Consultancy Services

Royal London Buildings
42-46 Baldwin Street
Bristol
BS1 1PN

Evaluation of CERT Demonstration Actions Related to Smart Meters – Analysis Methodology and Results

Final Report (11/02/2013)

Authors: Rachel Freeman, Phil Gait

11/02/2013

Contents

1.0 Introduction

This report presents the results of Sustain's analysis of data from RWE Npower's CERT Demonstration Action related to smart meters and associated products. The aim of the trial was to evaluate the carbon savings associated with three different products which were offered in conjunction with Smart Meters (SM)s and Customer Display Units (CDU)s, and the lifetimes of these savings. The products trialled were:

1. Time of Day Tariff Product (Timeout) – An electricity-only, time of day tariff, designed to better understand if customers will respond to price signals to move electricity demand away from peak periods.
2. Enhanced Web Service Product (Bounty) - An enhanced web-based information service that provides detailed consumption data for gas and electricity usage as well as links to energy efficiency information and messaging on energy efficiency.
3. Monthly Billed Product (Buttons) - A provision of monthly bills for both gas and electricity consumption to assess the impact that more frequent billing has on consumption.

The analysis task had the following aims:

- To establish an estimate of the annual carbon savings due to each of the products.
- To establish an estimate of the lifetime of the carbon savings due to each of the products.

The study was designed to establish the incremental CO_2 savings due to the products being trialled, above and beyond any change in CO_2 due to the provision of a SM and CDU. This design was based on the requirement for CERT Demonstration Actions to prove additionality when calculating savings: *Only if suppliers can demonstrate that features of their proposed demonstration action may produce additional carbon emissions reductions to those achieved by Government activity may the action be eligible as a demonstration action*[1]*.* As SMs and CDUs have been mandated for all UK households by the government, savings due to the product only would be additional.

1.1 Nomenclature

To help the reader understand this report, several terms are defined here:

- The groups that got a SM and CDU but not the product are referred to as the "control trial groups" (groups 3, 9, and 12) and the groups that got the SM and CDU and the product are referred to as the "product trial groups" (groups 1, 2, 5, 6, 7, 8, and 11).
- The term "during trial" refers to the period after the customer received a SM up to the end of the trial, and the term "pre-trial" refers to the period for which pre-SM data was available, which varied by household but was between 3 and 8 quarters.
- The term "pre-SM data" indicates data taken from an electromechanical meter that is read by a person (either an RWE npower employee or an occupant of the house) or estimated on average every quarter. The term "SM data" indicates data that was recorded digitally by a SM every half hour and then downloaded by RWE npower.
- The ghost control households were not made aware that their consumption data was being used in the trial and did not receive a SM, therefore all of their consumption is classified as "pre-SM" even though some of it is from the trial period.
- The term "savings" in this document indicates an incremental decrease in energy use or CO_2 emissions when comparing the product group with the control group. When the term "negative savings" is used, it means that there was a relative increase for the product trial groups during the trial period.
- The term URN stands for a Universal Reference Number which is a unique anonymous identifier for each household.
- The term "statistically significant" means that the result is statistically significantly different to zero. Whether a particular value is statistically significant depends on both its magnitude and its variability, so that due to the small sample sizes in this trial, small values will never be found to be statistically significant.

[1] Ofgem's document "Carbon Emissions Reduction Target (CERT) 2008-2011 Supplier Guidance - Version 2" (September 2009)

2.0 Methodology

2.1 Trial Design

The trial was run over a 16 month period, from August 2009 to January 2011, and in addition to the SM data from the during-trial period, up to a year's worth of pre-SM data from the pre-trial period was made available for the analysis. All of the control trial groups and all but two of the product trial groups got a CDU as well as a SM. The main characteristics of the different groups within the trial are shown in Table 1. The final size reflects the count of URNs for which SM data was available; for several trial groups the number of URNs with pre-SM data was less than this, and of these many of the URNs did not have a complete data set that covered all of pre-SM quarters. However, SM data was complete for almost all customers.

Table 1: Characteristics of the Trial Groups

Product	Group	Group Description	Target Size of Group	Final Size of Group	Smart meter?	CDU?
Time of Day (Timeout)	1	Happy Hour	225	171	Yes	Yes
	2	Timeout Regular	225	183	Yes	Yes
	3	Control Group	50	47	Yes	Yes
	4	Ghost Control Group	500	431	No	No
Enhanced Web Service (Bounty)	5	CDU / high communication	112	103	Yes	Yes
	6	CDU / low communication	113	118	Yes	Yes
	7	No CDU / high communication	112	106	Yes	No
	8	No CDU / low communication	113	88	Yes	No
	9	Control with CDU	50	51	Yes	Yes
	10	Ghost Control Group	500	449	No	No
Monthly Billing (Buttons)	11	Monthly billing group	450	355	yes	Yes
	12	Control group	50	75	yes	Yes
	13	Ghost Control Group	500	446	No	No

2.2 A Note on Regression Analysis

Regression analysis is a statistical tool for identifying the "footprint" of a treatment, or intervention, in a trial group's consumption data. It must distinguish this footprint from all of the other factors that affect household energy consumption and from the effects of running a trial, as described here.

The energy use of households is affected by five main **household characteristics**, the last two of which would be affected by behavioural changes:

1. Physical construction of the property and its level of insulation/building envelope efficiency
2. End-use equipment in the property and its efficiency (e.g. boilers, refrigerators, light fittings)
3. The time of year and ambient temperature (energy use rises as temperature decreases, lighting use increases as daylight decreases)
4. The desired level of comfort of residents (temperature settings, lighting levels, use of other electrical appliances)
5. The hours of use of appliances and heating (how much residents are at home at different times of day, length of time spent away from the home each year, number of and age of people in the household, etc.)

When looking at the analysis of trial data, there are four main **trial design** factors that can influence the results:

1. The accuracy of the data in terms of how contiguous it is, whether readings are estimates or actual readings, whether different data sources are mixed, and how closely aligned readings are with the time (week/month/quarter) in which the energy was used.
2. The size of the sample in each group
3. How well-matched the groups are that are being compared with each other (i.e. how different the households were before the trial started in terms of consumption patterns)
4. The length of time the data is available for from both pre-trial and during-trial periods.

Although the trial set-up and the regression modelling methods were designed to take into consideration these factors it is not possible to say definitively that the savings results, even when statistically significant, occurred as a direct result of the customer having the product. This is because there are so many additional unknown and un-monitored factors that can affect household energy use and the effect of the product will be a small percentage of total consumption (less than 10% in most cases). In other words, this analysis could not prove causation between the product and savings, but it did provide a good indication of the effects of the products.

2.3 Methodology to Calculate Carbon Savings

The monitoring... must be designed to enable suppliers to determine the carbon emissions reduction per annum of the measure. (Ofgem)

The approach taken during the study broadly followed the methodology proposed in Sustain's document "Evaluation of CERT Demonstration Actions Related to Smart Meters– Analysis Methodology" from June 2011. Several statistical modelling methods were used during the study, and as the analysis progressed, these methods were adapted to deal with some of the data-related issues that arose:

- Data from pre-trial and during-trial periods varied in its accuracy and its alignment to the seasons, with many of the data points being system-generated estimates rather than actual reads;
- Some of the product groups showed strong differences in consumption patterns when compared to their respective control group in the pre-trial period;
- There was a large difference in the availability and reliability of data, especially between pre-SM and SM data;
- The control groups were relatively small (the target size was 50);
- It was necessary to compare groups over a two to three year time period in which consumption rose and fell with the seasons.

Two basic statistical approaches were used during the study: cross-sectional (between groups in one time period) and longitudinal (between groups over time). After using several different modelling methods and comparing the results – including ANOVA, linear mixed model (LMM) regression on the quarterly data and LMM on SM monthly data – the method selected as producing the most reliable estimate of savings was a monthly difference-of-differences (DoD) method[2]. This method involved analysing the SM on a monthly basis in a LMM, analysing the quarterly pre-SM in a separate LMM model, and then combining the results with a DoD calculation. This analysing of the two data types separately avoided the problems encountered when comparing highly accurate SM data that is aligned with the seasons with averaged and smoothed out pre-SM data.

Final savings values were calculated by extracting the reduction in kWh in the calendar year 2010 when comparing monthly consumption in product groups with the control group, and then adjusting this amount with the existing differences between the product and control groups from the pre-trial period to get an adjusted kWh savings value. This kWh value was then converted to an amount of CO_2[3] savings and also to a percentage reduction in the product group's consumption in the year before the trial. Thus, the two inputs to the DoD calculation were:

1. The annual difference between control and product trial groups in the during-trial period, based on a LMM of monthly consumption data extracted from the SM data.
2. The annual difference between control and product groups in the pre-trial period based on a LMM of quarterly consumption data extracted from the pre-SM data.

The need for a DoD method is illustrated by the graph for Buttons Electric groups shown in Figure 1, which shows average electric consumption in kWh/day for the three Buttons groups – control, product,

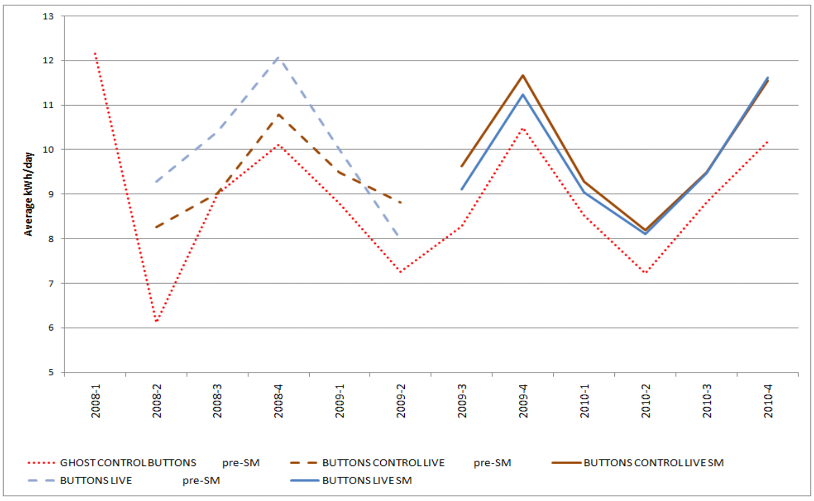
---

[2] DoD can be used where outcomes have been observed for two groups over two time periods, and where the treatment group gets the treatment in the second period but not in the first period. DoD 'removes biases in second period comparisons between the treatment and control group that could be the result of permanent differences between those groups, as well as biases from comparisons over time in the treatment group that could be the result of trends'. [7] A DoD is done by subtracting the average change in the control group from the average change in the treatment group over time. In this case, possible trends influencing energy consumption include weather and macroeconomic influences such as recession which affect householders' purchasing patterns.

[3] Conversion was done using coefficients taken from The Electricity and Gas (Carbon Emissions Reduction) Order of 2008, Schedule 3: 0.431 kg CO_2/kWh electric and 0.190 kg CO_2/kWh natural gas.

and ghost control. The product group had higher overall consumption before the trial started, but then switched to have lower consumption than the control group during the trial. Therefore, to assign the full savings to the product, the difference from the before the trial started must be added to the difference between the groups during the trial.

Figure 1: Illustration of the Need for a DoD – Buttons Electric Groups

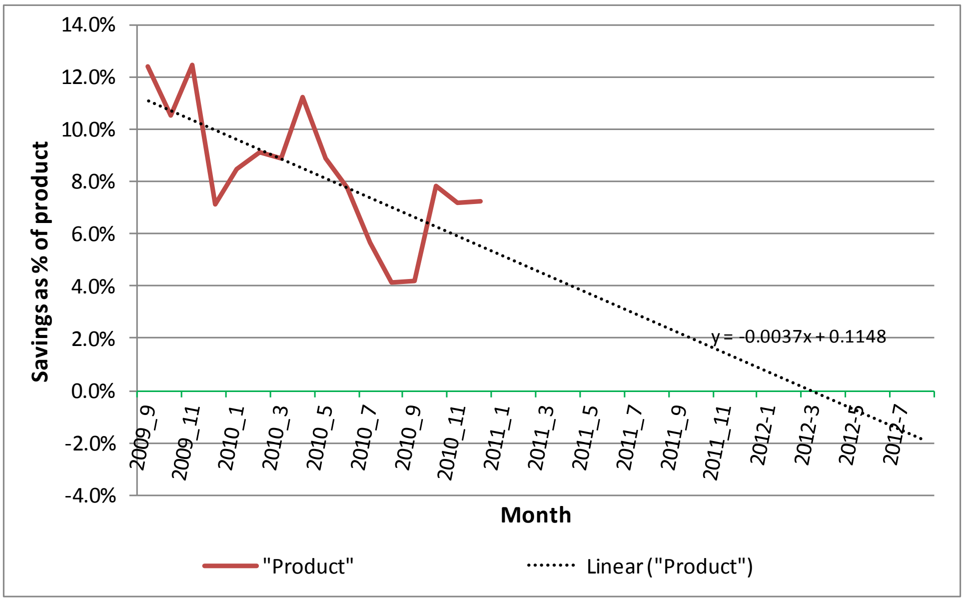


2.4 Methods to Estimate Savings Lifetimes

For behavioural measures it will be for the length of time the consumer reacts to the suppliers' interaction. (Ofgem)

The method used to estimate the product lifetime was based on a graph of adjusted monthly savings derived from the SM data. The values used were difference of differences values; the monthly percentage differences between control and product group(s) were adjusted with the annual average pre-trial percentage difference between product and control consumption. These adjusted differences were plotted on a graph and a trendline was added to the graph; the point at which the trendline crosses the X-axis was taken to be the last month of product effects. In other words, the graph gave the trend, or shape, of savings over time, and the trendline extended that shape to predict when the savings would reach zero. Trendlines were fitted by trying linear, logarithmic and polynomial versions and finding the one that fit the curve and also trended to zero. In some cases no trendline could be found that fit suitably and so a moving average trendline was used and then extended out to zero by hand. An example lifetime graph is shown in Figure 2. This graph has the “wedge” shape that we expected to see, based on evidence from previous studies that behavioural savings tend to decay over time. In this case the lifetime of the product is approximately 30 months.

Figure 2: Example Lifetime Graph with Wedge Shape – Buttons Electric



3.0 Results

3.1 Product Annual Energy Savings

Table 2 shows the final results for all of the product groups. The results are presented as the amount of kWh saved in the year 2010, as a percentage reduction in the product group consumption, and as the amount of CO_2 saved in the year 2010. The savings range from 0.9% to 20.4% of product group consumption, and all of the products show positive savings. Annual CO_2 savings range from 13 to 627 kg. The highest average CO_2 savings by product were achieved by the Buttons (monthly billing) groups, and the Timeout groups had the lowest. These results are for the year 2010, to account for possible teething problems with the data, and they will gradually reduce over the full lifetime of the product. They are mostly in the expected range, with the exception of Buttons gas which is higher than expected; in fact, this product group showed high savings in all of the different analysis methods used which indicates that it could be a reliable result.

Table 2: Summary of Product Energy and CO_2 Savings

Product	Product Trial Group	Annual Savings (2010)		
		kWh/year	% of Product kWh	kg CO2/year
Bounty Electric	CDU_HIGH	156	4.5%	67
	CDU_LOW	91	2.6%	39
	NOCDU_HIGH	95	2.7%	41
	NOCDU_LOW	96	2.5%	41
Bounty Gas	CDU_HIGH	1809	10.8%	344
	CDU_LOW	1286	7.0%	244
	NOCDU_HIGH	216	1.2%	41
	NOCDU_LOW	849	4.3%	161
Buttons Electric	Monthly Billing	271	8.4%	117
Buttons Gas	Monthly Billing	3300	20.4%	627
Timeout Electric	Happy Hour	29	0.9%	13
	Timeout Regular	443	12.8%	191

3.2 Product Lifetimes and Lifetime Savings

The trendline analysis produced estimates of lifetimes between 16 and 30 months, the highest lifetime being for Buttons and the lowest for Bounty electric CDU_LOW. In general, the Buttons groups showed a fairly clear wedge shape that trended to zero and longer lifetimes, while the other groups showed varied savings over time, but tending to be higher in winter months. Average lifetimes for the products are highest for Buttons at 30 months[4], followed by Timeout at 20 months and Bounty at 17 months. Other trials have shown a wide range of lifetimes, varying according to the interventions used and the characteristics of the households taking part, but in general lifetimes have not extended very far beyond the end of the trial. Therefore these results seem to be in line with other results, with the exception of Buttons.

The total lifetime savings were calculated according to the approximate shape of the trendline graphs. For Buttons gas and electric the shape can be estimated as a simple wedge shape. For the other products the majority of the savings occur in the first year, followed by a steep decline. Therefore the savings for non-Buttons products were calculated by adding the first year savings to estimated savings from the following year (or the portion of that year within the product lifetime). The lifetime savings calculations produced a wide range of savings, from 19 kg CO2 for Timeout Happy Hour to 783 kg CO2 for Buttons gas. This shows the importance of the rate of decline of savings when considering the lifetime savings. Gas savings were in general much higher than electric savings. The total lifetime savings calculation is an approximation from a trendline and therefore its accuracy is not very high; however, it gives an indication of the relative total impact of each product.

[4] The lifetime for Buttons gas could be over 30 months, based on trial evidence, but a value of 30 months has been used as a more reasonable estimate.

Table 3: Summary of Product Lifetime Estimates

Product	Product Trial Group	Final Month of Effect	Estimated Lifetime of Product in Months	Estimated Total Lifetime Savings in kg CO_2
Bounty Electric	CDU_HIGH	03-2011	18	84
	CDU_LOW	01-2011	16	46
	NOCDU_HIGH	03-2011	18	51
	NOCDU_LOW	02-2011	17	50
Bounty Gas	CDU_HIGH	04-2011	19	444
	CDU_LOW	02-2011	17	295
	NOCDU_HIGH	02-2011	17	50
	NOCDU_LOW	02-2011	17	195
Buttons Electric	Monthly Billing	03-2012	30	146
Buttons Gas	Monthly Billing	03-2012	30	783
Timeout Electric	Happy Hour	09-2011	24	19
	Timeout Regular	02-2011	17	230

4.0 Interpretation of Results and Conclusions

4.1 Benchmarking the Results

When compared to other smart meter trials, most of the results of this study are in the expected range – savings between 0% and 10% and lifetimes ending within a year of the study – although the results for Buttons gas and Timeout Regular electric were higher. However, there are differences between this trial and the studies we reviewed. The primary one is that other studies did not attempt to identify savings from a product (for example information on a website) *separately* from savings due to the provision of a smart meter and CDU. Therefore it could be argued that these savings are higher than expected; in other words, we would expect this magnitude of effect from providing a SM and CDU and an informational product, rather than from just the product.

To explore this further, an attempt was made to determine what the effects of the SM and CDU were, separate from the product. Because the ghost group were not provided with SMs during the trial, no accurate SM data for the ghost control was available and therefore it was not possible to get savings estimates for the control group from the SM data. However, the quarterly LMM did produce a coefficient related to the effect of having a SM. in both control and product groups. These values turned out to be mostly negative, as shown in Table 4, taken from the LMM quarterly model. We suspect that these negative values are due to the problems with comparing SM and pre-SM data, with pre-SM being much more smoothed out and not showing the seasonal peaks seen in the more accurate SM Data, and that these results are not a reliable picture of the effect of the SM and CDU. In addition, half of the Bounty groups had a SM but no CDU.

Table 4: Estimates of Savings Due to SMs

Product Group	Savings due to SM and CDU (kWh/year)
Bounty Electric	-355
Bounty Gas	-1641
Buttons Electric	-35
Buttons Gas	-10
Timeout Electric	18

Therefore, without a good estimate of the SM and CDU effect it was not possible to determine what the incremental effect from the product was, when compared to the effect of the SM and CDU.

4.2 Highlights and Uncertainties

The following points are of interest:

- Two of the Bounty product trial groups (NOCDU_HIGH and NOCDU_LOW) did not receive a CDU, but their consumption was compared with a control group that did receive a CDU. Evidence from previous trials points to the benefits of having a CDU in allowing households to become more aware of their energy use and manage it so we would expect the NOCDU groups to respond less well overall than the control group, even though they had the informational product on the website. In fact they did show the lowest savings in both electricity and gas.
- We suspect that there a self-selection bias exists in these results because although households were selected randomly, they all agreed to go on the trial voluntarily and so those with more interest in managing their energy use would have been more likely to sign up. This has implications for extrapolating the results to the wider population.
- The Buttons gas group showed high positive savings in all of the models that were run, including the ANOVA, the LMM and the DoD monthly model. This group also showed the longest lifetime of all the groups, indicating a very strong result for the product.
- There was an overall savings in both Timeout groups. It is not clear why shifting to a time of day tariff would encourage savings but a small conservation effect has been measured in other time of day tariff trials, mostly due to customers paying more attention to their energy use. A fuller explanation of the Timeout response could be gained through a detailed review of the survey results in combination with the consumption data, but this is out of the scope of this study.

4.3 Conclusions

The final results show that it is highly likely that the provision of the trialled products to households produced some energy savings in all of the product trial groups, and that the lifetime of these savings was for most products the trial period plus a few months, except for the monthly billing product which showed a lifetime approximation of the trial period plus a year. The results of this study show some clear patterns of savings for the different products, and these results could be used as pointers when rolling out products to a wider population, in terms of which products are most likely to produce lasting energy savings and what the magnitude of those savings would be. Specifically:

[1] The monthly billing products showed the highest consistency of savings when looking at gas, electricity, and lifetimes. These results show that customers receiving regular pricing information that they must act on (i.e. they have to make sure there is enough money in their bank account to pay this month’s bill) make bigger and more long lasting savings than those just getting detailed information about their usage. Total product lifetime savings for Buttons gas were almost double those of the best savings for Bounty gas products.

[2] The groups that got varying amount of information and tips on consumption (Bounty) showed a wide range of savings, from 1% to 11% of pre-trial consumption. This shows that there can be a wide variety of responses to information-only products.

[3] For the information-only product (Bounty), the groups that did not receive a CDU achieved the lowest savings compared to the other groups of the same fuel type, indicating the importance of having a physical display inside the home.

[4] The group that was on a time of use tariff showed considerable savings, but the group that was also offered a Happy Hour in which electricity was at a special low rate reduced their consumption by much less. This could be because Happy Hour households used the opportunity of the low tariff to fuel switch and use electricity instead of gas for heating, but further detailed analysis of half hourly data would be needed to determine this.

[5] The lifetime of the products was estimated using a trendline method and there is a high amount of uncertainty about their accuracy; however, they indicate that except for the Buttons products, savings due to the products would not last much longer than a few months after the trial period.

5.0 APPENDIX A – Background to the Analysis

5.1 CERT Demonstration Action Requirements

Ofgem's document "Carbon Emissions Reduction Target (CERT) 2008-2011 Supplier Guidance - Version 2" (September 2009) provides guidance on what obligated gas and electricity suppliers need to do to comply with the CERT target. Chapter 4 gives details on Demonstration Actions, and how Ofgem will assess the suppliers' activity under this mechanism. In this section key requirements and statements in the Ofgem document that are relevant to the project being proposed have been quoted (in italics) and then addressed.

Types of Demonstration Action

A demonstration action is an action which is reasonably expected to achieve a reduction in emissions, but for which a firm quantified carbon emission reduction cannot yet be attributed. Demonstration actions fall into three main categories:

1. *trialling a technology*
2. *trialling consumer reaction to a technology*
3. *trialling consumer behaviour in response to better information.*

For category three the document states that *the trial should provide customers with better information about their specific energy consumption, or more general information on energy consumption.* All of the product trial groups received better information through having a CDU showing consumption data taken from the SM, thus they would fit into category three. However, of the three products, only the Bounty product is purely informational; Buttons also provides monthly billing in addition to the information and Timeout provides two types of time of day tariff.

Demonstration Action Submissions

Paragraph 4.15 states that: *"As part of determining the carbon emissions reduction, it is necessary to determine the lifetime of any measure trialled as demonstration action. [...] For behavioural measures it will be for the length of time the consumer reacts to the suppliers' interaction. When reporting on the savings and lifetime in their report on the completed action, suppliers must provide reasonable justification of the lifetime."*

These statements indicate that in addition to savings for the first year of the trial, Sustain must provide an estimate of the lifetime of the measure, which is the length of time the energy consumption behaviour of the customer is expected to remain different when compared to before the trial started. Annual savings would be expected to gradually reduce in each of the years within the lifetime.

Additionality

Paragraph 4.17 states that: *In considering the suppliers' proposals for demonstration action Ofgem will apply the additionality criteria as it does for standard measures. For example, we expect the suppliers will consider trialling a range of behavioural measures under the demonstration route, particularly how consumers will react to the provision of information about their own consumption. Only if suppliers can demonstrate that features of their proposed demonstration action may produce additional carbon emissions reductions to those achieved by Government activity may the action be eligible as a demonstration action.*

This statement points to the need to identify savings from other energy efficiency measures implemented by residents during the trial period, especially through any government programmes such as CERT. This was achieved by extracting data on improvements made by participants during the trial period from the end of trial survey.

Demonstration Action Reports

Paragraph 4.26 states that: *Suppliers will be required to submit to Ofgem a comprehensive report assessing the impact of their action. This must contain the following:*

1. *Outline of the methodology of the action that was taken (including details of and significance level of sampling);*
2. *Brief summary of the results;*
3. *Interpretation of the results including potential for their repeatability on a larger scale; and*
4. *Conclusions drawn.*

Full results data should be provided as an annexe. This will not be published.

This report contains the full methodology and results data from the analysis of the trial data and an interpretation of the results. It does not include comments on the potential for the repeatability of the results on a larger scale as this is beyond the scope of the study.

5.2 Current Practice in Smart Meter Evaluation

Energy savings from smart meter trials are most often expressed in terms of a percentage reduction in energy use, compared to use before the trial started. There may also be an evaluation of the persistence of those savings over the medium or long term to establish a lifetime for the savings. Many different approaches to designing trials and establishing savings have been used, some of which have been more successful than others. Some trials produced no savings estimates as the results were inconclusive or not statistically significant. Most evaluations are based on a statistical analysis of consumption data, sometimes disaggregated by end use.

The European Smart Metering Guide [1] includes a best practices guide for ensuring valid estimates of energy savings are achieved and additional factors are taken into account, with the following points:

- Include a control group and trace the participant selection process.
- Combine electricity consumption data with data from questionnaires
- Involve a sociologist to check that relevant social and cultural issues have been taken into account
- Involve a statistician from the beginning of the planning process.
- Consider using statistical methods, e.g. regressions analyses, to document impact.
- Look for sub-groups with high savings – as a supplement to the study of average impact.
- Smart meters, ICT and communication systems may increase electricity demand. This additional load should be accounted for in the energy saving analysis of the field trial.

There are several uncertainties and outstanding issues related to how energy savings are established. These are explored in EPRI's research synthesis [2], which highlights the fact that a straightforward statistical test used to establish savings from feedback, as described above, does not provide a full picture because:

- It establishes association of savings with feedback, but not necessarily causation;
- It does not provide a detailed characterisation of how the savings were achieved by consumers; and
- There is uncertainty about how results of relatively small trials can be extrapolated to the general population as there are differences in how people respond to feedback due to demographics, local energy prices, and other as-yet unidentified influencing factors.

Analysing the energy bills customers to determine savings can be done in one of two main ways: pooling groups together and comparing their mean (cross sectional), or comparing pre and post for each individual household (longitudinal). The first method may require a larger sample size. According to Parker 'With a standard deviation in annual electricity consumption approaching half the mean (as seen in various utility data sets), the required sample size to reach 95% confidence interval can be large if pooled groups are used rather than pre and post analysis methods.' [3] Table 5 presents several methods that have been used to determine savings due to SMs and feedback devices.

Table 5: Methods Used for Determining Savings from Billing Records

Method	Description
Difference of differences (DoD)	The difference-of-differences statistic is the difference between the control and treatment groups in the change in their annual rate of kWh use across the pre- and post-treatment periods. Dividing the difference-of-differences statistic by the average energy use of the control group in the pre-treatment period gives the proportional reduction from the treatment. [4]
ANOVA (Analysis of Variance) Fixed Effects Model	The ANOVA fixed effects model compares the means and variances of three or more groups at the same time to deduce whether there is any difference in the means. If there is a verified difference, a Tukey's HSD test can determine what the differences are between all of the groups. The fixed effect model assumes that data comes from normally distributed populations.
Linear multivariate regression model	In this model, household energy use is a function of a variety of explanatory variables including: group (treatment vs. control); observation period (pre- or post- treatment); weather variables (HDD, CDD, humidity); demographics (e.g. number in household); and an error term reflecting variables not included in the available data.

Method	Description
Difference of differences (DoD)	The difference-of-differences statistic is the difference between the control and treatment groups in the change in their annual rate of kWh use across the pre- and post-treatment periods. Dividing the difference-of-differences statistic by the average energy use of the control group in the pre-treatment period gives the proportional reduction from the treatment. [4]
ANOVA (Analysis of Variance) Fixed Effects Model	The ANOVA fixed effects model compares the means and variances of three or more groups at the same time to deduce whether there is any difference in the means. If there is a verified difference, a Tukey's HSD test can determine what the differences are between all of the groups. The fixed effect model assumes that data comes from normally distributed populations.
Linear multivariate regression model	In this model, household energy use is a function of a variety of explanatory variables including: group (treatment vs. control); observation period (pre- or post- treatment); weather variables (HDD, CDD, humidity); demographics (e.g. number in household); and an error term reflecting variables not included in the available data.
Differenced linear fixed effects (DLFE) model	This is a fixed effects model in which a household fixed effects parameter captures all household-specific effects on energy use that do not change over time. It ensures that no bias exists in the model due to correlation between household characteristics and household assignment across the treatment and control groups. This method allows for the estimation of the effects of household characteristics on response to the treatment.
Fixed Effects Panel Data Model	Data that is available both across a group of customers and over time is known as 'panel' data'. With this type of data it is possible to control at the same time for differences across households as well as across periods in time. This done through the use of a fixed effects model which assumes that differences across customers can be explained by customer-specific intercept terms. Because the consumption data includes both pre and during trial periods, the pre-trial data can act as a control for the during-trial period.

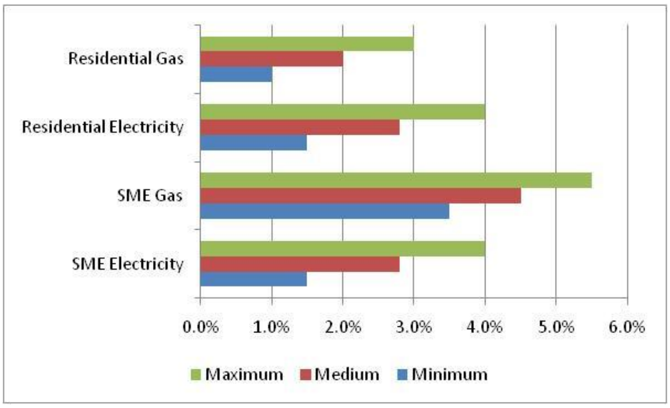
5.3 Benchmarking

In order to gauge the reasonableness of the results of this study, several sources of benchmarking data were accessed.

5.3.1 DECC

The UK DECC has established estimates of gross annual savings of gas and electricity due to smart metering with feedback for both SMEs [5] and residential [6] properties that range from 2% to 4.5%, including a sensitivity analysis, as shown in Figure 3 below. These estimates are lower than savings established through small smart meter trials due in part to the difficulty in extrapolating savings results from a small group of volunteer participants to the general population.

Figure 3: DECC Gross Annual Energy Savings with Sensitivity Analysis



Source: Data from [6] and [5]

5.3.2 Ofgem's Energy Demand Research Project

The Energy Demand Research Project (EDRP), carried out by a consortium run by AECOM, tested consumers' responses to different forms of information about their energy use. Four energy suppliers conducted trials of interventions directed at reducing domestic energy consumption with over 60,000 households, including 18,000 with SMs, between 2007 and 2010. Following are highlights of the results from the final report that are relevant to this study. [7]

- **Energy Efficiency Advice:** The particular combination of advice and historic feedback on consumption that EDF deployed (along with smart meters) reduced electricity consumption by 2.3% overall in the first in-trial year and this was statistically significant. When dual fuel and electricity-only customers were considered separately, the reduction was significant only for dual fuel customers (4.6% saving) and not for electricity-only customers (0.9% saving). Furthermore, the effect was persistent into the second in-trial year for both customer types combined (4.0% saving) and dual fuel customers (2.4% saving), and became statistically significant for electricity-only customers (5.0% saving).
- **Smart Meters and RTDs:** From the literature review, tentative estimates of what could be expected in the UK were that a base level effect of RTDs alone could be less than 3% electricity savings whereas supplementary interventions that increase engagement could double or triple the benefit. Gas savings tend to be of a similar order to those for electricity. In the EDRP savings from the mains RTD devices provided with SMs were generally 2-3% for electricity but higher in the EDF trials (4% overall but 7% for electricity-only customers). These effects were persistent. Only the Scottish Power trial showed no positive effect of RTDs with smart meters and this may be related to the fact that the trials were not presented to customers as smart meter trials.
- **Time of Use Tariff:** The literature shows clear evidence that time of use tariffs can shift consumption from the peak period, and often also bring about reductions in total energy consumption. However, the evidence is almost exclusively from studies in regions with different consumption patterns to the UK. The limited evidence from the UK suggested that only small reductions in overall electricity demand (3% or less) should be expected. EDRP tested TOU tariffs in combination with smart meters and did not provide convincing evidence of an overall reduction in demand.
- **Web-based interventions:** The literature shows the potential benefits of online services to reduce energy demand, but also shows that the potential is rarely realised. EDRP confirmed this, with neither of the suppliers who used web-based interventions seeing any energy savings. The trials also showed that a major reason for failure is likely to be lack of engagement with the web sites, not necessarily a lack of effect among those who use the sites.
- **Population Segmentation Effects:** EDF found clearly defined effects, with smaller households being more likely to save energy overall and to shift consumption from the evening peak period. There were also differences between electricity-only and dual fuel customers. These differences were not due to the electricity-only customers not having gas heating but they are otherwise difficult to explain with any certainty. E.ON findings clearly varied between customer strata. The customers selected because of their initial high consumption reduced their consumption whatever the intervention, which can be explained by “regression to the mean”, rather than the impact of any particular intervention. Behavioural changes were more weakly evidenced in this group. More interesting is the difference between E.ON’s ‘fuel poor’ (FP) and ‘not fuel poor’ (NFP) groups. The generally more positive response to interventions in FP most likely signifies their greater motivation to save money. SSE found significant effects of postcode (for electricity only) and Mosaic demographic group (for gas and electricity).

5.3.3 Other Smart Meter Trials

Benchmarks from several other similar trials to the RWE npower trial are shown in the table below, with some details of the trial design and key findings.

Table 6: Selected Smart Metering Studies and Results

Name of Study	Product Trialled	Savings Estimates
Focus on Energy PowerCost Monitor Study [8]	PowerCost Monitor from Blue Line Innovations: radio signal device that attaches to the electric meter, and a display device for the customer. Display shows amount of electricity and cost of the electricity.	1.4% for the whole group, 3.4% for the top 75% highest users, and 5.4% for those who said the feedback was useful. Billing analysis used.
Evaluation of the Massachusetts PowerCost Monitor Pilot Program [9]	Same as above	When the monitor was used: 2.9% of annual usage. All participants that received a monitor: 1.9% of annual usage Billing analysis used.

Name of Study	Product Trialled	Savings Estimates
Pilot Evaluation of Energy Savings and Persistence from Residential Energy Demand Feedback Devices in a Hot Climate [10]	Two year pilot evaluation of a low cost residential energy feedback system installed in twenty case study homes in Florida, with a follow up in the following year.	Average 7% savings, but reduced to 1% savings after two years. Monitors increased awareness of energy use but did not led to long-term changes. Those claiming to have made long term changes or that did retrofits were still making savings in the third year.
Home Energy Monitors: Impact Over the Medium-term [11]	Four-month trial of a new home energy monitor, and a follow-up study 11 months after the initial trial ended.	Average savings 7.8% after four months, but not sustained over the medium term (after 15 months). Final savings for the group that kept the meter but had not developed new habits was 3% and for the group that kept the meter and had developed new habits was 8%.
Reducing Household Energy consumption: A Qualitative and Quantitative Field Study [12]	Several types of energy information feedback were supplied to 120 households on a monthly basis. Printed information informed households on their energy consumption for the month compared with average consumption of similar households.	3.7% reduction, 2.48% reduction, and 10.7% increase in energy use for high, medium, and low users respectively. Key significant influencing factors were: environmental beliefs, predicted personal behaviour, the relative level of pre-trial energy use, and the number of occupants. Chi square and linear regression analysis used.

5.3.4 Conclusion

Based on all of these studies, we would expect a savings range of between 0% and 10%.

5.4 Assumptions

We established several assumptions at the beginning of this analysis planning which arose as a result of the way the trial was designed:

- The ghost control group will reflect changes in consumption in the general population due to weather and macroeconomic factors but will not be affected by the existence of the trial.
- Households within each of the three product groups (including the control and ghost control) have similar demographics and energy usage patterns and their consumption should be a normal distribution before the start of the trial.
- Control trial groups that get the SM and the CDU will respond by using less energy initially than the associated ghost control group.
- Product trial groups that get the SM, CDU and the product will respond by using less energy initially than the associated ghost and control trial groups.
- The differences between the product trial groups and control trial groups will change over time, most likely getting smaller until they reach zero or close to zero.

The analysis did not prove all of these to be true in the end.

5.5 Uncertainties Due to the Trial Design

There are several issues that have added to the uncertainty about the savings results which are to do with the design of the trial and the availability of data:

- The fact that quarterly consumption values from non-SM meters (pre-trial for all groups and during-trial for the ghost groups) were generated from read periods that were not aligned to standard quarter start and end dates, and some of these reads were estimates, meant that there were significant misalignment and rounding effects. This introduced additional variability in the data (by moving consumption between periods in a random way) and also a bias when actual consumption peaked or troughed in a quarter (which was particularly the case for winter gas)
- There were many gaps in the pre-SM data, meaning that not all of the pre-SM read periods were contiguous. This meant that sometimes the kWh/day value for only a part of the quarter

had to be used for the whole quarter because there were not enough read periods over the standard quarter, although quarters that had data for less than 50% of the period were eliminated.

- The fact that pre-SM data was generated by a different physical meter from the SM data, which was differently calibrated and read at a different frequency, means that there were differences in the quality and accuracy of the data. Although theoretically changes in actual usage at a particular household between pre-trial and during-trial periods would be incremental (i.e. not more than 20% at the most) the simultaneous change in metering technology complicates estimates of the change in consumption due to being on the trial by adding uncertainty related to how the data was collected. In other words, when more than one variable changes at the same time it makes it harder to differentiate the causation of a change.
- We have assumed that the SM data is more accurate than the pre-SM data; however, there may have been teething problems with the SMs that we do not know about.
- Some of the trial groups had a small number (less than 40) of participants in some of the quarters from the pre-trial period. This is especially true of the Timeout control and product trial groups. This will mean that the trial results have a much higher variability and are thus less likely to be found to be statistically significant.
- Identifying an estimate of change in consumption in the product trial groups does not necessarily mean that causation has been proven, i.e. that the changes can be directly attributed to the product. Further examination of the consumption data in combination with associated survey data could shed light on the causation; however this more detailed analysis is out of the scope of this study.

1.0 APPENDIX B – Data Cleaning and Data Statistics

1.1 Data Cleaning

Two basic types of consumption data were delivered to Sustain by RWE npower: smart meter (post installation) and pre-SM (pre installation or ghost control). Most of the data was assigned to the customer through a Unique Reference Number, or URN.

The final version of the data was delivered to Sustain on 19th April. The data consisted of:

1. 1834 DATA - survey data from all three waves, identified by Universal Reference Number (URN)
2. Previous Consumption Data for customers on the trials (pre-SM data)
3. SMRS MPAN Extracts 1 - 13 Matched - SM consumption and CO_2 data identified by URN
4. Ghost Control Consumption Data - ghost control data (pre-SM data)

Upon receipt of the data, Sustain performed the following data cleaning tasks:

1. Create a table of URNs and trial group associations: Although most of the data was identified by URN and also by the trial Group (1 to 12), not all records had both – some had only the URN or only the Group, and some had neither. Therefore the URNs and Group names were extracted from both the survey data and the SM and pre-SM data, and then combined to produce a master reference list of URNs and Group names. Some records were not identifiable and thus were not useable. If a record was identified to a group but had no URN, the data was put into the general pool of data for that group.
2. Extract only the daily gas and electric totals from the SM data and compile into one file, identifying all records by URN or Group or both
3. Reformat previous consumption pre-SM data and ghost control data, to create an average for each quarter; this was done with the following method:
    a) Sort data by URN then by service (electric or gas) then by reading date
    b) Calculate the days in the period between subsequent reads (read period) for reads that have a previous read for the same URN
    c) Re-calculate the advance and compare with the advance provided in the data. If the calculated advance and given advance don’t match, mark the read period as unusable (as we do not know how many days are in the read period).

d) Calculate average electric and gas use per day (kWh/day) in the read period for the useable periods.

e) Assign the average kWh/day value to each quarter that the read period extends over. Quarters are defined according to typical seasons (spring, summer, autumn, winter) as electric and gas use is markedly different during these periods, with the majority of gas being used in the winter quarter and electricity use also going up in winter, while the summer months are low-use periods for both gas and electric with little weather-related use. The table below show the definitions of the quarters, with final quarter including the last month for which data was received, January 2011.

Table 7: Definition of Quarters

Quarter	start date	end date	days
2008-1	01/03/2008	31/05/2008	92
2008-2	01/06/2008	31/08/2008	92
2008-3	01/09/2008	30/11/2008	91
2008-4	01/12/2008	28/02/2009	90
2009-1	01/03/2009	31/05/2009	92
2009-2	01/06/2009	31/08/2009	92
2009-3	01/09/2009	30/11/2009	91
2009-4	01/12/2009	28/02/2010	90
2010-1	01/03/2010	31/05/2010	92
2010-2	01/06/2010	31/08/2010	92
2010-3	01/09/2010	30/11/2010	91
2010-4	01/12/2010	28/02/2011	90

f) If more than one read period exists in a single quarter, combine these proportionally to give a final average kWh/day for each quarter.

g) If the read periods cover less than 50% of the quarter period, eliminate that quarter from the useable values.

4. Calculate quarterly averages in kWh/day for each URN from the SM data where there are data for at least 50% of the days in the quarter.

5. Combine pre-SM and SM quarterly data by URN, according to what quarter the customer was switched to a SM.

6. Identify outliers (i.e. customers with significantly higher or lower consumption than the average) in the pre-SM and the SM data. This was done because we are interested in the effect of the products on average consumers; evidence from other studies shows that very high and very low users generally respond differently to trials similar to this one, but as there are not enough of these customers to assess this behaviour we did not split the URNs into consumption tiers and do a tier-by-tier analysis. This was done by:

   a) Calculating the average kWh/day for each URN from the SM data, multiplying by 365 to get annual consumption, and then dividing that by the benchmark average from Ofgem for households[5]. URNs with less than a third or more than double that average were identified as outliers

   b) Calculating the average quarterly consumption for each URN from the pre-SM data, multiplying by 4 to get annual consumption, and then dividing that by the benchmark average from Ofgem for households. URNs with less than a third or more than double that average were identified as outliers

   c) Combining the two lists of outliers and removing them from both the pre-SM and the SM data sets.

[5] Ofgem currently defines typical annual domestic consumption as 20,500 kWh for gas and 3,300 kWh for standard electricity. Source: 'Revision of typical domestic consumption values', August 2010, OFGEM (www.ofgem.gov.uk/Markets/RetMkts/Compl/Consumption/Documents1/Review%20of%20typical%20domestic%20consumption%20values.pdf)

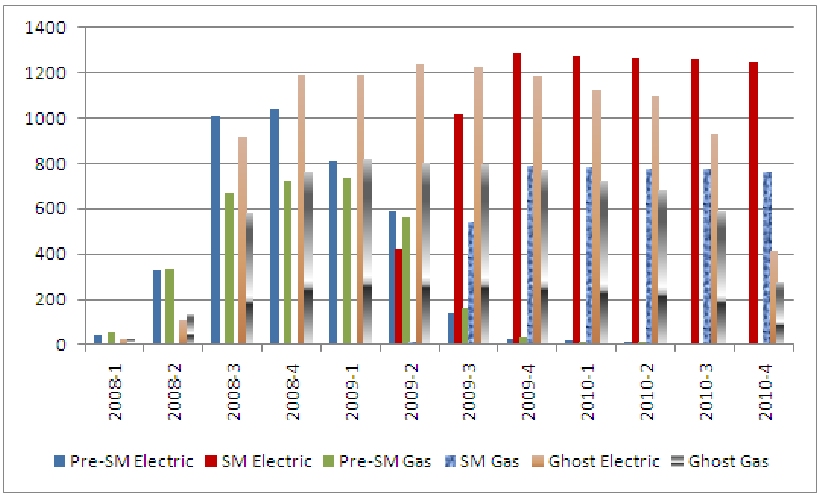
1.2 Basic Data Statistics

This section presents an overview of the data that was received.

1.2.1 Useable Data

Figure 4 shows the count of pre-SM, SM, and Ghost quarterly values for gas and electric consumption. As expected, the count of SM electric is the largest due to Timeout being only electric and SM data being more complete. There is also a consistently high amount of data for the ghost electric and ghost gas categories. Quarters 2009-2 to 2009-3 show both pre-SM and SM data, being the quarters during which customers were switched over to smart meters. Timeout customers were switched over to SM's mainly in 2009-2 and the rest mainly in 2009-3. The first two quarters of 2008 do not have enough data to be able to use them in the analysis.

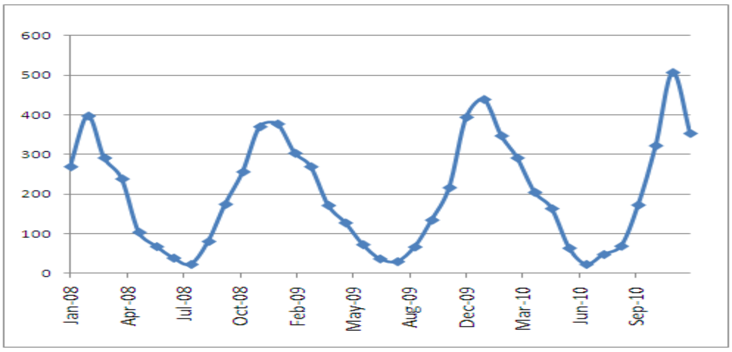
Figure 4: Count of Pre-SM and SM Data Points over Time



1.2.2 General Observations on Consumption

- Heating Degree Days (HDD)[6] data were downloaded from the Carbon Trust for the Midlands region for a base of 15.5°C. HDD values were higher in the winter of 2009-2010 than in the previous year, and higher again in the winter of 2010-2011 (up to January 2011), as shown below. The average HDD in the winter of 2010 was 9% higher that the winter of 2009, indicating that total gas use for heating and electric use for heating should be higher.

Figure 5: Heating Degree Days for the Midlands

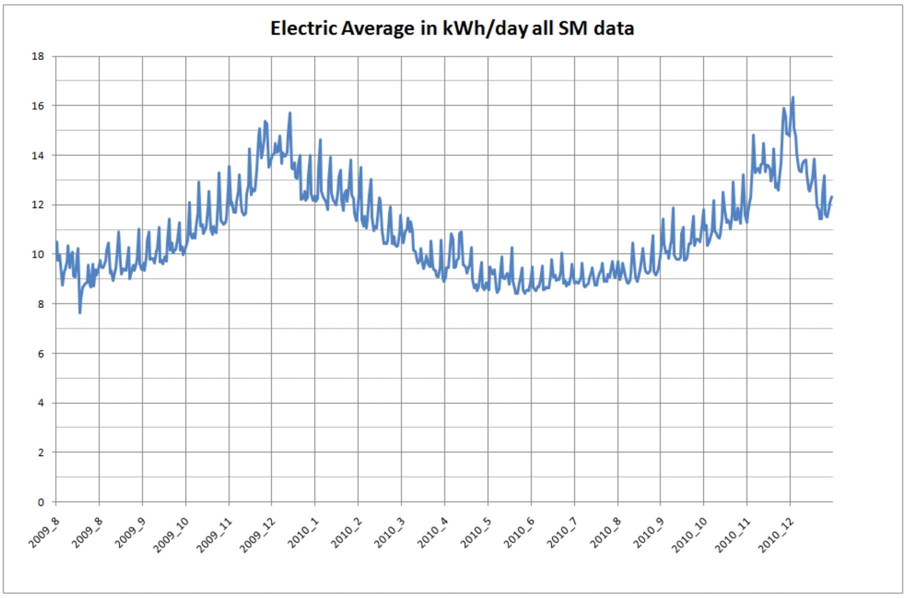


- Analysis of the SM data over all the trial groups shows that electricity use in the winter of 2010-2011 was slightly higher than in the winter of 2009-2010. The graph below shows daily averages for all the SM data combined. Electricity use is around 50% higher in winter than in summer – this is likely

[6] 'Heating degree days are a measure of the severity and duration of cold weather...a summation over time of the difference between a reference or 'base' temperature and the outside temperature. The base temperature is defined as the outside temperature above which the heating system in a building would not be required to operate. Published degree days in the UK are calculated to a base temperature of 15.5ºC for general use.' Degree Days for Energy Management, Carbon Trust, 2010

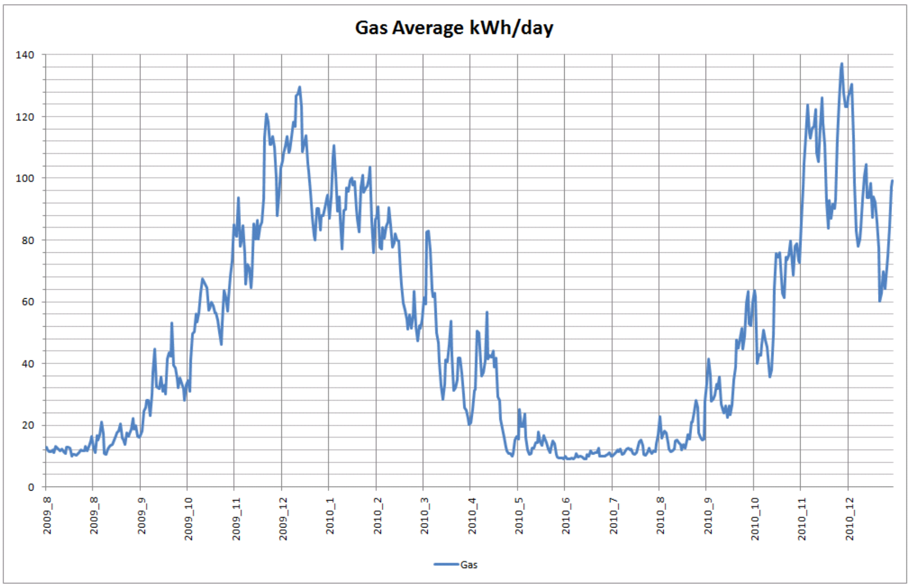
due to factors such as electric heating, more use of electric clothes dryers and longer lighting hours. Baseline use is around 9 kWh/day.

Figure 6: Average Electric Use During Trial



- A similar analysis of the gas data shows that winter gas consumption peaked at around 137 kWh/day in December 2010, compared to a peak of 129 kWh/day in 2009, an increase of 6%. This agrees with the expected rise due to the change in HDD values. The average use of gas in relation to outside temperature was almost the same for both winters (0.215 kWh/day/HDD). Baseline (non-weather-dependent) gas use was around 8 kWh/day.

Figure 7: Average Gas Use During Trial



1.3 Comparison of Pre-SM and SM Quarterly Averages

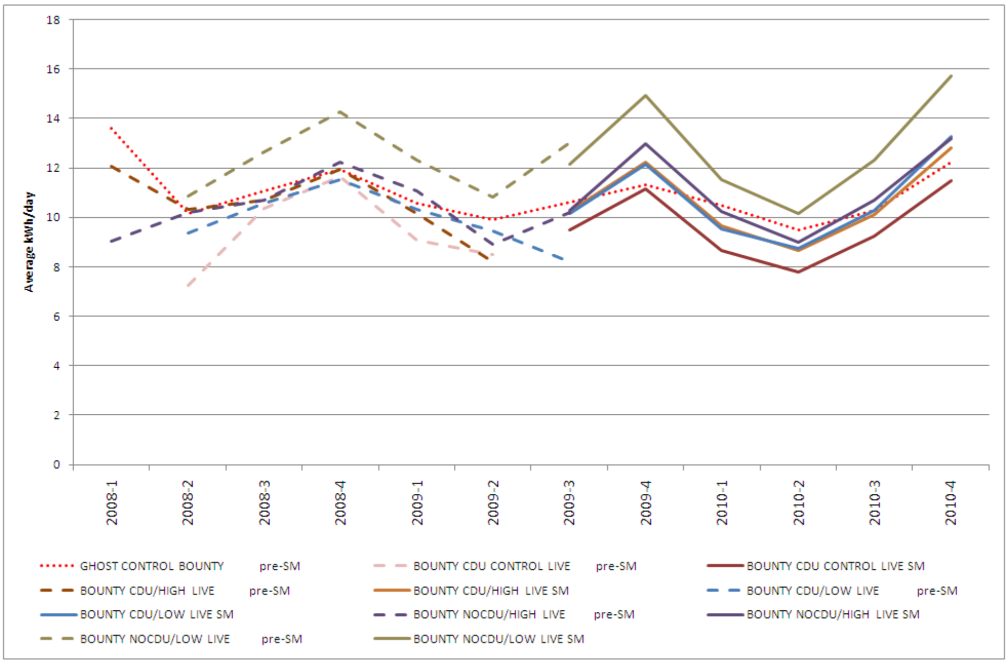
The graphs presented here show the 'raw' consumption data and differences between the groups may not reflect the savings due to the SM or due to the product as they are not adjusted to reflect differences between the groups before the trial started. However, they do give an indication of general trends over time and how well the pre-SM and SM data match each other for each group. Data points were only calculated when there were at least 5 data points for that quarter for the trial or product group, meaning that there are a few gaps in the data.

Timeout

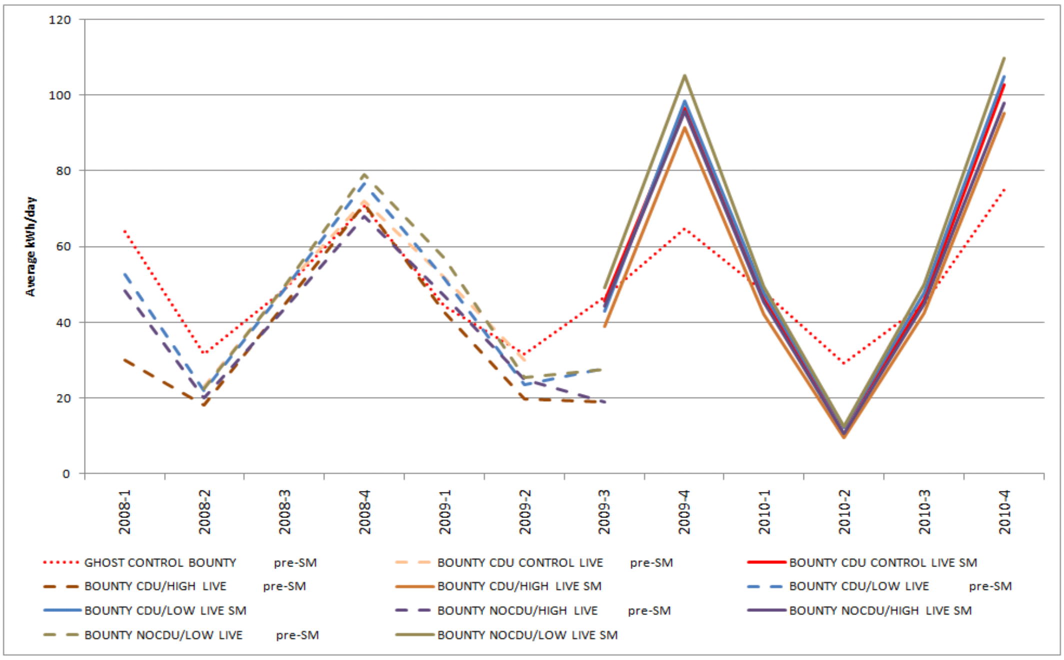
The timeout chart shows a good correlation between ghost control and trial groups before the start of the trial, except at the start of 2008, and a slightly higher ghost control usage than other groups after the trial has started. Very few data points were available in the pre-SM data for quarters 2009-1 and 2009-2.

Bounty

The chart for Bounty electric consumption shows a fairly good correlation between the ghost control and the trial groups. The SM product averages are all above the control.

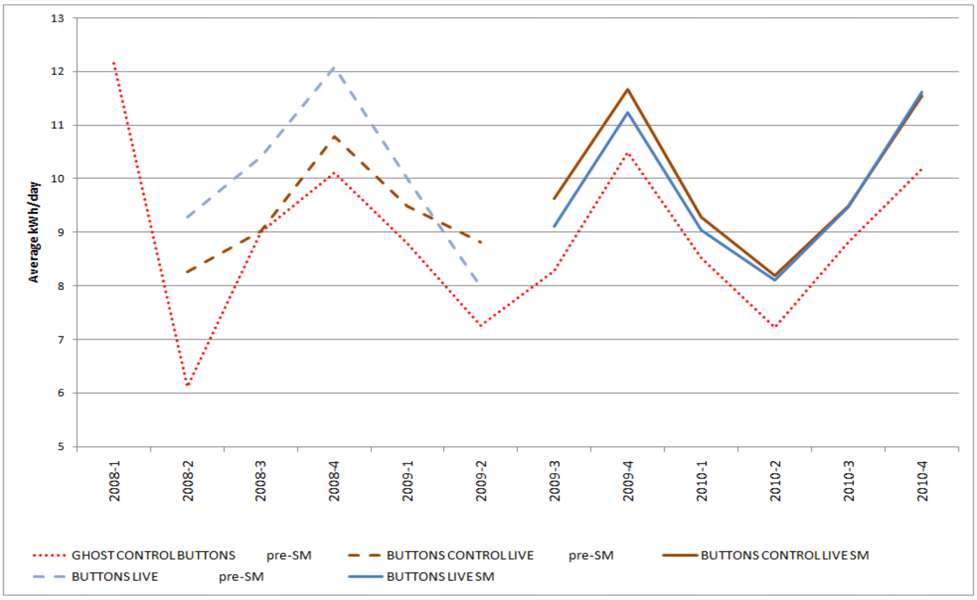


The chart for Bounty gas consumption shows the pre-SM data much closer to the ghost control than the SM data. The rounding effect of averaging the pre-SM data is large for this data set, with much higher peaks in winter and lower troughs in summer than the ghost control group.

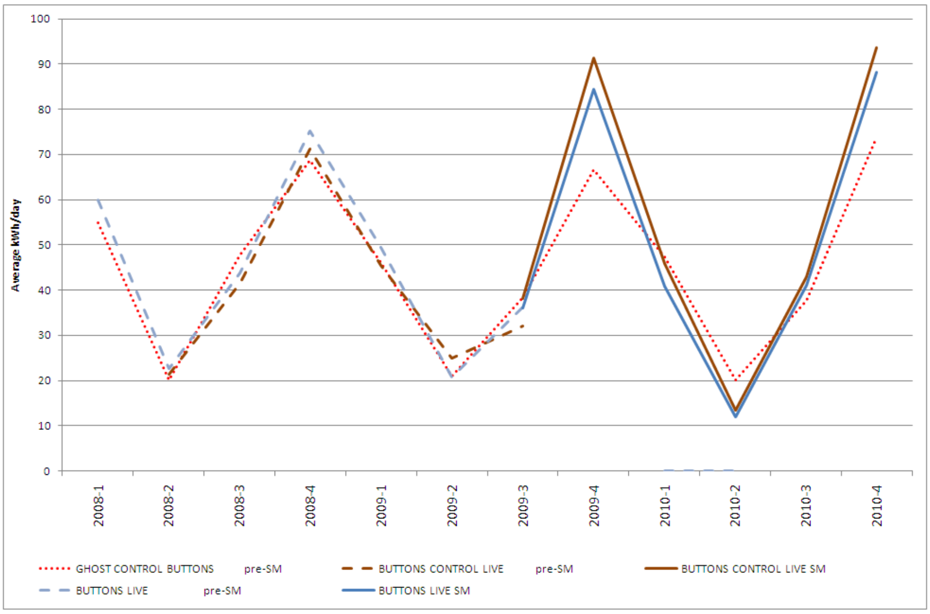


Buttons

The Buttons electric chart shows a fairly good correlation between the ghost group and both SM and pre-SM data. After the trial starts the control group and product trial group swap their order, indicating savings due to the product.

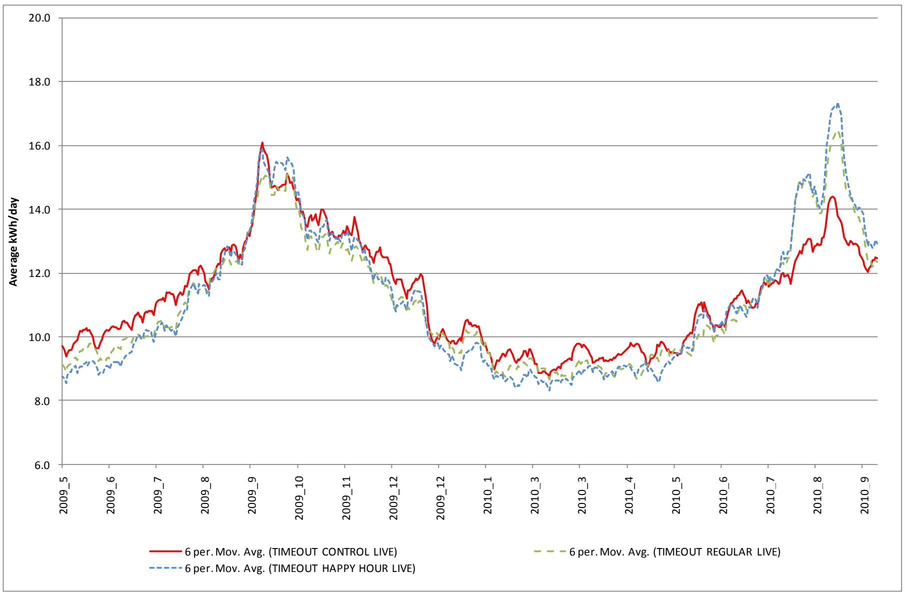


The Buttons gas chart also shows a fairly close match between the ghost control, SM and pre-SM data, with the control group and product trial groups again swapping places in the order of magnitude.



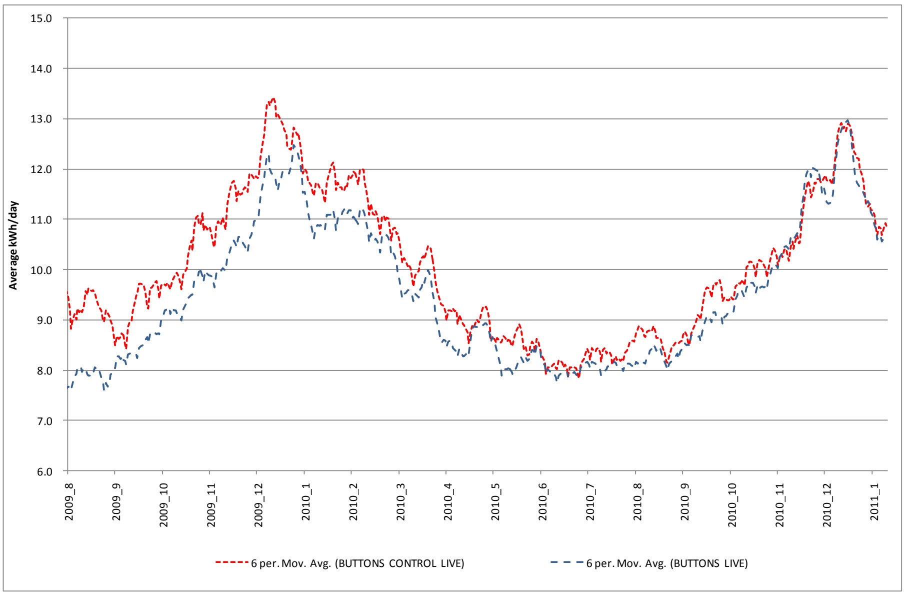
1.4 Monthly Averages from the SM Data

Timeout: Graphical analysis indicates that there were initial savings due to the product as the control group were using more than the product groups; however, over time there was a gradual switch and by the second winter the opposite case was true, and the control group were using less than the trial groups.

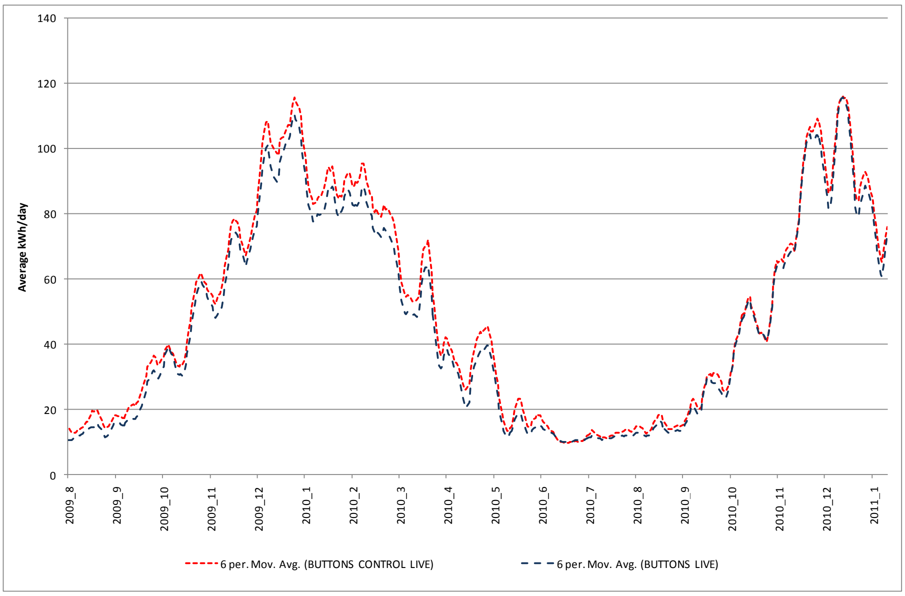


Buttons

The electric data for Buttons shows a similar pattern as for the Timeout groups, with the gap between control and treatment group narrowing over time. However, the control stays above the trial group to the end which is what would be expected.

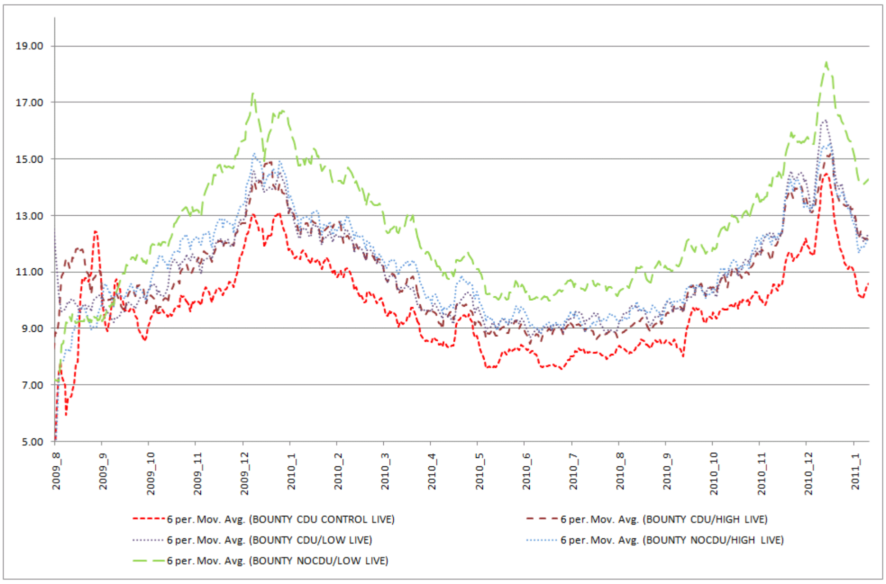


The gas data for Buttons shows that the control group and trial group are very close in their consumption, with the control group usage being slightly higher for most of the trial period.

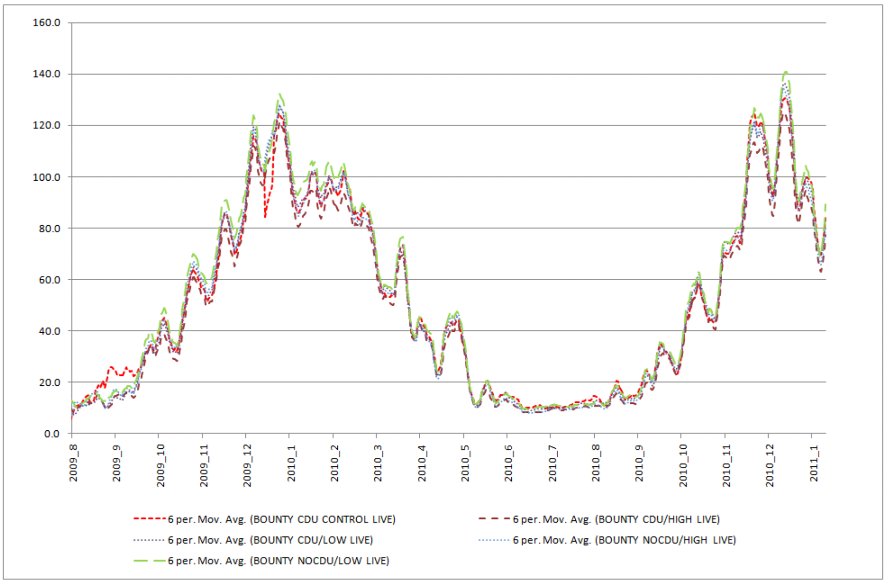


Bounty

The Bounty electric chart shows that the control group uses less than the other four groups for most of the trial period, which is unexpected. In addition, the group No CDU/Low Live used much more than all the other groups; however, this was also true during the pre-trial period.



The Bounty gas chart shows a similar picture to the electric groups, with the control group using less than the other groups for most of the trial period, although the difference is less.



2.0 APPENDIX C – Methodology and Results of the Data Modelling

This section provides details of the statistical modelling performed and the methodology followed in reaching the final results.

2.1 Cross-Sectional Analysis (ANOVA)

2.1.1 Theory

The ANOVA technique was used as a cross-sectional analysis to compare averages between the trial groups in two periods: pre-trial and during-trial. A one-way ANOVA model compares the means and variances of three or more groups at the same time to deduce whether there is any difference in the means and it assumes that data comes from normally distributed populations. Subsequently, Tukey's HSD test calculates the magnitude and statistical significance of the difference.

Using the results of the ANOVA, a Difference of Differences (DoD) analysis can determine the share of those differences that are not due to factors outside the scope of the trial. DoD can be used in situations where outcomes have been observed for two groups over two time periods, and where the treatment group gets the treatment in the second period but not in the first period. DoD 'removes biases in second period comparisons between the treatment and control group that could be the result of permanent differences between those groups, as well as biases from comparisons over time in the treatment group that could be the result of trends'. [13]

A DoD is done by subtracting the average change in the control group from the average change in the treatment group over time. In our case, the second period is the during-trial period, the control group is the ghost control group, and possible trends influencing energy consumption include weather and macroeconomic influences such as recession which affect householders' purchasing patterns.

The DoD comparison can be made as follows for each product from the ANOVA:

1. Savings due solely to the SM and CDU:

   (control group pre-trial consumption – during-trial consumption) – (ghost control group pre-trial consumption – during-trial consumption)

2. Savings due to the product and the SM combined:

   (product group pre-trial consumption – during-trial consumption) – (ghost control group pre-trial consumption – during-trial consumption)

3. Savings due solely to the product

   results of calculation 2 minus results of calculation 1

The ANOVA was done to provide an assessment of the matching of the customer groups. If they were well matched then most of the pre-trial differences should have been not significant (i.e. with the significance test at 95% we would expect to have around 5% of the comparisons to be significant). Similarly, if the sample sizes are sufficient and the data is accurate, we would expect most of the post tests to be significant. The ANOVA averages are highly influenced by the behaviour of large users, which will be a particular issue if the groups of consumers are not well matched.

2.1.2 Procedure

It was initially planned to do the ANOVA modelling on an annual basis but when the data was prepared for the model it was found that a quarterly time scale would be the best because:

1. The requirement that there was consumption data for at least 50% of the quarter (SM or pre-SM) meant that there were far fewer URNs with full and contiguous sets of data for a year than at first estimated, which would be a necessity to include the URN in the data set.
2. Because energy use is highly correlated with ambient temperature and the SM changeover varied by one or two quarters, it was not possible to align the full year of data along both quarter and SM groupings.

Therefore it was decided to perform an ANOVA for each quarter so that the weather effects would change equally for all groups and SM and pre-SM data could be separated. This also meant that all quarters for which there was data could be included and not just URNs with a full year of pre- and during-trial data. The quarters were assigned to either SM or pre-SM data categories. Ten runs of the ANOVA were performed, one for each product group.

2.1.3 ANOVA Results

The results presented here show the differences in energy consumption between ghost and control groups, and between ghost and product groups. A negative number indicates the control or product group was *higher* than the ghost group. Data is pre-SM for the quarters up to 2009-2 and SM for quarters 2009-3 and after, with the exception of Timeout for which there was no pre-SM data for quarter 2009-2. All ANOVA model results are in total kWh per quarter. Difference of difference (DoD) values are in kWh per year. The significance levels of the model results are shown as a percentage from 0 to 100% and any values with a significance of less than 95% are assumed to be not significantly different to zero. The DoD value for each product group is shown both as a kWh amount and as a percentage of the product group consumption – this value is the DoD in kWh/year divided by the average annual consumption in the product trial group during the pre-trial period.

2.1.3.1 Buttons

<u>Electric</u>

The difference between product and ghost group was significant in the pre-SM period and in two of the SM quarters, and the rest of the differences were not significant. This significant different before the trial started indicates a possible mismatch between ghost and product groups. Table 8 shows all the difference values and their significances, and the results of the DoD calculation. The DoD calculation between control and product showed positive savings due to the product of 319 kWh.

Table 8: Buttons Electric ANOVA Results

Quarter	Diff Between Control and ghost	Significance (control to ghost)	Diff Between Product and Ghost	Significance (product to ghost)
2008-3	-40.75	27%	-102.98	100%
2008-4	-69.87	52%	-163.40	100%
2009-1	-50.69	39%	-99.92	100%
2009-2	-85.12	86%	-66.39	97%
2009-3	-113.11	93%	-64.18	95%
2009-4	-136.83	97%	-79.71	97%
2010-1	-68.13	68%	-42.89	75%
2010-2	-107.39	98%	-106.22	100%
2010-3	-84.84	78%	-80.42	98%
2010-4	-148.94	93%	-132.53	99%

Between Product and Control Groups	BUTTONS
Annual pre-trial Diffs	-186
Annual Diffs during trial	132
Diff of Diffs	319
Savings as % of product	10%

Gas

The differences between product and ghost group were significant in the SM period, while the rest of the quarter differences were not significant. This is what we would have expected if the groups are well matched before the trial and there is an effect from the product. Table 9 shows all the difference values and their significances, and the results of the DoD calculation. The DoD between control and product showed high savings due to the product of 2520 kWh.

Table 9: Buttons Gas ANOVA Results

Quarter	Diff Between Control and ghost	Significance (control to ghost)	Diff Between Product and Ghost	Significance (product to ghost)
2008-3	382	53%	21	1%
2008-4	-195	11%	-487	89%
2009-1	120	7%	-173	38%
2009-2	107	13%	-4	0%
2009-3	57	2%	203	56%
2009-4	-2,512	100%	-1,760	100%
2010-1	23	0%	530	100%
2010-2	695	100%	751	100%
2010-3	-519	88%	-335	93%
2010-4	-2,263	100%	-1,637	100%

Between Product and Control Groups	BUTTONS
Annual pre-trial Diffs	-1057
Annual Diffs during trial	1463
Diff of Diffs	2520
Savings as % of product	15%

2.1.3.2 Bounty

Electric

The majority of Bounty electric comparisons by quarter were not significant in both pre-SM and SM quarters, except for four quarters for the group NO_CDU LOW. Table 10 shows all the difference values and their significances, and the results of the DoD calculation. The DoD showed positive savings for all of the groups except NOCDU_HIGH and savings ranged from between -16 to 85 kWh.

Table 10: Bounty Electric ANOVA Results

Quarter	Diff Control and ghost	Significance (control to ghost)	diff CDU HIGH & ghost	Significance (product to ghost)	diff CDU LOW & ghost	Significance (product to ghost)	diff NOCDU HIGH & ghost	Significance (product to ghost)	diff NOCDU LOW & ghost	Significance (product to ghost)
2008-3	-6.7	0%	-10.3	0%	12.1	0%	-5.7	0%	-82.4	41%
2008-4	47.3	1%	-66.0	21%	1.2	0%	-67.3	21%	-175.1	98%
2009-1	91.6	30%	4.0	0%	9.6	0%	-53.4	17%	-106.2	75%
2009-2	104.6	28%	75.4	42%	-1.5	0%	77.7	36%	18.4	0%
2009-3	49.5	2%	-19.5	0%	21.2	0%	-39.4	4%	-105.4	69%
2009-4	-55.6	6%	-96.5	72%	-78.8	56%	-137.0	97%	-227.5	100%
2010-1	76.4	34%	44.3	15%	45.2	19%	2.8	0%	-59.9	28%
2010-2	84.1	52%	52.3	32%	36.7	12%	26.1	2%	-19.8	0%
2010-3	19.5	0%	-0.7	0%	-5.5	0%	-16.7	0%	-96.2	70%
2010-4	-44.8	1%	-68.7	13%	-103.5	51%	-42.3	2%	-163.1	85%

Between Product and Control Groups	CDU HIGH	CDU LOW	NO CDU HIGH	NO CDU LOW
Annual pre-trial Diffs	-234	-215	-286	-582
Annual Diffs during trial	-174	-130	-302	-567
Diff of Diffs	60	85	-16	15
Savings as % of product	1.7%	2.4%	-0.5%	0.4%

Gas

The majority of Bounty gas comparisons by quarter were not significant in both pre-SM and SM periods, except for the winter quarters of 2009 and 2010. This large winter difference is likely partly due to the difference in accuracy between SM and pre-SM data, with the ghost group having a more smoothed out load shape compared to the SM data. Table 11 shows all the difference values and their significances, and the results of the DoD calculations. The DoD showed negative savings for two of the trial groups, and the savings ranged from -531 to 159 kWh. The negative savings for CDU_HIGH was unexpected as this should be the group with the highest savings.

Table 11: Bounty Gas ANOVA Results

Quarter	Diff Control and ghost	Significance (control to ghost)	diff BOUNTY CDU HIGH & ghost	Significance (product to ghost)	diff BOUNTY CDU LOW & ghost	Significance (product to ghost)	diff BOUNTY NOCDU HIGH & ghost	Significance (product to ghost)	diff BOUNTY NOCDU LOW & ghost	Significance (product to ghost)
2008-3	-251	1%	138	0%	-263	7%	108	0%	-121	0%
2008-4	-151	0%	-197	0%	-681	56%	-307	3%	-636	40%
2009-1	-465	11%	128	0%	-597	75%	-436	33%	-1065	100%
2009-2	840	82%	987	100%	566	86%	709	96%	561	80%
2009-3	96	0%	536	45%	219	2%	-14	0%	-213	1%
2009-4	-2854	100%	-2733	100%	-3402	100%	-3131	100%	-3444	100%
2010-1	138	0%	340	18%	-197	2%	-25	0%	-81	0%
2010-2	1556	100%	1820	100%	1526	100%	1675	100%	1600	100%
2010-3	-74	0%	101	0%	-368	31%	-178	2%	-311	16%
2010-4	-2586	100%	-1959	100%	-2888	100%	-2372	100%	-2931	100%

Between Product and Control Groups	CDU HIGH	CDU LOW	NO CDU HIGH	NO CDU LOW
Annual pre-trial Diffs	1083	-949	100	-1233
Annual Diffs during trial	1026	-790	-431	-1074
Diff of Diffs	-57	159	-531	159
Savings as % of product	-0.3%	0.8%	-2.8%	0.8%

2.1.3.3 Timeout

The majority of Timeout comparisons by quarter were not significant. There was no pre-SM data for quarter 2 of 2009. Table 12 shows all the difference values and their significances, and the results of the DoD calculation. The DoD was performed on quarters 3 to 1 only. The DoD showed positive savings for both groups and the savings ranged from 189 to 200 kWh.

Table 12: Timeout ANOVA Results

Quarter	Diff Between Control and ghost	Significance (control to ghost)	Diff Between HH and ghost	Signifi-cance (control to ghost)	Diff Between Regular and Ghost	Significance (product to ghost)
2008-3	0.54	0%	-20.6	6%	-86.73	94%
2008-4	4.34	0%	16.3	2%	-50.32	41%
2009-1	31.22	1%	-65.0	19%	-84.35	55%
2009-2	No Data	No Data	No Data	No Data	No Data	No Data
2009-3	-26.32	3%	-10.7	1%	-52.82	65%
2009-4	-11.98	0%	0.2	0%	-44.34	25%
2010-1	-29.30	4%	26.5	11%	-27.71	15%
2010-2	-59.07	32%	-16.0	4%	-54.38	74%
2010-3	-12.72	0%	-5.7	0%	-35.46	23%
2010-4	-70.96	14%	-102.0	67%	-144.49	94%

Between Product and Control Groups	Happy Hour	Timeout Regular
Annual pre-trial Diffs	-105	-257
Annual Diffs during trial	84	-57
Diff of Diffs	189	200
Savings as % of product	5.6%	5.7%

2.1.3.4 Summary of ANOVA Results

The ANOVA analysis was done by quarter and it compared all groups within a product – ghost, control and product groups. Most of the comparisons were not statistically significant, except for Buttons Gas during-trial and Buttons electric pre-trial, and a few other exceptions. Table 13 shows the final results from the ANOVA and the DoD calculations. All of the savings were positive except for the Bounty NOCDU_HIGH group which had negative savings for both gas and electricity. Savings expressed as a percentage of product group consumption in 2010 were in the expected range for most products, although fairly high for Buttons gas and electric. Timeout showed significant (above 5%) savings for both groups and Bounty was less than 2% in all but one cases.

Although we suspect a strong influence on these differences from the effect of combining pre-SM and SM data, when these differences are differenced together (between ghost and control, and between ghost and product) this reduces the problem, as both differences will see this same effect. However, there is still a large amount of uncertainty about the cause of these differences. In general, the fact that during-trial comparisons were not significant is an indicator that the response to the product was not significantly different to zero. This is true for all groups with the exception of Buttons Gas.

Table 13: Summary of ANOVA DoD Results

Product	Product Trial Group	Annual kWh Savings Product (Above Control)	Product Savings as % of Consumption
Bounty Electric	CDU_HIGH	60	1.7%
	CDU_LOW	85	2.4%
	NOCDU_HIGH	-16	-0.5%
	NOCDU_LOW	15	0.4%
Bounty Gas	CDU_HIGH	-57	-0.3%
	CDU_LOW	159	0.8%
	NOCDU_HIGH	-531	-2.8%
	NOCDU_LOW	159	0.8%
Buttons Electric	Monthly Billing	319	9.7%
Buttons Gas	Monthly Billing	2,520	14.5%
Timeout Electric	Happy Hour	189	5.6%
	Timeout Regular	200	5.7%

2.2 Longitudinal Analysis 1 – Linear Mixed Model Regression

2.2.1 Theory

The first technique for longitudinal analysis was a linear mixed model (LMM) regression on all of the pre-SM and SM quarterly data (ghost, control and product groups). SPSS documentation describes the MIXED model in this way: 'In a linear mixed-effects model, responses from a subject are modelled as the sum of so-called fixed and random effects. If the treatment affects the population mean, it is a fixed effect. If there is an effect associated with a sampling procedure, it is a random effect. Though the fixed effect is the primary interest in most studies or experiments, it is necessary to adjust for the covariance structure of the data. The MIXED procedure provides the tools necessary to estimate fixed and random effects in one model. It uses maximum likelihood (ML) and restricted maximum likelihood (REML) methods, which yield asymptotically efficient estimators for balanced and unbalanced designs, presenting an advantage over ANOVA methods in modelling real data, since data are often unbalanced.'

The mixed effects model can include a repeated measures term for when observations are correlated rather than independent as in time-series data, which is true in our case as quarterly consumption is related to the previous and succeeding quarters' consumption. The repeated measures calculation adjusts standard errors and significance tests for patterned (hence non-independent) relationships across time, which in our case is a series of quarter consumption values by URN. We expected the LMM to be better than a simple linear regression because it includes the time series effect of the consumption values and compares values by quarter rather than by putting all values together regardless of the time of year.

In the LMM energy consumption is a function of a variety of possible explanatory variables including: group (treatment or control); observation period (pre- or post- treatment); presence of a SM; quarter in which data was gathered; and an error term reflecting any influential variables not specified in the model. The formula for the linear regression is:

Consumption (kWh) = intercept + β_1*PRODUCT(S) + β_2*CONTROL + β_3*PRODUCT_SM(S) + β_4*SM_FLAG + β_5*[QUARTER_FLAG_N (2 to 11)] (1)

Where:

- PRODUCT(S) is 1 when the customer is in the product group(s) and 0 otherwise
- CONTROL is 1 when the customer is in the control group and 0 otherwise
- PRODUCT_SM(S) is 1 when the customer has a SM and is in the product group(s) and 0 otherwise
- SM FLAG is 1 when the customer has a SM and 0 otherwise (for both control and product groups)
- QUARTER_FLAG_N (N goes from 2 to 11) are 1 when the quarter is N and 0 otherwise
- Consumption is the total kWh in the quarter
- β_1 to β_5 are the coefficients produced by the model
- The intercept includes effects not included in the other terms

The coefficients produced by the model indicate the effect of that independent variable on the dependent variable (the consumption). The savings due to being in the product group, above and beyond the savings due to the SM, are given by the coefficient of the PRODUCT_SM flags (one for each of the product trial groups). To get product annual savings, this coefficient is multiplied by 4, as the time unit in the data is a quarter, and then by -1, because a negative coefficient indicates a drop in usage due to the product, which is a saving. Similarly, the annual effect of having a SM, in both the control and product groups, is given by multiplying the SM_FLAG coefficient by 4 and then -1.

The data used for the linear regression models was the same as used for the ANOVA except it was reformatted with each line being a single consumption value. Five models were run for the five product-fuel combinations and all groups within those products were included.

2.2.2 Results

2.2.2.1 Buttons

The Buttons LMM produced significant results for the gas product group savings but not for the electric group. Interpretation of the results from the models produced estimated savings due to the product when compared to the product group's consumption in the pre-trial period of 3.6% and 6.0% respectively, which are close to the expected range.

Table 14: Buttons LMM Results

Product	kWh/year	Significance	% of Product kWh	kg CO2
Buttons Electric	126	91%	3.6%	54
Buttons Gas	1072	98%	6.0%	204

2.2.2.2 Bounty

The Bounty LMM did not produce any significant results for both the gas and electric product groups. Interpretation of the results from the models produced estimated savings due to the product when compared to the product group's consumption of between 1% and 4%. Significance levels for the savings were higher for the CDU_HIGH group. The group with the highest savings was CDU_HIGH, which would be expected as they had the most interactive product. Gas savings were about on par with electric savings.

Table 15: Bounty Mixed Model Results

Product	Product Trial Group	kWh/year	Significance	% of Product kWh	kg CO2
Bounty Electric	CDU_HIGH	140	83%	4.0%	61
	CDU_LOW	92	65%	2.6%	40
	NOCDU_HIGH	35	27%	1.0%	15
	NOCDU_LOW	61	44%	1.6%	26
Bounty Gas	CDU_HIGH	681	54%	4.0%	129
	CDU_LOW	215	18%	1.1%	41
	NOCDU_HIGH	247	21%	1.4%	47
	NOCDU_LOW	613	49%	3.2%	116

2.2.2.3 Timeout

The Timeout LMM produced significant savings for the Regular group but not the Happy Hour group. Interpretation of the results from the model produced estimated savings due to the product when compared to the product group's consumption of 3.7% for Happy Hour and 6.7% for Timeout Regular. It was expected that Happy Hour households might save less as there was a financial incentive to use more electricity during certain hours of the day when there was a particularly cheap rate.

Table 16: Timeout Mixed Model Results

Product	Product Trial Group	kWh/year	Significance	% of Product kWh	kg CO2
Timeout Electric	Happy Hour	116	72%	3.7%	50
	Timeout Regular	220	97%	6.7%	95

2.2.2.4 Summary of Linear Mixed Model Analysis Results

The LMM models produced significant coefficients for two of the groups (Bounty Gas and Timeout Regular) and results in the expected range. All savings were positive. The findings from the LMM suggest that putting both SM and pre-SM data into a robust model that isolates fixed and random effects, and includes the effect of variables being correlated over time, produces a fairly good fit and reasonable value. These results compare fairly well to the ANOVA results in that the Buttons groups are the highest, Timeout is around 5% and Bounty groups have a wider range but generally lower savings values.

Table 17: LMM Results for All Products

Product	Product Trial Group	Annual Product Savings (first year of product)			
		kWh/year	Significance	% of Product kWh	kg CO2
Bounty Electric	CDU_HIGH	140	83%	4.0%	61
	CDU_LOW	92	65%	2.6%	40
	NOCDU_HIGH	35	27%	1.0%	15
	NOCDU_LOW	61	44%	1.6%	26
Bounty Gas	CDU_HIGH	681	54%	4.0%	129
	CDU_LOW	215	18%	1.1%	41
	NOCDU_HIGH	247	21%	1.4%	47
	NOCDU_LOW	613	49%	3.2%	116
Buttons Electric	Monthly Billing	126	91%	3.6%	54
Buttons Gas	Monthly Billing	1072	98%	6.0%	204
Timeout Electric	Happy Hour	116	72%	3.7%	50
	Timeout Regular	220	97%	6.7%	95

2.3 Longitudinal Analysis 2 – Linear Mixed Model on Pre-Trial Quarterly Data

In order to determine what differences existed between the groups before the trial started, the ghost, control and product quarterly data from quarters 2008-2 to 2009-2 were combined and analysed in a LMM (one model per product). These results were required for adjusting product savings calculated for the during-trial period with the SM data in Longitudinal Analysis 3.

The same basic formula was used as in Longitudinal Analysis 1 except that there were no SM flags, but there were flags indicating the group and the quarter (from 1 to 5). Not all quarters in all the models were statistically significant, but the majority were and data from all quarters was used to calculate the differences.

2.3.1 Results

The results of the pre-trial analysis are shown in Table 18 below. A negative difference indicates that the product consumption was higher than the control consumption. All of the product trial groups were higher than the control group except for Bounty gas CDU_HIGH. The Bounty electric groups were all significantly higher than the control group, which had been expected as this was seen in the data statistics charts of average kWh/day (see Appendix B). Both Buttons electric and gas product groups were higher than the control group, as were both Timeout groups.

Table 18: Differences between Control and Product Groups before Trial Start

Product	Product Trial Group	Difference to Control Group (kWh/year)	Difference as % of Product
Bounty Electric	CDU_HIGH	-244	-6.9%
	CDU_LOW	-214	-6.1%
	NOCDU_HIGH	-310	-8.7%
	NOCDU_LOW	-591	-15.3%
Bounty Gas	CDU_HIGH	984	5.8%
	CDU_LOW	-1086	-5.7%
	NOCDU_HIGH	-145	-0.8%
	NOCDU_LOW	-1431	-7.4%
Buttons Electric	Monthly Billing	-146	-4.2%
Buttons Gas	Monthly Billing	-978	-5.5%
Timeout Electric	Happy Hour	-47	-1.3%
	Timeout Regular	-293	-7.9%

2.4 Longitudinal Analysis 3 – Linear Mixed Model on Monthly SM Data

The ANOVA and LMM included all of the quarterly data from the trial groups within a product grouping. One of the benefits of this approach is that the ghost group data provides a consistent benchmark throughout the SM and pre-SM periods. However, all of the pre-SM quarterly data suffered from an averaging effect and a lack of consistency in the availability of the data. On the other hand, we have assumed that the SM data has a very high level of accuracy and the availability of data during the trial period was very high. Therefore, it was decided to analyse savings on a monthly basis with the SM data alone and then adjust the savings with the pre-SM data to account for the differences between groups that existed before the trial started. In other words, do a DoD on the results of the separate models.

The formula for the monthly model was similar to the previous model, but as there is no ghost data the trend flags just indicate what month the data is from and what group the URN is in. In other words, all URNs have SMs all of the time. The models were run for months from September 2009 to January 2011. Each unique value of the trend variable (from 1 to 17) produced a coefficient, indicating the effect of having the SM and being in the trial for that length of time.

The formula for the monthly LMM that was run is:

$$\text{Consumption (kWh)} = \text{intercept} + \text{fixed effects } [\beta_1{*}\text{TRENDCONTROL} + \beta_2{*}\text{TRENDPRODUCT}] + \text{random effects } (\beta_3{*}\text{URN}) \qquad (2)$$

Where:

- TRENDCONTROL is a value from 1 to 16 (total number of months for which there is SM data) indicating the number of months control group customers have had a SM, and 20 if not in the control group.
- TRENDPRODUCT is a value from 1 to 16 (total number of months for which there is SM data) indicating the number of months product group customers have had a SM and 20 if not in the product group(s)
- HDD is the total amount of HDD in the month
- Consumption is the total kWh in the month
- β_1 to β_3 are the coefficients produced by the model

The month was included as the repeated measure. Results were interpreted by subtracting the coefficients for the product groups from the coefficients for the control group. An alternative version used the kWh/HDD as the dependent variable, and the differences in coefficients were then multiplied by the HDD in each month to get the kWh savings.

To calculate savings from the products the annual differences between control and product from all months in 2010 were added up to give annual unadjusted savings. Then these savings were adjusted using the annual kWh differences between product and control groups from the pre-SM period (from the ANOVA). To get a percentage change in consumption the kWh savings value was divided by the average product group consumption for 2010.

2.4.1 Results

The model produced significant results for most of the months for each trend value for the product and control groups. The months ran from September 2009 to January 2011, but January 2011 proved to be not significant for most of the product groups.

2.4.1.1 Buttons

Electric Model

The monthly model for Buttons electric was significant for most of the months for the product group but not for four of the months for the control group. Savings were generally higher in the shoulder months of spring and autumn. The result is higher than both the ANOVA and Longitudinal Analysis 1.

Table 19: Buttons Electric Monthly Model Results

Product Group	Savings from Monthly Model (kWh/year)	Savings Adjusted by pre-Trial Differences (kWh/year)	Savings as % of Product
Buttons Electric	125	271	8.4%

Gas Model

The monthly model for Buttons gas was significant for all months up to January 2011, except for one month in the control group. The highest month for savings was March 2010. The result is higher than the ANOVA and Longitudinal Analysis 1.

Table 20: Buttons Gas Monthly Model Results

Product Group	Savings from Monthly Model (kWh/year)	Savings Adjusted by pre-Trial Differences (kWh/year)	Savings as % of Product
Buttons Gas	2,321	3,300	20.4%

2.4.1.2 Bounty

Electric Model

The monthly model for Bounty electric was significant for most of the months for all of the product groups except for December 2009 and January 2010. Table 21 shows the results of the monthly model before and after adjustment with the pre-trial differences, and the final result as a percentage of average product group consumption. The results are higher than the results from Longitudinal Analysis 1 but still within a reasonable range. We expected that the group CDU_HIGH would have the highest savings and this in fact turned out to be so; however, savings for the other groups that got the CDU, CDU_LOW, were about the same as the savings for the groups that did not get a CDU.

Table 21: Bounty Electric Monthly Model Results

Product Group	Savings from Monthly Model (kWh/year)	Savings Adjusted by pre-Trial Differences (kWh/year)	Savings as % of Product
CDU_HIGH	-88	156	4.5%
CDU_LOW	-123	91	2.6%
NOCDU_HIGH	-215	95	2.7%
NOCDU_LOW	-495	96	2.5%

Gas Model

The monthly model for Bounty gas was significant for all of the months for all of the product groups. The highest average savings over the four groups were in December 2010, and October and November 2009 showed average negative savings. The final adjusted savings were positive for all the groups, with the two groups that got the CDU showing much higher savings than the other two groups. Again, CDU_HIGH had the highest savings of all four groups.

Table 22: Bounty Gas Monthly Model Results

Product Group	Savings from Monthly Model (kWh/year)	Savings Adjusted by pre-Trial Differences (kWh/year)	Savings as % of Product
CDU_HIGH	2793	1809	10.8%
CDU_LOW	200	1286	7.0%
NOCDU_HIGH	71	216	1.2%
NOCDU_LOW	-581	849	4.3%

2.4.1.3 Timeout

The monthly model for Timeout was significant in months except February 2010. As a percentage of consumption, savings are generally higher in summer than in winter. The final result is higher than the quarterly LMM result for Timeout Regular but lower for Happy Hour.

Table 23: Timeout Monthly Model Results

Product Group	Savings from Monthly Model (kWh/year)	Savings Adjusted by pre-Trial Differences (kWh/year)	Savings as % of Product
Happy Hour	-18	29	0.9%
Timeout Regular	150	443	12.8%

2.4.1.4 Summary of Monthly SM Data Analysis Results

The results from a difference of differences approach based on separate SM and pre-SM models produced mostly significant coefficients within each model and results in the expected range; the results ranged from 29 to 3300 kWh/year. Buttons gas again had the highest savings of all the groups and in this case the savings are much higher than would be expected.

Although the monthly data is more accurate and reliable than the pre-SM data, the drawback of this approach is that there is no ghost control data to compare to, and so the savings must be adjusted with estimates of existing differences from the pre-trial period taken from a separate model on the pre-SM data. This step introduced some uncertainty into the results, especially as differences between groups change with the seasons and an average annual difference was used. However, we believe that these results are a good estimate of savings from the products.

Table 24: Monthly LMM Results for All Products

Product	Product Trial Group	Annual Savings		
		kWh/year	% of Product pre-trial kWh	kg CO2
Bounty Electric	CDU_HIGH	156	4.5%	67
	CDU_LOW	91	2.6%	39
	NOCDU_HIGH	95	2.7%	41
	NOCDU_LOW	96	2.5%	41
Bounty Gas	CDU_HIGH	1809	10.8%	344
	CDU_LOW	1286	7.0%	244
	NOCDU_HIGH	216	1.2%	41
	NOCDU_LOW	849	4.3%	161
Buttons Electric	Monthly Billing	271	8.4%	117
Buttons Gas	Monthly Billing	3300	20.4%	627
Timeout Electric	Happy Hour	29	0.9%	13
	Timeout Regular	443	12.8%	191

2.5 Combined Electric and Gas Model

When comparing gas and electric savings for the groups that had both fuels, four out of five groups had higher gas than electric savings and in three of the groups gas savings were over twice as much as electric savings as a percentage of consumption.

Table 25: Comparison of Savings for Electric and Gas Use by Group

Product Group	Electric Savings	Gas Savings
BOUNTY CDU_HIGH	4.5%	10.8%
BOUNTY CDU_LOW	2.6%	7.0%
BOUNTY NOCDU_HIGH	2.7%	1.2%
BOUNTY NOCDU_LOW	2.5%	4.3%
BUTTONS	8.4%	20.4%

There could be many reasons for this larger savings in gas compared to electricity including the following:

- Households found it easier to reduce operating hours or temperature at a central thermostat compared to paying attention to many different electrical appliances round the house.
- There was more willingness to reduce gas use because there was less inconvenience and loss of comfort associated with it – e.g. lighting, kitchen appliance, computer, and TV use are less likely to be restricted to save money or energy as they are considered an essential part of everyday life.
- There was a greater amount of wastage occurring in gas use compared to electricity use before the trial started;
- The highest proportion of household energy bills is for gas and so there was more monetary motivation to reduce gas consumption than electric.

These kinds of issues could be explored by examining the survey data but this is out of the scope of this study.

2.6 Significance of Final Results

Whether the results presented in Table 2 are statistically significant or not is a difficult question to answer because these values are the final result of several addition, subtraction and averaging operations on sets of monthly and quarterly regression results. For example, for Buttons Gas, the monthly regression results for period 1 for the Control and Product groups were -1,023 and -1,052, respectively. These are both high numbers compared with their Standard Errors (57 and 32, respectively), but the different between them, which is used to estimate the final savings number, is 29, which is of a similar magnitude to the standard errors of the individual values and so would not be significant. The significance of the lower value results, such as 0.9% for Timeout-Happy Hour and 1.2% for Bounty Gas NOCDU_High, are not different from higher values such as Buttons Gas.

An estimate of what confidence can be assumed for the results was done by looking at what percentage of individual results which have fed into the final results are significant. Assuming that the best possible confidence level would be when there are statistically significant results for all possible monthly (16 each for control and product groups) and quarterly (6 each for control and product groups) regression values, an overall percentage of confidence was calculated by dividing the maximum possible of significant values by the count of monthly or quarterly values that exist and are over 80% significant. Applying a level of 80% again for a cut off point of overall confidence, we find that we can be confident about all final values apart from Timeout Regular and Timeout Happy Hour. The reason for Timeout having a lower confidence is the lack of pre-SM data, with only three quarters significant. Table 26 shows the count of significant values and overall confidence levels.

Table 26: Estimates of Levels of Confidence in Final Results

		Number of Months or Quarters Existing and Significant				Estimate of Overall Confidence	
		Pre-SM Control	Pre-SM	SM Control	SM Trial	% of all periods significant	Significant Overall?
Bounty Electric	CDU_HIGH	4	6	15	13	85%	Yes
	CDU_LOW	4	6	15	13	85%	Yes
	NOCDU_HIGH	4	6	15	13	85%	Yes
	NOCDU_LOW	4	6	15	13	85%	Yes
Bounty Gas	CDU_HIGH	4	4	16	16	83%	Yes

	CDU_LOW	4	4	16	16	83%	Yes
	NOCDU_HIGH	4	4	16	16	83%	Yes
	NOCDU_LOW	4	4	16	16	83%	Yes
Buttons Electric	Monthly Billing	6	6	10	15	89%	Yes
Buttons Gas	Monthly Billing	6	6	15	16	98%	Yes
Timeout Electric	Happy Hour	3	4	15	15	76%	No
	Timeout Regular	3	4	15	15	76%	No

3.0 APPENDIX D – Methodology and Results of the Lifetime Analysis

3.1 Theory

The aim of the lifetime analysis was to identify in which month the differences between the control group and the product groups will cease to be different to zero. To estimate this lifetime a trendline method was used based on the results of Longitudinal Analysis 3. The theory is that the curve of savings will continue in the same direction once the trial is over, and thus a trendline will predict when those savings will cease.

Monthly differences between the product group(s) and the control group were plotted on a graph, expressed as a percentage of the product group consumption. These values were adjusted so that the average annual savings in 2010 agreed with the results of the quarterly LMM. A trendline was then added to the graph that extended out into the post-trial period until it hit the X-axis. Judgment was used to determine the lifetime in cases where this crossing of the X-axis was not achieved using a simple trendline.

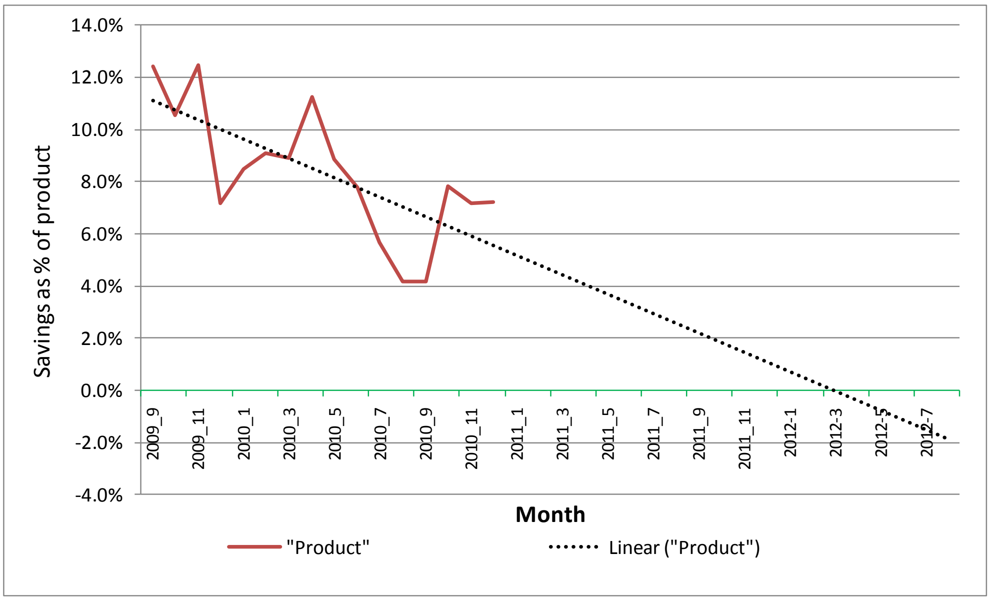
3.2 Results

3.2.1 Buttons

Electric

The Buttons electric graph shows a clear trend towards zero savings based on a linear trendline. The lifetime was estimated at 30 months, with the savings lasting four months after the end of the trial.

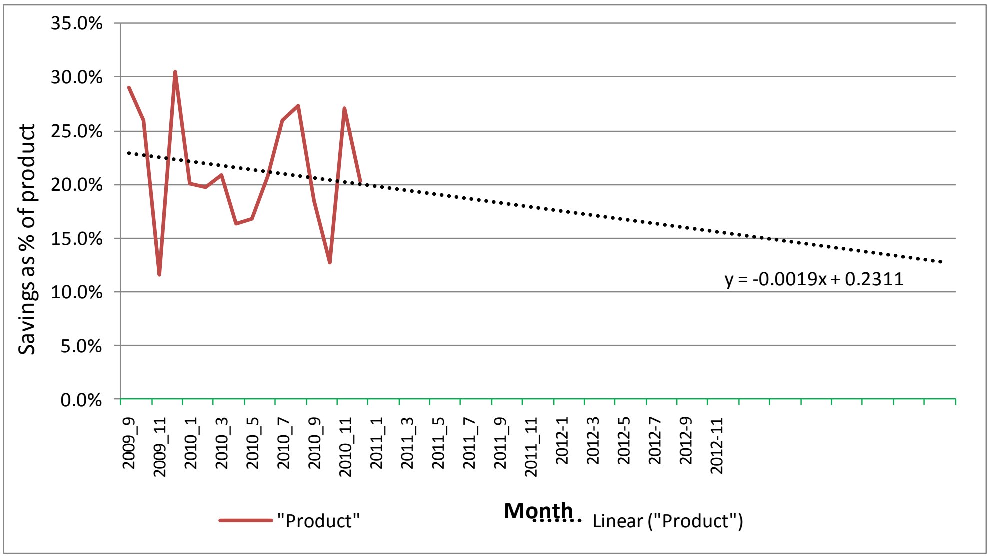
Figure 8: Buttons Electric Trendline



Gas

The Buttons gas graph also shows a trend towards zero savings based on a linear trendline, but at a much slower rate than the electric one. The timeline does not meet the X-axis until September 2019; however, as this is not a reasonable value for a lifetime we have set the lifetime to 2.5 years, or 30 months, the same as for Buttons electric[7].

Figure 9: Buttons Gas Trendline



3.2.2 Bounty

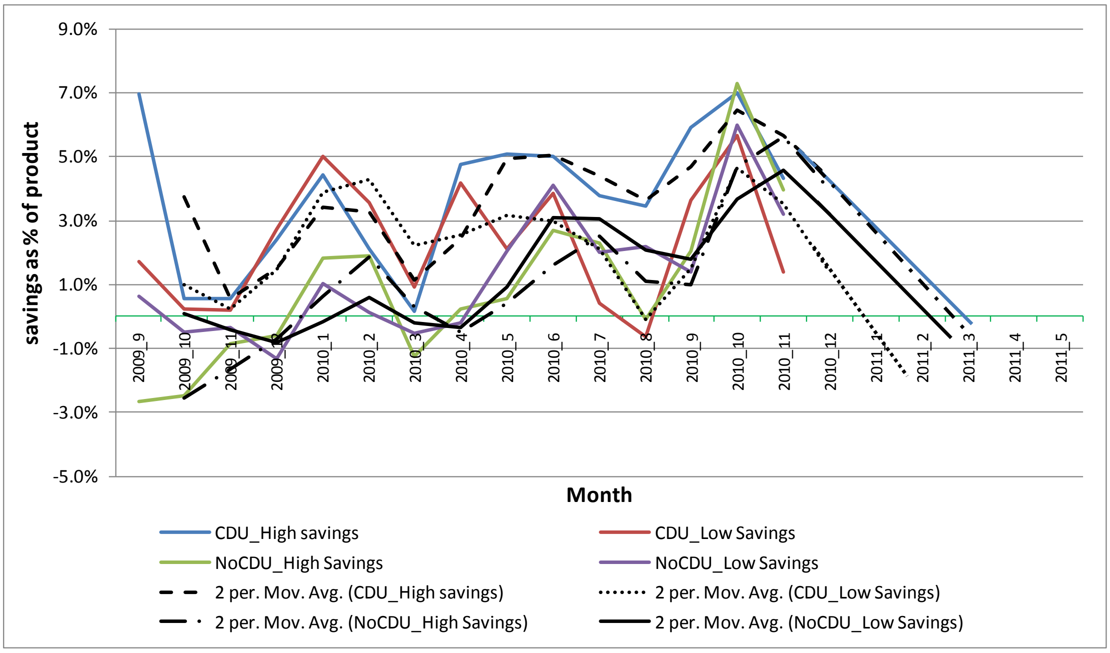
<u>Electric</u>

None of the Bounty electric groups showed a definite trend towards zero savings and so moving average trendlines were used and then extended out in time to meet the X-axis. The CDU_HIGH and NOCDU_HIGH group showed slightly longer lifetimes at 18 months, with the other two groups being 16 and 17 months in length and all groups ending shortly after the end of the trial. The shape of the savings lines are not clear wedge shapes as could have been expected but contain many oscillations as they trend downwards. Highest savings were seen in the winter months, although the differences between seasons are not large.

---

[7] The lifetime for Buttons Gas could be over 30 months, based on trial evidence.

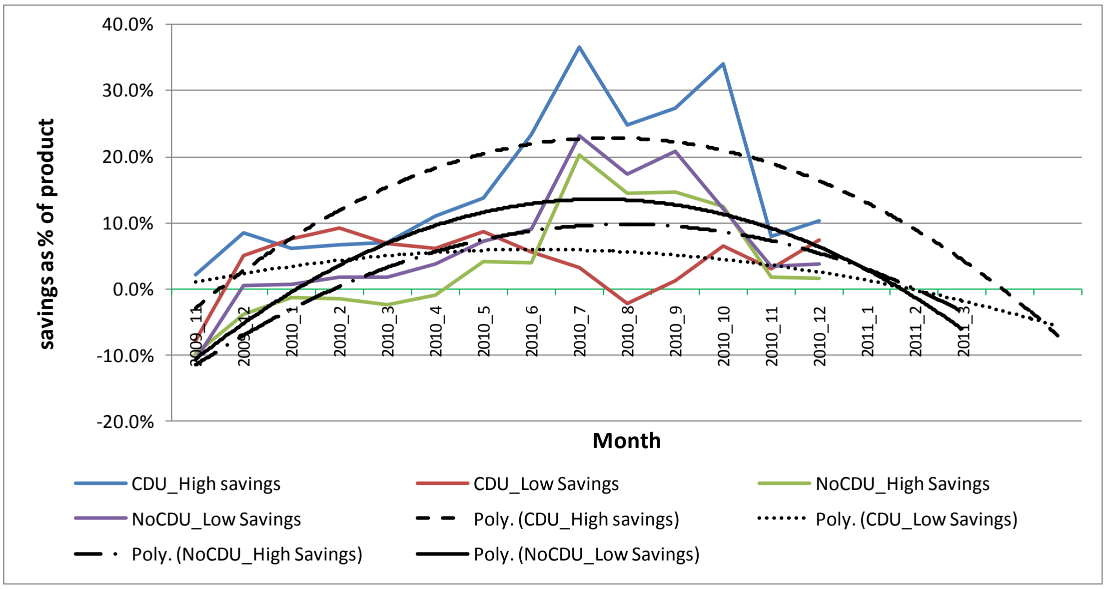
Figure 10: Bounty Electric Trendlines



Gas

All of the Bounty gas groups trended to zero savings with a polynomial trendline. Three out of the four groups showed lifetimes of one month more than the trial length at 17 months. The CDU_HIGH group was slightly longer at 19 months. Highest savings as a percentage of consumption were seen in the summer months when gas use would have been very low.

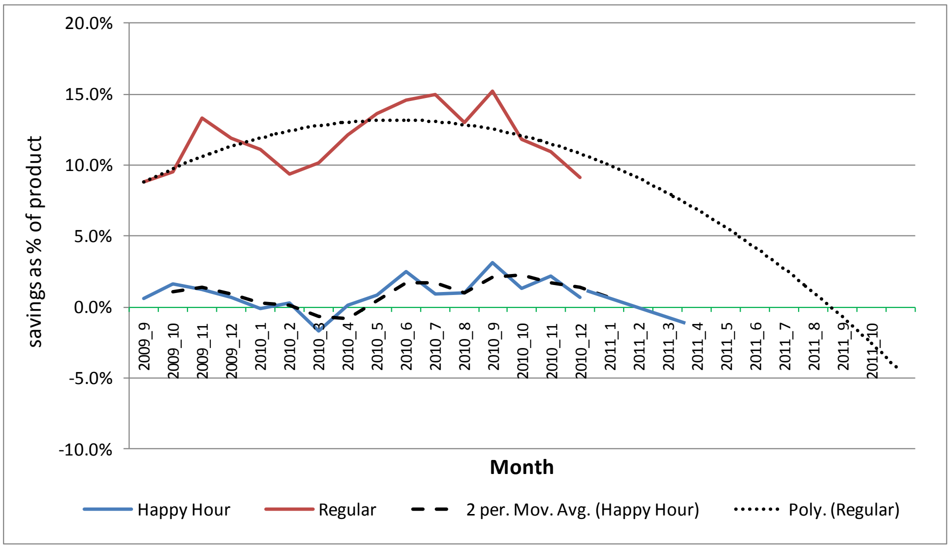
Figure 11: Bounty Gas Trendlines



3.2.3 Timeout

The lifetime for Timeout Regular was significantly longer than the one for Timeout Happy Hour. A polynomial trendline was used for Timeout Regular, which crossed the X axis after 24 months. The Happy Hour group did not show a clear trend to zero so a moving average trendline was used and then extended out to meet the X axis at 17 months. Savings were lowest in the spring months and highest in the autumn.

Figure 12: Timeout Trendlines

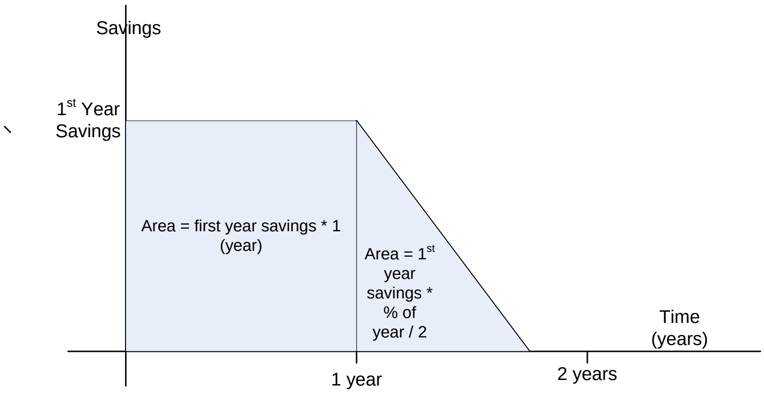


3.2.4 Lifetime Savings

The total lifetime savings were calculated according to the approximate shape of the trendline graphs. The total savings are equivalent to the area under the trendline. For Buttons gas and electric the shape can be estimated as a simple wedge shape. For the other products the majority of the savings occur in the first year, followed by a steep decline. Therefore the savings for non-Buttons products were calculated by adding the first year savings to estimated savings from the following year (or the portion of that year within the product lifetime). Savings from the 2nd year were calculated as if there is a straight line graph from the first year savings level to zero at the end of the lifetime, as follows:

Lifetime Savings = First Year Savings + First Year Savings * (Product Lifetime in years – 1) / 2

This can be visualised as (in the case of a lifetime of 1.75):



Lifetime savings for Buttons were calculated as a simple triangle:

Lifetime Savings = First Year Savings * Product Lifetime in years / 2

The lifetime savings calculations produced a wide range of lifetime energy savings, from 44 kWh for Timeout Happy Hour to 4125 kWh for Buttons gas. This shows the importance of the rate of decline of savings when considering the lifetime savings. Gas savings were in general much higher than electric savings. The total lifetime energy savings calculation is an approximation from a trendline and therefore its accuracy is not very high; however, it gives an indication of the relative total impact of each product.

Table 27: First Year and Lifetime Savings in kWh

Product	Product Trial Group	Savings for First Year (kWh)	Lifetime of Product in Years	Total Lifetime Savings (kWh)
Bounty Electric	CDU_HIGH	156	1.5	195
	CDU_LOW	91	1.3	106
	NOCDU_HIGH	95	1.5	119
	NOCDU_LOW	96	1.4	116
Bounty Gas	CDU_HIGH	1,809	1.6	2337
	CDU_LOW	1,286	1.4	1554
	NOCDU_HIGH	216	1.4	261
	NOCDU_LOW	849	1.4	1026
Buttons Electric	Monthly Billing	271	2.5	339
Buttons Gas	Monthly Billing	3300	2.5	4125
Timeout Electric	Happy Hour	29	2.0	44
	Timeout Regular	443	1.4	535

3.2.5 Summary of Lifetimes and Lifetime Savings

Table 28 shows a summary of the lifetimes for all of the product groups and the lifetime CO_2 savings. The lifetimes ranged from 16 to 30, with Buttons groups having the longest estimates on average and Bounty groups the shortest. The results show that savings from behavioural measures cannot be relied upon to be sustained indefinitely and that there is more to encouraging long term energy demand reduction than simply providing information. When these lifetime estimates are used to compute total lifetime savings for each product, the results indicate that the most effective product in terms of CO_2 saved was the monthly billing product.

Table 28: Summary of Product Lifetime Estimates in kg CO_2

Product	Product Trial Group	Projected Final Month of Effect	Estimated Lifetime of Product in Months	Total Lifetime Savings (kg CO_2)
Bounty Electric	CDU_HIGH	03-2011	18	84
	CDU_LOW	01-2011	16	46
	NOCDU_HIGH	03-2011	18	51
	NOCDU_LOW	02-2011	17	50
Bounty Gas	CDU_HIGH	04-2011	19	444
	CDU_LOW	02-2011	17	295
	NOCDU_HIGH	02-2011	17	50
	NOCDU_LOW	02-2011	17	195
Buttons Electric	Monthly Billing	03-2012	30	146
Buttons Gas	Monthly Billing	03-2012	30	783
Timeout Electric	Happy Hour	09-2011	24	19
	Timeout Regular	02-2011	17	230

4.0 APPENDIX E – Analysis of Survey Data

Three rounds of surveys were carried out. The first one (Wave 1) was conducted before the trial had started and customers were accepted onto the trial in June 2009, and the second one was conducted after customers had had the smart meter and product for several months, in February 2010. The third wave was carried out in January 2011. The questions on each wave of the survey varied somewhat due to the timing of the survey and the product the customer had. Table 29 below shows a summary of the data collected in all three waves.

Table 29: Summary of Surveys and Usefulness for the Analysis

Section Number	Section Title	Wave 1	Wave 2	Additional Data in Wave 3
1	Trial Sign Up and Meter Installation	No	Reasons for joining, performance of SM and CDU, contact with npower re problems.	
2	Attitudes Towards Energy Efficiency & Climate Change	Climate change, who is responsible for energy savings, source of energy savings advice, etc.	Same as Wave 1.	
3	Energy Usage & Monitoring	How often bill checked, knowledge of carbon footprint, bills/usage gone up/down.	Same as Wave 1.	
4	Perceived impact of SM and CDU and product specific features.	No	How often CDU viewed, usefulness of data, benefits of product features (varies by product)	
5	Energy Efficiency Measures Taken/Intending to take	Insulation, appliances, hot water and heating, general, microgeneration. Some capital and some non-capital	Same as Wave 1.	When capital upgrades were done, and if they were done through CERT or some other scheme.
6	Satisfaction with CDU and SM, and relationship with npower	Satisfaction with npower as supplier	If they will keep the SM/CDU, how satisfied with the SM/CDU, satisfaction with npower as supplier	
7	Respondent Profile	Size of household, bedrooms, tenure, type of property, children, gender, benefits receipt, type of heating.	Same as Wave 1, minus tenure, type of property, plus age group.	Any significant events in last 12 months such as being away for a significant time or working from home.

Data pertinent to the savings analysis was extracted from the final survey data set (Wave 3). The key reason for looking at the survey data was to determine if a significant number of participants had installed capital improvements in their houses. If this was the case, the savings from the capital improvements would be misrepresented as behavioural savings if not taken into account. In addition, households with electric heating or cooking could be identified as these would have a different mix of electricity and gas to the majority. Finally, households that made changes that would affect energy use, such as a change in the number of people living in the house, could be identified to account for this as changes in use would not be only due to conservation due to being in the programme. Not all of the participants answered the survey in each wave. The response rates are shown below.

Table 30: Response Rate for Surveys

Product	Wave 1	Wave 2	Wave 3	URNs with SM Data	Final survey % of URNs
Bounty	459	355	209	541	39%
Buttons	421	338	201	430	47%
Timeout	386	309	183	401	46%

Amongst other questions, respondents were asked whether they had installed any capital efficiency measures (prompted from a set list), and also whether those improvements had been done as a result of being on the trial. The data showed that only a small percentage of households made capital improvements they attributed to being on the trial, as shown below. Probably the highest impact measure would have been cavity wall insulation, in terms of affecting gas use, followed by dry lining or external wall insulation, replacing boilers, and external wall insulation. A household that did all three of these could have expected significant gas savings. The percentage of all households on the trial that made these changes will be less than the percentages shown in the table below by about ½ as the table shows the percentage of survey respondents and the % of all URNs that completed the third survey is less than ½ for all the product groups.

Table 31: Attributable Capital Improvements from Survey Data, Wave 3

	Bounty		Buttons		Timeout	
Capital Measure	Count Attributed to Trial	% of Respond-ents	Count Attributed to Trial	% of Respond-ents	Count Attributed to Trial	% of Respond-ents
Cavity wall insulation	6	2.9%	4	2.0%	7	2.0%
Draught Proofing	1	0.5%	5	2.5%	1	2.5%
Lagging of hot water tank	0	0.0%	2	1.0%	1	1.0%
Floor insulation	3	1.4%	3	1.5%	1	1.5%
Dry lining or external wall insulation	4	1.9%	6	3.0%	2	3.0%
Loft insulation	2	1.0%	3	1.5%	0	1.5%
Double glazing	0	0.0%	0	0.0%	0	0.0%
Bought a rated appliances or higher	10	4.8%	10	5.0%	13	5.0%
Got rid of old high energy appliances	13	6.2%	11	5.5%	13	5.5%
TRVs	7	3.3%	6	3.0%	3	3.0%
Other	5	2.4%	7	3.5%	2	3.5%
Low flow shower heads/taps	4	1.9%	6	3.0%	5	3.0%
Replace old boiler with new A-rated	13	6.2%	5	2.5%	5	2.5%
Solar thermal (hot water)	0	0.0%	0	0.0%	0	0.0%
Solar PV (solar panels)	0	0.0%	0	0.0%	0	0.0%
Micro wind turbine	0	0.0%	0	0.0%	0	0.0%
Ground source heat pumps	0	0.0%	0	0.0%	0	0.0%
Other	2	1.0%	0	0.0%	1	0.0%

Respondents were also asked about whether there had been demographic changes to their household. 12%, 6% and 10% of Bounty, Buttons and Timeout respondents, respectively, said there had been changes; the average change in household size was relatively small at an average of -5%, 13% and 10% respectively (number of persons in the household). Buttons and Bounty respondents were asked about gas appliances and around 5% said they did not have gas water heaters and 25% did not have gas cookers, so we can assume that they have electric ones.

Based on the relatively low level of capital improvements and changes in household members, no adjustments were made to the model results based on the survey data.

5.0 APPENDIX F – Original Trial Design

The trial has been designed with a total of 12 different groups, as shown in Table 32 below. Note that the size of the groups shown in the table, as defined at design stage, will not be the same as the final number of customers for which trial data will be supplied as some customers may have dropped out or been removed from the trial for some reason.

Table 32: Trial Groups, Consumption Data and Survey Data

Product type	Group	Group description	Target Size of Group	Smart meter?	CDU?	Advice / information provided?	Consumption Data	Survey Data (Wave 1/ Wave 2)
Time of Day (Timeout)	1	Choose and use	225	Yes	Yes	Yes (monthly tips to CDU)	Up to 24 months of pre-trial quarterly electricity readings. 12 months of post-installation ½ hourly electricity meter readings. Yearly kWh values (pre and post).	276/155
	2	Choose and use 2	225	Yes	Yes	Yes (monthly tips to CDU)		276/144
	3	Control Group	50	Yes	Yes	No		62/41
	4	Ghost Control Group	500	No	No	No	UP to 24 months of pre-trial quarterly electricity readings. 12 months of trial period quarterly electricity readings. Yearly kWh values (pre and post).	none
Enhanced Web Service (Bounty)	5	CDU / high communication	112	Yes	Yes	No - information and links on website	Up to 24 months of pre-trial quarterly gas and electricity readings. 12 months of post-installation ½ hourly gas and electricity meter readings. Yearly kWh values for gas and electricity (pre and post).	138/86
	6	CDU / low communication	113	Yes	Yes	No - information and links on website		142/100
	7	No CDU / high communication	112	Yes	No	Information and links on website only		139/82
	8	No CDU / low communication	113	Yes	No	Information and links on website only		131/79
	9	Control with CDU	50	Yes	Yes	No		62/39
	10	Ghost Control Group	500	No	No	No	Up to 24 months of pre-trial quarterly electricity and gas readings. 12 months of trial period quarterly electricity and gas readings. Yearly kWh values (pre and post).	None
Monthly Billing (Buttons)	10	Monthly billing group	450	yes	Yes	Yes (monthly tips to CDU)	Up to 24 months of pre-trial quarterly gas and electricity readings. 12 months of post-installation ½ hourly gas and electricity meter readings. Yearly kWh values for gas and electricity (pre and post).	550/328
	11	Control group	50	yes	Yes	No		64/36
	12	Ghost Control Group	500	No	No	No	Up to 24 months of pre-trial quarterly electricity and gas readings. 12 months of trial period quarterly electricity and gas readings. Yearly kWh values (pre and post).	None

6.0 APPENDIX G - References

1. Kester, J.C.P.E., M.J.G.E.I. Burgos, and P.J. (BEAMA), *Smart Metering Guide - Energy Saving and the Customer, Edition 2010*, 2009, European Smart Metering Alliance, sponsored by Intelligent Energy Europe.
2. Neenan, B., EPRI, J. Robinson, EPRI, and R.N. Boisvert, *Residential Electricity Use Feedback: A Research Synthesis and Economic Framework*, 2009, Electric Power Research Institute: Palo Alto, California, USA.
3. Parker, D., D. Hoak, and J. Cummings. *Pilot Evaluation of Energy Savings and Persistence from Residential Energy Demand Feedback Devices in a Hot Climate*. in *ACEEE Summer Study on Energy Efficiency in Buildings*. 2010. Florida Solar Energy Center.
4. Klos, M., Navigant Consulting, *Impact Evaluation of OPOWER SMUD Pilot Study*, 2009, Navigant Consulting for Sacramento Municipal Utility District: Boulder, Colorado.
5. DECC, *Impact Assessment of Smart / Advanced Meters Roll Out to Small and Medium Businesses*, D.f.E.a.C. Change, Editor 2009: London.
6. DECC, *Impact Assessment of a GB-wide Smart Meter Roll Out for the Domestic Sector*, D.f.E.a.C. Change, Editor 2009: London.
7. OFGEM, A., Building Engineering, *Energy Demand Research Project: Final Analysis*, OFGEM, Editor 2011, AECOM Building Engineering.
8. Kihm, S., Research Director, Energy Center of Wisconsin, K. Koski, Analyst, Energy Center of Wisconsin, and A. Mendyk, Project Manager, Energy Center of Wisconsin, *Focus on Energy – PowerCost Monitor Study, Final Report*, 2010, Energy Center of Wisconsin: Madison.
9. Siems, A., Opinion Dynamics. <*Siems Eval Massch PowerCost Monitor.pdf*>. in *ACEEE 2009 National Symposium on Market Transformation*. 2009. Washington DC.
10. Parker, D., D. Hoak, and J. Cummings, Florida Solar Energy Center, *Pilot Evaluation of Energy Savings and Persistence from Residential Energy Demand Feedback Devices in a Hot Climate*, in *ACEEE Summer Study on Energy Efficiency in Buildings* 2010.
11. van Dam, S.S., C.A. Bakker, and J.D.M. van Hal, *Home Energy Monitors: Impact Over the Medium-term.* Building Research & Information, 2010. **38**(5): p. 458-469.
12. Brandon, G., Centre for Economic Psychology, Department of Psychology Faculty of Humanities & Social Sciences, University of Bath and A. Lewis, *Reducing Household Energy Consumption: A Qualitative and Quantitative Field Study.* Journal of Environmental Psychology, 1999. **19**.
13. Woolridge, J. *What's New in Econometrics? Lecture 10, Tuesday, July 31st, 4.30-5.30 pm, Difference-in-Differences Estimation*. in *National Bureau of Economic Research (NBER) Summer Institute*. 2007.