Do Water Audits Work? *

Jesper Akesson
The Behavioralist

Robert W. Hahn
University of Oxford & Carnegie Mellon University

Rajat Kochhar
University of Chicago

Robert D. Metcalfe
Columbia University & NBER

April 25, 2025

Abstract

Water suppliers are showing greater interest in using different mechanisms to promote conservation. One such mechanism is conducting home water audits, which involves assessing water use and providing tailored suggestions for conserving water for residential customers. Yet, very little is known about the economic impacts of these water audits. This paper helps fill this gap by implementing a natural field experiment in the United Kingdom. The experiment randomly allocates 45,000 water customers to a control group or to treatment groups that receive different behavioral encouragements to take-up an online water audit. Our analysis yields three main findings. First, encouraging subjects to participate in an audit with financial incentives reduces household consumption by about 17 percent over two months. Furthermore, we find that the size of the financial incentive used to encourage conservation matters for take-up, but not conservation. Second, although there are substantial improvements in water conservation for some interventions, they do not appear to yield net benefits of more than £1 per person under various sensitivity analyses. We also implement a marginal value of public funds approach that considers benefits and costs and reach a similar conclusion. Third, we find that targeting high users could double the effectiveness of the financial incentive interventions.

*Acknowledgements: We thank Tom Andrewartha, Jonathan Baker, David Black, Daniel Brent, Jesse Buchsbaum, David Byrne, Keith Colquhoun, Jonathan Dennis, Dustin Garrick, Ludovica Gazzè, Robin Gilthorpe, Michael Greenstone, Karen Guz, Nathan Hendren, John List, Sethu Odayappan, Sheila Olmstead, Alex Plant, Louise Richardson, Mar Reguant, Matthias Rodemeier, Cathryn Ross, Ruth Scott, Ruozi Song, Dale Whittington, Casey Wichman, Crawford Winton, and Benjamin Wisniewski for helpful comments and guidance. Jesper Akesson received financial support from NWG 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 conservation has the potential to help address water shortages, which are projected to be severe in many parts of the world and could affect billions of people by 2050.[1] One increasingly common demand-side mechanism to help promote conservation is a residential water audit, which identifies behavioral and technological changes that could be made in the home, and provides tailored recommendations for conserving water. The U.S. Environmental Protection Agency considers water audits to be a an important first step in identifying and quantifying water uses and losses (Environmental Protection Agency, USA, 2013). And a large number of water suppliers have begun to promote and encourage audits as a means to conserve water (Sturm et al., 2015; Rupiper et al., 2022). However, very little is known about the efficiency and cost-effectiveness of audits, particularly at the household level.

This paper helps fill this gap by implementing a natural field experiment that encourages residential customers to take an online 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. We randomize the type of encouragement that customers receive using financial incentives, environmental appeals, moral suasion, and social comparisons. Our experimental design allows us to provide a short-run estimate of how different encouragements affect take up, how online water audits affect consumption, and the welfare implications of such interventions.

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 home water audits, online or in-person, on consumption. Second, most previous research on water conservation studies the effect of non-pecuniary interventions, such as moral suasion and social comparisons (Nauges and Whittington, 2019). We introduce several treatments that include both financial incentives and non-financial incentives, which allows us to compare the effectiveness of different kinds of treatments. 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; a second is a less traditional approach based on the marginal value of public funds (*MVPF*) (Hendren and Sprung-Keyser, 2020; Hahn et al., 2024).[2]

The experimental design involves sending letters to residential customers in the United Kingdom that encourage them to take a *do-it-yourself* online water audit. This self-audit consists of logging into the company's online water audit tool, answering questions on water use habits and

[1] He et al. (2021) suggest that the global urban population facing water scarcity is projected to increase from 0.93 billion people in 2016 to between 1.7 and 2.4 billion people in 2050. Furthermore, the number of large cities in all countries (population $>$ 1 million) facing water shortages is also projected to increase from 193 (37 percent) to 292 (56 percent) by 2050.

[2] The benefit-cost analysis in Section 4.2 considers overall benefits and costs of the experiment. The *MVPF* approach in Section 4.3 considers the impact of spending a marginal pound on after-tax benefits to producers and consumers, and compares this with the net cost to the entity providing the subsidy, which is typically the government.

home features, and receiving recommendations for reducing consumption. The online tool provided information on free water-saving devices offered by the utility, and 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 (*Vanilla*) received an encouragement letter that was in use by the water utility, Northumbrian Water Group (NWG), 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 (*Simplified*) received a simplified version of the letter sent to treatment group 1, which made the call to action more salient. The third treatment group (*Altruism*) received letters reminding them to save water in order to protect their local environment, while treatment group 4 (*Moral Cost*) was sent a letter comparing the household's consumption to that of their neighbors (*i.e.*, moral suasion). Treatment groups 5 and 6 (*Incentives*) received letters that provided different levels of monetary incentives (£10 and £15) to encourage completion of the home water audit.

We have three main results about the short-run effectiveness of water audits related to take-up, overall welfare, and the benefits of targeting. First, the interventions affect both the take-up of the audit and subsequent consumption. Relative to the *Vanilla* letter, all letters led to a significant increase in the take-up of the diagnostic for about two months, with the *Incentives* treatment having the maximum impact.[3] Specifically, the increase in the rate of take-up for households exposed to the *Incentives £10* treatment relative to the *Vanilla* group was 4.5 percentage points. That increase was 5.7 percentage points for households in the *Incentives £15* group. Because the impact of the two *Incentives* treatments are statistically different from each other, we calculate a price elasticity of audit demand to be 0.53. Thus, increasing the amount of the financial incentive could be a fruitful strategy to increase participation.

Next, we estimate the causal impact of audits on water consumption for metered households using an encouragement design with two-stage least squares. We require an instrumental variables (IV) design to estimate this impact because there is potential for self selection among the compliers, *i.e.*, households that complete the audit when receiving the encouragement letters. For the first stage, we estimate the impact of the randomized encouragement on the take-up of the audit. Though we use different combinations of treatment assignment as IVs, we focus mainly on the *Incentives* treatment which we believe is most likely to satisfy the exclusion restriction. This is because the *Incentives* treatment did not include an environmental or altruistic message and, therefore, could affect water consumption only through 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). Our analysis suggests that there is a reduction in consumption of about 17 to 18 percent (43 to 45 liters per day).[4] Limiting our focus to the

[3] We use the terms *audit* and *diagnostic* interchangeably.

[4] Such effect sizes are not different than those from the non-experimental literature on household water programs,

financial incentives intervention, we estimate the £15 treatment reduces consumption for metered households by 44 liters per day, while the £10 treatment 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 also consider external validity by examining how our results could generalize to customers who currently do not have metered water consumption, assuming that they were metered. Weighting each household by the inverse probability of being metered revises our estimates on water conservation from 16 to 14 percent of pre-treatment consumption, though the results still remain significant with our preferred specification. These effects persist for at least two months post-treatment.

As a complement to our results on conservation, we conducted a survey that examines possible factors or mechanisms that could help explain the increase in conservation. We find suggestive evidence that the conservation resulted from both behavioral changes (*e.g.*, shorter showers, detecting leaks, and turning off taps) and the installation of water-saving technologies.

The second main result relates to the overall welfare impacts of the intervention, which we find are close to zero. Net benefits per person typically range between -£1 to £1. Even though there are substantial improvements in water conservation for some interventions, they do not appear to yield net benefits of more than £1 per person unless we assume the conservation effects persist for at least a year and producer profits are not reduced as a result of conservation.[5]

Because our analysis quantifies greenhouse gas emission benefits associated with water conservation, and does not quantify other potentially important benefits, we estimate the monetary value of other benefits needed for net benefits to be zero.[6] We begin with an estimate of the social cost of carbon (SCC) of £241 per ton (United Kingdom Government, 2021).[7] Using this estimate, we find that a cubic meter of water conservation would need to yield other benefits of about £4.6, or four times the marginal price of water, for benefits to just equal costs.[8] In addition, we calculate that the social cost of carbon would need to be 2.8 times higher, or £910 per ton, for benefits to just equal costs.[9] Additional sensitivity analyses on the magnitude of water conservation needed to pass a benefit-cost test reveal that the impact needs to be three to seven times higher under

such as the pre-post analysis of Manouseli et al. (2019).

[5]We use intention-to-treat estimates for all our welfare calculations, and not the local average treatment effects.

[6]These emissions are associated with abstracting, pumping, treating and heating water, and treating and pumping waste water. Among these, heating water in the home is the most significant contributor (89 percent) to carbon emissions in the total water system (Reffold et al., 2008). Breakdown of emission estimates by component in the water supply-use-disposal system is presented in Table E.1.

[7]The SCC is a relevant metric for this exercise given water supply, use and disposal has a large carbon footprint.

[8]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).

[9]The official US government estimate for the SCC is $51 per ton (Interagency Working Group, US Government, 2021). The benefit-cost calculations look substantially worse if we apply this estimate.

different scenarios for the intervention to yield positive net benefits.

We do several other sensitivity analyses, including changing the social cost of carbon, assuming producer surplus is zero instead of negative, varying the long-run marginal cost, extending the time horizon for which conservation benefits accrue, and assuming the price of water better reflects the opportunity cost of scarce water. These sensitivity analyses support the finding that the net benefits are likely to be relatively small in per capita terms.

As an alternative to our benefit-cost framework, we also apply a marginal value of public funds approach (Finkelstein and Hendren, 2020; Hendren and Sprung-Keyser, 2020; Hahn et al., 2024). 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. While it is convenient to assume a lump sum tax will be used for analytical simplicity, it is not necessary.[10] We calculate the *MVPF* using both short-run and long-run marginal costs for the utility. Our results suggest that the *MVPF* in the base case is 0.18 using the short-run marginal cost, and increases to 0.26 using the long-run marginal cost.[11] The *MVPF* values under the different cost assumptions imply that the government would be spending $1 to generate less than 30 cents of benefits. However, assuming that the water conservation benefits last for a longer time period (*i.e.,* a year), and the utility breaks even, we find that the *MVPF* increases to 1.7. This implies that using the long-run case is much better in that $1 of net costs to the government generates more than a dollar in welfare benefits.

The third main result relates to targeting. Our aim was to explore whether targeting of the intervention could substantially improve its effectiveness (Allcott, 2011; Ayres et al., 2013; Ferraro and Miranda, 2013; Brent et al., 2015; Wichman et al., 2016; Knittel and Stolper, 2019; Brent et al., 2020; Gerarden and Yang, 2021; Baker, 2021). We consider the targeting of high users, who are defined as users with pre-treatment consumption higher than the median consumption. We find that targeting of high users that receive financial incentives roughly doubles the reduction in consumption (89 liters per day versus 44 liters per day) over the short run. This suggests that audits can be targeted to improve their efficiency.

Taking this analysis a step further, we ask whether targeting could pass a benefit-cost test. Similar to our analysis above, we find that targeting is not sufficient for benefits to exceed costs in the short-run, but can help in the long-run. Though targeting helps to improve cost-effectiveness by 47 percent (£3.4 per cubic meter versus £6.5 per cubic meter without targeting), we estimate that a cubic meter of water conservation would need to yield other benefits or reduced investment costs of at least £1.7 for the intervention to pass a benefit-cost test.[12]

The basic intuition behind our results can be explained simply. The short-run reductions in

[10] Recent work on audits in the energy area, discussed below, uses the assumption of lump sum transfers.

[11] Assuming a value of SCC equals the US government estimate of $51/ton, the MVPF values turn negative. Negative *MVPF* values mean that the government is spending resources to generate negative net benefits as measured by the sum of willingness to pay.

[12] The social cost of carbon in the base case would need to be 2 times higher (as compared to 3.8 times before) for benefits to just equal costs in the short run.

greenhouse gas emissions from water conservation are comparatively small, on the order of 1.6 tons for 65 days. And while the experimental cost per person is also relatively small, on the order of $1.7 per consumer (not including the producer surplus loss), this leads to a cost effectiveness (CE) of £950 per ton, which is much higher than most estimates for the SCC (Interagency Working Group, US Government, 2021; Rennert et al., 2022). If we assume our results persist for a year and utilities break even, the cost-effectiveness calculus looks more attractive (£150 per ton) because the benefits from conservation increase.

Our analysis builds on economic literature in both the water and energy sectors. In addition, it builds more generally on the literature on behavioral nudges, particularly related to social norm messaging. We consider these literatures in turn.

There are relatively few rigorous estimates of the economic impacts of audits on water use. To the best of our knowledge, Ansink et al. (2021) provide the only cost effectiveness assessment of water audits. They do not identify any natural field experiments that address online audits. Their research suggests that technology is more cost-effective than information provision by a factor of two for a water audit program in the United Kingdom. Our study differs from theirs in that we have experimental variation into the audit program and we focus on all households in a geographical area—not just above average water users.

There are several measures of the cost-effectiveness impacts of water conservation related to other interventions. These include studies on the impacts of metering, social norm messaging, subsidies for replacing garden landscapes, and the nature of the regulatory intervention. Ferraro and 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 and 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. Baker (2021) estimates that the cost-effectiveness of the *Cash-for-Grass* rebate program ranged from $0.8 to $1.0 per cubic meter. All these values are substantially lower than the estimates for our experiment, which range from $1.3 to $8.4 per cubic meter reduced.[13, 14]

Several interventions aimed at promoting water conservation have resulted in substantial reductions in water use with some having effect sizes comparable to those we find. Browne et al. (2021) disentangle the effect of different residential water conservation policies adopted by a utility

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

[14] One area that we do not address is spillovers that may occur due to water conservation, for example, in terms of energy use. This could, in many cases, increase the attractiveness of the interventions we study. Goetz et al. (2022) find that a hot water saving intervention targeted at households in Switzerland had persistent spillover effects on room heating energy consumption, as well as cold water consumption for dishwasher use and toilet flushing. Jessoe et al. (2021b) 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 per household, an increase of 39 percent. It is, however, not always the case that such energy spillovers are positive, *e.g.* Baker (2021) finds that the *Cash-for-Grass* program increased household energy use by 3 percent.

during the 2011-2017 California drought. They find large effect of rate changes (elasticity between .22 and .41)[15] and outdoor water schedule regulations (water use decreased by 21 to 24 percent). These findings are similar in magnitude to our result that participation in audits leads to a 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. Baker (2021) studies the impact on water usage of the *Cash-for-Grass* program, a water conservation effort in the Las Vegas area that subsidized conversions of lawn to desert landscape. The author finds that this residential outdoor water conservation program had a sizeable impact, reducing monthly average water usage by 19 to 21 percent. Thus, changes in price, enforcement policies, and subsidies lead to effect sizes in the same range as our results. The one exception is Browne et al. (2023), who implement a field experiment in California, randomly assigning visual or automated enforcement methods to detect water-use violations. Their effect sizes are relatively small, with automated enforcement decreasing water consumption by about 3 percent.

The research literature on energy conservation is substantial, and we make no attempt to provide a comprehensive review. Insightful examples include Allcott and Greenstone (2017) and Fowlie et al. (2018), who study the welfare impact of audits. Their results are similar 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 costs with the present discounted value of energy savings indicates the programs has an internal rate of return of -4.1 percent, while a revealed preference model finds that the programs reduce welfare by $0.18 per dollar of subsidy. Our finding of a negative *MVPF*, discussed in the welfare results section has a similar implication. The costs to the government of the intervention are higher than the social 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 implies that the costs outweigh the benefits.

Our study also relates more generally to the literature on behavioral nudges, particular involving social norms. There have been several experiments and quasi-experiments examining the implementation of social norm messaging (Ferraro and Price, 2013; Brent et al., 2015; Datta et al., 2015; Jaime Torres and Carlsson, 2016; Gillingham and Tsvetanov, 2018; Jessoe et al., 2021a; Brent and Wichman, 2022), peer effects (Bollinger et al., 2020), increased billing frequency (Wichman,

[15]The elasticity refers to the absolute value here.

2017), and other nudges (Tiefenbeck et al., 2018; Byrne and Goette, 2022). Nauges and Whittington (2019) provide a review of the literature on the impact of information treatment on water and energy use.[16] 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 and persistent reductions typically observed when the intervention includes social norm comparisons as opposed to interventions providing technical advice or raising awareness.[17] Our paper integrates social norm messaging with online audits in the (*Moral Cost* letter), allowing us to study the effect on diagnostic completion. We find that though the *Moral Cost* letter has a significant effect on take-up of the audit, the effect of the letter and the audit on consumption is relatively small — a 1.2 percent decline in consumption relative to pre-treatment consumption. This is somewhat lower than the impact of several interventions that use only social norms in related contexts.

The paper proceeds as follows. Section 2 provides details on the audit program and randomized trial. In Section 3, we describe our empirical strategy and present the results from the experiment. Possible mechanisms that could explain the results on water conservation are also explored in this section. 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

The United Kingdom (UK) is expected to face significant water scarcity challenges in the coming decades due to climate change and rising population. Environmental Agency, UK (2021) expects that climate change will result in hotter, drier summers, and less predictable rainfall, which could lead to increased drought risk and possible water shortages in the UK.[18] By 2050, the Environmental Agency, UK (2022) expects the gap between water availability and needs to reach 4 billion liters per day in England.

In response to these challenges, the UK Water Services Regulation Authority (Ofwat) proposed a three part approach to increase water resilience and decrease greenhouse gas emissions from the water and sewerage sector (Ofwat, 2022).[19] First, it set leakage reduction targets, with utilities

[16] Nauges and Whittington (2019) use illustrative calculations to argue that social norm messaging instruments may not pass a benefit-cost test, especially in low- and middle-income countries. Our results suggest that the same could hold true for high-income countries for certain kinds of behavioral interventions, such as audits. In contrast, Mansur and Olmstead (2012) suggest there could be potential welfare gains of switching from non-market to market-based regulation of water supply during periods of drought.

[17] See, *e.g.*, Allcott (2011); Ferraro et al. (2011); Ito et al. (2018); Brandon et al. (2019).

[18] The UK is expected to experience a fall in summer rainfall by approximately 15 percent by the 2050s, and by up to 22 percent by the 2080s (Environmental Agency, UK, 2021). This prediction is supported by the UK Centre for Ecology & Hydrology, whose forecasts for river flows and groundwater levels up to 2080 suggest a worsening water scarcity scenario (Hannaford et al., 2023).

[19] The operational activities of the water sector contribute about 1 percent of the UK's Greenhouse Gas (GHG) emissions (Ofwat, 2008). Adding hot water use in households for washing, bathing and cooking increases this estimate to 5 percent of total UK GHG emissions (Department for Environment, Food & Rural Affairs, 2008).

tasked with cutting leakage by 16 percent in the five years to 2025.[20] Second, Ofwat proposed increasing supply through schemes to recycle and reuse water, and developing new reservoirs. Finally, it encouraged water companies to help consumers reduce their water usage. Reducing demand was recognized as a big part of the solution in the short-run. Beginning in 2009, utilities were asked to produce water resource management plans with specific details on demand management ambition and outcomes (Ofwat, 2008). Water companies were expected to help customers reduce their usage, and consequently their greenhouse gas emissions, by bringing technological and behavioral changes.[21]

It is in this context that in 2018, Northumbrian Water Group commissioned Save Water Save Money (SWSM), a distributor of water efficiency products, to provide its online water audit tool for Northumbrian's customers.[22] The tool, hosted on SWSM's 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 NWG offers, and helped them book an in-home water audit if appropriate. The questionnaire on the platform took approximately ten minutes to complete.

NWG was interested in getting customers to take their online water audit, and understanding the impact of the audits on consumption. We were interested in helping NWG 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 economics and behavioral science. We used one of NWG's existing direct mailers as a template, and designed 5 new direct mailers (see the templates in Appendix H). The only difference between the five communications was the application of different behavioral science ideas.

We implemented a natural field experiment (Harrison and List, 2004) to test the effectiveness of the redesigned letters, and to understand how the SWSM platform influences water consumption. This field experiment included 44,757 NWG customers, spread across three post code areas.[23] 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 NWG had email contact details were also randomly allocated to groups that either received or did

[20] Water utilities have a legal obligation to address leakage, with under-performance penalized through environmental fines (Ofwat, 2018).

[21] With the 2008 Climate Change Act, the UK became the first country to have a legally binding long-term framework to cut emissions. In light of this, Ofwat expected water utilities to play a key part in cutting emissions and mitigating climate change (Ofwat, 2010). Utilities are required to report their annual operational emissions, and must adhere to performance commitments to achieve emission reduction targets (Ofwat, 2023). This is why our welfare analysis of the intervention in Section 4 considers GHG reductions caused by a fall in household water consumption as the chief component of the benefits derived.

[22] The water audit can be accessed at this url: https://www.getwaterfit.co.uk/questions/ (last accessed: July 08, 2023)

[23] There are 121 2-letter postcode areas across the UK (*e.g.*, YO for the city of York). These typically have hundreds of thousands of residents.

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.[24] Note that the field experiment included both metered (43 percent) and unmetered customers (57 percent).[25] Unmetered customers do not have a water meter attached to their houses and, therefore, water utilities, including NWG, do not have data on water consumption for such households. Consequently, their water bill is not based on the amount of water used.[26] They were included in the field experiment because even though such households will not save money by conserving water, promoting water conservation among this subgroup could help NWG with its demand management targets. Furthermore, because there is no data on water consumption for unmetered households, we can only study the impact of a treatment on the take-up of an audit for such households, but not the impact of the audit tool on water consumption.

There were six letter treatments. Treatment 1 (*Vanilla*) was NWG's initial letter, and 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 succinct 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^{th} percentile of consumption, whereas for the bottom 50^{th} percentile, 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 incentives were provided in the form of a voucher which could be used in various stores.[27]

The data used to randomize the trial participants and to measure outcomes came from three

[24] We are not aware of other studies of water audits that have estimated the effect of email reminders. However, there are previous field experiments estimating the impact of reminders on behavior (Castleman and Page, 2016; Karlan et al., 2016; Calzolari and Nardotto, 2017; Damgaard and Gravert, 2018; Rodríguez and Saavedra, 2019; Fishbane et al., 2020; Dai et al., 2021; Domurat et al., 2021; Zhang et al., 2023). See Appendix D.2 for the impact of reminders on increasing take-up of audit.

[25] About 50 percent of households in England have a water meter (United Kingdom Government, 2019b).

[26] The bill for unmetered customers consists of two components: a (*i*) fixed charge, which includes billing and customer service costs; and a (*ii*) charge based on the *rateable* value, an estimate of the property's expected yearly rent. The latter estimate is based on the UK's Valuation Office assessment. *Rateable* values were frozen in 1990. Further information is provided at https://www.ofwat.gov.uk/households/your-water-bill/unmetered/ (last accessed on 2024-07-08).

[27] The UK stores where the voucher could be redeemed are listed here: https://www.highstreetvouchers.com/gift/where-to-spend-love2shop-vouchers

anonymized sources: NWG'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.[28]

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 until October 2018. All direct mailers were posted on 8th December 2018, and email reminders were sent on 6th February 2019.

Table A.1 in the Appendix presents summary statistics on the observable characteristics of the households across treatment groups, and shows that the groups were balanced across these variables. We have data on whether the household was in a rural or urban area, whether they had a water meter, whether they provided NWG with an email, and their consumption before the experiment. Using an F-test of joint significance, we find that the differences across different treatment groups are not statistically significant at conventional levels. This suggests that the various treatments are balanced on pre-treatment observable variables. Though ideally we would like to balance on a host of socio-economic variables, such as building and yard characteristics like owner/renter, household income, education, length of occupancy, age of building and occupant, and the existence and size of lawns, NWG does not collect information on these variables. However, Cominola et al. (2023), based on a comprehensive literature review of the determinants of household water consumption, show that there is a strong correlation between these observable characteristics and water consumption. This makes pre-treatment water consumption a reasonable statistic for observable household characteristics.

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 using an intention-to-treat (ITT) analysis. Next, we use a LATE framework to measure the effect of completing the diagnostic on water conservation. Our analysis of consumption is limited to households with meters as these were the only subset of households for which we have water usage data. 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. Subsequently, we provide preliminary evidence on the mechanisms through which the audits enabled water conservation. Finally, we discuss data limitations and results from robustness checks. The effect of the email reminders on completing the audit is presented in Appendix D.2.[29]

[28]CRM is a tool to help manage and analyze customer interactions and data on websites.

[29]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

The total number of households that completed the audit (*compliers*) was 1,287, or 2.9 percent of the total sample.[30] To examine the effects of the behavioral interventions on the share of households that complete the diagnostic, we run the following regression:

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

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 additional average treatment effect (ATE) of the different letters, over and above the impact of the excluded treatment group, 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. 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. We present the results both with and without controls included in the regressions.

Table 1 presents the estimates from this regression equation. The excluded category is the control group in models (1) and (2), the *Vanilla* letter in models (3) and (4), and the *Simplified* letter in models (5) and (6).[31] Our results indicate that relative to the control group, all interventions led to a significant increase in take-up of the audit, with effect sizes ranging from 1.8 percentage points for the *Vanilla* treatment arm to 7.5 percentage points for the *Incentives 15* treatment arm. We also compare the impact of *Vanilla* treatment arms relative to other treatments, because NWG was planning to send the former letter irrespective of our intervention. We find that relative to the *Vanilla* treatment arm, all the letters increased the take-up of the diagnostic significantly, with the *Incentive* treatment arm performing the best. *Simplified* and *Altruism* letters increased take-up by 0.7 and 0.5 percentage points, respectively, relative to the *Vanilla* letter. The *Moral Cost* letter was more effective, with diagnostic completion higher by 1.6 percentage points in comparison to the *Vanilla* letter. However, the *Incentives* treatment resulted in the greatest impact. Within the *Incentives* treatment arm, the higher financial incentive of £15 had a marginally greater impact (5.7 percentage points versus 4.5 percentage points, $p < 0.01$). In percentage terms, this is equivalent to the *Incentives* treatment having a 240 to 300 percent greater impact than the *Vanilla* treatment.

Next, we change our reference group from *Vanilla* to *Simplified*, and remove all observations that received the former treatment from our sample. The effect of the *Altruism* letter becomes insignifi-

[30]The raw data from the field experiment on the number of households that completed the diagnostic, and how that differs across different treatment groups, and 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.

[31]The sample in models (3) and (4) excludes the control group, leading to 44,757 - 7,459 = 37,298 customers. The sample in models (5) and (6) excludes both the control and the *Vanilla* group, leading to a sample of 37,298 - 7,460 = 29,838 customers

Table 1: ATE Estimates of Letters on Diagnostic Completion

	Completed Diagnostic					
	Control		Vanilla		Simplified	
	(1)	(2)	(3)	(4)	(5)	(6)
Vanilla	0.018^{***}	0.018^{***}				
	(0.002)	(0.002)				
Simplified	0.025^{***}	0.025^{***}	0.007^{***}	0.007^{***}		
	(0.002)	(0.002)	(0.002)	(0.002)		
Altruism	0.023^{***}	0.023^{***}	0.005^{**}	0.005^{**}	-0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Incentives £10	0.063^{***}	0.063^{***}	0.045^{***}	0.045^{***}	0.039^{***}	0.038^{***}
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Incentives £15	0.075^{***}	0.075^{***}	0.057^{***}	0.057^{***}	0.050^{***}	0.051^{***}
	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
Moral Cost	0.034^{***}	0.034^{***}	0.016^{***}	0.016^{***}	0.009^{***}	0.009^{***}
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Intercept	0.000^{*}	-0.009^{***}	0.019^{***}	0.008^{***}	0.025^{***}	0.013^{***}
	(0.000)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
Controls	No	Yes	No	Yes	No	Yes
Observations	44,757	44,757	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 (ATE) 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 (3) and (4) exclude the observations in the control group, with the *Vanilla* letter comprising the reference treatment arm. Models (5) and (6) exclude the observations in the control and *Vanilla* groups, with the *Simplified* letter serving as the reference group. Models (2), (4) and (6) 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.

cant, indicating that it did not have a significantly different impact relative to *Simplified*. The impact of the *Incentives* and *Moral Cost* letters continues to be positive and significant, implying that these letters even outperformed the *Simplified* treatment arm in terms of their impact on take-up of the online audit. As before, the *£15 Incentives* treatment had a significantly higher impact than the *£10 Incentives* (5.0 percentage points versus 3.9 percentage points). In other words, the completion rate among households who got the *Incentives* treatment was higher by 160 to 200 percent relative to the *Simplified* treatment. The results do not differ when we control for whether a household is situated in a rural area or has a water meter. We can, therefore, conclude that behavioral interventions can help to promote the use of audit tools, with financial incentives being the most effective.

One question that arises is whether there is heterogeneity in who responds to different treatment arms, *i.e.*, do different treatments induce different people to sign up. We can test for heterogeneity based on three household characteristics: rural (vs. urban), metered (vs. unmetered), and pre-treatment water consumption. Specifically, we test if different treatment arms had varying effects on metered versus unmetered households, rural versus urban households, or high versus low water users.[32] For example, it is possible that the *Altruism* treatment incentivized both high and low users to take up the audit by appealing to an important environmental cause, whereas the *Simplified* treatment only encouraged high users. We provide a brief overview of the results here, and direct the reader to Appendix B.1 for details on the regression framework and tables.[33]

We have three main results with respect to heterogeneous effects on diagnostic completion. First, we find that metered customers always had a higher rate of diagnostic completion than the corresponding unmetered customers in each treatment group. The difference ranged from 2.2 percentage points for the *Vanilla* letter to 5.2 percentage points for *Moral Cost* letters. This result makes intuitive sense as the financial benefit of conserving water would only accrue to customers with meters, giving them a higher incentive to complete an audit.[34] Even then, the rate of diagnostic completion among unmetered customers was substantial, with 33 percent of total customers who completed the audit being unmetered (see Appendix Table A.2). This implies that unmetered customers also respond positively to behavioral interventions even when there may not be direct financial benefits. Therefore, to the extent that diagnostic completion can help with water conservation, unmetered customers could play a role in helping utilities meet their demand management targets. Second, within each treatment group, there were no significant differences in diagnostic completion rates between rural and urban households. Lastly, within both the *Vanilla* and *Simplified* treatment groups, metered high users were 2.2 percentage points more likely to complete the audit as compared to low users. However, the *Incentives*, *Altruism* and *Moral Cost* treatment did not have heterogeneous effects with respect to pre-treatment water consumption, implying both high and low users in those groups were equally likely to complete the audit. This provides suggestive evidence that without a call to an environmental cause or a financial incentive, low users may not be incentivized to conserve energy.

3.2 Effect of 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, *i.e.*, the ITT estimates. 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 (*compliers*). 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

[32] We define *high-use* households as metered consumers who had pre-treatment water consumption greater than the median of the pre-treatment consumption in the sample. The sample for this exercise only consists of metered households because data on pre- and post-treatment consumption was only available for this subset of customers.

[33] See Table B.1 for heterogeneous treatment effects of the interventions based on urban versus rural areas and metered versus unmetered customers. For heterogeneous treatment effects based on pre-treatment water consumption, see Table B.2.

[34] NWG charged unmetered customers based on a *rateable value*, and these customers got their bill once a year. See NWG (2020) for details.

impact on consumption of the direct encouragement and the take-up of the audit.

To estimate the effect of the treatment on consumption, we regress average daily water consumption post the treatment (y_i) on an indicator for whether the household was treated (T_i):

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

where T_i equals 1 if the household i received any treatment letter, and 0 if it was in the control group. Water consumption is measured in liters per day. 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 i. The vector of covariates, $\mathbf{X}_i$, consists of $Rural_i$ and an additional covariate, *Pre-Treatment Water Consumption*$_i$, which measures the average daily water usage of a household before the letters were sent. [35] For all regressions, variables related to water consumption are measured in liters per day.

The water consumption data that we obtained reduce the sample that we can use. First, we only have data on water usage for metered customers (42.9 percent, see Table A.1). This reduced our sample for the ITT analysis from 44,758 to 19,180 households. Second, for each metered household, we were provided with the last four readings of their water meter, including the date on which the reading was taken.[36] This implies we do not have data on monthly consumption, and can only calculate the average daily consumption of the household between two meter readings.[37] Furthermore, time periods between readings were not uniform across the metered households. This lack of uniformity meant that we lost a further 5,779 households for whom we either did not have pre- or post-consumption data because of a lack of readings for the respective time-period. Third, among the remaining households, the meter reading was reset for 1,461 customers (for *e.g.* due to a change of ownership), which left us with a final sample of 11,940 metered customers.[38] Finally, the latest meter reading for a majority of the sample was conducted by NWG in February 2019. As a consequence, we have, on average, 65 days of post-treatment water consumption for each household. Therefore, our results provide an estimate of the short-run impact of the audit.[39]

Though these data limitations reduce our sample, we believe our econometric identification is reasonable for several reasons. First, as shown in Table A.1, all treatment groups were balanced on the proportion of households with a water meter, pre-treatment water consumption among households, and the number of consumers in each treatment group who are in the top 50$^{\text{th}}$ per-

[35] None of the qualitative results presented in this paper are contingent on the inclusion or exclusion of household observable characteristics, suggesting that the randomization was done correctly.

[36] Water companies in the UK usually read each customer's meter twice a year (Ofwat, 2013).

[37] A better scenario would have been to have two readings pre-treatment and two readings post-treatment. This would provide us with two data points on average daily water consumption for each household. We could then have used these two data points to estimate the average impact of the treatment. However, NWG does not read meters for all households on the same date, with visits spread across the entire year; neither does it read the data at fixed intervals.

[38] Details on the computation of pre- and post-treatment water consumption are provided in Appendix F.1.

[39] NWG created a key that was used to match treatment assignment to water consumption. They deleted that key after the project ended in March 2019. Furthermore, they insisted they did not have the capacity to match on past water consumption when we asked them for that.

centile of consumption (*high-use* households). Second, 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 treatment groups. Third, we test for balance among treatment groups separately for the 11,940 metered customers in our consumption sample, results that are presented in Table A.3. Overall, 2 of the 20 differences reported in columns (2)-(5) are significant at the 5% level, and 2 are significant at the 10% level based on independent t-tests — as one would expect under random assignment. We also test the null hypothesis that household characteristics (*e.g.*, rural/urban area and pre-treatment consumption) do not predict participation in any treatment group using an F-Test for joint significance. As column (6) shows, we fail to reject the null for each of the treatment groups. Nevertheless, we control for the characteristics that are imbalanced, namely *rural* and *pre-treatment water consumption* in all our regressions. Finally, to address concerns about possible bias in sample selection due to the focus on metered households, we reweight our LATE estimates in Section 3.4 so that the metered sample matches the demographic composition of the general population of NWG customers.

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, except *Vanilla*, reduced water consumption, though results are statistically significant at the $p < 0.05$ level only for the *Incentives* group. [40] Column (1) provides the average treatment effect of receiving any letter on post-treatment consumption. Though the estimate is negative (-1.3 liters per day), it is not significantly different from 0. 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 average daily 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 (4.7 versus 3.5 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 3.0 liters per day under *Moral Cost* to 6.4 liters per day under *Incentives 15*. In percentage terms, this decrease amounts to between 1.2 percent and 2.5 percent of the average pre-treatment water consumption across all households, a small but economically meaningful impact.[41] 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.

[40] The *Vanilla* treatment group leads to a 1.7 liters per day increase in household water consumption, though the effect is not statistically significant. As a percentage of mean daily pre-treatment consumption (260 liters per day), it represents a 0.66 percent increase in water consumption, which is not economically meaningful.

[41] The relatively small decrease in consumption due to the *Moral Cost* letter (which also combined a social comparison message) stands in contrast to the literature (Ferraro and 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 (liters/day)			
	Control	**Control**	**Vanilla**	**Simplified**
	(1)	**(2)**	**(3)**	**(4)**
Treated	-1.324			
	(1.264)			
Vanilla		1.737		
		(1.609)		
Simplified		-1.484	-3.216^{**}	
		(1.634)	(1.598)	
Altruism		-1.482	-3.217^{**}	-0.032
		(1.624)	(1.587)	(1.613)
Incentives £10		-3.542^{*}	-5.287^{***}	-2.113
		(1.923)	(1.893)	(1.915)
Incentives £15		-4.685^{**}	-6.436^{***}	-3.295^{*}
		(1.919)	(1.889)	(1.909)
Moral Cost		-1.301	-3.035^{*}	0.155
		(1.592)	(1.555)	(1.581)
Intercept	8.798^{***}	8.814^{***}	11.038^{***}	9.761^{***}
	(1.614)	(1.613)	(1.662)	(1.841)
Controls	Yes	Yes	Yes	Yes
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 (ATE) 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 average daily water consumption of a household after the treatment date of December 8, 2018. 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 average daily water consumption of a household, in liters per day, before the treatment date of December 8, 2018.

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. Specifically, customers in the *£15 Incentives* group reduced their consumption by a significant 3.3 liters per day (1.3 percent of the average pre-treatment water consumption) relative to households in the *Simplified* group. In summary, the *Incentives* group experienced a significant reduction in their consumption relative to all comparison groups, while the other treatments had a significant negative impact only relative to the *Vanilla* arm. Again, it is important to note that these estimates represent the overall impact of the treatment letters, and so the numbers are bound to be small as we average across all households, many of which never completed the audit (*non-compliers*). We attempt to disentangle the effect of completing the audit in the following section.

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. This entails calculating the effect of the diagnostic on water consumption for households that completed the audit, *i.e.* the *compliers*. However, a local average treatment effect (LATE) estimated using a simple OLS could be biased because households that completed the audit may have unobservable differences with the *non-compliers*. To address this endogeneity, we employ an Instrumental Variable (IV) strategy using two stage least squares (2SLS).[42]

The first stage involves running the following regression:

$$Diagnostic\ Completion_i = \pi_0 + \sum_j \pi_{1j} Z_{ij} + \gamma \mathbf{X}_i + \upsilon_i \tag{3}$$

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 and combination of instruments vary depending on the specification, and the subscript j refers to the different instruments. π_{1j} 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 $Pre\text{-}Treatment\ Water\ Consumption_i$. γ is a vector of estimates of the impact of the household covariates, and υ_i is the error term. The second stage uses the predicted values from Equation (3), $Diagnostic\ \widehat{Completion}_i$, to run the following regression:

$$y_i = \alpha + \beta Diagnostic\ \widehat{Completion}_i + \eta \mathbf{X}_i + \epsilon_i \tag{4}$$

where y_i represents average daily post-treatment water consumption in liters per day, 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 similar results. As in the previous section, the sample for this exercise includes metered households for which we had both pre- and post-treatment water consumption data (11,940 households). 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 treatments. This model satisfies the *relevance* condition as letters do tend to increase adoption

[42] A comparison of observable characteristics between compliers and non-compliers is presented in Appendix A.7.

of the water audit tool (see Section 3.1). The estimates in column (1) suggest that completing the diagnostic led to a significant fall in consumption of 45 liters per day ($p < 0.01$) or 18 percent of average daily pre-treatment water consumption (255 liters per day, Table A.3).

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 restrict 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.[43] 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 affect water consumption directly. Our results under this specification indicate that completing the diagnostic reduces daily water consumption on average by 43 liters per day (17 percent; $p < 0.01$).

Finally, in column (3) we run a regression that we believe is even more likely to satisfy the *exclusion restriction*. We restrict our sample to *Incentives £10*, *Incentives £15*, and control groups, with the set of instruments now limited to the two *Incentive* treatments. In our opinion, this specification is more robust because the *Simplified* letter incentivized readers to download the audit as the same would help them "save water, energy and money" (see Appendix H for templates). It is plausible that this call to save resources led to subliminal changes in conservation behavior notwithstanding the effect from downloading the audit. Consequently, the presence of the *Simplified* group in the regression would violate the *exclusion restriction*. Model (3) circumvents this problem by only including customers that received the *Incentives* letter (which had no such message) or the control group. Given this, we feel more confident using model (3) as our preferred specification. The effect size is similar, and still significant despite the fall in sample size. Completing the diagnostic led to a average fall in post-treatment consumption by 44 liters per day (17 percent, $p < 0.01$).[44]

Our results suggest that there is a meaningful effect of completing the water audit tool on water consumption, ranging from 17 to 18 percent of pre-treatment consumption. However, the duration of this effect beyond 65 days is not known; nor do we know whether the audit may have stimulated the adoption of new water-saving technologies over time.

[43] 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 Table 1 shows, the letters do tend to increase the likelihood of completing the audit.

[44] We also estimated LATE using a sample of: *i*) only *Incentives 15* and control and; *ii*) *Incentives 10* and control. The respective *Incentives* treatments formed the IV in each regression. Results (not presented) show that the £15 treatment reduced average daily post-treatment consumption by a significant 44 liters per day, while the £10 treatment reduced consumption by a significant 43 liters per day (approximately 17 percent of average daily pre-treatment consumption in both cases). This suggests that the size of the subsidy for completing the audit may not be that important for water conservation, which differs from our result on take-up (see Section 3.1).

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

	Post-Treatment Water Consumption (liters/day)		
	(1)	(2)	(3)
Complete Diagnostic	-45.446^{***}	-43.442^{***}	-43.751^{***}
	(15.335)	(16.691)	(16.699)
Intercept	10.402^{***}	10.276^{***}	10.351^{***}
	(1.525)	(2.122)	(2.476)
Instruments	All Treatment	Incentives+Simplified	Incentives
F-stat in First Stage	39	66	102
Controls	✓	✓	✓
Observations	11,700	5,830	3,904

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 average daily water consumption of a household, in liters per day, post the treatment date of 08-Dec-2018. The data has been trimmed at 1 and 99 percentile of pre-treatment consumption. The regressor of interest is *Complete Diagnostic*, 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 six 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. The sample in model (1) includes all metered households for which we had both pre- and post-consumption data. The sample in model (2) consists of the *Incentives*, *Simplified* and control group, while the sample in model (3) includes only *Incentives* and the control 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 average daily water consumption of a household, in liters per day, before the treatment date of 08-Dec-2018.

Next, we examine whether there were any heterogeneous average treatment effects of the behavioral interventions on average daily post-treatment water consumption. Specifically, we test whether households with consumption in the top half of the distribution (*high-use* households) conserved more in absolute terms after receiving the letter.[45] This is relevant because if audits differ in terms of their impact across high and low use groups, it may be more effective to target a behavioral intervention based on this attribute. Details are provided in Appendix B.2.1, and we present a brief summary here. We find strong evidence that *high-use* households reduced their daily water usage significantly after the treatment, while the effect on *low-use* households is indistinguishable from 0, irrespective of the intervention. Receiving any letter reduced average daily consumption for *high-use* group by a significant 3.7 liters per day. Within treatment groups, the *Incentives 15* and *Incentives 10* had the highest impact for high-use consumers, with water consumption falling by 8.3 and 7.4 liters per day, respectively, in comparison to high users in the control group. This was followed by the *Simplified* and *Altruism* letter, with a significant reduction of 4.4 and 4.0 liters per day, respectively. The *Moral Cost* treatment tends to be significant for the high users only when the comparison group is high users in the *Vanilla* treatment, with post-treatment consumption reduc-

[45]The average daily pre-treatment water consumption among high and low use households was 360 and 140 liters per day, respectively.

ing for high users by 3.9 liters per day. Finally, the *Vanilla* treatment did not have any impact on conservation, and this is consistent across both high- and low-use households.

A related analysis for the LATE, presented in Appendix B.2.2, reveals a similar pattern even with respect to the impact of completing the audit. For high users, completion of the audit led to a large fall in average daily post-treatment water consumption, with effect sizes ranging from a significant 78 to 89 liters per day. This amounts to a 21 to 25 percent reduction compared to pre-treatment consumption for the high users. As before, the impact of completing the audit on low users, though negative, was not statistically different from zero. Thus, average treatment effects mask crucial heterogeneity in terms of which subgroup is driving the results. As our findings show, it is the users with above median consumption that are positively impacted.[46]

3.4 External Validity

The main results of our experiment are for metered households in NWG's service region. In this section we explore how the results could generalize to other NWG consumers who do not currently have meters. To evaluate this, we undertake a reweighting exercise that reduces the estimates of water savings due to diagnostic completion from 43 liters per day (17 percent) to between 36 and 42 liters per day (14-16 percent) depending on the specification.

The reweighting exercise is important because the sample used for estimating the effect sizes of the interventions consists solely of metered households (see discussion in Section 3.2). 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 NWG 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 almost all observable covariates for the metered households are balanced across the treatment groups (Table 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 NWG customers. We reweight the metered sample so that it looks like the general 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. As of December 2017, only 32 percent of NWG customers had a water meter, with the corresponding number for the United Kingdom at 50 percent (United Kingdom Government, 2017, 2019b). Though customers have a right to demand a water meter, they are not required to install one.[47] Water metering in the UK

[46]We also checked whether different interventions encouraged different categories of households (number of appliances, water use, frequency of usage, etc.) to take-up the water-audit tool. Results are presented in Appendix D.1.

[47]There are certain exceptions such as if the area is under severe water stress or the premises are not solely used as a home, among others (United Kingdom Government, 2019b).

has largely depended on whether the property is on a shared supply line, the condition of the pipework, the ability of the utility to find a suitable place to fit a meter, and whether the property accesses a communal water supply (Ofwat, 2013). Other reasons for the low metering statistics include concerns related to impacts on large families on low incomes. However, there are efforts currently underway to require water metering due to increased demand and frequency of droughts and floods (National Infrastructure Commission, 2023).

To implement the reweighting, we conduct the following four steps (similar to Stuart et al. (2011); Hahn and Metcalfe (2021)). 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 NWG: rural-urban classification, availability of email address, and residential post-code.[48] 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.[49] IPTW gives each household 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.

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 (3) (*Incentives £10* and *Incentives £15* group as IV), re-weighting reduces the point estimates of water conservation from 44 liters per day to 36 liters per day (14 per cent), and the coefficients still remain statistically significant. The reweighted LATE estimates in columns (1) and (2) are also very similar to the unweighted LATE in Table 3, with water usage reduced by 42 (16 percent) and 34 (13 percent) liters per day, though the latter effect size is not significant (*p-value*=11.4). 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.[50]

3.5 Mechanisms

Our estimates of the impact of the audit on post-treatment water consumption range between 36 to 44 liters per day, effect sizes that are relatively large compared to the impact of most other interventions studied in the literature.[51] Given the magnitude of these effects, a natural question arises as to the mechanisms through which the audits enabled water conservation. We provide *preliminary* evidence on the channels in this section, using an online follow-up survey which was

[48] We do not have data on household income or the number of household members, so we use the data on post codes as a proxy. Owner/renter status could also affect external validity because presumably those that rent would be less likely to perform the audit. NWG did not have access to this information. However, based on survey data discussed in Section 3.5, we find no evidence that homeowners have a significantly higher rate of diagnostic completion than renters.

[49] See Hahn and Metcalfe (2021) for weighting using *sub-classification*, which is a coarser method than IPTW.

[50] Metered and unmetered households may differ on unobservables, in which case the results may not generalize.

[51] Ansink et al. (2021) is one study that finds comparable water savings to ours.

Table 4: Reweighted LATE Estimates

	Post-Treatment Water Consumption (liters/day)		
	(1)	**(2)**	**(3)**
Complete Diagnostic	-41.634^{**}	-33.854	-36.391^{*}
	(20.607)	(21.454)	(21.572)
Intercept	7.819^{***}	7.125^{***}	7.954^{***}
	(1.790)	(2.405)	(2.783)
Instruments	All Treatment	Incentives+Simplified	Incentives
F-stat in First Stage	39	66	96
Controls	Yes	Yes	Yes
Observations	11,700	5,830	3,904

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 (LATE) 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 average daily water consumption of a household, in liters per day, post the treatment date of 08-Dec-2018. The data has been trimmed at 1 and 99 percentile of pre-treatment consumption. The regressor of interest is *Complete Diagnostic*, 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 six 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. The sample in model (1) includes all metered households for which we had both pre- and post-consumption data. The sample in model (2) consists of the *Incentives*, *Simplified* and control group, while the sample in model (3) includes only *Incentives* and the control 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 average daily water consumption of a household, in liters per day, before the treatment date of 08-Dec-2018.

administered to both control and treated customers for whom NWG had an email address (31 percent of the households, or 13,984 participants; see Table A.1). The survey was conducted in March 2019, four months after the initial direct mailers were sent to households, and asked questions relating to water use habits and customer attitudes towards the mailer and online audit, among other things.[52]

There are two potential mechanisms through which the audit can directly affect water conservation: behavioral changes and adoption of water-saving devices. We, thus, focus our analysis on questions in the survey that can shed light on whether households that completed the audit were different either in terms of their attitude towards water-saving devices or in terms of actions that could potentially reduce water usage.

We analyzed several different questions. To examine behavioral changes, we identify questions that ask households whether they currently take shorter showers; turn off the shower when shampooing and the tap when brushing teeth or shaving; check for dripping taps and leaks; avoid washing dishes under a running tap; and water the garden or yard less. To analyze whether adoption of water saving devices after the audit could have played a role, we examine questions that ask

[52]The follow-up survey contained 22 questions. The list of all the questions is provided in Appendix G.

customers whether they believe that water saving appliances are useful; requested water saving products from NWG; and currently use a water butt, which is a large container for collecting and storing rainwater.

Our analysis draws on the responses of households who completed the survey. The response rate for the follow-up survey was 6.2 percent (861 households). 156 participants of all households that received the survey completed both the survey and the audit.[53] Given the low response rate and the selection into who completes the follow-up survey, the findings we present below should be viewed as suggestive.

Our empirical specification takes the following form:

$$y_i = \alpha + \beta \mathit{Diagnostic\ Completion}_i + \gamma \mathbf{X}_i + \epsilon_i \quad (5)$$

where y_i is a dummy indicating the response to a survey question.[54] X_i is a control vector, consisting of $Rural_i$ and $Meter_i$. The coefficient of interest is β, which would indicate if answers to the survey questions differed depending on whether households completed the diagnostic or not.[55]

We have three results.[56] First, we find evidence that the diagnostic encouraged water conservation. Households that completed the diagnostic were 6.5 percentage points more likely to be currently trying to reduce their water consumption. Furthermore, such customers were also 12 percentage points more likely to believe that they were well informed about ways to save water.

Second, our results indicate that completing the audit did lead to changes in habits, with households more likely to engage in certain water conservation activities. Specifically, customers that completed the online diagnostic were significantly more likely — in varying magnitude — to take shorter showers, turn off showers while shampooing and taps when brushing or shaving, check for leaks, and not wash dishes under a running tap.

Third, there is evidence of significant differences in attitudes towards water saving products. People that completed the diagnostic were (*i*) 6.2 percentage points more likely to believe that water saving appliances are useful; (*ii*) 26 percentage points more likely to have requested or looking forward to order water saving products from NWG, and; (*iii*) 10 percentage points more likely to use a water butt, a fixture recommended by the audit as a means of water conservation.

The sample for the preceding analysis includes all households that completed the survey, regardless of whether they were in the treatment or the control group. Our identification comes from comparing responses of households that completed the diagnostic with responses of households that did not. A more natural comparison group for households that received the letters and completed the diagnostic are customers in the control group, none of whom received the letter or completed the diagnostic. As a robustness check, we reran the analysis using only the survey

[53]Summary statistics on survey completion are provided in Table A.4.

[54]For *e.g.*, on the question of whether a household was taking shorter showers, we would code *Yes* as 1 and *No* as 0.

[55]We considered using an IV strategy similar to the one employed in Section 3.3. However, it would not address the problems associated with selection into completing the survey.

[56]This is based on 22 different regressions, which are available on request.

respondents from the control group as a comparison. Reassuringly, the results are very similar in terms of the direction of the effects, their magnitude, and significance. The only difference arises in three cases (out of the 22 tested), namely whether customers who completed the diagnostic were more likely to (*i*) turn off showers when shampooing; (*ii*) check for dripping taps and turn them off, and; (*iii*) check for leaks and repair them. Even in these three cases, the effect sizes are similar to the initial specification, but not significant, an issue that could be due to power.

Finally, we also investigate differences in survey responses between unmetered and metered households who completed the audit. The large number of completed audits among unmetered customers was surprising as unmetered households are not charged based on their water usage and, therefore, have little incentive to interact with tools that promote water conservation. Our analysis involves estimating the differences in answers to survey questions between metered and unmetered households that completed the audit. Empirically, this involves estimating the interaction term between $Diagnostic\ Completion_i$ and $Meter_i$ in Equation (5) and removing the control vector $\mathbf{X}_i$. The only significant difference we observe relates to attitudes towards pecuniary savings from water conservation, with metered households that completed the diagnostic being 19 percentage points more likely to value money savings on water bills as compared to unmetered households. This suggests that financial savings arising from lower water use were not the primary driver of sign-ups among unmetered households. For every other survey question, there were no significant differences between metered and unmetered customers who completed the audit. This suggests that unmetered customers who completed the audit may have implemented water conservation activities such as taking shorter showers and turning off taps when not in use. However, as noted earlier, we cannot test the impact of diagnostic completion, and subsequent changes in behavior and attitudes, on water consumption for the unmetered households because NWG does not have usage data on this group. The limited survey evidence does suggest that the effect sizes could be similar for metered and unmetered households that took the audit.

In summary, an analysis of the follow-up survey suggests that both behavioral changes and technology adoption were important mechanisms underlying water conservation.[57] More research is needed to understand the potential importance of the mechanisms identified here. However, note that the mechanisms only affect the welfare calculation insofar as the assumed duration of the treatment effects persists, which might differ across behavioral changes and technology adoption (Allcott and Rogers, 2014; Brandon et al., 2017). Since the results from the survey suggest some technology adoption, we perform robustness checks to project the water conservation effects beyond the short-term.

[57] While selection into the survey may explain these results, we did not find significant differences in attitudes towards environment and water conservation between customers who completed the audit and customers who did not. In addition, people who completed the diagnostic were not more or less likely to: value monetary savings on water bills; save water only if it helps them save money; and save water only if rest of their community does so.

3.6 Data Limitations and Robustness Checks

The household level water consumption data provided by NWG has certain limitations, which could potentially bias our results. In this section, we discuss these limitations and consider different robustness checks to address concerns with identification.

3.6.1 Measurement Error Caused by Lack of Clean Pre-Post Delineation of Consumption

As discussed in Appendix F.1 on calculating pre- and post-treatment consumption data, we have four meter readings for each household. The difference between two successive meter readings provides the water consumption in the corresponding time period. However, these meter readings are not cleanly delineated into pre- and post-treatment time periods. For illustrative purposes, consider a household that has a meter reading on 1st November 2018 followed by another meter reading on 28th December 2018. The total consumption between these two time periods, calculated by taking the difference between the meter readings, would include 38 days of pre-treatment consumption (1st November 2018 - 8th December 2018) and 20 days of post-treatment consumption (9th December 2018 - 28th December 2018). Given that the readings used to calculate post-treatment consumption include a median time of four months prior to the treatment date, there are valid concerns about using such a long pre-treatment period to attribute causality to the treatments.

Before we proceed to the robustness checks, it is important to note that the contamination from the pre-treatment period in our measure of post-treatment consumption is not likely a major concern because measurement error here would be biased against finding any effect. For example, if the number of days post-treatment in the meter reading is 10% of the number of days in the pre-treatment period, then we are attributing water consumption from a large number of pre-treatment days, when there should be no effect, to the post-treatment period. This will lead to reduced estimate of the true effect size. The fact that we estimate a significant drop in consumption post-treatment suggests that a cleaner sample would offer stronger results.

Nevertheless, we address this concern by running weighted regressions estimating the ATE and LATE of letters and diagnostic completion on water consumption, respectively. The weights correspond to the fraction of days post-treatment included in the meter readings used to calculate post-treatment consumption, divided by total number of days between the two readings. For example, if the household had a reading on 8th December 2018, then 100 percent of the post period is *treated* and so this household gets a weight of 1. Another household may have a reading on 30th July 2018 followed by another reading on 9th December 2018, and would receive a weight of 1/133, where 133 is the number of days between 30th July and 9th December.

We first provide more transparency regarding how the meter readings used to calculate post-treatment water consumption are distributed across the different months. This provides context on where most of the weights used in the weighted regression might lie. Table F.2 in Appendix F.1 presents this using a 17×5 matrix, with rows representing the month of initial reading used to calculate the post-treatment consumption, while the column names represent the month of the latest reading used for the same calculation. For example, there were 7,871 households for which post-treatment consumption was calculated using the meter reading in August 2018 and February 2019. Our analysis reveals that for the majority of the households, the initial meter reading used for calculating the post-treatment consumption was three to five months prior to the treatment date (July to September 2018), while their final reading was one to three months after the treatment (January to March 2019). The weighted average for the entire sample of the proportion of treated days in post-treatment observations was 37 percent.

Second, we use the weights derived to calculate the weighted ATE of letters on post-treatment water consumption. Details are provided in Appendix C.1, and we present a brief summary here. Reassuringly, the results are very similar to the unweighted ATE regressions presented in Table 2. With reference to the control group, all treatment arms except *Vanilla* experienced a fall in average daily consumption; however, only the *Incentives 15* treatment arm experienced a statistically significant decrease of 5.8 liters per day (as opposed to 4.7 liters in the unweighted regression). When we exclude the control group, and the *Vanilla* letter becomes the omitted category, the drop in consumption is significant across all remaining categories, with the decrease in consumption ranging from 4.1 liters per day under *Moral Cost* to 8.5 liters per day under *Incentives 15*. This is in contrast to the unweighted estimates of 3.0 and 6.4 liters per day for *Moral Cost* and *Incentives 15*, respectively. Finally, when we drop both the control group and the *Vanilla* group, and the *Simplified* letter becomes the reference category, only the £15 financial incentive remains significant with consumption falling by 4.5 liters per day on average. In summary, the results are qualitatively similar to the unweighted regressions in terms of direction, and larger in terms of magnitude. This is expected as weighting reduces the downward bias that could be introduced.

Third, we present the weighted LATE estimates of the impact of diagnostic completion on water consumption in Table C.2. Again, results are robust to weighting. For our preferred specification with the *Incentive* treatments as IV, we find that completing the diagnostic led to an average fall in post-treatment consumption by 45 liters per day (18 percent, $p < 0.05$) as opposed to the unweighted LATE estimate of 44 liters per day. As was the case with the weighted ATE results, the weighted LATE estimates across all specifications are in the same direction as the unweighted LATE results (Table 3), significant, and of a slightly higher magnitude.

As a final robustness check, we run the ATE analysis on the subset of households that had a meter reading just before the letters went out (so those in November and early December).[58] This offers the cleanest sample with a clear pre- and post-treatment period. We consider the subset of houses that had at least two meter readings after 1st November 2018, including necessarily after the treatment date of 8th December 2018. However, this drastically reduces our sample, from 11,700 households to 128 households. The results, presented in Table C.3, are revealing. Though the direction of the results is similar across different treatment arms and specifications, the magnitude of the coefficients increases dramatically. With reference to the control group, the £15 monetary

[58] An ATE analysis on the sample of households with meter readings within a particular time frame requires us to assume that the timing of the meter readings is essentially random.

incentives led to a statistically significant decrease of 51 liters per day, or 19.9 percent. This number is almost an order of magnitude higher compared to the estimate of 4.7 liters per day reduction in consumption for the *Incentives 15* group with the unweighted regression for the entire sample. If we drop the control group and compare the effect sizes to the *Vanilla* letter, the drop in consumption is significant across all remaining categories except *Simplified*, with the decrease in consumption ranging from 42 liters per day (16 percent) under *Incentives 10* to 74 liters per day (29 percent) under Incentives 15.

In summary, all of the robustness checks addressing potential measurement error indicate that the initial estimates of the impact of letters and diagnostic completion on water conservation represent an underestimate. The true impact may be much larger. Because an underestimate would bias our net benefit calculations downward, we perform a wide array of sensitivity analyses for the welfare calculations in Section 4.

3.6.2 Seasonality

Water consumption usually displays strong seasonality, with irrigation during summer months the largest source of household water use.[59] Ignoring this seasonality may bias our results. As discussed previously, the consumption data shared by NWG is not at the monthly level, and so the conventional way of controlling for seasonality by including month fixed effects is not possible. However, we argue that seasonality is not a big source of concern for our results.

Two points are worth noting. First, the regressions estimate the difference in post-treatment consumption between consumers who received and did not receive the letters. As long as there are no differences between treatment and control groups in the impact of seasonality distribution of meter readings across months, our results should not be biased. Given that households were randomized into different groups after balancing them on pre-treatment observables, we do not expect any significant differences in the impact of seasonality across the treatment arms.

Second, we can verify whether the distribution of meter readings across months differed between treatment and control. To test this, we run four different OLS regression where the regressand is a dummy variable that equals 0 if the customer is assigned to the control group, and equals 1 for customers assigned to any treatment group. The independent variables are a dummy for all the months in a year. We run this regression separately for each of the four meter readings to test if the seasonality concern manifests in any of the readings conducted by NWG. If any of the month coefficients across these regressions is significant, it would suggest that meter reading in that particular month predicts participation in the treatment group — which likely biases our results. We find that none of the month coefficients in any of the four regressions are significant, indicating that the month of the meter reading does not predict participation in the treatment groups.[60] Therefore, we are reasonably confident that the treatment group was not being affected by seasonality

[59] This is another reason why the experiment likely identifies a lower bound. Consumption during the summer months is higher due to irrigation needs, and the treatment effects during those months may have potentially been higher.

[60] This is based on 4 different regressions, which are available on request.

differently from the control group.

3.6.3 Spillovers

Spillovers of treatments effects to the control group could also cause the estimated effects to be biased. However, again, the spillover would tend to bias our results downward. To see this, assume that the control group is influenced by their neighbors in the treatment group, and they adopt water saving methods after being inspired from the advice offered through the letter or audit to treated households. That would imply the treatment effects would be biased downward. Potential conservation effects associated with the treatments are thus likely to be higher than we have estimated.[61]

3.6.4 Inclusion of Reminders Treatment Group in Sample

Households who had not completed the audit by 6th February 2019, *i.e.* 60 days after the first treatment on 8th December 2018, were randomly allocated to groups that either received or did not receive an email reminder. Inclusion of this group in our analysis may lead to concerns as to whether we are estimating the singular effect of a first letter treatment or dynamic effects of the audit program including the reminders. However, note that we are only estimating the short-run effect of the intervention, with only 65 days of post-treatment consumption data on average. Therefore, given that the reminders were sent almost at the end of the our study period, our estimated effect of the intervention (letters or audit) on consumption is being primarily driven by households who completed the audit pre-reminders. In fact, for the two monetary interventions that had the largest impact on both take-up and conservation, only 7.0 (£10) and 4.7 (£15) percent of the household who completed the audit were in the reminder treatment group. Furthermore, including the reminders treatment group would further downward bias our results, as their post-treatment time period is minuscule and these households would not have had sufficient time to make changes to their water consumption.

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. Note that for all welfare calculations, we use the intention-to-treat estimates in Section 3.2 as a measure of water conservation, and not the local average treatment effect estimates discussed in Section 3.3.

We have three main findings. First, the cost effectiveness of these interventions does not appear to be attractive relative to interventions studied by other researchers for promoting water

[61] We do not expect spillover impacts, if any, to increase the post-treatment consumption of the control group.

conservation. Second, the per capita net benefits of the intervention are close to zero. Net benefits typically range between plus or minus $1 per person under a wide range of assumptions. Using a marginal value of public funds approach yields similar findings, implying government investment in such interventions may be an unattractive strategy. Third, both cost effectiveness and welfare improve if it is assumed that benefits accrue over a year and producers just break even. Details on the parameters used in our welfare analysis and additional sensitivity analyses are presented in Appendix E.

4.1 Cost Effectiveness

There are many definitions of cost effectiveness. Here, we use different measures that correspond to direct resource cost and foregone profits. Our purpose is to consider the cost effectiveness of the interventions in our field experiment (Section 4.1.1) and then compare it with other interventions in the literature (Section 4.1.2).[62]

4.1.1 Cost effectiveness of the natural field experiment

We measure four categories of costs: the cost of sending letters, the direct cost of the incentives, the lost producer surplus associated with the decline in production, and the value of time in filling out the survey. Effectiveness is measured by the per capita reduction in water consumption. Our base case is the *Incentive £10* treatment which, along-with the *Incentive £15* treatment, was the only intervention that resulted in a significant reduction in water consumption among treated households (see Table 2). We measure its effectiveness relative to not sending out a letter, and to sending out the *Vanilla* letter.[63] Dividing total cost by effectiveness yields our cost effectiveness estimate. Results are presented in Table 5.

We describe the parameters for the base case (column 1) below.[64] The cost of mailing represents the postal cost of sending letters to 1,020 participants (the sample size in the *Incentive £10* group for which we had both pre- and post-treatment consumption data) 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 direct cost of incentives refers to the pecuniary transfer to the customers who completed the diagnostic. 85 households from the 1,020 participants in the *Incentives 10* treatment group completed the audit and received £10, yielding a direct cost of £850. The incentive costs are included in the calculations because only the *Incentives* group experienced a drop in water con-

[62] We do not attempt to harmonize different measures of cost effectiveness that have been used in this area. This could be a useful exercise but is beyond the scope of this paper. For additional discussion of this issue in the context of climate change cost effectiveness measures, see Hahn et al. (2024).

[63] We also run a sensitivity analysis with the *Incentive £15* treatment in Appendix E.2. In all cases, even though the water consumption was higher (4.7 liters per day as opposed to 3.5 liters per day with the *Incentives £10* treatment), the marginal gain is not worth the extra £5 of incentives.

[64] For a list of all the parameters and their sources, see Table E.1 in Appendix E.

Table 5: Different Measures of Cost Effectiveness

Case **Parameter**	Base Case (1)	No Producer Surplus Loss (2)	Vanilla Letter (3)	Targeting High Users (4)	Duration: 1 Yr & No PS Loss (5)
Cost of Mailing	420	420	0	200	420
Direct Cost of Incentive	850	850	850	380	850
Producer Surplus Loss	200	0	300	200	0
Time Cost	68	68	0	30	68
[A]: Total Cost (in £)	1,500	1,300	1,100	810	1,300
[B]: Effectiveness (in m^3)	240	240	350	230	1,300
Cost Effectiveness (£/m^3)	6.5	5.7	3.3	3.4	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.

sumption post-treatment when compared to usage in the control group.[65] The *Producer Surplus Loss* is defined as the total loss in net revenue (*i.e.*, revenue minus cost) caused by water savings.[66] 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 (see discussion in Section 3.2). 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 (refer to Table 2), 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 (Office for National Statistics, UK, 2021).[67] The sum of these items gives a total cost of £1,500. To calculate effectiveness, we multiply the per capita reduction in water consumption relative to the case of no letter (3.5 liters per day for 65 days; see Table 2) with the number of people in the £10 incentive group, which gives 240 cubic meters. Dividing the total cost by effectiveness gives us a cost effectiveness estimate of £6.5 per cubic meter for the base case.

The other four cases are variations on the base case. They lead to cost effectiveness numbers that range between £1.0 and £5.7 per cubic meter. The first variation labeled *No Producer Surplus*

[65] Considering either the *Simplified* and *Altruism* treatment groups as the base case will eliminate these costs, but since there was no meaningful impact of these treatments on water consumption relative to consumption in the control group (Table 2), the effectiveness will likely be small.

[66] We include changes in producer surplus here because they represent a real cost to producers and society under certain conditions. For a more in-depth discussion of different definitions of cost-effectiveness, see Hahn et al. (2024).

[67] 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.

Loss (column 2) sets producer surplus losses to zero. This yields a cost effectiveness of £5.7 per cubic meter, which is a 13 percent decline relative to the base case.[68] 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 and 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 (column 3). We do this analysis because NWG planned to send out this letter to their customers without our intervention. The cost effectiveness falls to £3.3 per cubic meter, a decline from the base case by 50 percent. The decline results from the reduction in mailing and time costs to zero, and the increase in water savings per household.[69]

The third variation targets only high users (column 4), 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 7.4 liters per household per day (see Appendix B.2.1 for details). The more than doubling of water conservation also substantially improves cost effectiveness from £6.5 per cubic meter in the base case to £3.4 per cubic meter in the targeted case.[70] This suggests that targeting could be an important strategy for improving cost effectiveness and increasing net benefits, which is consistent with other studies (Ferraro and Price, 2013; Ferraro and Miranda, 2013; 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 along with setting producers surplus losses equal to zero (column 5).[71] It is reasonable to assume that regulators periodically adjust prices to cover costs, which would imply producer surplus losses from water conservation would tend toward zero in the long run. In addition, various studies (Bernedo et al., 2014; Ansink et al., 2021) find a long-term persistent impact of their interventions, going up to 6 years in some instances.[72] If we assume the benefits last for a year, and prices adjust to eliminate producer surplus losses, cost effectiveness decreases from £6.5 in the base case to £1.0 per cubic meter, or by 84 percent. This improvement in cost effectiveness relative to the base case arises because the reduction in water use increased and producers did not incur losses. Because we are estimating a plausible lower bound for conservation (see Section 3.6.1), we also do a sensitivity analyses whereby we calculate by how much daily water consumption needs to decrease for cost effectiveness to be £1 per cubic meter, the lowest CE value which was calculated

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

[69] *Incentives 10* treatment leads to water savings of 5.3 liters per day in comparison to the *Vanilla* treatment, as opposed to 3.5 liter per day earlier in comparison to the control group. See Section 3.2 for details.

[70] A similar calculation for the £15 intervention reveals that cost effectiveness is reduced by 44 percent.

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

[72] Brandon et al. (2017) demonstrate that the social norm home energy reports have a persistent effect on energy consumption for up to two years after the discontinuation of the reports.

under the most optimistic scenario of no producer surplus losses and conservation effects lasting for a year. Our analysis reveals that daily water conservation would need to be 6.6 times higher, or 23 liters per day, for cost effectiveness to fall to this level. As discussed in Section 3.6.1, eliminating the measurement error is expected to increase the effect sizes by 2.7 times, at most. The cost effectiveness under such a scenario would only rise to £2.5 per cubic meter. Therefore, cost-effectiveness remains unattractive even under the most bullish scenarios.

In Appendix E.2, we run a similar analysis for the £15 intervention. Under all variations, both costs and effectiveness increase, but effectiveness increases by less than the costs. The result is that, for the base case, the effectiveness of the £15 intervention is £7.8 per cubic meter, 19 percent higher than the £10 intervention. Similarly, for all other scenarios, the £15 incentive is between 22 to 53 percent higher than the corresponding variation in the £10 incentive.

Finally, we also measure the cost-effectiveness in terms of costs incurred to reduce a tonne of CO_2 emissions. Details are presented in Appendix E.3, but our results suggest that cost effectiveness numbers range between £150 and £950 per tonne of CO_2 emissions reduced. These numbers are higher than almost all estimates of the social cost of carbon, implying the costs outweigh the policy benefits — at least based on short-run effects.

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 6.[73]

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 and Miranda (2013) to $8.1 per cubic meter in Ansink et al. (2021). Second, depending on the assumptions, our study estimates appear to fall toward the center or right tail of the distribution 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 and 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. Also, 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).

[73]These studies generally define costs in terms of direct costs, but the definitions also include forgone revenues in certain cases.

Table 6: 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
This Study (NWG)	1.3 - 8.4	Online audit with £10 incentive	Yes	0.24	$1.6 per hh	$0.3 per hh
Ferraro and Price (2013)	0.12 - 0.17	Social norm letter	Yes	700	$1.2 per hh	Forgone Revenues: $1.5mn to $1.6mn
Ferraro and 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
Baker (2021)	0.78 - 0.97	Turf replacement subsidy	Yes	4,900 - 6,100 per year	*Total Rebate Outlay*: $59mn; *Total Admin. Cost*: $13mn	Total Out-of-pocket conversion costs: $60mn

Notes: Notes: *hh* refers to household. Details of the calculation for this table are provided in Table E.4 and Table E.5. 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 and Price (2013), Ferraro and 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 a simple question: how large do other benefits (*i.e.*, those not quantified in our analysis) need to be under different scenarios to just offset costs that we estimate? Other benefits could include the private opportunity cost of water, ecosystem benefits, and 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.[74] Non-carbon greenhouse gas emissions have been converted to CO_2-equivalents for use in our analysis.[75] The carbon footprint numbers for the water supply, use, and the disposal system have been obtained from the Environment Agency, a leading public body for the environment in England and Wales (Reffold 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.[76] The benefits from the intervention, B, are then $-V\Delta g$. The costs, C, are given by losses in producer surplus plus the direct incremental costs of the experiment, E. The producer surplus losses can be represented by the difference in the price of water, p, and the marginal cost of water, c (presumed to be constant), multiplied by the change in water consumption, Δg. That is, the producer surplus losses are $(p-c)\Delta g$. The direct costs, E, include the cost of mailing letters.[77] They also include the cost of the financial incentives.[78] We can now estimate net benefits as follows:

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

Equation (6) does not include a measure of consumer surplus.[79] This is because we assume that consumers who switch are just as well off after receiving the subsidy and taking the audit as

[74] Precisely because the reduction in GHG emissions is only a subset of the overall benefits from water conservation, we do a bounding exercise later to calculate how high other benefits need to be to offset the estimated costs of the intervention.

[75] 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) (Reffold et al., 2008).

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

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

[78] We assume the financial incentives are financed through a lump sum tax. It would be straightforward to add a marginal deadweight loss associated through taxation if this were not the case, and it would not materially change our qualitative findings.

[79] 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.

they were before.[80] If consumers are actually better off after switching, then the measure of $B-C$ is an underestimate of net benefits. We explore this issue in a sensitivity analysis in Appendix E.5.4 in which we assume that the cost savings from water conservation and the financial incentive provide net benefits to the consumer.

In what follows, we also do not explicitly include other benefits from water conservation, which may be substantial, but for which we do not have precise estimates.[81] We do, however, include these benefits in a bounding calculation described below. These benefits include possible savings from reduced capital and operating costs associated with expanded supply (Maddaus, 2011), or the value of conservation in areas that may experience scarcity (Baker, 2021). 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.[82]

Estimates for the various parameters in Equation (6) are shown in Table 7 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 NWG (NWG, 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.[83] LRMC, on the other hand, is the sum of marginal operating and marginal capacity costs.[84] We consider LRMC in our sensitivity analysis because it allows us to evaluate potential long-run benefits of conservation. The LRMC was calculated based on the annualized cost of the last major water resource investment undertaken by NWG – expanding Abberton reservoir in 2009 – and equals £0.54 per cubic meter.[85]

We consider five 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.[86]

[80] We are applying the envelope theorem. See, *e.g.*, Finkelstein and Hendren (2020).

[81] We asked NWG to provide a willingness to pay for a cubic meter increase in water conservation, but they were unable to furnish a value.

[82] See Bishop and Weber (1996) for a more extended discussion on the impact of demand reduction on water utilities and the environment.

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

[84] Marginal capacity costs are defined as costs associated with investments as a result of an incremental increase in demand.

[85] 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 NWG did not have an estimate.

[86] Results with the *£15 Incentive* treatment as the base case are presented in Appendix E.5.3. We use the two incentive treatments for all welfare analysis because those interventions are the ones that resulted in an economically significant reduction in water consumption.

The second case sets producer surplus losses to zero. This assumption is made to explore what would happen if the regulatory authority revised the rate setting process in response to capacity expansion plans.[87]

The third case uses the *Vanilla* letter as the benchmark with the *£10 Incentive*. The rationale for considering a different benchmark is that NWG was going to send out the *Vanilla* letter anyway. Using this benchmark means that there is not incremental cost of sending letters by mail. The fourth case focuses on targeting high-users, defined as households who consume above the median pretreatment consumption threshold. The fifth case assumes the impact of the intervention lasts for a year, in addition to eliminating producer surplus losses. For each case, we also specify the number of people affected, or N, which is used to estimate the per capita net benefits. 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 £241 per metric ton of CO_2 (in 2020 £) (United Kingdom Government, 2021).[88]

Table 7 shows that the measured benefits fall short of the measured costs in four out of five cases under both the cost structures. The only scenario with a positive net-benefit is the final one where the effect lasts for a year and there is no producer surplus loss.[89] For all five cases, the range of net benefits is roughly from -£1 to +£1 per person, suggesting that the net benefits of the intervention are small when only considering the climate change benefits.

The benefits considered up until now only include the reductions in greenhouse gas emissions because estimates of other benefits are not readily available. Still, it is possible to do a bounding analysis and explore what other benefits (or costs) would need to be for net benefits to be zero. It is also possible to consider other estimates in the literature for key parameters, such as the opportunity cost of water, and compare these with the findings of our bounding analysis.

A key parameter is the opportunity cost of water. Unfortunately, this number is not available from NWG or a water market in the UK. In the absence of such a market, we rely on three estimates for the opportunity cost of scarce water. First, we use an estimate from Baker (2021), who approximates a value of \$0.06 per gallon in 2014 dollars (equivalent to £1.4 per cubic meter in 2020 £).[90] Note, however, that this estimate is from Nevada, the driest state in the United States. The opportunity cost in the relatively wetter regions served by NWG would likely be lower. A second estimate can be derived from a water market in Spain, a water-stressed industrialized country in Europe, to arrive at a marginal private value of water. Palomo-Hierro et al. (2015) list details of

[87] Water utilities in the UK are regulated as part of a legally binding long-term framework to cut emissions (Ofwat, 2023). Thus, they may be required to undertake conservation programs that result in short-term loss of profits.

[88] We explore the sensitivity of our results to the SCC in Appendix E.5.2 where we assume an SCC estimate of \$51 per ton of CO_2e based on the recommendations of the Interagency Working Group, US Government (2021).

[89] We get the same values for net benefits under SRMC and LRMC for cases with no producer surplus loss (columns (2) and (5))because the marginal cost is only used for calculating the losses to the utility due to conservation. If these losses are assumed away, the values under SRMC and LRMC are the same.

[90] Baker (2021) approximates the scarcity value of water based on an estimate from Edwards and Libecap (2015). The latter study uses agriculture-to-urban water rights sales in the Truckee river basin in Nevada to assess the value of scarce water.

Table 7: Benefit-Cost Analysis

Case Parameter	Units	Base Case (£10 Incentive) (1)	No Producer Surplus Loss (2)	Vanilla Letter as Benchmark (3)	Targeting High Users (4)	Duration: 1 Yr & No PS Loss (5)
V	$£/m^3$	1.7	1.7	1.7	1.7	1.7
p	$£/m^3$	1.3	1.3	1.3	1.3	1.3
Δg	m^3	-240	-240	-340	-230	-1,300
$-V\Delta g$	£	390	390	570	390	2,200
E	£	1,300	1,300	850	580	1,300
N	*integer*	1,020	1,020	1,020	484	1,020
***Panel A* (*SRMC*)**						
c	$£/m^3$	0.44	0.44	0.44	0.44	0.44
$(p-c)\Delta g$	£	-200	0	-290	-200	0
$B-C$ (Equation (6) above)	£	-1,100	-880	-570	-390	920
$(B-C)/N$	*£/ capita*	-1.1	-0.86	-0.56	-0.80	0.90
Breakeven Other Benefits $= -(B-C)$	£	1,100	880	570	390	-920
Breakeven Other Benefits / Δg	$£/m^3$	4.6	3.7	1.7	1.7	-0.70
Breakeven Other Benefits / GHG benefits	*multiple*	2.8	2.3	1.0	1.0	-0.42
Increase in Water Savings	*multiple*	6.8	3.3	3.1	3.1	0.58
Months Effect Needs to Last	*months*	15	7	7	7	-
***Panel B* (*LRMC*)**						
c	$£/m^3$	0.98	0.98	0.98	0.98	0.98
$(p-c)\Delta g$	£	-72	0	-110	-71	0
$B-C$ (Equation (6) above)	£	-950	-880	-390	-260	920
$(B-C)/N$	*£/ capita*	-0.93	-0.86	-0.38	-0.54	0.90
Breakeven Other Benefits $= -(B-C)$	£	950	880	390	260	-920
Breakeven Other Benefits / Δg	$£/m^3$	4.1	3.7	1.1	1.1	-0.70
Breakeven Other Benefits / GHG Benefits	*multiple*	2.4	2.3	0.68	0.68	-0.42
Increase in Water Savings	*multiple*	4.1	3.3	1.8	1.9	0.58
Months Effect Needs to Last	*months*	9	7	4	4	-

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 Table E.1 for details on parameters used for welfare calculations.

several formal intra- and inter-basin lease contracts, including their prices, signed over the past two decades in Spain. The prices for water transfer across all these leases ranged from £0.16 per cubic meter to £0.28 per cubic meter (2020 £), which is approximately 80 percent lower than the opportunity cost of scarce water in Nevada, USA. A third estimate we consider as a proxy for private value of water is the cost of desalination. Water-scarce regions may rely on this technology as a substitute for fresh water (Gude, 2017; World Bank, 2019). As of 2019, the average cost range of desalinated water was between \$0.5 to \$1.5 per cubic meter, *i.e.* £0.41 to £1.21 per cubic meter in 2020£(Cosín, 2019).

We compare these estimates for the opportunity cost of water with the bounding calculations in Table 7. As shown in Panel A (SRMC) and B (LRMC) of the table, we would need other benefits to be between £1.1 and £4.6 per cubic meter in order to break-even. Our opportunity cost of water estimates range from £0.16 in Spain to £1.4 per cubic meter in Nevada, USA. Using the lower end of the range for the opportunity cost of water would not be enough to change the sign of the net benefits in any of the cases, while using the upper end of the range would change the sign in selected cases (*e.g.*, when long-run marginal cost is the relevant measure of cost and *i*) we target high users, or *ii*) there are no costs to sending letters.)

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 240 percent to £830 per ton, and to £910 per ton using the SRMC. If we include the opportunity cost of scarce water among other benefits, the SCC numbers would still need to be in the range of £630 to £700 depending on whether we use LRMC or SRMC, respectively. These numbers are higher than most estimates for the SCC.[91]

As discussed in Section 3.6, we identify a plausible lower bound of the treatment effects due to potential measurement error and spillovers. In addition, since we only have consumption data for 65 days post treatment, the short time frame of accrued benefits could bias net benefits downward. To address these concerns, we conduct two further sensitivity analyses. First, we measure how high water savings need to be to pass a benefit-cost analysis, provided effects last only for 65 days. Second, we ask how long the effects would need to last under the current conservation estimates for the intervention to yield positive benefits. The penultimate row in each panel of Table 7 provides a measure of the conservation magnitudes needed. Under SRMC, the water savings would need to be at least 3 to 7 times higher. This amounts to a 10 percent decline in consumption for the average household under the base case (24 liters per day), as opposed to the negative 1.5 percent average treatment effect we currently estimate. This multiple is lower for LRMC, with a decline in water consumption required to be 2 to 4 times higher for net benefits to be positive.[92] Next, assuming the effect sizes stay the same, the last row of each panel indicates that the time period of the effects will need to be at least 15 months under the base case for net benefits to be positive. This reduces to 9 months under LRMC as producer surplus losses are lower.

There is also uncertainty regarding the estimate of long-run csts. Specifically, one may be concerned that the data on the LRMC from current projects do not reflect the expected LRMC under a climate change scenario where costs could be higher and will probably exceed the current water price. Therefore, as a sensitivity check, we ask how high the LRMC needs to be such that net benefits are at least 0 under the base case. Assuming benefits last for 65 days, under the base case scenario, the LRMC would need to £5.1 per cubic meter as opposed to the 0.98 currently. This represents a 500 percent increase in the long-run marginal cost.

The preceding analysis assumes that consumer private benefits from taking the audit are neg-

[91] In Appendix E.5, we explore a number of other sensitivities, including varying the relationship between price and costs, varying the LRMC, and assuming the transfer from utility to consumers represents a net benefit social benefit.

[92] Note that the weighted average for the entire sample of the proportion of treated days in post-treatment observations was 37 percent (see Section 3.6.1. Therefore, assuming that the true effect sizes are 2.7 times higher (100/37), the net benefits would still not be greater than 0 under any of the SRMC scenarios which do not originally pass the BCA.

ligible. In the appendix, we relax this assumption and develop an upper bound on consumer benefits. We define a plausible upper bound as the monetary savings in water plus the value of the incentive. In the base case scenario with £10 incentive and benefits lasting for two months, we find per capita net benefits become slightly positive, increasing from negative £1.1 per person in Table 7 to £0.073 per person. Assuming persistent benefits lasting a year and no producer surplus loss, the net benefits increase to £3.4 per person from £0.9 per person. This suggests that under extreme assumptions, there may be modest net benefits from conservation when only considering climate-related benefits. Further details are presented in Appendix E.5.4.

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 and Sprung-Keyser, 2020; Finkelstein and Hendren, 2020; Hahn et al., 2024).[93] In general, the higher the benefits and the lower the net cost to the government, the more attractive the intervention is.

We introduce some notation to clarify our estimation procedure. Define after-tax benefits as WTP or willingness to pay net of private costs; and define G as the net cost to the government. The 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 by the government 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

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

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).[94]

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

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

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.[95]

[93] Here, we are implicitly assuming that the program is paid for by the government or by a central authority that will need to raise revenues to pay for the program. For other examples, see Hendren and Sprung-Keyser (2020).

[94] We ignore taxes on water for the consumer in this analysis in the interest of simplicity.

[95] A more complete analysis might consider rebound effects. For example, if water conservation resulted in a reduction in the price of water, which led to greater use in other activities and more pollution, this would need to be accounted for. In this case, we think such effects are not likely to be important. See Hahn et al. (2024) for an analysis of how to include rebound effects in MVPF calculations.

The formula for the MVPF is thus:

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

Table 8: MVPF Calculations

Case Parameter	Base Case (£10 Incentive) (1)	No Producer Surplus Loss (2)	Vanilla Letter as Benchmark (3)	Targeting High Users (4)	Duration: 1 Yr & No PS Loss (5)
Panel A **(*SRMC*)**					
$Cost$	0.44	0.44	0.44	0.44	0.44
WTP	0.18	0.31	0.40	0.39	1.7
G	1.0	1.0	1.1	1.1	1.0
$MVPF = \frac{WTP}{G}$	0.18	0.31	0.38	0.37	1.7
Panel B **(*LRMC*)**					
$Cost$	0.98	0.98	0.98	0.98	0.98
WTP	0.26	0.31	0.58	0.57	1.7
G	1.0	1.0	1.0	1.0	1.0
$MVPF = \frac{WTP}{G}$	0.26	0.31	0.57	0.56	1.7

Notes: This table computes the MVPF for the three scenarios described in Table Table 7 using Equation (9). *Panel A* shows the results for the short-run marginal cost. *Panel B* shows the results for long-run marginal cost. The values for V and p are the same as those in Table 7. See Table E.1 in Appendix for details on parameters used for welfare calculations.

Table 8 summarizes five MVPF calculations. It mirrors the net benefit calculations in Section 4.2. For the short-run marginal cost scenario, the MVPF ranges from 0.18 to 1.7. The only scenario where the MVPF is greater than 1 is when we assume there are no producer surplus losses and the benefits last for a year (column (5)). This analysis is similar to our benefit-cost analysis in that it suggests the investment may not be worth making unless other benefits not included here are significant, or the conservation effects persist for a year. Using the LRMC instead of the SRMC (Panel B) increases the after-tax benefits due to a reduction in the producer surplus loss. The MVPF, though positive under all scenarios, still remains small and less than 1 for four of the five cases.[96]

As with the benefit-cost analysis, the MVPF values depend on the value of the SCC. Equation (7) shows that increasing the social cost of carbon, which is proportional to V, would increase the MVPF. For example, if we consider the current US administration 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, US Government, 2021), the MVPF is negative for the first four scenarios, and positive but less than 1 for the scenario that assumes long-run benefits and no producer surplus loss. The

[96]MVPF analysis for the £15 intervention is presented in Appendix E.5.6.

negative sign here arises because WTP is negative and the net cost to the government is positive. Further details are presented in Appendix E.5.5.

Similarly, increasing V to 6.2 in the case of the SRMC would mean that *WTP*, and hence *MVPF*, were 1 using the other base case assumptions.[97] This means that the "return" (or WTP) on a government investment of £1 is £1. As other scholars have noted, there are several other climate investments that yield higher *MVPFs* (Hahn et al., 2024).

We can also analyze the change in *MVPF* if we include the opportunity cost of water. Denoting this by ϕ, the *WTP* in Equation (7) will now be expressed as:

$$WTP = ((1 - t_c)(p - c) - V - \phi)\frac{dg}{dn} \tag{10}$$

We also do a sensitivity analysis that measures the impact on the MVPF when we add in the scarcity value of water (£1.4 per cubic meter), and find that the effect is small. The *MVPF* increases but still remains below 1 for all scenarios that do not assume that conservation benefits last for a long time. It equals 0.44 in our base case of *£10 Incentive*, and is just below 1 (0.98) in the case of *Vanilla* letter. Importantly, the *MVPF* is much greater than 1, at 3.2, for the case with persistent conservation impact up to one year and no producer surplus loss. This implies that in areas with scarce water, conservation programs may be fruitful provided the effects can last for a long time and utilities are able to recover their losses quickly. With *LRMC*, the *MVPF* increases, with the value in the base case now equaling 0.52. However, in the long-run — with the exception of the base case and the scenario with no producer surplus losses and only short-run benefits — the *MVPF* is greater than 1. Thus, the potential benefits of the policy are greater than the net cost of the policy for the government under a larger range of scenarios. This arises because a higher marginal cost reduces the loss in after-tax unit profits, and taking into account scarcity value of water increases the environmental gain. Both factors increase the *WTP* in Equation (7).

In conclusion, in line with our analysis of net benefits, the *MVPF* increases when there are no producer losses and we extend the period for which benefits accrue. In cases where we take into account the opportunity cost of water, it exceeds one which would mean the benefits from the policy would exceed the net cost to the government if the marginal value of water is high. In all other cases, unless the effects persist for a long time, the government may not find it worthwhile to spend resources on conserving water using the interventions discussed here.

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.

[97]This amounts to an SCC of £910 per ton of CO_2e, which is 380 percent higher than the current estimate of £241 per ton.

5 Conclusion

Water suppliers and regulators are showing greater interest in assessing different mechanisms to encourage conservation. One approach that is being used is water audits, which offer customers recommendations on how they could reduce their water consumption.

This paper uses a natural field experiment to explore the cost effectiveness and economic efficiency of online water audits. We have three main findings. First, financial incentives have a strong, positive impact on the take-up of online water audits in the short run. Additionally, the size of the financial incentive used to encourage conservation matters, as reflected in the positive correlation between the size of the incentive and the probability of completing the audit. Furthermore, encouraging subjects to participate in an online water audit with financial incentives reduces household consumption by about 17 percent. These findings suggest that it may be useful to introduce different levels of the subsidy in future experiments to help optimize social welfare or better meet a decision-maker's objectives. We also find some evidence that the large magnitudes of water conservation among treated households are being driven by both behavioral changes and technology adoption. Second, notwithstanding these improvements in water conservation, the per capita net benefits of the intervention are close to zero under a wide range of assumptions. 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. This suggests that further experiments targeting particular groups could help improve social welfare.

There are several areas for future research that we think could be fruitful. First, it would be useful to compare the cost-effectiveness of audits with behavioral interventions related to price salience. For example, households often may not know the marginal price they pay for water, especially in a non-linear price setting. An interesting behavioral intervention could involve correcting the biased beliefs and testing if it helps with water conservation and reducing bias (Rodemeier and Löschel, 2022).

Second, 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 6, which reviews behavioral economics research in this area, reveals how little we know about the cost effectiveness of different interventions. Decision-makers in charge of water conservation may find it helpful to know something about the likely costs and effectiveness of the 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 and Sprung-Keyser (2020) and Hahn et al. (2024) developed and compared several estimates of *MVPF*s in the education, health, labor, and climate change areas, it could be useful to undertake a similar exercise for water interventions. Such analyses could also help inform equity and efficiency trade-offs (Cardoso and Wichman, 2022; Wichman, 2023).

Third, it would be very useful to try to quantify some of the other benefits associated with water conservation in monetary terms, such as the willingness to pay for greater reliability of supply. 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).

Finally, 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, Hunt, (2011), "Social Norms and Energy Conservation," *Journal of Public Economics*, *95* (9-10), 1082–1095.

_ **and Judd B. Kessler**, (2019), "The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons," *American Economic Journal: Applied Economics*, *11* (1), 236–276.

_ **and Michael Greenstone**, (2017), "Measuring the Welfare Effects of Residential Energy Efficiency Programs," *Technical Report*, *National Bureau of Economic Research.*

_ **and Todd Rogers**, (2014), "The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation," *American Economic Review*, *104* (10), 3003–3037.

Ansink, Erik, Carmine Ornaghi, and Mirco Tonin, (2021), "Technology vs Information to Promote Conservation: Evidence from Water Audits," *Tinbergen Institute Discussion Paper 2021-014/VIII.*

Ayres, Ian, Sophie Raseman, and Alice Shih, (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.

Baker, Jonathan E., (2021), "Subsidies for Succulents: Evaluating the Las Vegas Cash-for-Grass Rebate Program," *Journal of the Association of Environmental and Resource Economists*, *8* (3), 475–508.

Bennear, Lori S., Jonathan M. Lee, and Laura O. Taylor, (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, Maria, Paul J. Ferraro, and Michael Price, (2014), "The Persistent Impacts of Norm-Based Messaging and their Implications for Water Conservation," *Journal of Consumer Policy*, *37* (3), 437–452.

Bernheim, B. Douglas and Dmitry Taubinsky, (2018), "Behavioral Public Economics," *Handbook of Behavioral Economics: Applications and Foundations 1*, *1*, 381–516.

Bishop, Daniel B. and Jack A. Weber, (1996), "Impacts of Demand Reduction on Water Utilities," *American Water Works Association.*

Bollinger, Bryan, Jesse Burkhardt, and Kenneth T. Gillingham, (2020), "Peer Effects in Residential Water Conservation: Evidence from Migration," *American Economic Journal: Economic Policy*, *12* (3), 107–33.

Brandon, Alec, John A List, Robert D Metcalfe, Michael K Price, and Florian Rundhammer, (2019), "Testing for crowd out in social nudges: Evidence from a natural field experiment in the market for electricity," *Proceedings of the National Academy of Sciences*, *116* (12), 5293–5298.

__ , **Paul J. Ferraro, John A. List, Robert D. Metcalfe, Michael K. Price, and Florian Rundhammer**, (2017), "Do the Effects of Nudges Persist? Theory and Evidence from 38 Natural Field Experiments," *Technical Report*, *National Bureau of Economic Research.*

Brent, Daniel A. and Casey J. Wichman, (2022), "Do Behavioral Nudges Interact with Prevailing Economic Incentives? Pairing Experimental and Quasi-Experimental Evidence from Water Consumption," *Working Paper 22-2, Resources for the Future (RFF).*

__ , **Corey Lott, Michael Taylor, Joseph Cook, Kimberly Rollins, and Shawn Stoddard**, (2020), "What Causes Heterogeneous Responses to Social Comparison Messages for Water Conservation?," *Environmental and Resource Economics*, *77* (3), 503–537.

__ , **Joseph H. Cook, and Skylar Olsen**, (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.

Browne, Oliver R., Ludovica Gazze, and Michael Greenstone, (2021), "Do Conservation Policies Work? Evidence from Residential Water Use," *Environmental and Energy Policy and the Economy*, *2* (1), 190–225.

__ , __ , __ , **and Olga Rostapshova**, (2023), "Man vs. Machine: Technological Promise and Political Limits of Automated Regulation Enforcement," *Technical Report*, *National Bureau of Economic Research.*

Butera, Luigi, Robert Metcalfe, William Morrison, and Dmitry Taubinsky, (2022), "Measuring the Welfare Effects of Shame and Pride," *American Economic Review*, *112* (1), 122–68.

Byrne, David P. and Lorenz Goette, (2022), "Information, Incentives, and Goal-Setting: A Field Experiment in Water Usage," *Working Paper.*

Calzolari, Giacomo and Mattia Nardotto, (2017), "Effective reminders," *Management Science*, *63* (9), 2915–2932.

Cardoso, Diego S. and Casey J. Wichman, (2022), "Water Affordability in the United States," *Water Resources Research*, *58* (12).

Castleman, Benjamin L and Lindsay C Page, (2016), "Freshman year financial aid nudges: An experiment to increase FAFSA renewal and college persistence," *Journal of Human Resources*, *51* (2), 389–415.

Cominola, Andrea, Lucy Preiss, M Thyer, Holger R Maier, Peter Prevos, RA Stewart, and Andrea Castelletti, (2023), "The Determinants of Household Water Consumption: A Review and Assessment Framework for Research and Practice," *npj Clean Water*, *6* (1), 11.

Cosín, Carlos, January (2019), "The Evolution of Rates in Desalination (Part I)," *Blog Post, Smart Water Magazine.* Available at: https://smartwatermagazine.com/blogs/carlos-cosin/evolution-rates-desalination-part-i, Last Accessed: 2023-07-12.

Dai, Hengchen, Silvia Saccardo, Maria A Han, Lily Roh, Naveen Raja, Sitaram Vangala, Hardikkumar Modi, Shital Pandya, Michael Sloyan, and Daniel M Croymans, (2021), "Behavioural nudges increase COVID-19 vaccinations," *Nature*, *597* (7876), 404–409.

Damgaard, Mette Trier and Christina Gravert, (2018), "The hidden costs of nudging: Experimental evidence from reminders in fundraising," *Journal of Public Economics*, *157*, 15–26.

Datta, Saugato, Juan Josï, Laura Zoratto, Oscar Calvo-Gonzï, Matthew Darling, and Karina Lorenzana, (2015), "A Behavioral Approach to Water Conservation: Evidence from Costa Rica," *World Bank Washington, DC, USA.*

Department for Environment, Food & Rural Affairs, Great Britain, (2008), "Future Water: The Government's Water Strategy for England," Vol. 7319, *The Stationery Office.*

Domurat, Richard, Isaac Menashe, and Wesley Yin, (2021), "The role of behavioral frictions in health insurance marketplace enrollment and risk: Evidence from a field experiment," *American Economic Review*, *111* (5), 1549–1574.

Edwards, Eric C. and Gary D. Libecap, (2015), "Water Institutions and the Law of One Price," *Handbook on the Economics of Natural Resources*, p. 442.

Environmental Agency, UK, Oct (2021), "Lack of Water Presents Existential Threat, Says Environment Agency Chief," *Press Release, GOV.UK.* Available at: https://www.gov.uk/government/news/lack-of-water-presents-existential-threat-says-environment-agency-chief, Last Accessed: 2023-07-13.

_, May (2022), "Review of England's Emerging Regional Water Resources Plans," *Research and Analysis, GOV.UK.* Available at: https://www.gov.uk/government/publications/review-of-englands-emerging-regional-water-resources-plans, Last Accessed: 2023-07-13.

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). Available at: https://www.exchangerates.org.uk/, Last Accessed: 2022-04-30.

Ferraro, Paul J. and Juan José Miranda, (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.

— **and Michael K. Price**, (2013), "Using Nonpecuniary Strategies to Influence Behavior: Evidence from a Large-Scale Field Experiment," *Review of Economics and Statistics*, *95* (1), 64–73.

Ferraro, Paul J, Juan Jose Miranda, and Michael K Price, (2011), "The Persistence of Treatment Effects with Norm-Based Policy Instruments: Evidence from a Randomized Environmental Policy Experiment," *American Economic Review*, *101* (3), 318–22.

Finkelstein, Amy and Nathaniel Hendren, (2020), "Welfare Analysis Meets Causal Inference," *Journal of Economic Perspectives*, *34* (4), 146–67.

Fishbane, Alissa, Aurelie Ouss, and Anuj K Shah, (2020), "Behavioral Nudges Reduce Failure to Appear for Court," *Science*, *370* (6517).

Fowlie, Meredith, Michael Greenstone, and Catherine Wolfram, (2018), "Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program," *The Quarterly Journal of Economics*, *133* (3), 1597–1644.

Garrick, Dustin E., Jim W. Hall, Andrew Dobson, Richard Damania, R. Quentin Grafton, Robert Hope, Cameron Hepburn, Rosalind Bark, Frederick Boltz, Lucia De Stefano et al., (2017), "Valuing Water for Sustainable Development," *Science*, *358* (6366), 1003–1005.

Gerarden, Todd D. and Muxi Yang, (2021), "Using Targeting to Optimize Program Design: Evidence from an Energy Conservation Experiment," *Energy, COVID, and Climate Change, Online-June*, pp. 7–9.

Gillingham, Kenneth and Tsvetan Tsvetanov, (2018), "Nudging Energy Efficiency Audits: Evidence from a Field Experiment," *Journal of Environmental Economics and Management*, *90*, 303–316.

Goetz, Alexander, Harald Mayr, and Renate Schubert, (2022), "One Thing Leads to Another: Evidence on the Scope and Persistence of Behavioral Spillovers," *Working Paper, Available at SSRN 3919454.*

Gude, Veera Gnaneswar, (2017), "Desalination and Water Reuse to Address Global Water Scarcity," *Reviews in Environmental Science and Bio/Technology*, *16* (4), 591–609.

Hahn, Robert W. and Robert D. Metcalfe, (2021), "Efficiency and Equity Impacts of Energy Subsidies," *American Economic Review*, *111* (5), 1658–88.

Hahn, Robert W, Nathan Hendren, Robert D Metcalfe, and Ben Sprung-Keyser, (2024), "A Welfare Analysis of Policies Impacting Climate Change," *NBER Working Paper no. 32728.*

Hanemann, W. Michael et al., (2006), "The Economic Conception of Water," *Water Crisis: Myth or Reality*, *61*, 74–76.

Hannaford, Jamie, Jonathan D. Mackay, Matthew Ascott, Victoria A. Bell, Thomas Chitson, Steven Cole, Christian Counsell, Mason Durant, Christopher R. Jackson, Alison L. Kay et al., (2023), "The Enhanced Future Flows and Groundwater Dataset: Development and Evaluation of Nationally Consistent Hydrological Projections based on UKCP18," *Earth System Science Data*, *15* (6), 2391–2415.

Harrison, Glenn W. and John A. List, (2004), "Field Experiments," *Journal of Economic Literature*, *42* (4), 1009–1055.

He, Chunyang, Zhifeng Liu, Jianguo Wu, Xinhao Pan, Zihang Fang, Jingwei Li, and Brett A Bryan, (2021), "Future Global Urban Water Scarcity and Potential Solutions," *Nature Communications*, *12* (1), 4667.

Hendren, Nathaniel and Ben Sprung-Keyser, (2020), "A Unified Welfare Analysis of Government Policies," *The Quarterly Journal of Economics*, *135* (3), 1209–1318.

Interagency Working Group, US Government, (2021), "Technical Support Document: Social Cost of Carbon, Methane, and Nitrous Oxide Interim Estimates Under Executive Order 13990," *Technical Report*, *White House.* Available at: https://www.whitehouse.gov/wp-content/uploads/2021/02/TechnicalSupportDocument_SocialCostofCarbonMethaneNitrousOxide.pdf, Last Accessed: 2023-07-13.

Ito, Koichiro, Takanori Ida, and Makoto Tanaka, (2018), "Moral suasion and economic incentives: Field experimental evidence from energy demand," *American Economic Journal: Economic Policy*, *10* (1), 240–267.

Jessoe, Katrina, Gabriel E Lade, Frank Loge, and Edward Spang, (2021), "Residential Water Conservation during Drought: Experimental Evidence from Three Behavioral Interventions," *Journal of Environmental Economics and Management*, *110*, 102519.

_, Gabriel E. Lade, Frank Loge, and Edward Spang, (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.

Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman, (2016), "Getting to the top of mind: How reminders increase saving," *Management Science*, *62* (12), 3393–3411.

Knittel, Christopher R. and Samuel Stolper, (2019), "Using Machine Learning to Target Treatment: The Case of Household Energy Use," *Technical Report*, *National Bureau of Economic Research.*

List, John A., (2022), "The Voltage Effect: How to Make Good Ideas Great and Great Ideas Scale," *Currency.*

Maddaus, William O., (2011), "Realizing the Benefits from Water Conservation," *Journal of Contemporary Water Research and Education*, *114* (1), 2.

Mann, Patrick C., Robert J. Saunders, and Jeremy J. Warford, (1980), "A Note on Capital Indivisibility and the Definition of Marginal Cost," *Water Resources Research*, *16* (3), 602–604.

Manouseli, Despina, SM Kayaga, and Roy Kalawsky, (2019), "Evaluating the Effectiveness of Residential Water Efficiency Initiatives in England: Influencing Factors and Policy Implications," *Water Resources Management*, *33*, 2219–2238.

Mansur, Erin T. and Sheila M. Olmstead, (2012), "The Value of Scarce Water: Measuring the Inefficiency of Municipal Regulations," *Journal of Urban Economics*, *71* (3), 332–346.

Marsden Jacob Associates, (2004), "Estimation of Long Run Marginal Cost (LRMC)," *Queensland Competition Authority.*

National Infrastructure Commission, October (2023), "The Second National Infrastructure Assessment," *Technical Report*, *National Infrastructure Commission.* Available at: https://nic.org.uk/app/uploads/Final-NIA-2-Full-Document.pdf, Last Accessed: 2024-07-14.

Nauges, Céline and Dale Whittington, (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.

NWG, (2009), "Northumbrian Water Limited Annual Report and Financial Statements," *NWG Financial Statements 2008/09.*

—, (2020), "Northumbrian Water Limited Household Charges Scheme," *NWG Charges Scheme 2019/20.*

—, (2021), "Northumbrian Water Limited Annual Report and Financial Statements," *NWG Financial Statements 2020/21.*

Office for National Statistics, UK, (2021), "Annual Survey of Hours and Earnings." Data retrieved from https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/ashe1997to2015selectedestimates, Last Accessed: 2023-07-12.

Ofwat, July (2008), "Preparing for the Future — Ofwat's Climate Change Policy Statement," *Technical Report*, *Ofwat.* Available at: https://www.ofwat.gov.uk/wp-content/uploads/2015/10/pap_pos_climatechange.pdf, Last Accessed: 2023-07-12.

—, July (2010), "Playing Our Part—Reducing Greenhouse Gas Emissions in the Water and Sewerage Sectors Supporting Information," *Technical Report*, *Ofwat.* Available at: https://www.ofwat.gov.uk/wp-content/uploads/2015/11/prs_inf_emissionssup.pdf, Last Accessed: 2023-07-12.

—, December (2013), "Water Meters – Your Questions Answered: Information for Household Customers," *Technical Report*, *Ofwat.* Available at: https://www.ofwat.gov.uk/wp-content/uploads/2015/11/prs_lft_101117meters.pdf, Last Accessed: 2023-07-22.

—, June (2018), “PN 22/18: Thames Water’s Failure to Tackle Leakage Results in £65m Package for Customers,” *Corporate Release, Ofwat.* Available at: https://www.ofwat.gov.uk/pn-22-18-thames-waters-failure-tackle-leakage-results-65m-package-customers/, Last Accessed: 2023-07-12.

—, July (2022), “Safeguarding the Water Supply for the Future,” *Blog Post, Ofwat.* Available at: https://www.ofwat.gov.uk/rapid-and-safeguarding-the-uks-water-supply-for-the-future/, Last Accessed: 2023-07-12.

—, February (2023), “Consultation on PR24 Operational Greenhouse Gas Emissions Performance Commitments Definitions,” *Technical Report, Ofwat.* Available at: https://www.ofwat.gov.uk/wp-content/uploads/2023/02/GHG_PCs_definition_consultation_February_2023.pdf, Last Accessed: 2023-07-12.

Palomo-Hierro, Sara, José A. Gómez-Limón, and Laura Riesgo, (2015), “Water Markets in Spain: Performance and Challenges,” *Water*, *7* (2), 652–678.

Reffold, Emma, Feiferi Leighton, Fida Choudhury, and Paul S. Rayner, (2008), “Greenhouse Gas Emissions of Water Supply and Demand Management Options,” *UK Environmental Agency Science Report, SC070010.*

Reguant, Mar, (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.

Rennert, Kevin, Frank Errickson, Brian C Prest, Lisa Rennels, Richard G Newell, William Pizer, Cora Kingdon, Jordan Wingenroth, Roger Cooke, Bryan Parthum et al., (2022), “Comprehensive Evidence Implies a Higher Social Cost of CO2,” *Nature*, *610* (7933), 687–692.

Rodemeier, Matthias and Andreas Löschel, (2022), “Information Nudges, Subsidies, and Crowding Out of Attention: Field Evidence from Energy Efficiency Investments,” *Working Paper, Available at SSRN 4213071 (September 8, 2022).*

Rodríguez, Catherine and Juan E Saavedra, (2019), “The persistent effects of youth savings reminders: Experimental evidence from text-message campaigns in Colombia,” *Journal of Development Economics*, *139*, 135–156.

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 https://www.royalmail.com/sites/royalmail.com/files/2020-12/Marketing_services_and_general_correspondence_price_guide_1_January_2021_v2.pdf, Last Accessed: 2023-07-12.

Rupiper, Amanda, Joakim Weill, Ellen Bruno, Katrina Jessoe, and Frank Loge, (2022), "Untapped Potential: Leak Reduction is the Most Cost-Effective Urban Water Management Tool," *Environmental Research Letters*, *17* (3), 034021.

Stuart, Elizabeth A., Stephen R. Cole, Catherine P. Bradshaw, and Philip J. Leaf, (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, Reinhard, Katherine Gasner, and Lucy Andrews, (2015), "Water Audits in the United States: A Review of Water Losses and Data Validity," *Water Research Foundation.*

Tiefenbeck, Verena, Lorenz Goette, Kathrin Degen, Vojkan Tasic, Elgar Fleisch, Rafael Lalive, and Thorsten Staake, (2018), "Overcoming Salience Bias: How Real-Time Feedback Fosters Resource Conservation," *Management Science*, *64* (3), 1458–1476.

Torres, Mónica Marcela Jaime and Fredrik Carlsson, (2016), "Social Norms and Information Diffusion in Water-saving Programs: Evidence from a Randomized Field Experiment in Colombia," *Environment for Development Discussion Paper-Resources for the Future (RFF)*, (16-05).

United Kingdom Government, December (2017), "Water Meters: Question for Department for Environment, Food and Rural Affairs," *Written Questions, Answers and Statements.* Available at: https://questions-statements.parliament.uk/written-questions/detail/2017-12-21/121058, Last Accessed: 2024-07-14.

—, (2019), "Environmental Reporting Guidelines: Including Streamlined Energy and Carbon Reporting Guidance," *Technical Report*, *HM Government London.* Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/850130/Env-reporting-guidance_inc_SECR_31March.pdf, Last Accessed: 2023-07-12.

—, December (2019), "Water Meters," *Constituency Casework, House of Commons Library.* Available at: https://commonslibrary.parliament.uk/water-meters/, Last Accessed: 2023-07-12.

—, September (2021), "Valuation of Greenhouse Gas Emissions: For Policy Appraisal and Evaluation," *Policy Paper.* Available at: https://www.gov.uk/government/publications/valuing-greenhouse-gas-emissions-in-policy-appraisal/valuation-of-greenhouse-gas-emissions-for-policy-appraisal-and-evaluation#introduction, Last Accessed: 2024-07-23.

US Bureau of Labor Statistics, (2021), "All Urban Consumers (CPI-U)," *Washington, DC: US Department of Labor, Bureau of Labor Statistics.*

West, Jeremy D., Robert W. Fairlie, Bryan E. Pratt, and Liam Rose, (2021), "Automated Enforcement of Irrigation Regulations and Social Pressure for Water Conservation," *Technical Report*, *National Bureau of Economic Research.*

White, Vinn, (2016), "Revised Departmental Guidance on Valuation of Travel Time in Economic Analysis," *Office of the Secretary of Transportation, US Department of Transportation.* Available at: https://www.transportation.gov/office-policy/transportation-policy/revised-departmental-guidance-valuation-travel-time-economic, Last Accessed: 2023-07-12.

Whittington, Dale, W. Michael Hanemann, Claudia Sadoff, and Marc Jeuland, (2009), "The Challenge of Improving Water and Sanitation Services in Less Developed Countries," *Now Publishers Inc.*

Wichman, Casey J, (2017), "Information Provision and Consumer Behavior: A Natural Experiment in Billing Frequency," *Journal of Public Economics*, *152*, 13–33.

Wichman, Casey J., (2023), "The Unequal Burdens of Water Scarcity," *Nature Water*, *1* (1), 26–27.

_, Laura O. Taylor, and Roger H. Von Haefen, (2016), "Conservation Policies: Who Responds to Price and Who Responds to Prescription?," *Journal of Environmental Economics and Management*, *79*, 114–134.

World Bank, (2019), "The Role of Desalination in an Increasingly Water-Scarce World," *World Bank.*

Zhang, C Yiwei, Jeffrey Hemmeter, Judd B Kessler, Robert D Metcalfe, and Robert Weathers, (2023), "Nudging Timely Wage Reporting: Field Experimental Evidence from the US Supplemental Security Income Program," *Management Science*, *69* (3), 1341–1353.

Appendix

Table of Contents

The Appendix is divided into eight sections. Appendix A provides balance tables for the initial direct mailer treatment and the reminder treatment. It also provides statistics on diagnostic completion across the two experiments. Appendix B provides additional results on heterogeneity of treatment effects — both diagnostic completion and water conservation — based on pre-treatment water consumption. Appendix C presents results from a set of robustness checks meant to address concerns relating to a lack of clear pre-post delineation in meter readings and measurement error in post-treatment consumption estimates. Appendix D discusses three additional results. First, we compare the characteristics of households who complete the diagnostic versus those who did not. Second, we estimate the effect of the reminders on diagnostic completion. Finally, we analyze how households interact with the reminder emails, and how this differed across treatment arms. Appendix E 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. Appendix F sheds light on the measurement of pre- and post-treatment water consumption data using an illustrative example. It also provides information on the monthly distribution of meter readings. Appendix G lists the questions in the online follow-up survey that was administered in March 2019 to NWG customers that had an email address. Finally, in Appendix H, we provide samples of the different letters and reminders that were sent to households.

A Baseline Balance Tests and Summary Statistics

A.1 Balance Tables for Initial Direct Mailers

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 for the initial direct mailer treatment. 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 NWG had an email address, 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. It is important to note that water consumption data was only available for metered customers and, therefore, columns (6) to (9) pertain to the sample with meters. In this regard, columns (6) and (7) report the number of metered households living in rural areas who provided NWG with an email id. Finally, columns (8) and (9) provide balance on pre-treatment water consumption, and how many consumers within each treatment group were 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. Predominantly, we find that the covariates for the treatment arms are not significantly different from the covariates in the control group. The only significant differences (at 10 percent) are: *a*) metered *Vanilla* households have a lower probability (67 percent versus 70 percent in the control group) of living in rural areas, and *b*) fewer metered customers in the

Incentives 10 group (41 percent versus 44 percent in the control group) had registered their email ids with NWG. We, therefore, control for these covariates in our regressions.

Column (10) 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.

A.2 Statistics on Diagnostic Completion

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 (in contrast to the proportion of metered customers), and this is consistent across all treatment groups.

A.3 Balance Tables for Customers with Consumption Data

Table A.3 provides a measure of the balance on observed covariates across different treatment groups for the initial direct mailer treatment. It is limited to the sample for which we have consumption data. Refer to Section 3.2 and Appendix F.1 for details on sample selection. Column (1) reports the number of people with consumption data in each treatment group. Columns (2) and (3) provide the percentage of population with water consumption data living in a rural area, and for whom NWG had an email address, respectively. Columns (4) and (5) 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. In a majority of the cases, we find that the covariates for the treatment arms are not significantly different from the covariates in the control group. The only significant differences at 5 percent are: *a*) *Vanilla* households have a lower probability (62 percent versus 66 percent in the control group) of living in rural areas, and *b*) fewer Top 50% customers in the *Incentives 15* group (46 percent versus 51 percent in the control group). The significant differences at 10 percent are: *a*) *Altruism* households have a lower probability (63 percent versus 66 percent in the control group) of living in rural areas, and *b*) *Moral Cost* group has a higher average daily pre-treatment water consumption than the control group (262 liters per day versus 252 liters per day in the control

Table A.1: Baseline Balance Across Treatment Groups for Initial Direct Mailers

	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-Treatment 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	.701	.313	.324	0.685	0.434	.258	.499	
		(.002)	(.002)	(.002)	(.002)	(.003)	(0.004)	(.003)	(.004)	
Control	7,459	.427	.706	.312	.320	0.695	0.441	.257	.499	
		(.006)	(.005)	(.005)	(.005)	(0.008)	(0.010)	(.007)	(.010)	
Vanilla	7,460	.428	.696	.316	.327	0.674	0.433	.266	.498	{.296}
		(.006)	(.005)	(.005)	(.005)	(0.008)	(0.010)	(.007)	(.010)	
		[.887]	[.206]	[.638]	[.339]	[0.066]	[.581]	[.360]	[.921]	
Simplified	7,460	.428	.703	.313	.321	0.692	0.435	.256	.511	{.501}
		(.006)	(.005)	(.005)	(.005)	(0.008)	(0.010)	(.007)	(.010)	
		[.848]	[.697]	[.920]	[.852]	[.779]	[0.709]	[.855]	[.427]	
Altruism	7,460	.429	.699	.314	.324	0.685	0.442	.253	.500	{.486}
		(.006)	(.005)	(.005)	(.005)	(0.008)	(0.010)	(.007)	(.010)	
		[.797]	[.355]	[.809]	[.591]	[.386]	0.900	[.661]	[.966]	
Incentives 10	3,789	.436	.699	.320	.332	0.685	0.409	.248	.482	{.252}
		(.008)	(.007)	(.008)	(.008)	(0.011)	(0.014)	(.010)	(.014)	
		[.356]	[.488]	[.372]	[.189]	[.445]	[0.062]	[.451]	[.300]	
Incentives 15	3,670	.423	.698	.303	.324	0.678	0.440	.254	.472	{.499}
		(.008)	(.008)	(.008)	(.008)	(0.012)	(0.014)	(.010)	(.015)	
		[.670]	[.409]	[.313]	[.633]	[.219]	[0.985]	[.752]	[.121]	
Moral Cost	7,459	.430	.701	.309	.323	0.683	0.432	.263	.513	{.656}
		(.006)	(.005)	(.005)	(.005)	(0.008)	(0.010)	(.007)	(.010)	
		[.741]	[.519]	[.684]	[.700]	[.270]	[.529]	[.557]	[.373]	
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 (see Section 3.2 and Appendix F.1 for more details). Columns (8) and (9) provide the balance on pre-treatment water consumption and *Top 50% Consumers*. The latter represents the number of households within each treatment group who are in the top 50^{th} percentile of pre-treatment water consumption across the entire distribution of households. Column (10) 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 taking a value of 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 (% of Customers)	Metered Customers (% of Customers)	Metered Customers who Completed Audit (% 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 corresponding 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 corresponding treatment group are reported in parenthesis.

group). We, therefore, control for these covariates in our regressions.

Column (6) 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. We fail to reject the null hypothesis that all coefficients are 0, except for the variable *Top 50% consumers*. We control for this variable in our regressions.

A.4 Summary Statistics for Households who Completed Survey

Table A.4 provides summary statistics on which households received the follow-up survey in March 2019, the number of households who completed the survey, and the overlap between diagnostic and survey completion. This information is provided both for the entire population, and by treatment group. Column (1) reports the total number of customers in each treatment group, with the numbers in parenthesis representing the percentage of people in the treatment group relative to the total number of participating households. Column (2) reports the number of customers within each treatment group who received the survey, both in absolute terms, and as a percentage of total households in the respective treatment arm. Recall that surveys were only sent to households for whom NWG had email contact details. Columns (3) and (4) report the number of people who completed the survey and diagnostic, respectively. Columns (6) and (7) provide information on the number of households who completed either only the survey, or only the diagnostic. Columns (5) and (8), on the other hand, report the number of people who completed both the survey and diagnostic, or neither the survey and the diagnostic, respectively. Finally, all numbers in parenthesis in columns (3) to (8) represent the respective absolute numbers as a proportion of households in the group who received the follow-up survey.

A.5 Balance Tables for Customers in Reminder Treatment

This section provides details on balance between reminder treatment and control groups across various pre-treatment covariates. Table A.5 presents the results. Column (1) reports the number of people within each initial direct mailer treatment group who were randomized into receiving and not receiving a reminder. Columns (2) through (4) provide the percentage of the population with a water meter, living in a rural area, and for whom NWG had water data. Note that we do not check for balance on whether NWG had email contact information available because the reminders could only be sent to households with an email address. Therefore, by construction, NWG had email

Table A.3: Baseline Balance Across Treatment Groups (Households with Consumption Data)

	Number of Customers	Lives in Rural Area	Provided an Email	Pre-Treatment Consumption (m^3/day)	Top 50% Consumers	F-test of Joint Significance
	(1)	(2)	(3)	(4)	(5)	(6)
All Customers	11,940	.641 (.004)	.435 (.005)	.255 (.002)	.499 (.005)	
Control	1,962	.656 (.011)	.441 (.011)	.252 (.004)	.505 (.011)	
Vanilla	2,017	.624 (.011) [.034]	.436 (.011) [.771]	.260 (.004) [.201]	.501 (.011) [.807]	{.156}
Simplified	1,966	.659 (.011) [.857]	.430 (.011) [.505]	.257 (.004) [.380]	.510 (.011) [.775]	{.770}
Altruism	2,010	.628 (.011) [.065]	.448 (.011) [.640]	.254 (.004) [.779]	.501 (.011) [.796]	{.376}
Incentives 10	1,040	.650 (.015) [.723]	.411 (.015) [.110]	.251 (.006) [.921]	.476 (.016) [.129]	{.155}
Incentives 15	968	.633 (.015) [.218]	.435 (.016) [.760]	.244 (.006) [.325]	.459 (.016) [.018]	{.105}
Moral Cost	1,977	.634 (.011) [.137]	.434 (.011) [.687]	.262 (.004) [.096]	.510 (.011) [.741]	{.101}
F-test of Joint Significance		{.134}	{.620}	{.233}	{.095}	

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. This balance table only includes customers for which we had usable water consumption data, *i.e.* 11,970 customers (see Section 3.2 and Appendix F.1 for more details). 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. *Top 50% Consumers* (column (5)) represents the number of households within each treatment group who are in the top 50^{th} percentile of pre-treatment water consumption across the entire distribution of households. Robust standard errors are reported in parenthesis throughout. Column (6) 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 taking a value of 1 for customers assigned to treatment group j, and the independent variables are the variables in columns (2) to (5). The p-values reported in the last row are from the F-test of joint significance of the treatment dummies in each column regression.

Table A.4: Statistics on Survey and Diagnostic Completion

	Total Customers (% of All) (1)	Received Survey (% of Total) (2)	Completed Survey (% of Received) (3)	Completed Diagnostic (% of Received) (4)	Completed Survey & Diagnostic (% of Received) (5)	Completed only Survey (% of Received) (6)	Completed only Diagnostic (% of Received) (7)	Not Complete Survey or Diagnostic (% of Received) (8)
All Customers	44,757	13,984	862	781	156	706	625	12,497
		(31.2)	(6.2)	(5.6)	(1.1)	(5.0)	(4.5)	(89.4)
Control	7,459	2,328	152	1	0	152	1	2,175
	(16.7)	(31.2)	(6.5)	(0.0)	(0.0)	(6.5)	(0.0)	(93.4)
Vanilla	7,460	2,354	132	94	18	114	76	2,146
	(16.7)	(31.6)	(5.6)	(4.0)	(0.8)	(4.8)	(3.2)	(91.2)
Simplified	7,460	2,334	149	113	22	127	91	2,094
	(16.6)	(31.3)	(6.4)	(4.8)	(0.9)	(5.4)	(3.9)	(89.7)
Altruism	7,460	2,341	144	122	24	120	98	2,099
	(16.6)	(31.4)	(6.2)	(5.2)	(1.0)	(5.1)	(4.2)	(89.7)
Incentives £10	3,789	1,214	77	143	26	51	117	1,020
	(8.4)	(32.0)	(6.3)	(11.8)	(2.1)	(4.2)	(9.6)	(84.0)
Incentives £15	3,670	1,111	85	145	35	50	110	916
	(8.2)	(30.3)	(7.7)	(13.1)	(3.2)	(4.5)	(9.9)	(82.4)
Moral Cost	7,459	2,302	123	163	31	92	132	2,047
	(16.6)	(30.9)	(5.3)	(7.1)	(1.3)	(4.0)	(5.7)	(88.9)

Notes: This table provides details on the two phases of the natural field experiment. The first phase involved randomizing six different letter treatments to 44,757 NWL customers, with direct mailers posted on 8th December 2018. The second phase involved randomizing a reminder treatment for households who did not complete the audit up until 6th February 2019, and for whom NWL had email contact details. Column (1) provides details on the number of customers in the control and each of the six different treatment groups for the first phase of the field experiment. The percentages in each row denote the proportion of total NWL customers in the respective group. Column (2) lists the number of households who completed the diagnostic in each group. The percentages represent the proportion of people in the group who completed the audit relative to the total number of households who completed the audit. Column (3) represents the number of households for which NWL had email contact details. The percentages represent the proportion of people in that specific treatment or control group for which NWL had contact details. Column (4) provides the number of households for which NWL did not have contact details but who had completed the audit before 6th February 2019. Column (5) lists the number of households for which NWL had contact details (part of column (3)), but who had completed the audit before 6th February 2019. The percentages in column (5) and (6) represent the proportion of completed audits relative to the total number of completed audits in that group. Columns (6) and (7) list the number of households in the reminder treatment and control group by letter type. Therefore, the number of households who were a part of the reminder experiment (11,031) would be the total of column (5) subtracted from the total of column (3), not including the control group in either case (control group was not part of the reminder randomization). The percentages represent the proportion of households in the two groups relative to the total number of households in the first phase of the randomization. Column (8) represents the number of people who completed the audit after 6th February 2019, but were not in the reminder treatment group. Column (9) represents the number of consumers who were in the reminder treatment group and completed the audit post 6th February 2019. Finally, the percentages in the last two columns represent the proportion of completed audits relative to the total number of completed audits in that group.

contact details for the entire population in the reminder experiment. We also check for balance within the sub-sample of customers with meters. Note that water consumption data was only available for metered customers and, therefore, columns (5) through (7) pertain to the sub sample with meters. In this regard, columns (5) reports the number of metered households living in rural areas. Finally, columns (6) and (7) provide balance on pre-treatment water consumption, and how many consumers within each treatment group were in the top 50th percentile of water consumption distribution for the sample in the reminder experiment.

We calculated the p-value on t-test of equality of means with *Vanilla* group and the same is reported in brackets.[98] We find that almost all the covariates for the different reminder treatment arms are not significantly different from the covariates in the *Vanilla* reminder group. The only significant difference (at 5 percent) is metered *Simplified* households have a higher probability (69 percent versus 65 percent in the *Vanilla* group) of living in rural areas. We, therefore, control for this covariate in our regressions.

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 not receive the reminder within the initial treatment group (control), and it takes a value of 1 for customers assigned to receive the reminder (treatment).[99] Note that the assignment we refer to here is the allocation to treatment and control within each initial direct mailer treatment. For *e.g.*, 1,168 households within the *Vanilla* group were assigned to receive the reminder treatment, while 1,111 households were assigned to not receive one (see Table A.6 for details). A significant F-test would represent that within each initial direct mailer treatment group, covariates can predict whether or not households received or did not receive a reminder. We fail to reject the null hypothesis that all coefficients are 0. 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 of all coefficients not being significantly different from 0.

A.6 Statistics on Households in Reminder Treatment

Customers for which NWG had email contact details were randomly allocated to groups that either received or did not receive an email reminder. The reminder emails followed the same themes as the initial direct mailers that the customers received. These reminders were sent on 6th February 2019, two months after the direct mailers were first sent out. Table A.6 provides details on the sample chosen for the reminder treatment, the proportion of people who completed the audit prior to and post the reminders, and how that differed by treatment.

[98] The control group in the initial direct mailer experiment was not a part of the subsequent reminder experiment.

[99] The regressors include rural/urban classification, pre-treatment consumption, and dummy for whether the household had pre-treatment consumption greater than the median.

Table A.5: Baseline Balance Across Treatment Groups: Reminder Treatment

	Number of Customers	Has a Water Meter	Lives in Rural Area	Water Consumption Data Available	Lives in Rural Area (Metered h/h)	Pre-Treatment Consumption (m^3/day)	Top 50% Consumers	F-test of Joint Significance
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All Customers	11,661	.586 (.005)	.666 (.004)	.449 (.005)	0.657 (.006)	.269 (.006)	.499 (.007)	
Vanilla	2,355	.586 (.010)	.658 (.009)	.448 (.010)	0.647 (0.013)	.283 (.013)	.487 (.015)	{.393}
Simplified	2,334	.587 (.010) [.898]	.676 (.009) [.204]	.447 (.010) [.916]	0.686 (0.013) [.028]	.257 (.013) [.175]	.493 (.015) [.781]	{.111}
Altruism	2,342	.596 (.010) [.446]	.669 (.009) [.447]	.456 (.010) [.580]	0.659 (0.013) [.517]	.261 (.013) [.250]	.518 (.015) [.147]	{.315}
Incentives £10	1,214	.568 (.014) [.325]	.671 (.014) [.459]	.423 (.014) [.155]	0.657 (0.018) [.663]	.274 (.019) [.712]	.516 (.022) [.284]	{.608}
Incentives £15	1,111	.594 (.015) [.635]	.670 (.014) [.503]	.472 (.015) [.199]	0.645 (0.018) [.951]	.255 (.019) [.219]	.460 (.022) [.316]	{.507}
Moral Cost	2,305	.580 (.010) [.703]	.658 (.009) [.973]	.451 (.010) [.872]	0.641 (0.013) [.750]	.277 (.014) [.759]	.510 (.016) [.285]	{.705}
F-test of Joint Significance		{.658}	{.762}	{.298}	{.168}	{.667}	{.248}	

Notes: Robust standard errors from OLS regressions are in parenthesis.
P-value on t-test of equality of means with *Vanilla* 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 the reminder experiment (treatment + control) within each direct mailer group. Note that the control group in the direct mailer experiment was not a part of the reminder experiment. Columns (2) to (4) 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 *Vanilla* group. Robust standard errors are reported in parenthesis throughout. Columns (5) reports balance on rural/urban location for a specific sub sample: *metered* households (see Section 3.2 and Appendix F.1 for more details). Columns (6) and (7) provide the balance on pre-treatment water consumption and *Top 50% Consumers*. The latter represents the number of households within each treatment group who are in the top 50^{th} percentile of pre-treatment water consumption across the entire distribution of households. 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 in the reminder experiment within direct mailer group j, and taking a value of 1 for customers assigned to reminder treatment group within direct mailer group j. The independent variables are the variables in columns (5), (6) and (7). The sample only includes observations for which we have consumption data available, and therefore *Water Consumption Data Available* (column (4)) 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 (5) to (7), in which case the sample only includes observations for which we have consumption data available.

Table A.6: Summary Statistics on Initial Direct Mailers and Reminder Treatment

	Number of Customers (% of Total) (1)	Completed Audit (% of Completed) (2)	Provided an Email (% of Column (1)) (3)	No Email & Completed Before Reminder (% of Column (2)) (4)	Email & Completed Before Reminder (% of Column (2)) (5)	Reminder Treatment (% of Column (1)) (6)	Reminder Control (% of Column (1)) (7)	Completed After Reminder & Not Treated (% of Column (2)) (8)	Completed After Reminder & Treated (% of Column (2)) (9)
Control	7,459	3	2,328	1	1	-	-	1	-
	(16.6)	(0.2)	(31.2)	(33.3)	(33.3)			(33.3)	
Vanilla	7,460	140	2,355	43	76	1,168	1,111	3	18
	(16.6)	(10.9)	(31.6)	(30.7)	(54.3)	(15.7)	(14.9)	(2.1)	(12.9)
Simplified	7,460	189	2,334	73	85	1,097	1,152	3	28
	(16.6)	(14.7)	(31.3)	(38.6)	(44.9)	(14.7)	(15.4)	(1.6)	(14.8)
Altruism	7,460	176	2,342	53	97	1,141	1,104	1	25
	(16.6)	(13.7)	(31.4)	(30.1)	(55.1)	(15.3)	(14.8)	(0.6)	(14.2)
Incentives £10	3,789	242	1,214	98	126	564	524	1	17
	(8.4)	(18.8)	(32.0)	(40.5)	(52.1)	(14.9)	(13.8)	(0.4)	(7.0)
Incentives £15	3,670	278	1,111	132	132	493	486	1	13
	(8.2)	(21.6)	(30.3)	(47.5)	(47.5)	(13.4)	(13.2)	(0.4)	(4.7)
Moral Cost	7,459	259	2,305	92	114	1,100	1,091	3	50
	(16.6)	(20.1)	(30.9)	(35.5)	(44.0)	(14.8)	(14.6)	(1.2)	(19.3)
Total	**44,757**	**1,287**	**13,989**	**492**	**631**	**5,563**	**5,468**	**15**	**149**
	(100.0)	**(100.0)**	**(31.2)**	**(38.2)**	**(49.0)**	**(14.9)**	**(14.7)**	**(1.0)**	**(11.8)**

Notes: This table provides details on the two phases of the natural field experiment. The first phase involved randomizing six different letter treatments to 44,757 NWL customers, with direct mailers posted on 8th December 2018. The second phase involved randomizing a reminder treatment for households who did not complete the audit up until 6th February 2019, and for whom NWL had email contact details. Column (1) provides details on the number of customers in the control and each of the six different treatment groups for the first phase of the field experiment. The percentages in each row denote the proportion of total NWL customers in the respective group. Column (2) lists the number of households who completed the diagnostic in each group. The percentages represent the proportion of people in the group who completed the audit relative to the total number of households who completed the audit. Column (3) represents the number of households for which NWL had email contact details. The percentages represent the proportion of people in that specific treatment or control group for which NWL had contact details. Column (4) provides the number of households for which NWL did not have contact details but who had completed the audit before 6th February 2019. Column (5) lists the number of households for which NWL had contact details (part of column (3)), but who had completed the audit before 6th February 2019. The percentages in column (5) and (6) represent the proportion of completed audits relative to the total number of completed audits in that group. Columns (6) and (7) list the number of households in the reminder treatment and control group by letter type. Therefore, the number of households who were a part of the reminder experiment (11,031) would be the total of column (5) subtracted from the total of column (3), not including the control group in either case (control group was not part of the reminder randomization). The percentages represent the proportion of households in the two groups relative to the total number of households in the first phase of the randomization. Column (8) represents the number of people who completed the audit after 6th February 2019, but were not in the reminder treatment group. Column (9) represents the number of consumers who were in the reminder treatment group and completed the audit post 6th February 2019. Finally, the percentages in the last two columns represent the proportion of completed audits relative to the total number of completed audits in that group.

Column (1) provides details on the number of customers in the control and each of the six different treatment groups for the initial field experiment that involved sending direct mailers to households. The percentages in each row denote the proportion of total NWG customers in the respective group. Column (2) lists the total number of households who completed the diagnostic in each group, either pre- or post-reminders. The percentages represent the proportion of people in the group who completed the audit relative to the total number of completed audits across all groups. Column (3) represents the number of households for which NWG had email contact details. The percentages represent the proportion of people in that specific group for which NWG had contact details. It ranged from 30.3 to 32.0 percent. Note that this is also the population of households who could potentially be in the reminder treatment as reminders were sent via email, entailing email contact details a pre-requisite.

Column (4) provides the number of households for which NWG did not have contact details but who had completed the audit before 6^{th} February 2019. Column (5) lists the number of households for which NWG had contact details, but who had completed the audit before 6^{th} February 2019. Since this group had already completed the diagnostic before reminders were sent out, they were not included in the subsequent randomization. The percentages in column (5) and (6) represent the proportion of completed audits relative to the total number of completed audits in that group. Across all groups, approximately 80 to 95 percent of the audits had already been completed before reminders were sent.

Columns (6) and (7) list the number of households in the reminder treatment and control group by letter type. The percentages represent the proportion of households in the two groups relative to the total number of households in the first phase of the randomization. The numbers indicate that around 26 to 28 percent of each group was a part of the reminder experiment. Approximately half of this number got the reminder, while the other half did not. Note that the number of households who were a part of this experiment (11,031) would be the difference between the total population for which NWG had email contact details and the population for which NWG had contact details but who had completed the audit before 6^{th} February 2019 (total of column (5) subtracted from the total of column (3)). Also, the control group in the initial direct mailer experiment was not a part of the reminder experiment and, therefore, we remove this sub-group to arrive at the final sample.

Finally, column (8) represents the number of people who completed the audit after 6^{th} February 2019, but were not in the reminder treatment group. Column (9), on the other hand, represents the number of consumers who were in the reminder treatment group and completed the audit post 6^{th} February 2019. Comparing these two groups would indicate the effectiveness of the reminders, an exercise which we conduct in Appendix D.2. The percentages in the last two columns represent the proportion of audits relative to the total number of completed audits in each group.

A.7 Comparison Between Compliers and Non-Compliers

Table A.7 provides summary statistics on observable characteristics for households that completed the audit (*Compliers*) and households that did not (*Non-Compliers*). This information is provided for the entire population and by treatment group. We expect compliers and non-compliers to be different from each other, and summary statistics reflect this.

Columns (1) and (2) report the number of compliers and non-compliers, respectively. The percentage of households within each treatment group that complied and did not not comply are reported in brackets. The percentage of compliers ranged from 7.6 percent in the *Incentives 15* group to almost 0 percent in the control group. Columns (3) to (12) report the mean value of each customer characteristic, derived from OLS regressions of the characteristic of interest on a series of dummy variables for each treatment group (except row 1 which reports the average for the entire sample). The regressions were run separately for compliers and non-compliers.

The excluded (comparison) group in these regressions is the control group.

Columns (3) and (4) represent the proportion of compliers and non-compliers who had a water meter. The percentage of complier households with a water meter is consistently higher than the corresponding percentage for non-compliers across all treatment groups. Columns (5)-(12) report similar percentages for compliers and non-compliers for other customer characteristics, including whether they live in rural areas, whether they provided NWG with an email, their pre-treatment water consumption, and whether they are in the top 50$^{\text{th}}$ percentile of pre-treatment water usage across the entire distribution of households. The results highlight the differences between households who completed the diagnostic and households that did not. Compliers are less likely to live in rural areas, more likely to have provided an email to NWG, have slightly lower pre-treatment water consumption, and are also less likely to be in the top 50$^{\text{th}}$ percentile of consumers in terms of pre-treatment usage.

In conclusion, these results indicate that compliers were very different from non-compliers. This is expected given the motivation to complete the audit varies. It is also the primary reason why we use an IV strategy to estimate the impact of completing the diagnostic on water conservation. Running a simple OLS would likely lead to biased estimates as households who completed the diagnostic could be different across both observables and unobservables. We, thus, use our LATE analysis to obtain unbiased estimates.

B Heterogeneity Analysis

Targeting households based on some pre-treatment covariates may be a more cost-effective intervention for utilities if there is heterogeneity in response to the letters. This would allow the utility to only target the subgroup of households who are most likely to respond to the treatment. In our field experiment, heterogeneity can manifest in two domains: (*a*) diagnostic completion, and; (*b*) post-treatment water consumption. Analysis of heterogeneity in diagnostic completion is pre-

Table A.7: Comparison Between Compliers and Non-Compliers

	Number of Customers		Has a Water Meter		Lives in Rural Area		Provided an Email		Pre-Treatment Consumption (m^3/day)		Top 50% Consumers	
	Compliers	Non-Compliers	Compliers	Non-Compliers	Compliers	Non-Compliers	Compliers	Non-Compliers	Compliers	Non-Compliers	Compliers	Non-Compliers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
All Customers	1,287	43,470	.668	.421	.664	.702	.608	.304	.239	.259	.492	.500
	[2.9]	[97.1]	(.013)	(.002)	(.013)	(.002)	(.014)	(.002)	(.006)	(.003)	(.019)	(0.004)
Control	3	7,456	1	.427	.667	.706	.333	.312	.187	.258	.333	.500
	[0.0]	[99.9]	(.000)	(.006)	(.273)	(.005)	(.273)	(.005)	(.052)	(.008)	(.274)	(0.010)
Vanilla	140	7,320	.729	.422	.614	.698	.671	.309	.249	.267	.544	.497
	[1.9]	[98.1]	(.034)	(.006)	(.040)	(.005)	(.041)	(.005)	(.017)	(.007)	(.056)	(0.010)
Simplified	189	7,271	.704	.421	.651	.704	.598	.305	.233	.257	.549	.509
	[2.5]	[97.5]	(.034)	(.006)	(.034)	(.005)	(.035)	(.005)	(.014)	(.007)	(.047)	(0.010)
Altruism	176	7,284	.676	.423	.665	.700	.693	.305	.225	.254	.447	.503
	[2.4]	[97.6]	(.033)	(.008)	(.036)	(.005)	(.037)	(.005)	(.015)	(.007)	(.052)	(0.010)
Incentives £10	242	3,547	.562	.427	.698	.699	.591	.302	.236	.250	.466	.483
	[6.4]	[93.6]	(.030)	(.008)	(.030)	(.008)	(.031)	(.008)	(.015)	(.010)	(.049)	(0.015)
Incentives £15	278	3,392	.579	.410	.680	.700	.522	.285	.225	.257	.442	.476
	[7.6]	[92.4]	(.028)	(.008)	(.028)	(.008)	(.029)	(.008)	(.013)	(.011)	(.044)	(0.015)
Moral Cost	259	7,200	.795	.416	.649	.703	.633	.297	.259	.264	.512	.513
	[3.5]	[96.5]	(.029)	(.006)	(.029)	(.005)	(.030)	(.005)	(.011)	(.007)	(.039)	(.011)

Notes: Robust standard errors from OLS regressions are in parenthesis.
All data was provided by Northumbrian Water Limited. Columns (1) and (2) report the number of customers who completed (*Compliers*) and did not complete the online diagnostic (*Non-Compliers*), respectively. Percentage of households within each treatment group which complied and did not not comply are reported in brackets. Columns (3) to (12) report the mean value of each customer characteristic, derived from an OLS regression — run separately for *Compliers* and *Non-Compliers* — 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. Top 50% Consumers (column (11) and (12)) represents the number of households within each treatment group who are in the top 50^{th} percentile of pre-treatment water consumption across the entire distribution of households. Robust standard errors are reported in parenthesis throughout.

sented in Appendix B.1, while heterogeneity in post-treatment water consumption is discussed in Appendix B.2.

B.1 Heterogeneity in Diagnostic Completion

We test for heterogeneity in diagnostic completion based on three household characteristics: rural, metered, and pre-treatment water consumption. Our regression framework is as follows:

$$y_i = \alpha + \sum_j \beta_j T_{ij} + \phi\, Z_i + \sum_j \eta_j\, Z_i \times T_{ij} + \gamma \mathbf{X}_i + \epsilon_i \tag{11}$$

where, y_i is a dummy variable that equals 1 if household *i* completed the water audit, and 0 otherwise. T_{ij} is a dummy that equals 1 if household i received treatment j, where j refers to the different treatment groups. Depending on the regression, Z_i refers to one of the three different pre-treatment covariates of interest: *Meter$_i$* is a binary variable that equals 1 if household i has a water meter; *Rural$_i$* is a dummy that equals 1 if household i resided in a rural area, and; *High-Use$_i$* which is also a dummy and equals 1 if household i had a pre-treatment water consumption greater than the median pre-treatment consumption in the sample. γ is a vector of estimates for the different dummy controls, represented by $\mathbf{X}_i$, for household i. These controls will vary according to the regression and include *Meter$_i$*, *Rural$_i$* and *Pre-Treatment Water Consumption$_i$* in liters per day. Finally, ϵ_i is the error term. Results for heterogeneity in diagnostic completion based on status of metering as well as place of residence (urban/rural) are presented in Table B.1.

We find that the metering status had an impact on diagnostic completion, but there was no differential impact of whether the household was in a rural or urban area. 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. Models (1) and (3) check for heterogeneity based on whether households had a meter installed. Note that in order to compute the differential impact of a treatment (T_{ij}) based on a covariate Z_i, we need to sum up the coefficients ϕ and η_j in Equation (11), and then check if they were significantly different from 0. For example, if we wanted to check whether the *£10 Incentives* treatment had a different impact for metered versus unmetered customers, we add up the coefficient on Z_i (metering status of household i), *i.e.* ϕ, and the coefficient on the interaction term between *£10 Incentives* and Z_i, which is η_j (where j refers to the *£10 Incentives* treatment). We find that for all treatment groups and notwithstanding the excluded group, metered customers were always significantly more likely to complete the diagnostic relative to unmetered customers. Specifically, for customers treated with the *Vanilla* letters, the probability of completing the diagnostic was 2.3 percentage points (100 percent) higher if the household had a meter installed. The highest difference was for the *Moral Cost* letter, where metered customers were 5.2 percentage points (330 percent) more likely to complete the diagnostic (column 1). Models (2) and (4) show regression results when we analyze whether the location of the household in an urban or rural area impacted their interaction with the diagnostic. Our results indicate that there was no impact of household location on the probability of take-up of the audit. This is true irrespective of whether the excluded group is the *Vanilla* or the *Simplified* letter.

Table B.1: Heterogeneity in Diagnostic Completion Based on Household Characteristics

	Completed Diagnostic			
	Vanilla		Simplified	
	(1)	(2)	(3)	(4)
Rural	-0.005^{**}	-0.006	-0.004^{*}	-0.006
	(0.002)	(0.004)	(0.002)	(0.004)
Meter	0.023^{***}	0.033^{***}	0.028^{***}	0.036^{***}
	(0.003)	(0.002)	(0.004)	(0.002)
Simplified	0.004^{*}	0.006		
	(0.002)	(0.005)		
Altruism	0.004^{**}	0.003	0.000	-0.004
	(0.002)	(0.005)	(0.002)	(0.005)
Incentives £10	0.041^{***}	0.040^{***}	0.036^{***}	0.034^{***}
	(0.005)	(0.008)	(0.005)	(0.008)
Incentives £15	0.046^{***}	0.057^{***}	0.042^{***}	0.050^{***}
	(0.005)	(0.009)	(0.005)	(0.009)
Moral Cost	0.004	0.017^{***}	-0.001	0.011^{*}
	(0.002)	(0.005)	(0.002)	(0.005)
Rural × Simplified		0.000		
		(0.006)		
Rural × Altruism		0.003		0.003
		(0.005)		(0.006)
Rural × Incentives £10		0.006		0.006
		(0.009)		(0.010)
Rural × Incentives £15		0.001		0.000
		(0.010)		(0.010)
Rural × Moral Cost		-0.002		-0.002
		(0.006)		(0.006)
Meter × Simplified	0.006			
	(0.005)			
Meter × Altruism	0.001		-0.005	
	(0.005)		(0.005)	
Meter × Incentives £10	0.010		0.004	
	(0.009)		(0.009)	
Meter × Incentives £15	0.026^{***}		0.020^{**}	
	(0.010)		(0.010)	
Meter × Moral Cost	0.029^{***}		0.023^{***}	
	(0.006)		(0.006)	
Intercept	0.012^{***}	0.008^{***}	0.016^{***}	0.014^{***}
	(0.002)	(0.003)	(0.002)	(0.004)
Controls	Rural	Meter	Rural	Meter
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 ATE estimates of different behavioural interventions on the take-up of the online diagnostic tool (Equation (11)). 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. The model names reflect the reference group for each regression. The regressors of interest are *Meter* and *Rural*. The former equals 1 if the household has a water meter attached to it, while *Rural* is a dummy variable that equals 1 if the household is located in a rural area. 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.

Next, we analyze if pre-treatment water consumption interacted with the treatment letters to affect the probability of completing the diagnostic. Our sample for this exercise is limited to metered households as data on pre- and post-treatment water consumption was only available for households that had a meter installed (see discussion in Section 3.2). Results are presented in Table B.2. As before, the excluded group in models (1) and (2) is the *Vanilla* letter, while the excluded group in models (3) and (4) is the *Simplified* letter, with both control and *Vanilla* removed from the sample.

We detect heterogeneous effects based on pre-treatment usage, but only for customers who were sent the *Vanilla* and *Simplified* letters. First, we use Equation (1) to study the effect of the letters on diagnostic completion, but limiting the sample to only metered customers with water consumption data both pre- and post-treatment. This is presented in models (1) and (3). We find that relative to *Vanilla*, all letters except *Altruism* had a significant impact on diagnostic completion. The *Simplified* and *Moral Cost* letters increased the probability of completing the diagnostic by 1.7 and 3.8 percentage points, respectively, relative to the *Vanilla* group. Similar to the pooled regression with both metered and unmetered customers (Section 3.1), the largest effect was observed for the *Incentives* treatment, with the £10 and £15 treatment increasing take-up by 5.0 and 7.4 percentage points (110 percent and 164 percent greater than the impact of the *Vanilla* letter), respectively. In column 3, we find that the *Incentives* and *Moral Cost* treatment had a significantly higher impact relative to the *Simplified* treatment as well. *£15 Incentives* treatment increased take-up by 5.7 percentage points more than *Simplified*, while *Incentives 10* was slightly lower at 3.3 percentage points (89 and 51 percent higher). Models (2) and (4) use Equation (11) to understand whether, conditional on having a meter, the response of high users to the treatment was different from the response of the low users. The intercept represents the impact of the reference group treatment on low users. We find that high users who received the *Vanilla* treatment were 2.2 percentage points more likely to complete the audit relative to low users in the *Vanilla* group. A similar result is seen for high use households in the *Simplified* treatment group, who were 2.2 percentage points more likely to complete the audit as compared to low use households in the *Simplified* group. However, there was no such heterogeneity observed for the *Incentives*, *Altruism* and *Moral Cost* treatment group, implying both high and low users in those groups were equally likely to complete the audit. When we remove the *Vanilla* treatment group from our sample, we find similar results. The *Simplified* treatment had a larger impact on high users (2.2 percentage points higher), while the effects for high users in other treatment arms was not significantly different from 0. The latter results have interesting implications on how to incentivize water conservation among households. Both *Vanilla* and *Simplified* groups did not appeal to an environmental or an altruistic cause, and neither did they offer a financial incentive. In such cases, it may be difficult to incentivize low users to take action. Of course, whether low users should even be targeted is another policy question.

Table B.2: Heterogeneity in Diagnostic Completion Based on Pre-Treatment Usage

	Completed Diagnostic			
	Vanilla		Simplified	
	(1)	(2)	(3)	(4)
High-Use		0.022^{**}		0.022^{*}
		(0.010)		(0.012)
Simplified	0.017^{***}	0.017^{*}		
	(0.006)	(0.009)		
Altruism	0.008	0.015^{*}	-0.010	-0.002
	(0.006)	(0.008)	(0.007)	(0.009)
Incentives £10	0.050^{***}	0.059^{***}	0.033^{***}	0.042^{***}
	(0.010)	(0.013)	(0.010)	(0.014)
Incentives £15	0.074^{***}	0.089^{***}	0.057^{***}	0.072^{***}
	(0.011)	(0.015)	(0.011)	(0.016)
Moral Cost	0.038^{***}	0.040^{***}	0.021^{***}	0.022^{**}
	(0.007)	(0.010)	(0.008)	(0.011)
High-Use × Simplified		0.000		
		(0.013)		
High-Use × Altruism		-0.014		-0.014
		(0.012)		(0.013)
High-Use × Incentives £10		-0.018		-0.018
		(0.019)		(0.020)
High-Use × Incentives £15		-0.030		-0.030
		(0.022)		(0.022)
High-Use × Moral Cost		-0.002		-0.003
		(0.014)		(0.015)
Intercept	0.045^{***}	0.043^{***}	0.064^{***}	0.061^{***}
	(0.006)	(0.007)	(0.008)	(0.009)
Controls	Yes	Yes	Yes	Yes
Observations	9,778	9,778	7,802	7,802

Notes: Robust standard errors are in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.
All regressions report the ATE estimates of different behavioural interventions on the take-up of the online diagnostic tool (Equation (1) or Equation (11)).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. The sample for all the regressions presented in this table include metered households for which both pre- and post-treatment water consumption was available (see Section 3.2 and Appendix F.1 for details). 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. 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. 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 litres/day, before the treatment date of 08-Dec-2018.

B.2 Heterogeneity in Post-Treatment Water Consumption

In this section, we show that the average treatment effect of the letters on water consumption was greater for high water users, and the result holds with the LATE estimate of the impact of the online audit. The results 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 are more likely to be incentivized to conserve energy.

B.2.1 Average Treatment Effect

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

$$y_i = \alpha + \sum_j \beta_j T_{ij} + \phi \textit{ High-Use}_i + \sum_j \eta_j \textit{ High-Use}_i \times T_{ij} + \gamma \mathbf{X}_i + \epsilon_i \tag{12}$$

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. *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 *Pre-Treatment Water Consumption*$_{\text{i}}$ in liters per day. Finally, ϵ_i is the error term. If the letters incentivized households with higher pre-treatment usage to conserve more water, we would expect the sum of β_j and η_j to be negative and significant.

Table B.3 presents the results. The coefficient on *Treated* represents the average impact of a letter on post-treatment water consumption for low users. The average impact for the high users is the sum of *Treated* and *High-Use* $\times$ *Treated*. With reference to the untreated high users in the control group, the interventions reduced water consumption by 3.7 liters per day for 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 none of the treatments had any significant impact on consumption for the low use households. However, for *Simplified*, *Altruism* and *Incentive* letters, treated high use households reduced their consumption significantly as compared to untreated high use households in the control group. Consumption was lower by 4.4 liters per day for *Simplified* high users, 4.0 liters for *Altruism* high users, and 7.4 and 8.3 liters per day for *Incentive 10* and *Incentive 15* high users, respectively. This heterogeneity persists even when we change our reference group to *Vanilla* in column (3). Relative to treated high users in the *Vanilla* group, treated high users across all other treatments reduce their consumption significantly. Effect sizes vary from 3.9 liters per day with *Moral Cost* high users to 9.0 liters per day with *Incentives 15* high users. The heterogeneity, however, dissipates when our reference group changes to *Simplified* in column (4), but the effect sizes are still negative and large for the *Incentives* group. 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 or *Vanilla* group.

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

	Post-Treatment Water Consumption (liters/day)			
	Control (1)	Control (2)	Vanilla (3)	Simplified (4)
High-Use	9.716***	9.729***	7.328***	5.361*
	(2.702)	(2.701)	(2.822)	(2.918)
Treated	1.193			
	(0.933)			
High-Use × Treated	−4.891*			
	(2.503)			
Vanilla		2.896**		
		(1.203)		
Simplified		1.471	−1.429	
		(1.230)	(1.229)	
Altruism		1.207	−1.691	−0.323
		(1.181)	(1.178)	(1.209)
Incentives £10		0.430	−2.471*	−1.036
		(1.451)	(1.450)	(1.473)
Incentives £15		−0.945	−3.845***	−2.441*
		(1.450)	(1.448)	(1.472)
Moral Cost		0.727	−2.170*	−0.790
		(1.218)	(1.216)	(1.245)
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	Yes	Yes	Yes	Yes
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 (11)). 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.

B.2.2 Local Average Treatment Effect

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

$$y_i = \alpha + \beta \; \textit{Completed Diagnostic}_i + \phi \; \textit{High-Use}_i + \eta \; \textit{High-Use}_i \times \; \textit{Completed Diagnostic}_i + \gamma \mathbf{X}_i + \epsilon_i \quad (13)$$

where *Completed Diagnostic*$_i$ is an indicator for whether household i completed the audit or not. α represents the daily average post-treatment consumption for low users who did not complete the audit. The sum of α and β represents the daily consumption for low users who completed the diagnostic. β can, thus, be interpreted as the impact of completing the audit on water consumption for low users. The sum of α and ϕ is the average post-treatment consumption for high use households who did not complete the audit. The coefficient of interest is η, which represents the additional impact of completing the diagnostic for 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, *High-Use*$_i \times T_i$, just the IV for *Completed Diagnostic* interacted with *High-Use*$_i$. The results are presented in Table B.4.

For all specifications, the coefficient on *Completed Diagnostic* is negative but insignificant. This implies that the effect of completing the diagnostic for low users, though negative, was not significantly different from zero. Notably, the coefficient on the interaction term is negative and significant for all specifications, and also much higher than the coefficients in Table 3 where we do not distinguish between high and low use households. Completing the audit had a significant impact on water conservation for high users, with average daily post-treatment consumption falling in the range of 78 to 89 liters per day. This savings amounts to a 21 to 25 percent reduction compared to average daily pre-treatment consumption for the high users. These results imply that audits had a far greater impact on high use households than low use households. Thus, the online audit incentivized high use households to conserve more water, while the effect on low use households was negligible.

C Robustness Checks

The lack of clear pre-post delineation in meter readings and measurement error in post-treatment consumption estimates could bias our results. We perform a series of robustness checks to alleviate these concerns. First, we present weighted ATE estimates of the impact of letters on post-treatment consumption in Appendix C.1, followed by weighted LATE estimates of the effect of diagnostic completion on water conservation in Appendix C.2. The weights are equivalent to the fraction of days post-treatment included in the meter readings used to calculate post-treatment consumption, divided by total number of days between the two readings. This is followed by ATE estimates of the impact of treatment on consumption but for a restricted sample. We restrict the observations to the subset of houses that had a meter reading just before the letters went out (so those in November

Table B.4: LATE Estimates of Heterogeneous Treatment Effects

	Post-Treatment Water Consumption (liters/day)		
	(1)	**(2)**	**(3)**
Complete Diagnostic	-12.637	-5.097	-3.622
	(10.745)	(11.459)	(11.486)
High-Use	9.851^{***}	13.475^{***}	14.419^{***}
	(2.651)	(3.420)	(3.781)
High-Use × Complete Diagnostic	-65.567^{**}	-82.943^{**}	-85.422^{**}
	(32.642)	(36.576)	(36.987)
Intercept	10.292^{***}	10.460^{***}	10.324^{***}
	(1.603)	(2.311)	(2.759)
Instruments	All Treatment	Incentives+Simplified	Incentives
F-stat in First Stage	20, 17	34, 30	52, 49
Controls	Yes	Yes	Yes
Observations	11,700	5,830	3,904

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 average daily water consumption of a household, in liters per day, post the treatment date of December 08-Dec-2018. The data has been 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 six 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. The sample in model (1) includes all metered households for which we had both pre- and post-consumption data. The sample in model (2) consists of the *Incentives*, *Simplified* and control group, while the sample in model (3) includes only *Incentives* and the control 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 average daily water consumption of a household, in liters per day, before the treatment date of 08-Dec-2018.

and early December). This offers us the cleanest sample with a clear pre- and post-period. Results from this exercise are presented in Appendix C.3.

C.1 Weighted ATE Estimates of Letters on Post-Treatment Consumption

We augment the analysis relating to the ATE estimates of the impact of the treatment on water consumption, presented in Section 3.2, by using a weighted sample. Each observation is weighted by the fraction of number of days post-treatment included in the two meter readings used to calculate the post-treatment consumption estimates divided by the number of days between the two readings. A larger weight implies that a sizeable proportion of the days between the meter readings used to compute post-treatment consumption fell in the time period beyond 8$^{\text{th}}$ December 2018, the treatment date. We follow the same empirical specification as in Equation (2) and supplement it by adding weights to each observation.

Results, presented in Table C.1, are very similar to the unweighted ATE estimates. The effect of receiving any letter on consumption is presented in column (1), while the heterogeneity results are reported in columns (2)-(4). We find evidence that all behavioral interventions, except *Vanilla*, reduced water consumption, though results are statistically significant at the $p < 0.01$ level only for the *Incentives 15* group. Column (1) provides the average treatment effect of receiving any letter on post-treatment consumption. Though the estimate is negative (-1.1 liters per day), it is not significantly different from 0. 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 average daily consumption after letters were sent out; however, only the £15 monetary incentives led to a statistically significant decrease. This is in contrast to the unweighted regressions where even the £10 monetary treatment led to a statistically significant fall. Also, the magnitude of fall in consumption is higher, with consumption in the *Incentives 15* group falling by 5.8 liters per day as opposed to 4.7 liters per day earlier.

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 4.1 liters per day under *Moral Cost* to 8.5 liters per day under *Incentives 15*. In percentage terms, this decrease amounts to between 1.6 percent and 3.3 percent of the average pre-treatment water consumption across all households. As with the previous specification, the magnitude of fall is greater for every treatment arm (except *Incentive 10*) when compared to the unweighted regressions in Section 3.2. 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. Therefore, similar to the unweighted regressions, 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. Specifically, customers in the *£15 Incentives* group reduced their consumption by a significant 4.5 liters per day (1.8 percent of the average pre-treatment water consumption, as against 1.3 percent with the unweighted specification) relative to households in the *Simplified* group. In summary, the *Incentives 15* group experienced a significant reduction in their consumption relative to all comparison groups, while the other treatments had a significant negative impact only relative to the *Vanilla* arm. Reassuringly, all the weighted ATE estimates are similar in direction but larger in magnitude compared to the unweighted estimates indicating that the measurement error was downward biasing our effect sizes.

Table C.1: Weighted ATE Estimates of Letters on Post-Treatment Consumption

	Post-Treatment Water Consumption (liters/day)			
	Control	**Control**	**Vanilla**	**Simplified**
	(1)	**(2)**	**(3)**	**(4)**
Treated	-1.114			
	(1.446)			
Vanilla		2.728		
		(1.811)		
Simplified		-1.416	-4.147^{**}	
		(1.825)	(1.739)	
Altruism		-1.608	-4.337^{**}	-0.215
		(1.828)	(1.742)	(1.755)
Incentives £10		-2.277	-5.007^{**}	-0.899
		(2.201)	(2.132)	(2.146)
Incentives £15		-5.778^{***}	-8.507^{***}	-4.458^{**}
		(2.155)	(2.085)	(2.094)
Moral Cost		-1.359	-4.087^{**}	0.036
		(1.784)	(1.695)	(1.709)
Intercept	7.935^{***}	7.955^{***}	10.650^{***}	9.030^{***}
	(1.788)	(1.784)	(1.828)	(1.997)
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 ATE estimates of different behavioural interventions on post-treatment water consumption (Equation (2)). The dependent variable for all models is *Post-Treatment Water Consumption*, a continuous variable that measures the water consumption of a household, in litres/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 in Model (1), *Treated*, is a dummy variable that equals 1 for all households who 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 *Meter*, *Rural*, and *Pre-Treatment Consumption* as controls. Both *Rural* and *Meter* are dummies that equal 1 if the household is located in a rural area, or if it has a water meter attached to it, respectively. *Pre-Treatment Consumption* is a continuous variable that measures the water consumption of a household, in litres/day, before the treatment date of 08-Dec-2018. Each observation was weighted by the fraction of days post-treatment included in the meter readings used to calculate post-treatment consumption, divided by total number of days between the two readings.

C.2 Weighted LATE Estimates of Diagnostic Completion on Post-Treatment Consumption

We re-run the LATE analysis in Section 3.3 using a weighted sample. As in Appendix C.1, each observation is weighted by the fraction of number of days post-treatment included in the two meter readings used to calculate the post-treatment consumption estimates divided by the number of days between the two readings. We use different combinations of instruments for our LATE estimates, all of which give similar results. The results are presented in Table C.2 below.

Table C.2: Weighted LATE Estimates of Diagnostic Completion on Post-Treatment Consumption

	Post-Treatment Water Consumption (liters/day)		
	(1)	**(2)**	**(3)**
Complete Diagnostic	-49.801^{***}	-45.250^{**}	-45.397^{**}
	(17.753)	(19.050)	(19.036)
Intercept	9.880^{***}	10.191^{***}	10.639^{***}
	(1.714)	(2.453)	(2.924)
Instruments	All Treatment	Incentives+Simplified	Incentives
F-stat in First Stage	38	67	102
Controls	✓	✓	✓
Observations	11,700	5,830	3,904

Notes: Robust standard errors are in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.
All regressions report the LATE 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) includes dummies for all the different treatment arms. The IV in Model (2) includes dummies for *Incentives 10*, *Incentives 15*, and *Simplified* treatment arms. The IV in Model (3) is *Incentives 10* and *Incentives 15*, while the IV in Model (4) is only *Incentives 15*. 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. Each observation was weighted by the fraction of days post-treatment included in the meter readings used to calculate post-treatment consumption, divided by total number of days between the two readings.

The model in column (1) uses all the letters as instruments. The estimates suggest that completing the diagnostic led to a significant fall in consumption of 50 liters per day ($p < 0.01$; as opposed to the unweighted LATE estimate of 45 liters per day) or 20 percent of average daily pre-treatment water consumption. In column (2), we restrict the sample to the following four groups:

Incentive £10, *Incentive £15*, *Simplified* and the control group. The instrument set is now a vector of 3 instruments, namely *Incentives £10*, *Incentives £15*, and *Simplified* groups. Our results under this specification indicate that completing the diagnostic reduces daily water consumption on average by 45 liters per day (18 percent; $p < 0.05$), or 2 liters per day higher than the LATE estimate with the unweighted sample. Finally, in column (3) we restrict our sample to *Incentives £10*, *Incentives £15*, and control groups, with the set of instruments now limited to the two *Incentive* treatments. This is our preferred specification as it is most likely to satisfy the exclusion restriction as none of the treatments contained any inducement to an environmental or altruistic cause and, therefore, should not affect water consumption directly. The effect size is similar, and still significant despite the fall in sample size. Completing the diagnostic led to a average fall in post-treatment consumption by 45 liters per day (17 percent, $p < 0.01$), which is slightly higher than the unweighted LATE estimate of 44 liters per day.

In conclusion, our weighted LATE estimates suggest that the measurement error arising from lack of clear delineation between pre- and post-treatment consumption downward biases our estimates, and correcting for the same through weighting each observation increases the magnitude of the impact, both for the treatment as well as diagnostic completion.

C.3 ATE Estimates of Letters on Post-Treatment Consumption for Restricted Sample

In our final robustness check, we estimate the ATE of the letters on the sample of households that had a meter reading very close to the treatment date, *i.e.* November and early December, 2018. This offers us the cleanest sample with a clear pre- and post-treatment delineation. However, this inclusion criteria leads to a drastic drop in the number of observations from 11,700 to 128 (1.1 percent), which affects statistical power. Results are presented in Table C.3.

The direction of the estimates does not change but, importantly, the size of the effects increases by an order of magnitude. Column (1) provides the average treatment effect of receiving any letter on post-treatment consumption. Though the estimate is negative (-16 liters per day), it is not significantly different from 0. With reference to the control group, all treatment arms except *Vanilla* experienced a fall in average daily consumption after letters were sent out (column (2)); however, only the £15 monetary incentives led to a statistically significant decrease. Also, the magnitude of fall in consumption is remarkably higher, with consumption in the *Incentives 15* group falling by 51 liters per day as opposed to 4.7 liters per day with the unweighted and unrestricted sample.

When we exclude the control group, and the *Vanilla* letter becomes the omitted category (column (3)), the drop in consumption is significant across all categories except *Simplified*, with the decrease in consumption ranging from 42 liters per day under *Incentives 10* to 74 liters per day under *Incentives 15*. In percentage terms, this decrease amounts to between 16 percent and 29 percent of the average pre-treatment water consumption across all households. As with the previous specification, the magnitude of fall is greater for every treatment arm when compared to the unweighted regressions in Section 3.2.

Finally, dropping both the control group and the *Vanilla* group with the *Simplified* letter as the reference category (column (4)) preserves the direction of the effects but the statistical significance is lost, most likely due to noisy estimates arising from a very low observation count (95 observations). In summary, the *Incentives 15* group experienced a significant reduction in their consumption relative to control and *Vanilla* comparison groups, while the other treatments (except *Simplified*) had a significant negative impact only relative to the *Vanilla* arm. Reassuringly, all the ATE estimates with even the smaller, restricted sample are similar in direction, and much larger in magnitude compared to the unweighted and unrestricted estimates. Though we lose statistical power due to sample restriction, the results indicate that the measurement error has muted the true effect size and our results should be considered a lower bound.

D Additional Results

Continuing with our theme of targeting, this section presents three additional results. First, in Appendix D.1, we use the data submitted by households during diagnostic completion to study the characteristics (number of appliances, water use, frequency of usage, etc.) of homes that completed the audit, and how that differs across different treatment arms. Second, in Appendix D.2, we estimate the impact of the reminders on diagnostic completion, and the effectiveness of reminders in different treatment arms in incentivizing households to complete the audit. Finally, in Appendix D.3, we analyze how different treatment arms interacted with the reminders. Using data on who opened the reminder, who clicked on the link to complete the diagnostic, and who unsubscribed from any future emails, we can identify which behavioral interventions were successful in promoting user engagement.

D.1 Characteristics of Households that Complete the Diagnostic

We find that different behavioral interventions influenced different sets 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 D.1 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 unable to monitor their daily consumption were more likely to complete the diagnostic if offered monetary rewards. Furthermore, the average number of basins and toilets

Table C.3: ATE Estimates of Letters on Post-Treatment Consumption: Restricted Sample

	Post-Treatment Water Consumption (liters/day)			
	Control	**Control**	**Vanilla**	**Simplified**
	(1)	**(2)**	**(3)**	**(4)**
Treated	-16.019			
	(23.718)			
Vanilla		23.889		
		(30.115)		
Simplified		-13.816	-37.088	
		(25.901)	(22.374)	
Altruism		-18.171	-42.579^{*}	-2.911
		(28.637)	(24.458)	(19.361)
Incentives £10		-18.641	-41.884^{*}	-4.797
		(25.921)	(22.332)	(18.082)
Incentives £15		-50.790^{*}	-73.797^{***}	-37.424
		(30.026)	(27.469)	(23.266)
Moral Cost		-24.729	-48.692^{**}	-10.246
		(26.093)	(21.578)	(16.870)
Intercept	38.650^{*}	38.726^{*}	56.596^{***}	33.441^{*}
	(21.528)	(21.224)	(19.460)	(17.487)
Controls	✓	✓	✓	✓
Observations	128	128	110	95

Notes: Robust standard errors are in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.
All regressions report the ATE estimates of different behavioural interventions on post-treatment water consumption (Equation (2)). The dependent variable for all models is *Post-Treatment Water Consumption*, a continuous variable that measures the water consumption of a household, in litres/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 sample only includes the subset of houses that had a meter reading just before the letters went out, specifically post 01-Nov-2018. 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 who 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 *Meter*, *Rural*, and *Pre-Treatment Consumption* as controls. Both *Rural* and *Meter* are dummies that equal 1 if the household is located in a rural area, or if it has a water meter attached to it, respectively. *Pre-Treatment Consumption* is a continuous variable that measures the water consumption of a household, in litres/day, before the treatment date of 08-Dec-2018.

Table D.1: 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 test statistic and significance stars from a joint orthogonality test of equality of means between the six treatment groups.

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. Finally, there were also differences in the type of residences that completed the different audits. Households that completed the audit influenced by the *Incentives* treatment were less likely to reside in detached residences, and more likely to reside in semi-detached residences.

D.2 Effect of Reminders on Diagnostic Completion

Customers that provided NWG with email contact details (31 percent of total sample or 13,989 households; see Table A.1) 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 6th February 2019. Details on the sample, baseline balance and diagnostic completion rates for the reminder experiment are provided in Appendix A.5 and Appendix A.6. The reminder emails followed the same themes as the initial direct mailers that the customers received. For example, a household in the *Simplified* treatment arm in the initial experiment, if selected to receive a reminder, would receive one with the same *Simplified* theme. This allows us to estimate the impact of receiving a reminder on completing the audit. We run the following regression to estimate the effect of reminders:

$$y_i = \alpha + \phi R_i + \sum_j \beta_j T_{ij} + \sum_j \pi_j R_i \times T_{ij} + \gamma \mathbf{X}_i + \epsilon_i \tag{14}$$

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 equals 1 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 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, conditional on households not completing the audit before 6th February 2019. Our main coefficient of interest is π_j which is the estimate on the interaction term. It represents the additional effect of reminders belonging to the j^{th} treatment group on diagnostic completion. The sum of ϕ and π_j represents the difference in diagnostic completion rates for people who did or did not receive the reminder, conditional on being in the j^{th} treatment group. ϵ_i signifies the error term. Note that the control group in the initial direct mailer experiment was excluded from this exercise. Table D.2 presents the results.

In column (1) of Table D.2, we estimate the direct impact of any reminder on the likelihood of completing the diagnostic.[100] To do so, we modify Equation (14) and run the model without the

[100] See Appendix D.3 for an analysis of how households interacted with the reminders, *i.e.*, how the content of the

Table D.2: ATE Estimates of Reminders on Diagnostic Completion

	Completed Diagnostic		
		Vanilla	Simplified
	(1)	(2)	(3)
Reminder	0.026^{***}	0.015^{***}	0.026^{***}
	(0.002)	(0.004)	(0.005)
Simplified		-0.000	
		(0.000)	
Altruism		0.002	0.002
		(0.001)	(0.001)
Incentives £10		0.000	0.000
		(0.000)	(0.000)
Incentives £15		-0.000	0.000
		(0.000)	(0.000)
Moral Cost		0.000	0.001^{**}
		(0.000)	(0.000)
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	Yes	Yes	Yes
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 (14)). 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.

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, on average across all treatment groups, reminders increased the likelihood of completing the diagnostic by 2.6 percentage points as compared to the population 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. Within the *Vanilla* group, the population which received the reminders had a 1.5 percentage points higher probability of completing the audit. With reference to the effect of the reminders on the *Vanilla* group (ϕ), reminders to the *Moral Cost* group have the highest additional impact (π_j) of 3.0 percentage points, while the magnitude of impact for *Incentives £10* and *Simplified* groups is also significant (1.5 and 1.1 percentage points, respectively). Notably, the impact of the *Moral Cost* reminder is significantly different from the impact of these other two treatments at the 10 percent significance level. In the final specification in column (3), the omitted group is *Simplified*, with both control and *Vanilla* groups excluded from the sample. We find that a reminder to the *Simplified* group increases the likelihood of completing the audit by 2.6 percentage points relative to *Simplified* households who did not receive the reminder. Again, *Moral Cost* reminders tend to do significantly better, with reminders in this category having an additional significant impact of 1.9 percentage points, but the other reminders had no additional impact. In summary, we find that reminders were effective in increasing take-up of the audit across all treatment groups, with the impact on *Moral Cost* group significantly higher than the others.

D.3 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. Here we analyze the interaction of households with reminders, and refer the reader to Appendix D.2 for an analysis of the impact of reminders on take-up. 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 (see Appendix A.6 for details). 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 D.3. *Vanilla* treatment arm forms the excluded category in all the columns. The intercept, therefore, refers to the percentage of people in the *Vanilla* reminder treatment group who opened or clicked on the reminder (columns (1) and (2), respectively) or who unsubscribed from any future emails (column (3)).

As compared to the *Vanilla* reminder, all reminders, except *Altruism*, had a positive and sig-

reminder affected the probability of opening the reminder emails, clicking on the link to the audit tool, or simply unsubscribing from future emails.

Table D.3: 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	Yes	Yes	Yes
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.

nificant effect on the probability of opening the reminder. 43 percent of *Vanilla* households who received the reminder ended up opening the email. This number increases to 50 percent or more for the *Incentives* treatment. The email appealing to an altruistic motive, however, was opened considerably fewer times (37 percent). Moreover, almost no 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 (5.6 percent of *Moral Cost* households clicked on the reminder as opposed to 2.8 for the *Vanilla* households). Finally, 1.7 percent of the *Vanilla* households unsubscribed from the emails on receiving the reminder, and this hold true across all other treatments except *Incentives 10* where the probability of unsubscribing was 0.9 percent, significantly lower than *Vanilla* treatment.

E Welfare Calculations

This section is divided into five parts. First, in Appendix E.1, we describe the value and sources for all the parameters used in the welfare calculations. Appendix E.2 changes the base case for the cost effectiveness calculations from the *£10 Incentive* intervention in Section 4.1.1 to the *£15 Incentive* intervention. Appendix E.3 presents cost effectiveness calculations, measured by the costs incurred to reduce a tonne of CO_2 emissions. Appendix E.4 provides details on how we calculated cost effectiveness for different studies in the literature so as to enable comparison with our estimates. Finally, in Appendix E.5, we perform a number of sensitivity checks for the benefit cost analysis and *MVPF* analysis presented in Section 4.2 and Section 4.3, respectively.

E.1 Parameters

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

E.2 Cost Effectiveness of the *£15 Incentives* Intervention

As in Section 4.1.1, we measure four categories of costs: the cost of sending letters, the direct cost of the incentives, the lost producer surplus associated with the decline in production, and the value of time in filling out the survey. Effectiveness is measured by the per capita reduction in water consumption. Our base case is the *Incentive £15* treatment which, along-with the *Incentive £10* treatment, was the only intervention that resulted in a significant reduction in water consumption among treated households (see Table 2). We measure its effectiveness relative to not sending out a letter, and to sending out the *Vanilla* letter. Dividing total cost by effectiveness yields our cost effectiveness estimate. Results are presented in Table E.2.

We describe the parameters for the base case (column 1) below. The cost of mailing represents the postal cost of sending letters to 954 participants (the sample size in the *Incentive £15* group

Table E.1: Parameters and Sources

Variable	Unit (1)	Value (2)	Source (3)	Notes (4)
Consumer Price of Water	£/m^3	1.3	NWG (2020)	
Short-Run Marginal Cost	£/m^3	0.44	NWG (2021)	
Long-Run Marginal Cost	£/m^3	0.98	NWG (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	NWG (2021)	
Emissions/Ml of Sewage Treated	kgCO2e/MI	520	NWG (2021)	Includes Scope 3 (other indirect) GHG emissions. For details, see United Kingdom Government (2019a)
Emissions/MI from Household Water Use	kgCO2e/MI	6,200	Reffold et al. (2008)	
Social Cost of CO_2	£/tCO_2e	241	United Kingdom Government (2021)	Assumes a 3.25 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	NWG (2021)	
UK Median Wage	£/hour	14	Office for National Statistics, UK (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 United Kingdom Government (2019a), 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.

Table E.2: Different Measures of Cost Effectiveness: £15 Incentive Intervention

Case	Base Case	No Producer Surplus Loss	Vanilla Letter	Targeting High Users	Duration: 1 Yr & No PS Loss
Parameter	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Cost of Mailing	390	390	0	180	390
Direct Cost of Incentive	1,500	1,500	1,500	630	1,500
Producer Surplus Loss	250	0	330	200	0
Time Cost	82	82	0	34	82
[A]: Total Cost (in £)	2,300	2,000	2,000	1,000	2,000
[B]: Effectiveness (in m^3)	290	290	390	240	1,600
Cost Effectiveness (£/m^3)	7.8	6.9	5.0	4.4	1.2

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.

for which we had both pre- and post-treatment consumption data) 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. The direct cost of incentives refers to the pecuniary transfer to the customers who completed the diagnostic. 103 households from the 954 participants in the *Incentives 15* treatment group completed the audit and received £15, yielding a direct cost of £1,500. The *Producer Surplus Loss* is defined as the total loss in net revenue (*i.e.*, revenue minus cost) caused by water savings. 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 4.7 liters per day per household (refer to Table 2), the producer surplus loss over the 65-day period is £250. 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 (Office for National Statistics, UK, 2021). The sum of these items gives a total cost of £2,300. To calculate effectiveness, we multiply the per capita reduction in water consumption relative to the case of no letter (4.7 liters per day for 65 days; see Table 2) with the number of people in the £15 incentive group, which gives 290 cubic meters. Dividing the total cost by effectiveness gives us a cost effectiveness estimate of £7.8 per cubic meter for the base case.

The other four cases are variations on the base case. They lead to cost effectiveness numbers that range between £1.2 and £6.9 per cubic meter. The first variation labeled *No Producer Surplus Loss* (column 2) sets producer surplus losses to zero. This yields a cost effectiveness of £6.9 per cubic meter, which is a 11 percent decline relative to the base case.

The second variation changes the benchmark for comparison from the control group to the

Vanilla letter (column 3). The cost effectiveness falls to £5.0 per cubic meter, a decline from the base case by 36 percent.

The third variation targets only high users (column 4), 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 4.7 to 8.3 liters per household per day (see Appendix B.2.1 for details). The cost effectiveness is reduced by 44 percent as a result, from £7.8 per cubic meter in the base case to £4.4 per cubic meter.

The fourth variation considers the impact of a change in duration of the persistence of the effects due to the intervention along with setting producers surplus losses equal to zero (column 5). If we assume the benefits last for a year, and prices can adjust to eliminate producer surplus losses, cost effectiveness decreases from £7.8 in the base case to £1.2 per cubic meter, or by 84 percent.

E.3 Cost Effectiveness: Cost Per Tonne of CO_2 Reduced

In this section, we present results on the cost effectiveness of our intervention. However, unlike Section 4.1, where cost effectiveness was measured in terms of costs that need to be incurred to reduce consumption of water by a cubic meter, here we measure it in terms of costs incurred to reduce a tonne of CO_2 emissions. As before, we measure four categories of costs: the cost of sending letters, the direct cost of the incentives, the lost producer surplus associated with the decline in production, and the value of time in filling out the survey. Effectiveness is measured by the per capita reduction in CO_2 emissions. 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 (column 3). Dividing total cost by effectiveness yields our cost effectiveness estimate. Results are presented in Table E.3.

The parameters for the base case (column 1) are similar to our exercise in Section 4.1. The only difference is the measurement of effectiveness. To calculate effectiveness, we multiply the per capita reduction in water consumption relative to the case of no letter (3.5 liters per day for 65 days, see Section 3.2) with the number of people in the £10 incentive group (for whom we have both pre- and post-treatment consumption data), which equals 240 cubic meters. From Table E.1, we know that emissions per mega Liter (ML) from water supply, sewage treatment, and household water use equal 140, 520 and 6,200 kilograms (kg), respectively. Summing across all the three categories gives us total emissions from water usage, which amounts to 6,900 $kgCO_2e/ML$. We convert this to tonnes of CO_2e/m^3, and multiply by the total water savings to arrive at 1.6 tonnes of CO_2e emissions reduced due to the *Incentives 10* intervention. Dividing the total cost by effectiveness gives us a cost effectiveness estimate of £950 per tonne of CO_2 emissions for the base case.

The other four cases are variations on the base case. They lead to cost effectiveness numbers that range between £150 and £830 per tonne of CO_2 emissions. The first variation labeled *No Producer Surplus Loss* (column 2) sets producer surplus losses to zero, yielding a cost effectiveness of £830 per tonne of CO_2 emissions, which is a 13 percent decline relative to the base case. The second

Table E.3: Cost Effectiveness: Cost Per Tonne of CO_2 Reduced

Case **Parameter**	**Base Case** **(1)**	**No Producer Surplus Loss** **(2)**	**Vanilla Letter** **(3)**	**Targeting High Users** **(4)**	**Duration: 1 Yr & No PS Loss** **(5)**
Cost of Mailing	420	420	0	200	420
Direct Cost of Incentive	850	850	850	380	850
Producer Surplus Loss	200	0	300	200	0
Time Cost	68	68	0	30	68
[A]: Total Cost (in £)	1,500	1,300	1,100	810	1,300
Total Water Savings(m^3)	240	240	350	230	1,300
[B]: CO_2 Reduced (tCO_2e)	1.6	1.6	2.4	1.6	9.1
Cost Effectiveness (£/tCO_2e)	950	830	480	500	150

Notes: This table shows how the cost effectiveness changes using different assumptions. Cost effectiveness is measured in terms of pounds per tonne of CO_2 emissions reduced. It is computed as the total cost divided by the effectiveness (A/B). See text for details on the various cases, and Table E.1 for details on the value and sources of different parameters.

variation changes the benchmark for comparison from the control group to the *Vanilla* letter (column 3). The cost effectiveness falls to £480 per tonne of CO_2 emissions, a decline from the base case by 50 percent. The third variation targets only high users (column 4), 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 7.4 liters per household per day (see Appendix B.2.1 for details). The cost effectiveness is reduced by 47 percent as a result, from £950 per tonne of CO_2 emissions in the base case to £500 per tonne of CO_2 emissions.[101] The fourth variation considers the impact of a change in duration of the persistence of the effects due to the intervention, coupled with the elimination of producer surplus losses (column 5). If we assume the benefits last for a full year, this directly impacts the quantity of water conserved and, consequently, the CO_2 emissions avoided. Emissions reduced increase by a factor of 5.7 (9.1 v/s 1.6 tonne of CO_2 emissions), and cost effectiveness decreases from £950 in the base case to £150 per tonne of CO_2 emissions, or by 84 percent.

Finally, we 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 less than the costs. The result (not shown in the table) is that the effectiveness of the £15 intervention is £1,100 per tonne of CO_2 emissions, 19 percent higher than the £10 intervention.

[101] A similar calculation for the £15 intervention reveals that cost effectiveness is reduced by 44 percent.

E.4 Cost Effectiveness Calculations for Various Conservation Studies

Table 6 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 E.4. 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 cost effectiveness.

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).[102] The details of the calculations are presented in Table E.5.

E.5 Sensitivity Checks on Benefit-Cost Analysis and *MVPF* Analysis

This section explores a number of sensitivity checks related to the benefit-cost analysis and *MVPF* analysis presented in Section 4.2 and Section 4.3, We consider changes in the SCC, LRMC, changing the base case to the *£15 Incentive* intervention, and assuming transfers from the utility to consumers represent a net social benefit.

E.5.1 Varying the LRMC

We consider the impact of substantially underestimating the LRMC (see Footnote 85 for why that may be the case). 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 in the base case still remain negative, though become smaller in magnitude. For example in the *£10 Incentive* case with short-run benefits and a producer surplus loss, net benefits go from from -£950 to -£6.1. In per capita terms, this amounts to a change from -£0.93 per person to a small, but still negative, -£0.0059 per person. If the intervention is targeted to high users, per capita net benefits go from -£0.54 to £1.4.

E.5.2 Benefit-Cost Analysis for £10 Incentive Intervention using Current US SCC Estimates

The benefit-cost analysis in Section 4.2 assumes an SCC of £241 per ton of CO_2 (United Kingdom Government, 2021). However, the US estimate is much lower, at $51 per ton of CO_2 or approxi-

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

Table E.4: 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 E.1 for parameters

Table E.5: Cost Effectiveness Calculations for Other Studies

Paper	Population / Bound	\$ per 1000 gallons reduced (Year of Paper)	\$ per 1000 gallons reduced (2020)	Cost effectiveness (\$/m3)
	(1)	(2)	(3)	(4)
Bennear et al. (2013)	Lower Bound	7.33	8.3	2.2
	Upper Bound	26	29	7.6
Ferraro and Miranda (2013)	All Households	0.37	0.41	0.11
	High-Use Households	0.20	0.22	0.06
Ferraro and 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 and Miranda (2013) refer to households who both have above median consumption and own their homes. High-use households in Ferraro and Price (2013) refer to households who have above median consumption.

mately £40 per ton of CO_2 emitted in £2020 (Interagency Working Group, US Government, 2021). In this section, we consider how the benefit-cost numbers change if we use the US estimate of the SCC instead of the UK estimate.

Before explaining the results, it is useful to highlight one key point. The price for water when we assume an SCC of \$51 per ton of CO_2 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 (equivalent to V in Equation (6) above and equals £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 + £0.27). 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 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.

As before, we consider five 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 sets producer surplus losses to zero. The third uses the *Vanilla* letter as the benchmark with the *£10 Incentive*. The fourth focuses on targeting high-users. The fifth case assumes the impact of the intervention lasts for a year, in addition to eliminating producer surplus losses.

Table E.6: Simple Benefit-Cost Analysis: Using SCC of \$51/ton of CO_2

Case Parameter	Units	Base Case (£10 Incentive) (1)	No Producer Surplus Loss (2)	Vanilla Letter as Benchmark (3)	Targeting High Users (4)	Duration: 1 Yr & No PS Loss (5)
V	$£/m^3$	0.27	0.27	0.27	0.27	0.27
p	$£/m^3$	1.3	1.3	1.3	1.3	1.3
Δg	m^3	-240	-240	-340	-230	-1,300
$-V\Delta g$	£	64	64	94	64	360
E	£	1,300	1,300	850	580	1,300
N	*integer*	1,020	1,020	1,020	484	1,020
Panel A **(*SRMC*)**						
c	$£/m^3$	0.44	0.44	0.44	0.44	0.44
$(p-c)\Delta g$	£	-200	0	-290	-200	0
$B-C$ (Equation (6) above)	£	-1,400	-1,200	-1,000	-710	-910
$(B-C)/N$	£/ *capita*	-1.4	-1.2	-1.0	-1.5	-0.89
Breakeven Other Benefits $= -(B-C)$	£	1,400	1,200	1,000	710	910
Breakeven Other Benefits / Δg	$£/m^3$	6.0	5.1	3.0	3.0	0.69
Breakeven Other Benefits / GHG benefits	*multiple*	22	19	11	11	2.5
Panel B **(*LRMC*)**						
c	$£/m^3$	0.98	0.98	0.98	0.98	0.98
$(p-c)\Delta g$	£	-72	0	-110	-71	0
$B-C$ (Equation (6) above)	£	-1,300	-1,200	-860	-590	-910
$(B-C)/N$	£/ *capita*	-1.3	-1.2	-0.84	-1.2	-0.89
Breakeven Other Benefits $= -(B-C)$	£	1,300	1,200	860	590	910
Breakeven Other Benefits / Δg	$£/m^3$	5.4	5.1	2.5	2.5	0.69
Breakeven Other Benefits / GHG Benefits	*multiple*	20	19	9.2	9.2	2.5

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 Table E.1 for details on parameters used for welfare calculations.

Table E.6 shows that the measured benefits fall short of the measured costs in all five scenarios under both the cost structures, but that the net costs per capita are small, on the order of £0.89 to £1.5 per person. With the LRMC, the measured benefits are slightly higher, albeit still negative (but small) on a per capita basis. The second to last row of each panel in the table shows that other benefits would need to be between £0.69 to £6.0 per cubic meter for the total benefits to just offset the total costs. The final row for each of panels shows that other benefits would need to be anywhere from 2.5 to 22 times as great as carbon emission benefits for benefits to justify costs.

We also run a sensitivity analysis to understand how incorporating the opportunity cost of water, with estimates ranging from £0.16 in Spain to £1.4 per cubic meter in Nevada, USA, change our results. We find that under the assumptions that *i*) conservation effects persist for a at least a year; *ii*) utilities can increase prices to compensate for loss of revenue from conservation (no producer surplus loss), and; *iii*) opportunity cost of scarce water is high, benefits exceed costs. This can be seen in column (5), where other benefits need to be £0.69 per cubic meter to break-even, which is lower than the opportunity cost of scarce water in a few places. For all other scenarios in columns (1) to (4), this does not hold true. Even in the most conservative case (other benefits in order to break even equivalent to £2.5 per cubic meter, and an opportunity cost of water equivalent to £1.4 per cubic meter), we would still need other benefits, not including scarce value of water, to be £1.1 per cubic meter.

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. If we include the opportunity cost of scarce water among other benefits, the SCC numbers would still be required to be in the range of $800 to $900 depending on whether we use SRMC or LRMC, respectively. These numbers are much higher than most estimates for the SCC.

E.5.3 Benefit-Cost Analysis for £15 Incentive Intervention

The benefit-cost analysis in the text (Section 4.2) focused on the net benefits with respect to the *£10 Incentive* intervention. We now present a similar analysis for the *£15 Incentive* intervention which also resulted in significant water conservation (see Section 3.2). As with the £10 intervention, we consider five different cases for estimating net benefits associated with the SRMC and the LRMC.

The first scenario uses the base case with the *£15 Incentive*, and it is compared to the case of no letter. The second scenario sets producer surplus losses to zero. The third uses the *Vanilla* letter as the benchmark with the *£15 Incentive* letter. The fourth variation focuses on targeting high-users, defined as households who consume above the median pre-treatment consumption threshold. The final variation assumes the impact of the intervention lasts for a year, in addition to eliminating producer surplus losses.

Table E.7 shows that the measured benefits fall short of the measured costs in four out of five scenarios under both the cost structures. The only scenario with a positive net-benefit is the final one where the effect lasts for a period of a year and there is no producer surplus loss. The net costs per capita under the first four scenarios are small, on the order of £1.1 to £1.8 per person. With the LRMC, the measured benefits are slightly higher, albeit still negative (but small) on a per capita basis for the first four scenarios. For each scenario, the *£10 Incentive* intervention in Section 4.2 has a more favorable benefit-cost ratio as compared to the *£15 Incentive* intervention.

The second to last row of each panel in the table shows that other benefits would need to be between £2.1 to £5.9 per cubic meter (as opposed to between £1.1 to £4.6 per cubic meter with the *£10 Incentive* intervention) for the total benefits to just offset the total costs for the four scenarios where net benefits are negative. The final row of the panels shows that other benefits would need to be anywhere from 1.2 to 3.5 times (versus 0.68 to 2.8 times with the *£10 Incentive* intervention) as great as carbon emission reduction benefits to justify costs provided the intervention only leads to short-run benefits.

Including the opportunity cost of scarce water as a potential benefit in the four scenarios with

Table E.7: Simple Benefit-Cost Analysis: £15 Incentive Intervention

Case / Parameter	Units	Base Case (£15 Incentive) (1)	No Producer Surplus Loss (2)	Vanilla Letter as Benchmark (3)	Targeting High Users	Duration: 1 Yr & No PS Loss
V	$£/m^3$	1.7	1.7	1.7	1.7	1.7
p	$£/m^3$	1.3	1.3	1.3	1.3	1.3
Δg	m^3	-290	-290	-390	-240	-1,600
$-V\Delta g$	£	480	480	650	390	2,700
E	£	1,900	1,900	1,500	800	1,900
N	*integer*	954	954	954	437	954
***Panel A* (*SRMC*)**						
c	$£/m^3$	0.44	0.44	0.44	0.44	0.44
$(p-c)\Delta g$	£	-250	0	-330	-200	0
$B-C$ (Equation (6) above)	£	-1,700	-1,500	-1,200	-620	770
$(B-C)/N$	£/ *person*	-1.8	-1.5	-1.3	-1.4	0.81
Breakeven Other Benefits $= -(B-C)$	£	1,700	1,500	1,200	620	-770
Breakeven Other Benefits / Δg	$£/m^3$	5.9	5.0	3.1	2.6	-0.47
Breakeven Other Benefits / GHG benefits	*multiple*	3.5	3.0	1.9	1.6	-0.28
***Panel B* (*LRMC*)**						
c	$£/m^3$	0.98	0.98	0.98	0.98	0.98
$(p-c)\Delta g$	£	-89	0	-120	-72	0
$B-C$ (Equation (6) above)	£	-1,500	-1,500	-1,000	-490	770
$(B-C)/N$	£/ *person*	-1.6	-1.5	-1.1	-1.1	0.81
Breakeven Other Benefits $= -(B-C)$	£	1,500	1,500	1,000	490	-770
Breakeven Other Benefits / Δg	$£/m^3$	5.3	5.0	2.6	2.1	-0.47
Breakeven Other Benefits / GHG Benefits	*multiple*	3.2	3.0	1.6	1.2	-0.28

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 Table E.1 for details on parameters used for welfare calculations.

negative net benefits is also not sufficient for the benefits to exceed costs. This is because other benefits needed to break even are in the range of £2.1 to £5.9 per cubic meter, which are much higher than the most liberal estimates of the opportunity cost of scarce water in the literature (for *e.g.*, £1.4 per cubic meter in Nevada used by Baker (2021)). Even in the most conservative case (other benefits in order to break even equivalent to £2.1 per cubic meter, and an opportunity cost of water equivalent to £1.4 per cubic meter), we would still need other benefits, not including scarce value of water, to be £0.7 per cubic meter.

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 320 percent to £1,000 per ton, and to £1,100 per ton using the SRMC (as opposed to £830 and £910 per ton, respectively, with the *£10 Incentive* intervention). If we include the opportunity cost of scarce water among other benefits, the SCC numbers would still be required to be in the range of £810 to £890 per ton of $C0_2$ depending on whether we use LRMC or SRMC, respectively. These numbers are much higher than most estimates for the SCC.

E.5.4 Benefit-Cost Analysis for £10 Incentive Intervention: Upper Bound

In our benefit-cost analysis in Section 4.2, we have assumed that consumers are just as well off as they were before they switched to taking up the audit. 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"). Therefore, a reasonable assumption could be that consumer private benefits from taking the audit are non-negligible. Unfortunately, we do not have information on the extent to which such people benefited. What we can do, though, is estimate an upper-bound on the benefits. A plausible upper bound on how much better off they would be is to assume their welfare increases by the amount of the incentive and the private savings in water consumption.[103] Note that we are assuming the cost of the audit and consumer surplus loss from consuming less water is zero (*e.g.*, from technology improvements or simple reductions in consumption). Denoting the incentives as I and the private savings as $p\Delta g$, our new welfare bounding equation becomes:

$$\begin{aligned} \text{Net Benefits} &= B - C \\ &= -V\Delta g + (p - c)\Delta g - E - p\Delta g + I \end{aligned} \tag{15}$$

Results are presented in Table E.8.We consider five different cases for estimating net benefits associated with the SRMC and the LRMC. The first scenario uses the base case with the *£10 Incentive*, and it is compared to the case of no letter. The second scenario sets producer surplus losses to zero. The third uses the *Vanilla* letter as the benchmark with the *£10 Incentive* letter. The fourth variation focuses on targeting high-users, defined as households who consume above the median pre-treatment consumption threshold. The final variation assumes the impact of the intervention lasts for a year, in addition to eliminating producer surplus losses.

Our results indicate that the measured benefits exceed the measured costs in all of the five scenarios under the SRMC structure, but that the net benefits per capita are small, on the order of £0.073 to £3.4 per person. With the LRMC (*Panel B*), the measured benefits for the interventions are slightly higher, albeit still small on a per capita basis (£0.20 to £3.4 per person). Therefore, using upper-bound estimates for the benefits does make the net benefits positive, but they remain small. However, as noted, we are not quantifying any other benefits from water conservation, and including them along with the opportunity cost of scarce water in this analysis will make the intervention more attractive. Thus, extreme assumptions on both consumer and ecosystem benefits may change our conclusion about the intervention not being welfare improving in a majority of

[103] An alternative would be to estimate the consumer surplus loss associated with the decrease in water consumption. Using such an approach does not change our qualitative findings about the effectiveness of this intervention, though it would decrease consumer benefits as defined in this bounding case.

Table E.8: Simple Benefit-Cost Analysis: Upper Bound

Case Parameter	Units	Base Case (£10 Incentive) (1)	No Producer Surplus Loss (2)	Vanilla Letter as Benchmark (3)	Targeting High Users	Duration: 1 Yr & No PS Loss
V	£/m^3	1.7	1.7	1.7	1.7	1.7
p	£/m^3	1.3	1.3	1.3	1.3	1.3
Δg	m^3	-240	-240	-340	-230	-1,300
$p\Delta g$	£	300	300	440	300	1,700
$-V\Delta g$	£	390	390	570	390	2,200
E	£	1,300	1,300	850	580	1,300
I	£	850	850	850	380	850
N	*integer*	1,020	1,020	1,020	484	1,020
Panel A **(*SRMC*)**						
c	£/m^3	0.44	0.44	0.44	0.44	0.44
$(p-c)\Delta g$	£	-200	0	-290	-200	0
$B-C$ (Equation (15) above)	£	75	270	720	290	3,500
$(B-C)/N$	£/ *person*	0.073	0.27	0.71	0.60	3.4
Breakeven Other Benefits $= -(B-C)$	£	-75	-270	-720	-290	-3,500
Breakeven Other Benefits / Δg	£/m^3	-0.32	-1.2	-2.1	-1.3	-2.6
Breakeven Other Benefits / GHG benefits	*multiple*	-0.19	-0.70	-1.3	-0.75	-1.6
Panel B **(*LRMC*)**						
c	£/m^3	0.98	0.98	0.98	0.98	0.98
$(p-c)\Delta g$	£	-72	0	-110	-71	0
$B-C$ (Equation (15) above)	£	200	270	910	420	3,500
$(B-C)/N$	£/ *person*	0.20	0.27	0.89	0.86	3.4
Breakeven Other Benefits $= -(B-C)$	£	-200	-270	-910	-420	-3,500
Breakeven Other Benefits / Δg	£/m^3	-0.86	-1.2	-2.6	-1.8	-2.6
Breakeven Other Benefits / GHG Benefits	*multiple*	-0.52	-0.70	-1.6	-1.1	-1.6

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 Table E.1 for details on parameters used for welfare calculations.

the cases. Future work could address whether people who responded to a financial nudge were substantially better off if they changed their behavior (Bernheim and Taubinsky, 2018; Butera et al., 2022).

E.5.5 MVPF Analysis using Current US SCC Estimates

This section runs an *MVPF* analysis for the £10 incentive, exactly as presented in Section 4.3, but uses the US SCC estimate of \$51 per ton of CO_2 (Interagency Working Group, US Government, 2021) as opposed to the UK SCC of £241 per ton of CO_2 United Kingdom Government (2021). Results are presented in Table E.9.

For the short-run marginal cost scenario (Panel A), the *MVPF* ranges from -0.16 to 0.28. The negative sign arises because the *WTP* is negative and the net cost to the government is positive. The only scenarios under which the *MVPF* is positive, albeit less than 1, are when we assume away

Table E.9: MVPF Calculations: Using SCC of \$51/ton of CO_2

Case Parameter	Base Case (£10 Incentive) (1)	No Producer Surplus Loss (2)	Vanilla Letter as Benchmark (3)	Targeting High Users (4)	Duration: 1 Yr & No PS Loss (5)
Panel A (SRMC)					
$Cost$	0.44	0.44	0.44	0.44	0.44
WTP	-0.076	0.051	-0.17	-0.17	0.28
G	1.0	1.0	1.1	1.1	1.0
$MVPF = \frac{WTP}{G}$	-0.074	0.051	-0.16	-0.16	0.28
Panel B (LRMC)					
$Cost$	0.98	0.98	0.98	0.98	0.98
WTP	0.0049	0.051	0.011	0.011	0.28
G	1.0	1.0	1.0	1.0	1.0
$MVPF = \frac{WTP}{G}$	0.0048	0.051	0.011	0.010	0.28

Notes: This table computes the MVPF for the three scenarios described in Table Table 7 using Equation (9). *Panel A* shows the results for the short-run marginal cost. *Panel B* shows the results for long-run marginal cost. The values for V and p are the same as those in Table 7. See Table E.1 in Appendix for details on parameters used for welfare calculations.

any producer surplus losses (columns (2) and (5)). This analysis suggests the investment may not be worth making unless other benefits not included here are significant, or the conservation effects can persist for an extended period. 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 under all scenarios, but still remains small and less than 1.

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 in the case of the SRMC would mean that *WTP*, and hence *MVPF*, were zero using the other base case assumptions.[104] We can also analyze the change in *MVPF* if we include the opportunity cost of water.

We run a sensitivity analysis to measure the impact of increasing the *WTP* when we add in the scarcity value of water (£1.4 per cubic meter), and find that the *MVPF* turns positive even in the case of the SRMC. It equals 0.18 in our base case of *£10 Incentive*, and 0.38 in the case of *Vanilla* letter. Importantly, the *MVPF* is greater than 1 for the case with persistent conservation impact up to one year and no producer surplus loss. This implies that in areas with scarce water, conservation programs may be fruitful provided the effects can last for a long time and utilities are able to recover their losses quickly. Similarly, the *MVPF* for the LRMC increases, with the value in the base and *Vanilla* case now equaling 0.26 and 0.56, respectively. Again, even in the long-run — with the exception of the scenario with the long-term benefits coupled with no producer surplus losses — the *MVPF* is below 1. Thus, the net cost of the policy for the government is higher than the potential benefits.

[104] This amounts to an SCC of \$130 per tonne of CO_2e.

In conclusion, in line with our analysis of net benefits, the *MVPF* increases when there are no producer losses and we extend the period for which benefits accrue. In cases where we take into account the opportunity cost of water, it exceeds one, which would mean the benefits from the policy would exceed the net cost to the government if the marginal value of water is high. In all other cases, unless the effects persist for a long time, the government may not find it rewarding to spend resources on conserving water using the interventions discussed in our paper, with the caveat that the SCC is low.

E.5.6 MVPF Analysis for *£15 Incentive* Intervention

This section applies an *MVPF* approach to assessing benefits and costs for the *£15 Incentive* intervention (for the *£10 Incentive* intervention, see Section 4.3). Table E.10 summarizes five *MVPF* calculations. It mirrors the *MVPF* calculations for the *£10 Incentive* intervention. For the short-run marginal cost scenario, *MVPF* ranges from 0.14 to 1.4. The only scenario under which the MVPF is greater than 1 is when we assume away any producer surplus losses and conjecture that benefits last for at least a year (column (5)). This analysis is similar to our benefit-cost analysis in that it suggests the investment may not be worth making unless other benefits not included here are significant, or the conservation effects can persist for an extended period. Using LRMC instead of SRMC increases the after-tax benefits due to a fall in producer surplus loss. The MVPF, though positive under all scenarios as before, still remains small and less than 1 for four of the five cases.

Table E.10: MVPF Calculations: £15 Incentive Intervention

Case Parameter	Base Case (£15 Incentive) (1)	No Producer Surplus Loss (2)	Vanilla Letter as Benchmark (3)	Targeting High Users (4)	Duration: 1 Yr & No PS Loss (5)
Panel A (SRMC)					
$Cost$	0.44	0.44	0.44	0.44	0.44
WTP	0.15	0.25	0.25	0.29	1.4
G	1.0	1.0	1.0	1.0	1.0
$MVPF = \frac{WTP}{G}$	0.14	0.25	0.24	0.27	1.4
Panel B (LRMC)					
$Cost$	0.98	0.98	0.98	0.98	0.98
WTP	0.21	0.25	0.36	0.41	1.4
G	1.0	1.0	1.0	1.0	1.0
$MVPF = \frac{WTP}{G}$	0.21	0.25	0.35	0.41	1.4

Notes: This table computes the MVPF for the three scenarios described in Table Table 7 using Equation (9). *Panel A* shows the results for the short-run marginal cost. *Panel B* shows the results for long-run marginal cost. The values for V and p are the same as those in Table 7. See Table E.1 in Appendix for details on parameters used for welfare calculations.

We can also analyze the change in *MVPF* if we include the opportunity cost of scarce water.

The impact of increasing the *WTP* when we add in the scarcity value of water (£1.4 per cubic meter) is small. *MVPF* increases but still remains below 1 for all scenarios that do not assume that conservation benefits last for a long time. It equals 0.35 in our base case of £15 Incentive, and is just below 1 (0.86) in the case of targeting high-users. In contrast, the MVPF is 2.6 for the case with conservation for one year coupled with an assumption of no producer surplus loss. With LRMC, the MVPF increases, with the value in the base case now equaling 0.42. However, unlike the £10 case, in the long-run — with the exception of the case with long-run benefits and no producer surplus losses — the MVPF is not greater than 1 for any other scenario. Thus, the potential benefits of the policy are greater than the net cost of the policy for the government under a larger range of scenarios with the £10 intervention than the £15 intervention.

F Details Regarding Water Consumption Data

NWG provided us with data on meter readings for each household, and not their daily or monthly water consumption. Therefore, we provide details on the steps undertaken to calculate water consumption, as well as the distribution of meter reading dates. Appendix F.1 provides an illustrative example of the format of the meter readings data and how we computed average daily water consumption pre- and post-consumption for each household. The lack of clear pre-post delineation between meter readings raises concerns about introducing bias in our estimates. We correct this using weights in our regression, with weights equivalent to the fraction of days post-treatment included in the meter readings used to calculate post-treatment consumption, divided by total number of days between the two readings. Appendix F.2 provides more transparency in terms of the distribution of meter readings across different months, so as to shed light on where most of the weights lie.

F.1 Calculating Pre- and Post-Treatment Water Consumption

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

Table F.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

The consumption data from NWG consisted of a series of four meter readings for each household. Each meter reading includes the date of the reading and its corresponding value. For example, 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 the treatment were taken before 01-Jan-2010, the household was dropped as the date is too far back in time to accurately 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 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.

F.2 Distribution of Meter Readings

The distribution across months of the meter readings used for calculating the post-treatment consumption is presented in Table F.2. The row names represent the month of initial reading used to calculate the post-treatment consumption, while the column names represent the month of the latest reading used for the same calculation. For example, the value of 116 in the grid cell with the row name *Aug-2018* and column name *Jan-2019* indicates that there were 116 observations for which post-treatment consumption (after 8$^{\text{th}}$ December 2018) was calculated by taking the difference between meter readings in August 2018 and January 2019. In this case, the weight given to the household would range from 24/154 to 54/154. The former case arises when the meter readings are on 1$^{\text{st}}$ August 2018 and 1$^{\text{st}}$ January 2019 (24 is the number of days between 09$^{\text{th}}$ December 2018 and 1$^{\text{st}}$ January 2019), while the latter case arises when the meter readings are on 31$^{\text{st}}$ August 2018 and January 31$^{\text{st}}$ 2019 (54 is the number of days between 09$^{\text{th}}$ December 2018 and 31$^{\text{st}}$ January 2019). The denominator is the number of days between the two readings.

Table F.2 reveals that for majority of the households, the initial meter reading used for calculating the post-treatment consumption was around July to October 2018, while their final reading

Table F.2: Monthly Distribution of Readings Used to Calculate Post-Treatment Consumption

Month of Latest Reading	09th-31st Dec-2018	Jan-2019	Feb-2019	Mar-2019	Apr-2019	Total
Month of Initial Reading	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Mar-2019	0	0	0	0	2	**2**
Feb-2019	0	0	0	21	11	**32**
Jan-2019	0	0	5	6	7	**18**
09th-31st Dec-2018	0	1	12	3	3	**19**
Nov to 8th Dec-2018	0	11	26	11	9	**57**
Oct-2018	0	7	32	2	376	**417**
Sep-2018	0	6	46	495	3	**550**
Aug-2018	0	116	7,871	88	7	**8,082**
Jul-2018	0	1,860	29	5	3	**1,897**
Jun-2018	0	11	10	6	0	**27**
May-2018	0	3	0	0	0	**3**
Apr-2018	0	2	7	1	5	**15**
Mar-2018	0	3	17	44	6	**70**
Feb-2018	0	10	183	21	3	**217**
Jan-2018	0	77	7	10	1	**95**
Dec-2017	0	1	15	9	2	**27**
Prior to Dec-2017	0	61	44	59	8	**172**
Total	**0**	**2,169**	**8,304**	**781**	**446**	**11,700**

Notes: The table shows the distribution across months of the meter readings used for calculating the post-treatment consumption. The row names represent the month of initial reading used to calculate the post-treatment consumption, while the column names represent the month of the latest reading used for the same calculation.

was in the four month period between January to April 2019. Further, the top four rows indicate that for 72 households, both meter readings used for calculating post-treatment consumption were after the treatment date. The weights assigned to such households equals 1 because 100 percent of the post period is treated.

G Follow-up Survey

Below we list the questions in the online follow-up survey that was administered in March 2019 to NWG customers that had an email address.

1. Do you own or rent your home?
2. Which of the following best describes your home?
   - Detached house, cottage or bungalow
   - Flat
   - Semi-detached house
   - Terraced house
3. Do you have a garden and/or a front lawn?
4. How many people (including you) live in your home?
5. Do you live with your family/partner or in shared accommodation?
6. To what extent is it important for you to save money on your water bills?
   - Extremely important
   - Very important
   - Of slight importance
   - Neither important nor unimportant
   - Not important at all
7. To what extent are you concerned about the environment?
8. Please indicate whether you agree or disagree with the following statements:
   - I will only save water if it helps to lower my utility bills
   - I will only save water if the rest of my community does so
   - I do not believe that water saving appliances are useful
   - I will only save water if it is required by regulations
   - I know what I would have to do in order to save more water
   - I am currently using water wisely
9. Are you currently trying to reduce the amount of water that your household consumes?

10. Please indicate what you are doing or planning on doing to save water:
    - Have shorter showers
    - Turn off the shower when shampooing etc.
    - Check for dripping taps and turn them off
    - Turn off the tap when brushing teeth
    - Turn off the tap when shaving
    - Don't wash dishes under a running tap
    - Request water saving products from Northumbrian Water
    - Check for leaks and repair them
    - Use a water butt
    - Water the garden less
    - Encourage friends and family to save water
    - Other, please specify
11. Northumbrian Water sent out a circular mailer about the aqKWa Savings Engine™ in December of last year. Do you remember looking through the mailer?
12. Please indicate which factors made you look through the mailer:
    - I thought that the mailer looked appealing
    - I opened it to find out more about how to save money and water
    - I open all mailers
    - I thought it was something else
    - Other, please specify
13. Why did you not look through the mailer?
    - I thought it was spam
    - I intended to open it, but I didn't get around to it
    - I don't open mailers
    - I don't remember receiving a mailer
    - I usually throw away my mail
    - Other, please specify
14. Did you visit the Savings Engine website that was mentioned in the mailer?
15. Please indicate which factors encouraged you to visit the Savings Engine website that was mentioned in the mailer
    - I was curious
    - I wanted to know how to save money
    - I wanted to know how to save water

- It reminded me about the importance of saving water
- I wanted to claim the voucher
- Because so many other people have used it
- I was interested in requesting free water saving products
- I want to do my bit to help protect our local environment
- I like using online tools and platforms

16. Please indicate how you felt about the Savings Engine?
    - Easy to navigate
    - Fun
    - Informative
    - Useful
    - Worthwhile

17. Please indicate which of the following suggestions you have implemented or are planning on implementing:
    - Have shorter showers
    - Turn off the shower when shampooing etc.
    - Check for dripping taps and turn them off
    - Turn off the tap when brushing teeth
    - Turn off the tap when shaving
    - Don't wash dishes under a running tap
    - Request water saving products from Northumbrian Water
    - Check for leaks and repair them
    - Use a water butt
    - Water the garden less
    - Encourage friends and family to save water
    - Other, please specify

18. Have you noticed any changes in how you use water since completing the Savings Engine?

19. How likely would you be to recommend the Savings Engine overall to your friends and family? (1 is *Not at all* and 10 being *Very likely*)

20. Please indicate whether you had any trouble throughout the process of:
    - Receiving the mailer
    - Completing the Savings Engine
    - Taking action on the recommendations
    - Other, please specify

21. Do you have any suggestions on improving the Savings Engine™?

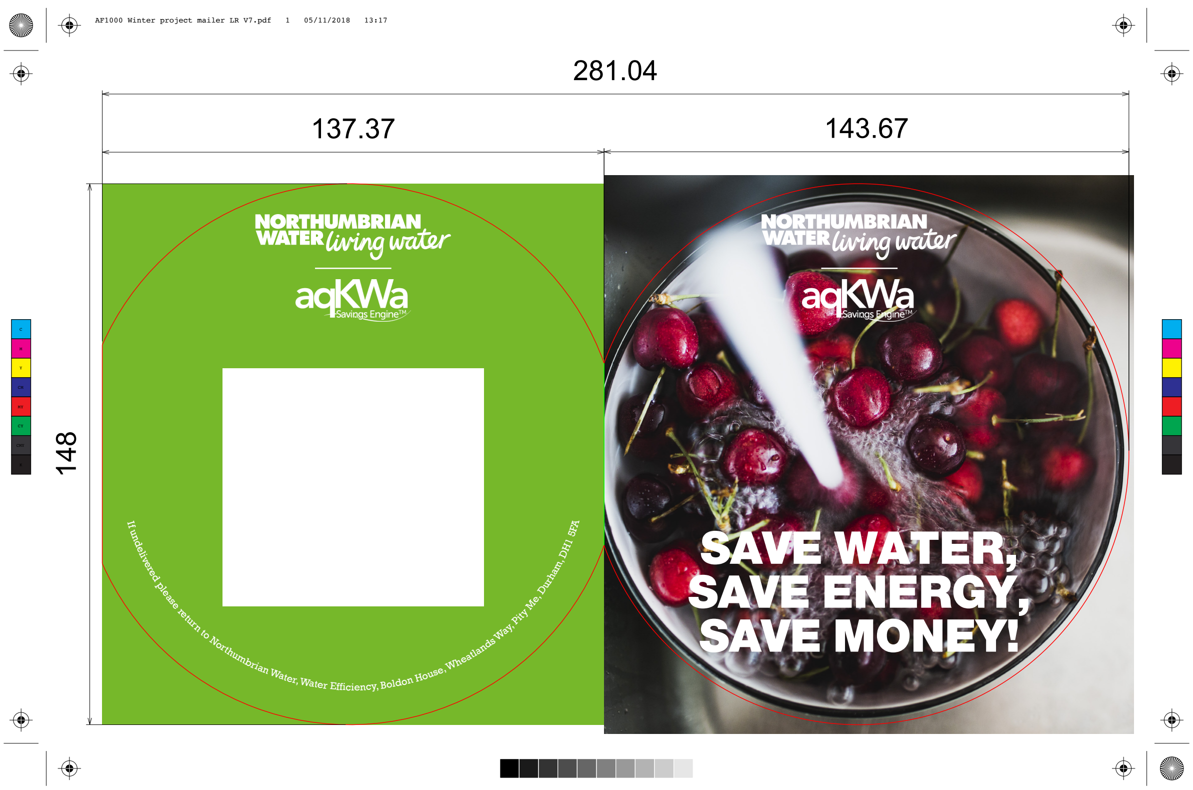
22. Why did you not visit the website?

    - I want to, but haven't gotten around to it
    - I didn't think that it would be worth my time
    - I didn't know how to visit the website
    - I'm not interested in saving more water
    - I don't feel like I need to save more on my utility bills
    - The mailer wasn't that attractive
    - I felt targeted by the mailer
    - I don't think the platform would actually help me save water or money
    - It wasn't a priority at the time

H Sample Letters

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

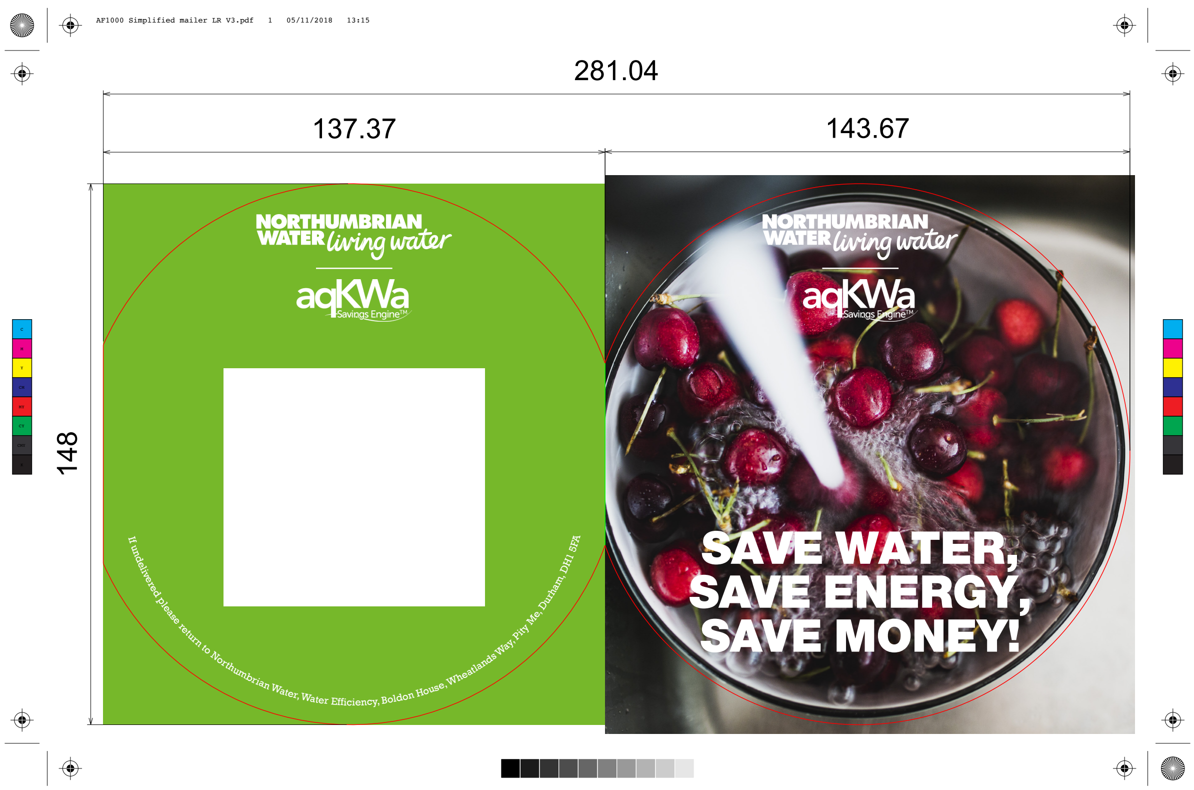
Figure H.1: Vanilla (Status Quo) Mailer



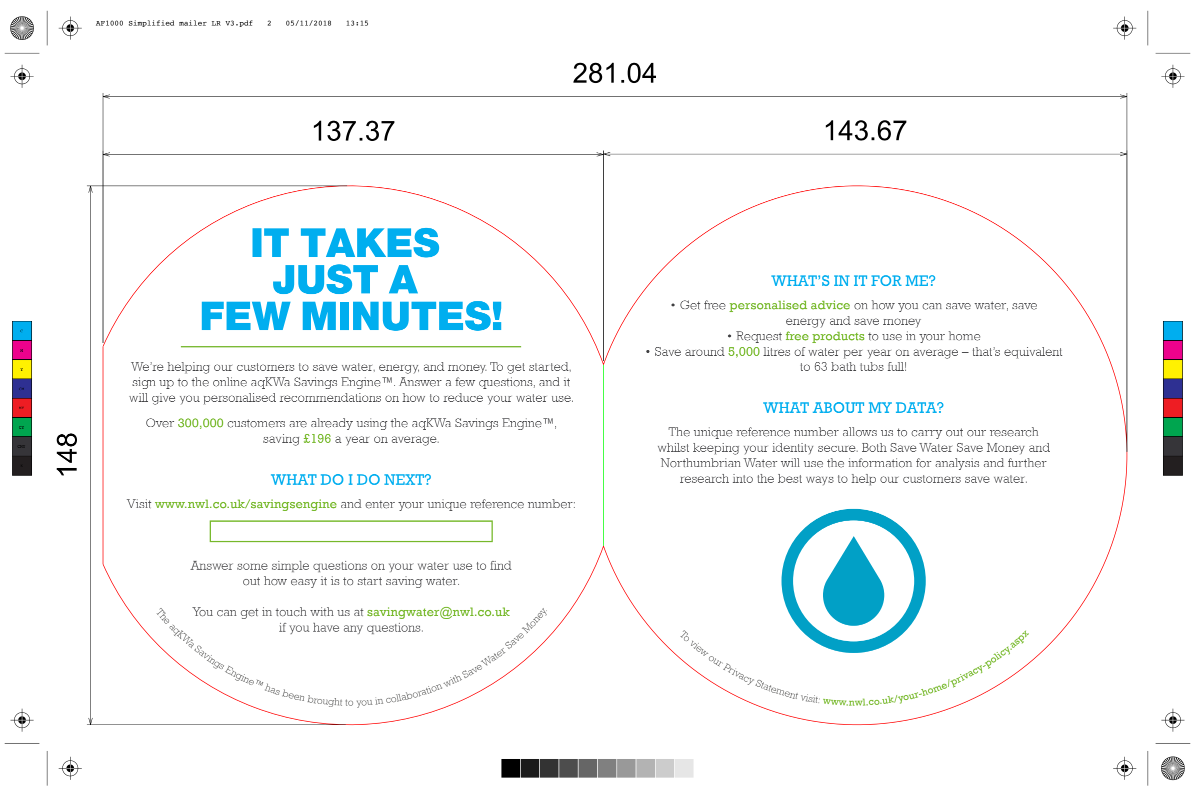
(a) Back and Front Page

(b) Inside the Mailer

Figure H.2: Simplified Mailer

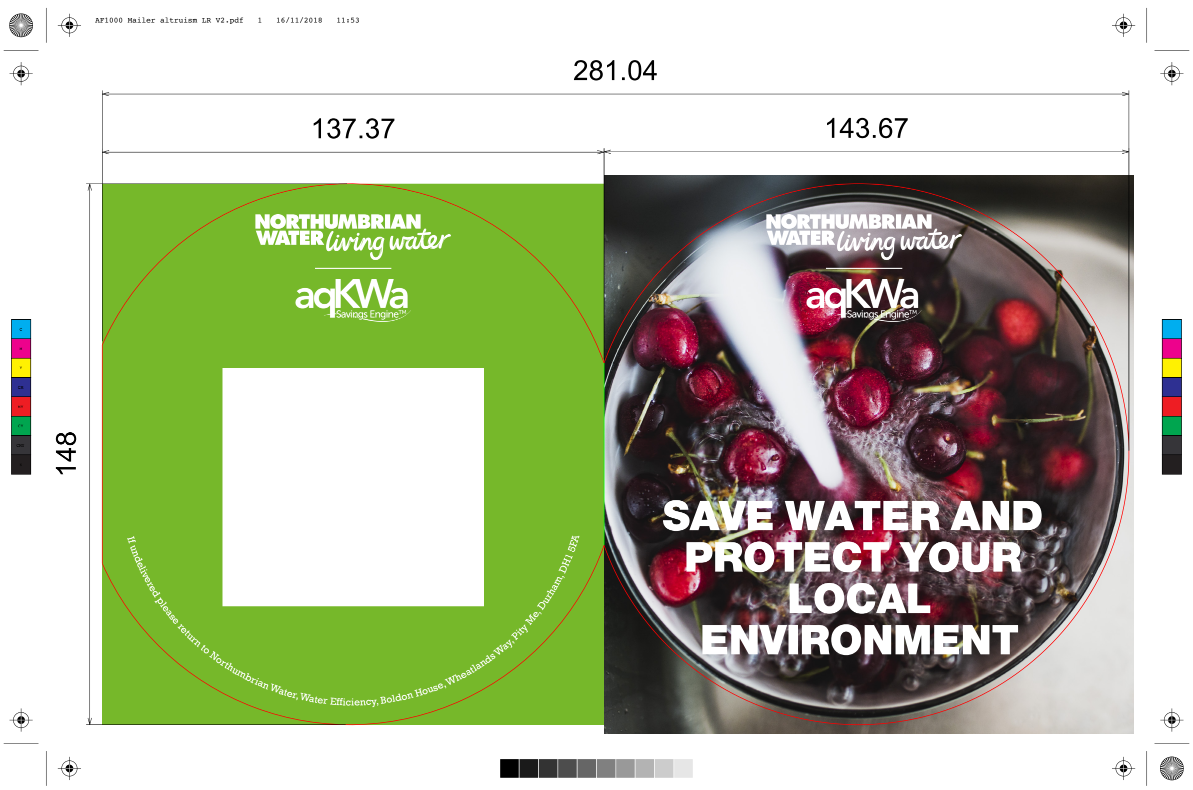


(a) Back and Front Page



(b) Inside the Mailer

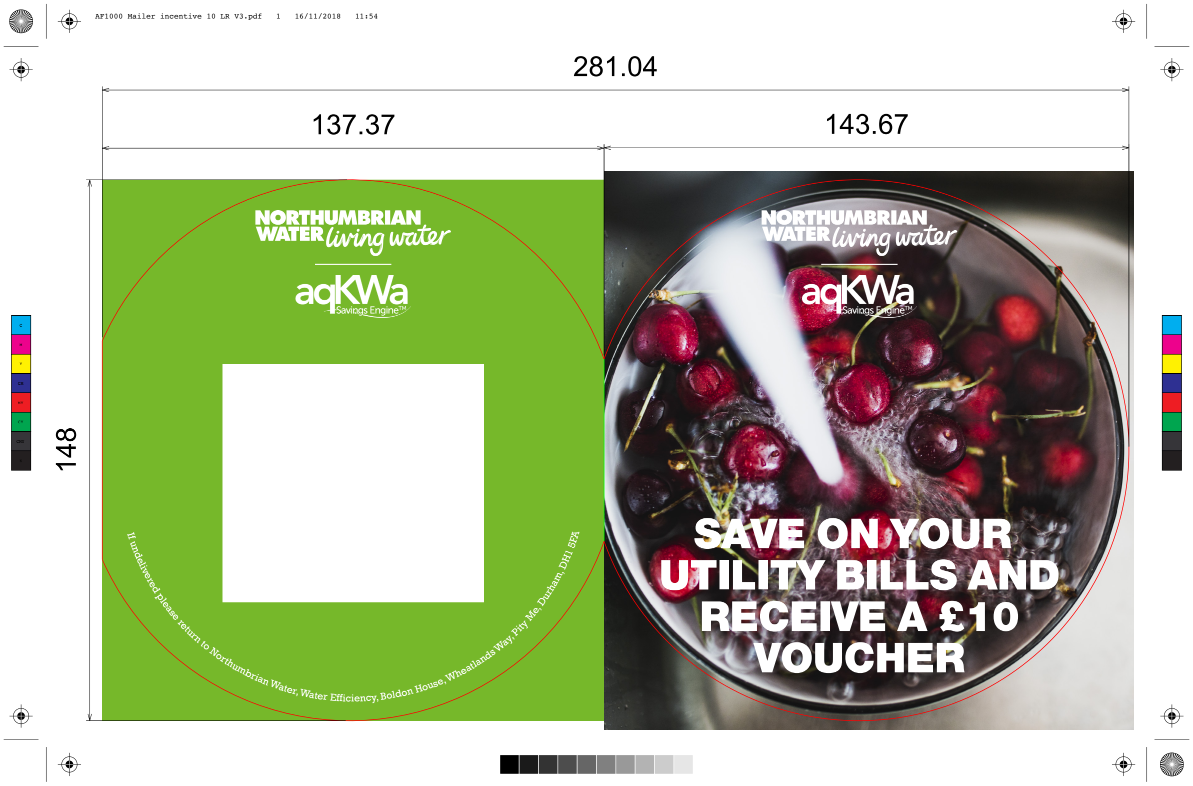
Figure H.3: Altruism Mailer



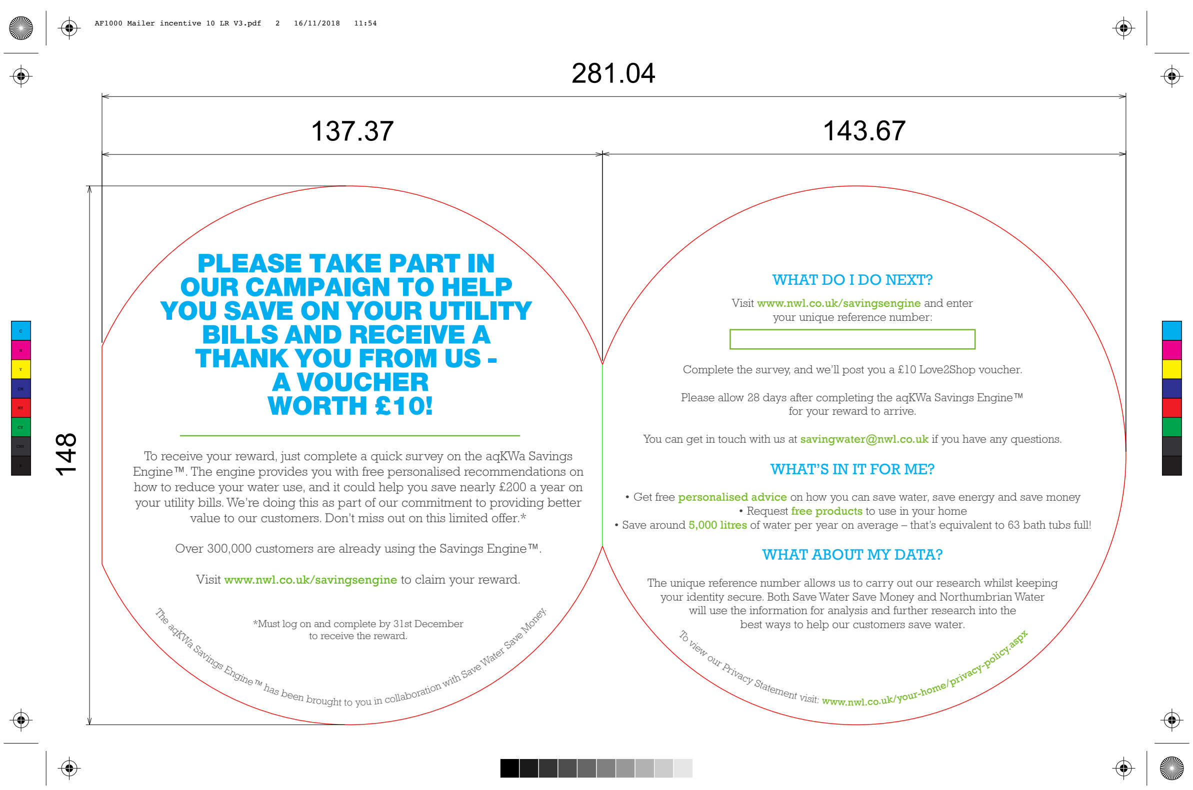
(a) Back and Front Page

(b) Inside the Mailer

Figure H.4: £10 Incentive Mailer

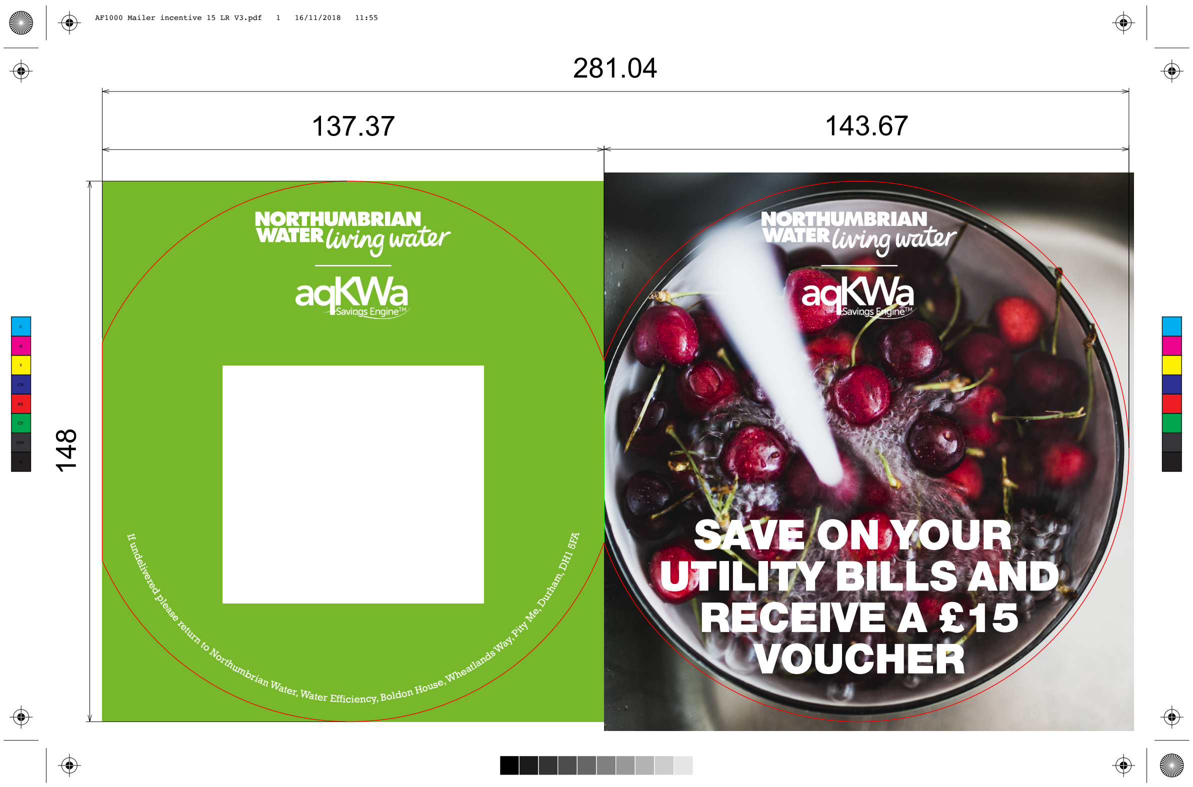


(a) Back and Front Page

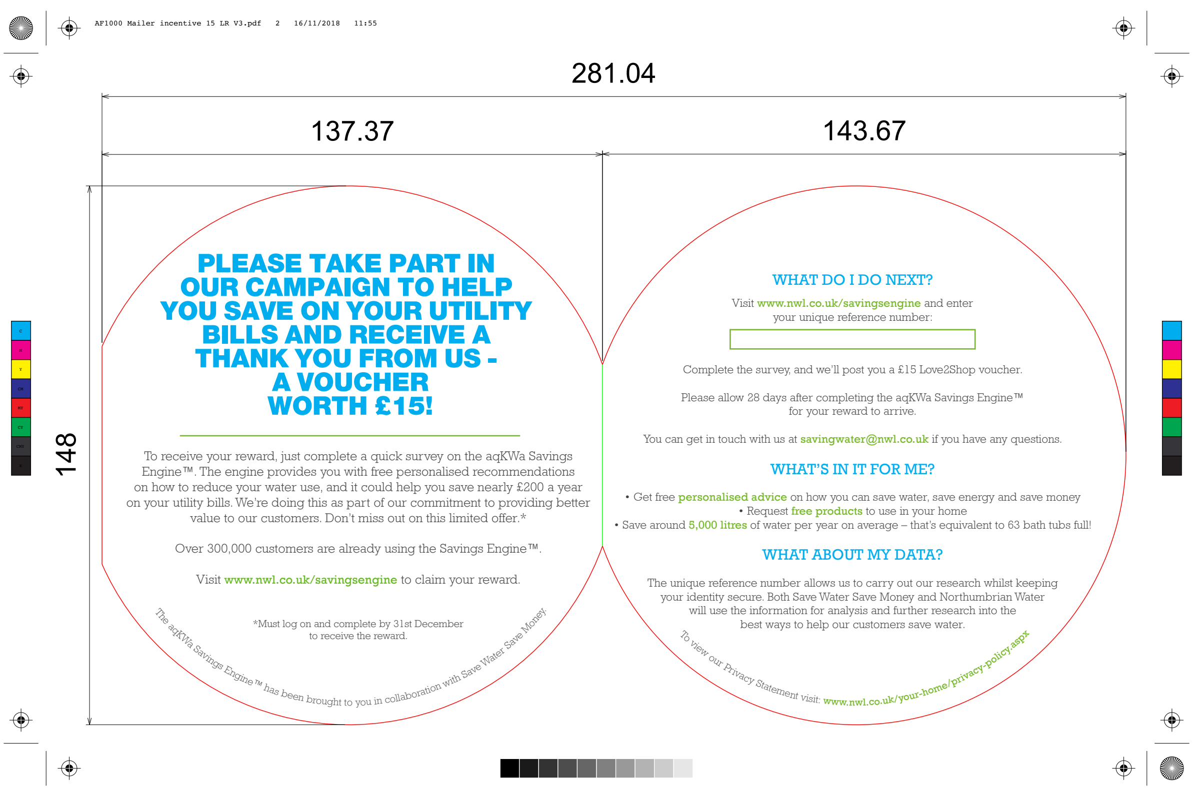


(b) Inside the Mailer

Figure H.5: £15 Incentive Mailer

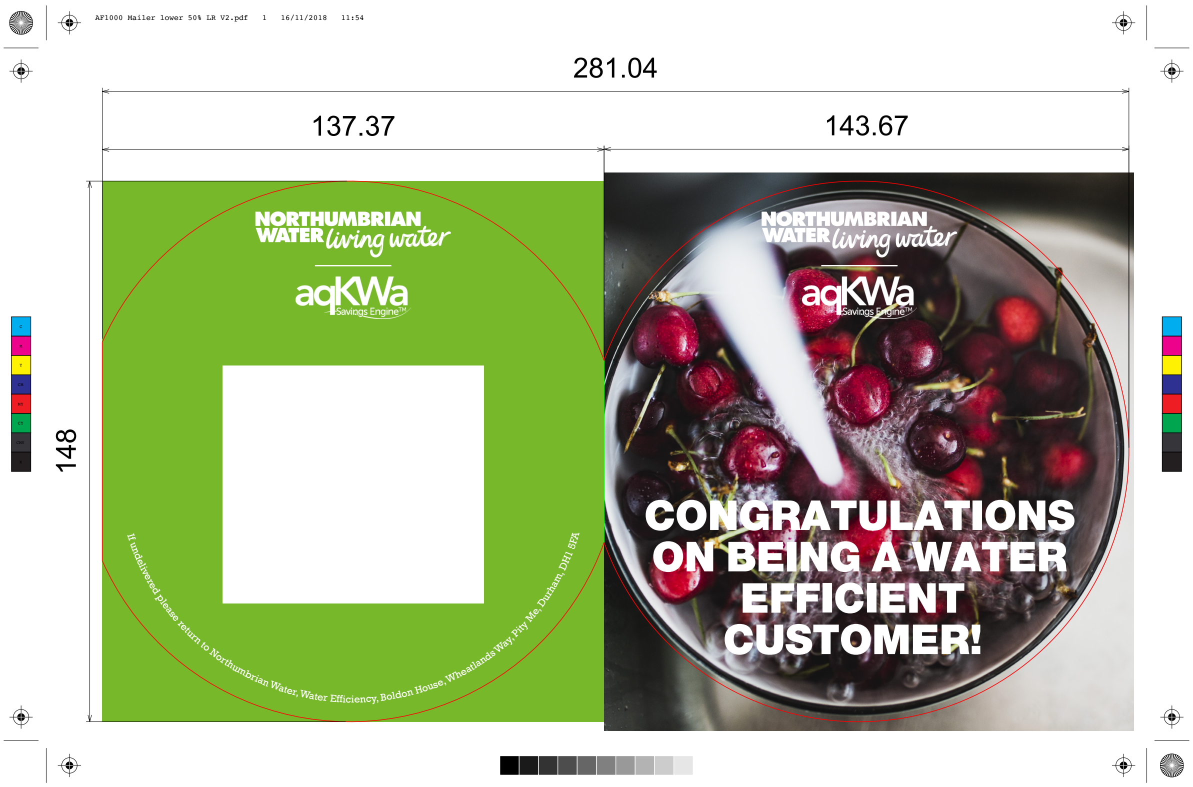


(a) Back and Front Page



(b) Inside the Mailer

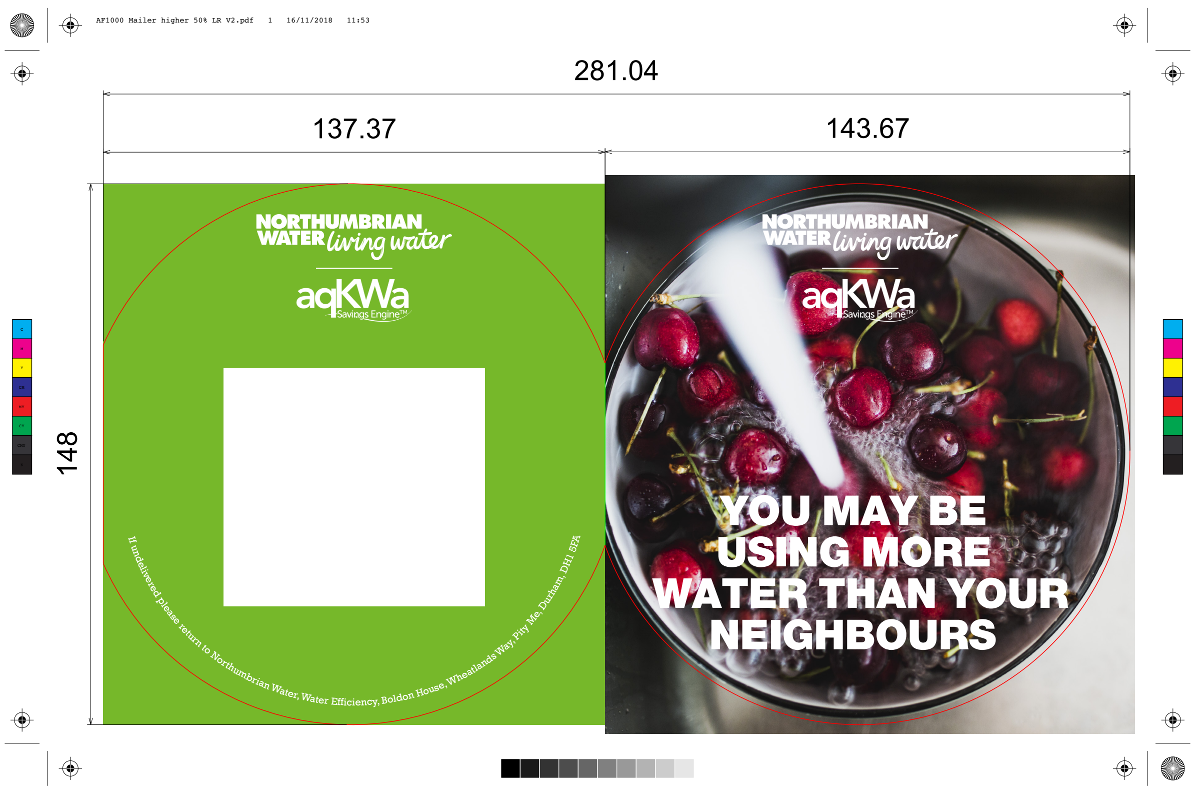
Figure H.6: Moral Cost Mailer - Lower 50%



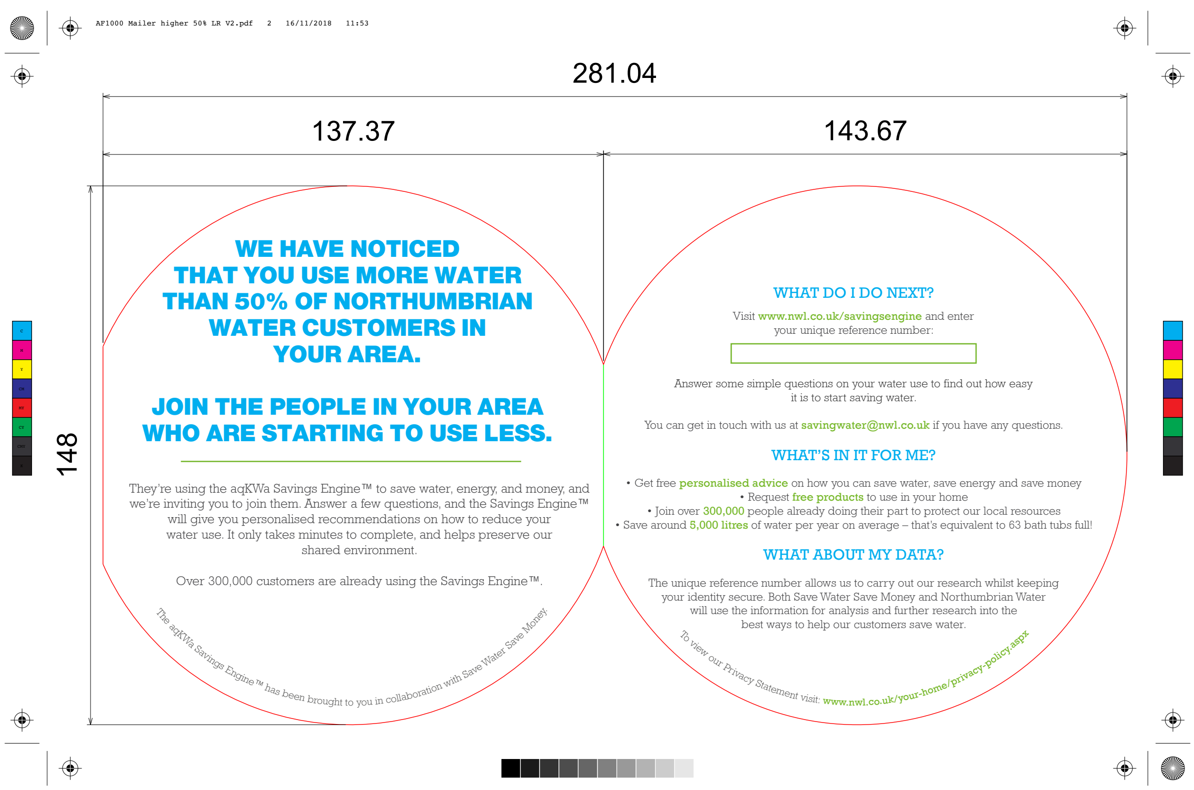
(a) Back and Front Page

(b) Inside the Mailer

Figure H.7: Moral Cost Mailer - Higher 50%

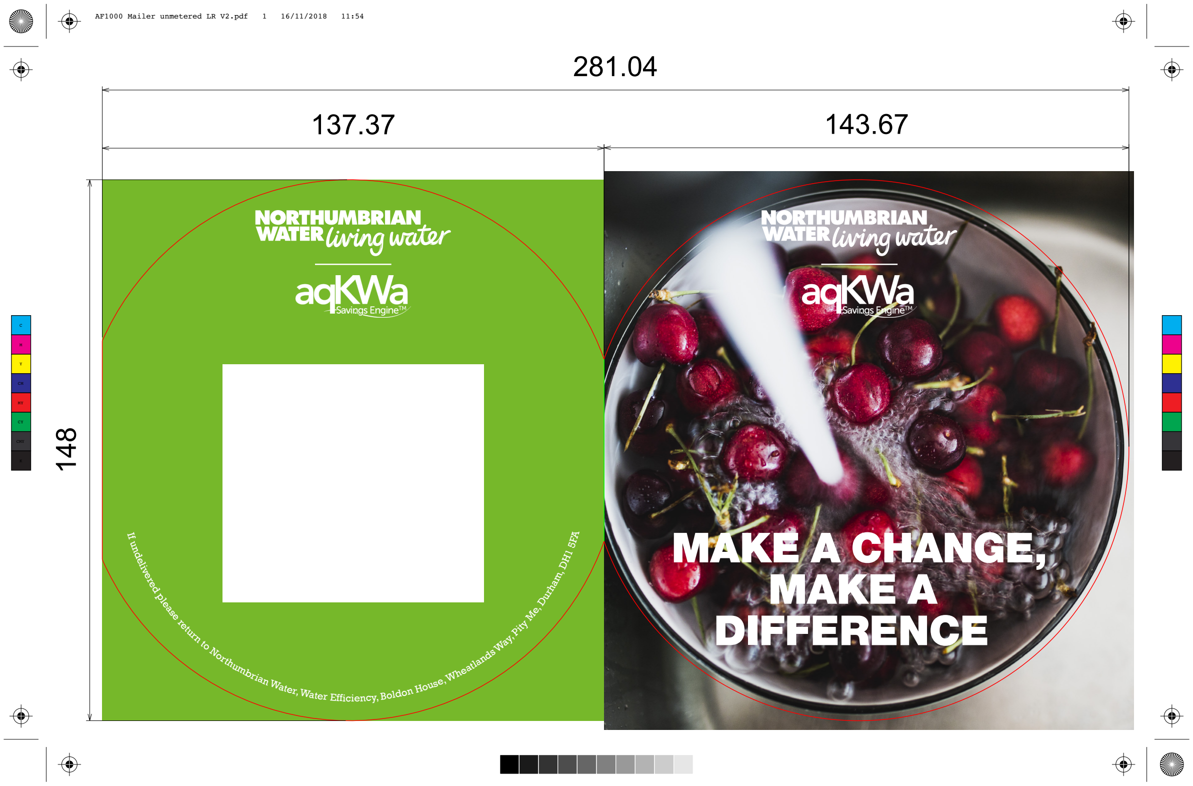


(a) Back and Front Page

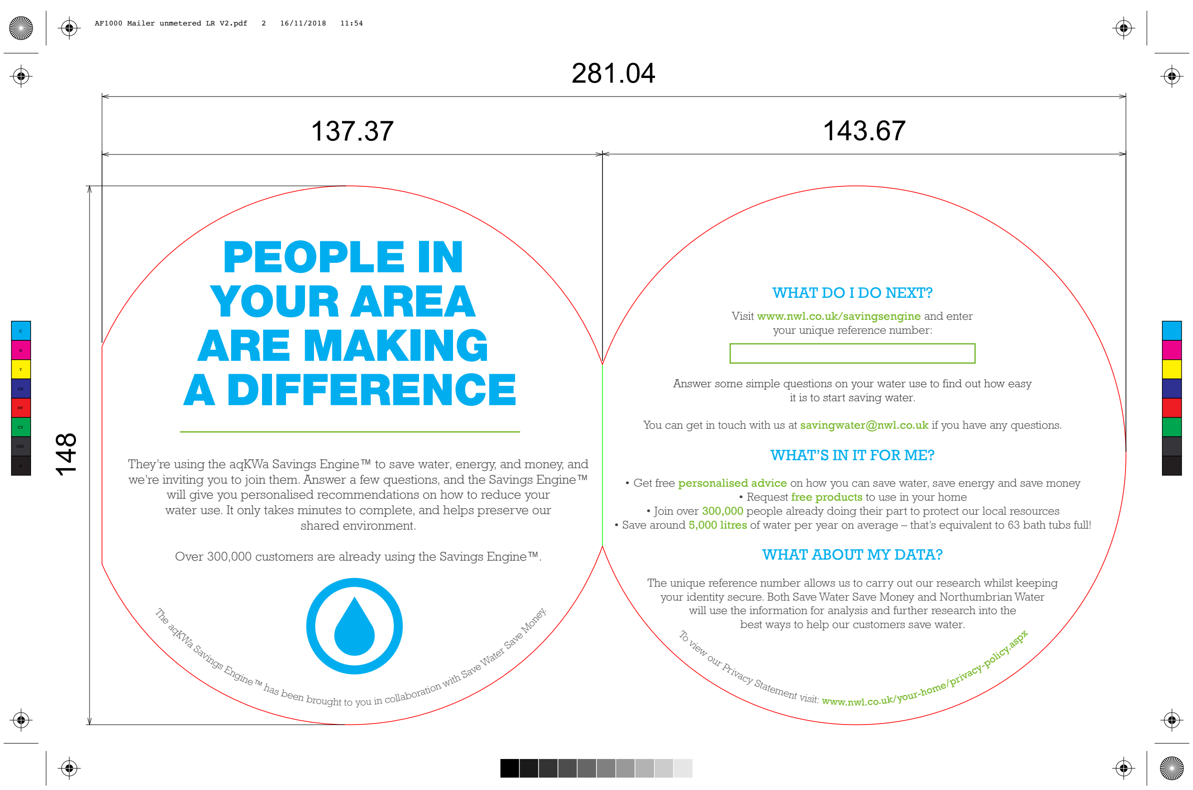


(b) Inside the Mailer

Figure H.8: Moral Cost Mailer - Unmetered



(a) Back and Front Page



(b) Inside the Mailer

Figure H.9: 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