to the audit tool, or simply unsubscribed. Results from this analysis are presented in Table A.6.

As compared to the *Vanilla* reminder, all reminders, except *Altruism*, had a positive and significant effect on the probability of opening the reminder. The email appealing to an altruistic motive, however, was opened considerably fewer times. Moreover, fewer households clicked on the diagnostic link after opening the email if it belonged to the said category. Surprisingly, the *Moral Cost* reminder resonated positively, with greater participation in the audit as compared to households who received the *Vanilla* reminder. Finally, receiving a reminder with a monetary incentive was the only intervention which reduced the probability of unsubscribing from future emails.

B Welfare Calculations

B.1 Parameters

The different parameters used in the welfare calculations are specified in Table B.1, along with their units and sources.

Table A.6: ATE Estimates of Letters on Interaction with Reminders

	Opened Reminder (1)	Clicked Reminder (2)	Email Unsubscribed (3)
Simplified	0.060^{***}	0.009	-0.001
	(0.021)	(0.009)	(0.004)
Altruism	-0.059^{***}	-0.039^{***}	-0.002
	(0.021)	(0.006)	(0.004)
Incentives 10	0.056^{**}	0.017	-0.008^{**}
	(0.026)	(0.011)	(0.003)
Incentives 15	0.072^{***}	0.011	-0.002
	(0.027)	(0.011)	(0.005)
Moral Cost	0.068^{***}	0.028^{***}	-0.000
	(0.021)	(0.010)	(0.004)
Intercept	0.432^{***}	0.028^{***}	0.017^{***}
	(0.019)	(0.008)	(0.004)
Controls	✓	✓	✓
Observations	5,563	5,563	5,563

Notes: Robust standard errors are in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.
All regressions report the average treatment effect estimates of different behavioral interventions on how customers interacted with the reminders. Dependent variables, all dummy variables, are presented as column names. *Opened Reminder* refers to if the household clicked the email and were shown its content. *Clicked Reminder* means that the household clicked the link to the audit tool within the reminder. *Email Unsubscribed* refers to a situation where the household unsubscribed from receiving any further reminder emails from NWL. The reference group in each model is the Vanilla group. The data for each regression includes only households who had not completed the diagnostic by 06-Feb-2019, and had received an email reminder. All models include the dummy variables *Meter* and *Rural* as controls. The former equals 1 if the household has a water meter attached to it, and the latter equals 1 if the household is located in a rural area.

Table B.1: Parameters and Sources

Variable	Unit (1)	Value (2)	Source (3)	Notes (4)
Consumer Price of Water	£/m^3	1.3	NWL Charges Scheme (2020)	
Short-Run Marginal Cost	£/m^3	0.44	NWL Financial Statements (2021)	
Long-Run Marginal Cost	£/m^3	0.98	NWL Financial Statements (2009, 2021)	Marginal operating cost of £0.44/m^3 (SRMC) and marginal capacity cost of £0.54/m^3 from Abberton Reservoir Water Resource Scheme
Emissions/Ml of Water Supply	kgCO2e/MI	140	NWL Financial Statements (2021)	
Emissions/Ml of Sewage Treated	kgCO2e/MI	520	NWL Financial Statements (2021)	Includes Scope 3 (other indirect) GHG emissions. For details, see HM Government (2019)
Emissions/MI from Household Water Use	kgCO2e/MI	6,200	Emma et al. (2008)	
Social Cost of CO2	$/tCO2e	51	Interagency Working Group, USG (2021)	Assumes a 3 percent discount rate
Time to Complete Diagnostic	Minutes	7	Field Experiment	Mean of all households
Cost of Posting a Letter	£/letter	0.41	Royal Mail (2021)	Standard tariff in 2020-21 for orders containing less than 2,500 items
Corporate Tax Rate	percentage	0.19	NWL Financial Statements (2021)	
UK Median Wage	£/hour	14	ASHE (2021)	
Weight on Leisure Time	percentage	0.5	White (2016)	
$ to £ conversion		0.78	*Exchange Rates UK* (2020)	Average 2020 rate

Notes: 1MI equals 10^6 liters. For emissions from wastewater treatment, NWL provides two measures of GHG intensity ratios: flow to full treatment, and water distribution input. Based on the Environmental Reporting Guidelines published by HM Government (2019), the latter intensity ratio takes into account *Scope 3* emissions, which are defined as emissions that are a consequence of the utility's actions, which occur at sources which the utility does not own or control and which are not classed as Scope 2 emissions.

B.2 Cost Effectiveness Calculations

Table 7 in the main text provides a comparison of the cost effectiveness with other studies in the literature. Calculations related to the comparison with Ansink et al. (2021) are presented below in Table B.2. The cost effectiveness calculations in their paper do not lend themselves easily to comparison with our numbers, and therefore, we provide a summary of our calculations below. *Panel A* shows the total water savings from the information and technology arm for all the months in the one year following the treatment. In other words, it provides a measure of the effectiveness against which costs need to be compared. *Panel B* shows the calculations related to total costs. Subsequently, we divide the costs in *Panel B* by the effectiveness in *Panel A* to arrive at the CE.

For studies other than Ansink et al. (2021), there was a cost effectiveness number specified, but in dollars ($) per gallon. The same has been converted to dollars ($) per cubic meter and in 2020 dollars (as opposed to dollars in the year of publishing) for comparison. The inflation adjustment used price data from US Bureau of Labor Statistics (2021)[42]. The details of the calculations are presented in Table Table B.3.

C Calculation of Pre- and Post-Treatment Water Consumption

We now provide a detailed description of the computation of consumption data for different households. To help illustrate the format of the data shared by NWL, and our data cleaning process, we use some randomly generated data in Table C.1

The consumption data from NWL consisted of a series of four meter readings for each household. Each meter reading includes the date of the reading, and its corresponding value. Thus, Readdate 1 represents the date of the earliest reading for the household in our data set, while Readdate 4 represents the date of the latest reading. All households for which we did not have at least one reading before and after the treatment date (*i.e.* 08-Dec-2018) were dropped from the sample. Readings for different households were taken at different times, and therefore, Readdate 1 for Unique ID 1 could be very different from Readdate 1 for Unique ID 2. Pre-treatment water consumption was calculated by differencing the two readings immediately prior to the treatment date. In the example, pre-treatment consumption for Unique ID 1 is the difference between Read 2 and Read 1, whereas the pre-treatment consumption for Unique ID 2 is the difference between Read 3 and Read 2. If either of the two readings immediately prior to treatment were taken before 01-Jan-2010, the household was dropped as the date is too far back in time to accurately

[42]Consumer Price Index for All Urban Consumers (US city average series for all items)

Table B.2: Cost Effectiveness in Ansink et al. (2021)

Panel A

Month	Reduction due to Information (liters/day/hh) (1)	Reduction due to Technology (liters/day/hh) (2)	Total Reduction due to Information (cubic meters) (3)	Total Reduction due to 1 Device (cubic meters) (4)
Month 1	-46	-6	-13,011	-1,403
Month 2	-42	-6	-11,977	-1,407
Month 3	-39	-6	-11,171	-1,341
Month 4	-38	-4	-10,813	-1,053
Month 5	-31	-6	-8,702	-1,345
Month 6	-28	-5	-7,958	-1,240
Month 7	-26	-5	-7,286	-1,276
Month 8	-22	-6	-6,232	-1,387
Month 9	-19	-7	-5,523	-1,570
Month 10	-17	-7	-4,765	-1,564
Month 11	-15	-7	-4,205	-1,713
Month 12	-14	-7	-3,928	-1,717
A: Total Water Conserved in 1 Year (m^3)			**-95,570**	**-17,016**

Panel B

Variable	Unit (1)	Information Component (2)	Technology Component (3)
Cost	£/hh (column 2) £/device (column 3)	30	13.5
B: Total Cost	£	284,880	107,685
Cost Effectiveness	**£/m^3**	**3.0**	**6.3**
Cost Effectiveness	**\$/m^3**	**3.8**	**8.1**

Notes: Total number of households in the study were 9,496. For calculating the reduction due to 1 device, the percentage of h/h's with no water saving devices (16 percent) were removed from the sample. Reductions due to information and technology component are sourced from Appendix Table A of Ansink et al. (2021). Total water conserved in 1 year is the sum of water reductions across all the 12 months. Cost of information component calculated as the product of time taken per audit (1.5 hours) and average hourly labor cost of £20/hour (as assumed by the authors). Cost of technology component includes cost of one device (£9 per device) plus delivery costs per household (£4.5). Total Cost calculated as per household cost multiplied by total number of households. Total number of households in the case of technology component adjusted for percentage of households with no water saving devices. For conversion rate from £ to $, see Table B.1 for parameters

Table B.3: Cost Effectiveness Calculations for Other Studies

Paper	Population / Bound (1)	$ per 1000 gallons reduced (Year of Paper) (2)	$ per 1000 gallons reduced (2020) (3)	Cost effectiveness ($/m3) (4)
Bennear et al. (2013)	Lower Bound	7.33	8.3	2.2
	Upper Bound	26	29	7.6
Ferraro & Miranda (2013)	All Households	0.37	0.41	0.11
	High-Use Households	0.20	0.22	0.06
Ferraro & Price (2013)	All Households	0.58	0.65	0.17
	High-Use Households	0.42	0.47	0.12
Bernedo et al. (2014)	All Households	0.24	0.26	0.07
Brent et al. (2015)	Lower Bound	1.7	1.9	0.50
	Upper Bound	2.6	2.9	0.75

Notes: 1000 gallons equals 4.5 cubic meters. For all studies, the cost effectiveness was converted to 2020 values based on the cumulative inflation rate between the year the study was published and 2020. The inflation adjustment used price data from US Bureau of Labor Statistics (2021). High-use households in Ferraro & Miranda (2013) refer to households who both have above median consumption and own their homes. High-use households in Ferraro & Price (2013) refer to households who have above median consumption.

Table C.1: Format of Consumption Data

Unique ID	Readdate 1	Read 1	Readdate 2	Read 2	Readdate 3	Read 3	Readdate 4	Read 4
1	2017-02-21	7438	2018-02-23	7585	2018-12-24	7864	2019-04-20	7986
2	2016-11-03	1184	2017-07-27	1379	2018-07-19	1674	2019-01-14	1803

measure consumption in the present period.

Post treatment water consumption was the difference between the two most recent readings. Most of the households only had a single reading post treatment, and therefore, post consumption in that case would be the difference between the reading post treatment and the reading immediately prior to the treatment. For example, post consumption for both Unique ID 1 and Unique ID 2 would be the difference between Read 3 and Read 4, but Readdate 3 in case of Unique ID 2 was prior to the treatment date.

The difference between any two readings gives us the water consumption in cubic meters during the time interval obtained by differencing the two corresponding reading dates. To standardize this measure across all households, the difference between any two readings was divided by the number of days between the respective readings to obtain average water consumption in cubic meters per day. Finally, this measure was multiplied by a 1000 to obtain water consumption in liters per day.

D Sample Letters

The letters and the reminder emails sent to the different treatment groups by NWL to their customers are presented below.

Figure D.1: Vanilla (Status Quo) Mailer

Back page / font page

Inside the mailer

Figure D.2: Simplified Mailer

Back page / font page

Inside the mailer

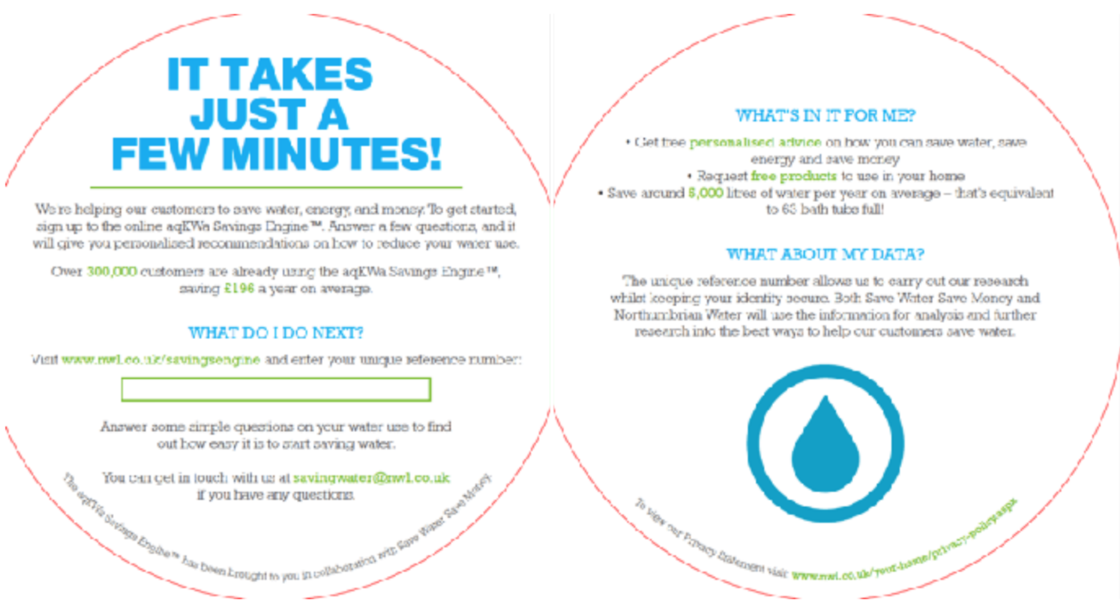

Figure D.3: Altruism Mailer

Back page / font page

Inside the mailer

Figure D.4: £10 Incentive Mailer

Back page / font page

Inside the mailer

Figure D.5: £15 Incentive Mailer

Figure D.6: Moral Cost Mailer

Back page / font page

Inside the mailer

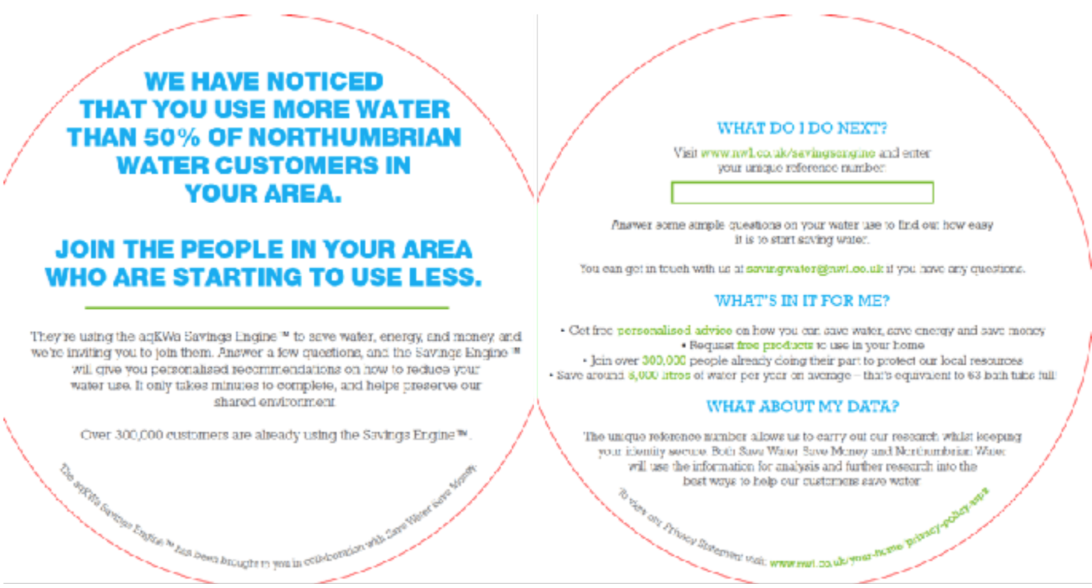

Figure D.7: Reminder Email

January 2019

View Online / Unsubscribe

NORTHUMBRIAN WATER *living water*

Dear [First Name],

HELP KEEP YOUR LOCAL WATER SUPPLY HEALTHY!

We're inviting you to try the aqKWa Savings Engine^{TM} as a part of our campaign to save water and protect the local environment.

Over 300,000 customers are already using the aqKWa Savings Engine^{TM}, saving tens of thousands of litres of water a year. By joining up and doing your bit, you can help keep our water supply healthy and promote a sustainable future.

The Savings Engine^{TM} gives you personalised recommendations on how to reduce your water use.

All you have to do is answer a few questions.

What do I do next?

Visit **www.nwl.co.uk/savingsengine** and enter your unique reference number

SAMPLECODE

Answer some simple questions on your water use to find out how easy it is to start saving water.

You can get in touch with us at **savingwater@nwl.co.uk** if you have any questions.

What's in it for me?

- Get **free personalised advice** on how you can save water, save energy and save money
- Request **free products** to use in your home
- Save around **5,000 litres** of water per year on average – that's equivalent of 63 bath tubs full!

What about my data?

The unique reference number allows us to carry out our research whilst keeping your identity safe and secure. Both Save Water Save Money and Northumbrian Water will use the information for analysis and further research into the best ways to help our customers save water.

Sincerely,

[Signature goes here]

[Name of NWL representative]
[Position of NWL representative]

Northumbrian Water, Customer Centre, PO Box 300, Durham, DH1 9WQ

Northumbrian Water Limited, a company registered in England and Wales with registration number 2366703 whose registered office address is Northumbria House, Abbey Road, Pity Me, Durham DH1 5FJ.

t: 0345 733 5566 | w: www.nwl.co.uk | Privacy Policy