Visible Energy Trial:

Report for OFGEM

December, 2010

Authors: Dr Michael Nye
Dr George D Smith
Dr Tom Hargreaves
Professor Jacquie Burgess

Table of Contents:

1 Introduction

The Visible Energy Trial (VET) was a year-long study that analysed the impact of three different types of In Home Display on the energy consumption behaviour of 275 homes across East Anglia. It was jointly funded by Carbon Connections[1], Green Energy Options (GEO)[2] and British Gas. It was conducted in collaboration with SYS Consulting Ltd[3] and the School of Environmental Sciences at the University of East Anglia and was approved by Ofgem as a CERT Demonstration Action. This report summarises the design, conduct and results derived from the trial.

2 The Trial

2.1 Objectives of the trial

There were 4 high-level objectives for the trial, each with 3 sub-objectives:

1. To prove the technology:
    a. How the individual communications technologies perform
    b. Prove the long term performance and reliability of the systems
    c. Establish the effectiveness different data management strategies.

2. To compare the performance and functionality of the displays:
    a. Identify the initial and long term response to the different displays
    b. Assess their relative effectiveness in terms of consumer engagement
    c. Quantify the value of individual functions and features

3. Identify and quantity the benefits:
    a. Identify the actions and behaviours prompted by the displays
    b. Establish the durability of behavioural effects
    c. Quantify the reductions achieved

4. Explore the value of the data:
    a. Statistical analysis of time series to identify 'normal' usage patterns
    b. Identification of appliance (type) signatures from time series data
    c. Market segmentation based on clustering of time series

2.2 The design of the trial

This was an ambitious trial as it used three prototype displays being developed in parallel, each of which introduced significant levels of new, untried functionality:

- The SOLO is a Real Time Display that introduced back-lit colour graphics, cumulative consumption and the concept of a daily target.

- The DUET extended the SOLO functionality to provide real time monitoring and control of 6 individual electrical appliances plus real time feedback on central heating and hot water boiler performance (oil, gas or propane).

[1] The Carbon Connections Development Fund is managed by the University of East Anglia and funded by the Higher Education Council for England. Its aim is to support organisations with innovative projects in carbon reduction.
[2] Green Energy Options design and manufacture energy monitoring devices for home, schools and business. The aim of the devices is to 'make energy visible'.
[3] SYS Consulting is a wholly owned consulting subsidiary of the University of East Anglia specialising in data analysis and data mining.

- The TRIO is an advanced energy application that runs on a dedicated micro-computer and provides detailed information on the electricity consumption of the whole house, all circuits and six individual appliances. It also provides detailed information on boiler functions and home temperatures.

The trial involved 4 groups of homes. Three groups, each made up of 75 homes, received one of the three types of display. The fourth group of 50 homes was designated a Non-Display group and were fitted with a TRIO system but without access to the display or any data. This group was intended to serve as a comparator against factors outside of the control of the trial (such as advertising, price adjustments and changes in economic conditions) that could affect energy use patterns in all four groups.

All three display groups are made up self-selecting volunteers. The SOLO Group were recruited mainly from Housing Associations with an emphasis on those in the Priority Group They were provided with the display for free via the relevant Housing Association. The DUET and TRIO groups were recruited across East Anglia through general advertising and from the CRED community, a UEA initiative to encourage a self-selecting group of people to pledge to reduce their carbon footprint through various voluntary actions. In order to assess the business viability of the more advanced displays, DUET volunteers were required to pay £75 and TRIO users £250.

Volunteers for the trial were interviewed over the telephone by a market research company prior to final selection. These interviews were conducted with the aims of collecting relevant demographic and contextual information about trialists, gauging levels of interest/ preference for display type, and screening out households with electric central heating. The demographic breakdown of the display groups (SOLO, DUET and TRIO) is summarised below:

2.2.1 Socio-demographic:

- The largest group (~40%) of households recruited were 2 person households. 1,3 and 4 person households make up a further ~50% collectively. Nearly 50% of those recruited reported that they did not have any children living at home.
- Most of the households recruited for the DUET and TRIO groups describe themselves as middle class or higher (~85% A, B, C1 social grade).
- The majority of respondents (~55%) report household incomes greater than £30k. Around 60% of households with a reported income less than £30k represent the SOLO group (recruited through a housing association).

2.2.2 Contextual/ infrastructure:

- The majority (>90%) of households recruited for the DUET and TRIO groups own their homes. Note that this is not the case for the Non-Display and SOLO groups as these were recruited through a housing association.
- The largest group (~35%) of trialists live report living in detached houses. Semi-detached is the next largest grouping (~20%).
- Over 55% of households recruited live in a house that was built prior to 1965.
- About 60% of recruited households indicated that they had already made energy saving or other ‘environmentally friendly’ investments such as gas condensing boilers, double glazing or water butts.

2.3 Coordination and technical issues

This was an ambitious trial as it used three prototype displays being developed in The original aim was to start all three groups in parallel, to install the systems over as short a period as possible and then run the trial for a full year with a minimum of intervention. Considerable technical difficulties were encountered getting all three systems involving 10 individual products ready together. This delayed the start of the trial by over a year. These delays hindered recruitment with some people leaving as a result, but replacements were found and the trial fully populated.

The TRIO systems were installed by GEO staff. This proved more difficult than expected, primarily because significant effort was required to co-ordinate visits across the region, particularly with regard to working families. The installation was carried out by two people - a qualified electrician and an IT installer - and took approximately 3 hours including training. Approximately 15% of sites could not be completed in one visit due to a mixture of electrical safety concerns and router issues. In contrast, the SOLOs and DUETs were self installed with little difficulty other than identifying the correct boiler pipes to be monitored.

Early analysis of received data provided considerable insight into the functioning of the systems themselves over extended periods and helped to identify a number of design defects. As a consequence the TRIOs received a number of system updates over the web during the trial and both the SOLO and DUET systems required firmware upgrades half way through the trial. This was achieved through a combination of returning units and site visits as selected by the trialist.

The combination of these delays meant that the start of the trial became considerably extended. Indeed, the heating element of the TRIO system was not installed until well over halfway through the trial and, unfortunately, after the end of the main heating season. Installations started in May 2009 and extended to September 2009. The trial completed at the end of July 2010.

Other technical findings/ issues are summarised as follows:

- PowerLine communications proved to be long-range-reliable but limited by bandwidth.
- ZigBee communications (private protocol, point to point) were more complex, less reliable but had more than adequate data rates.
- Many problems were experienced with broadband routers:
  - there are many differing types and ages of router,
  - several people changed their provider and/or router during the trial
  - many people turn their routers off for the majority of the time
  - considerable support effort was expended resolving broadband issues many of which were not down to the GEO equipment but nevertheless required action by GEO to identify the cause.
- Unless properly managed, loss of ZigBee communications can cause excessive battery usage as the sensor attempts to regain contact.
- Few problems were reported with gaining access to and connecting the CT clips to the meter tails.

2.4 Data collection

Near real time consumption data (15 minute epochs) were recorded for all 4 groups of trialists. Data for TRIO and the Non-Display groups were squirted automatically to

GEO Servers. Data from the SOLO and DUET systems was saved to an SD card, which the user could then transfer to PC for upload.

Data collection for the SOLO and DUET groups proved to be a challenge with only 50% of trialists downloading data from their SD Cards despite many efforts to encourage them to do so. Whilst data was received from all of the TRIO systems there proved to be several significant gaps in the data itself due in part to the volume of data and memory limitations on the TRIO system itself.

In addition to consumption data, on-line surveys using Survey Monkey were conducted halfway through the trial and on completion. A series of one-to-one interviews with a group of about one dozen trialists was also performed alongside the survey.

2.5 Data cleansing

As a result of the technical problems with the early versions of the systems, many of the systems reported series of NULL values, and sometimes even negative values, for daily usage as a result of system resets or slippage. Where possible, efforts were made to complete these gaps without compromising the data, (i.e. not making estimates). However, this still left a large number of systems that had to be dropped from the longitudinal analysis because of the large proportions of such records.

A further problem encountered was the lack of sufficient overlap in the days monitored. SOLO and DUET users, who had to regularly upload their data from the SD card, appeared to lose interest in doing so, hence the amount of data from these systems dramatically reduced around spring 2010.

3 Methodology of analysis

Analysis to support the objectives of the trial was performed in several distinct quantitative and qualitative stages and formats:

3.1 Quantitative analysis:

- *Measurement of extent of change.* This analysis assessed the extent to which different groups of trialists changed their energy use patterns compared to a Non-Display group.
- *Assessment of factors contributing to behaviour change and the importance of feedback*. This analysis was applied to the two online surveys conducted midway, and one month after the trial had finished. Key questions considered were: what level of information is needed; what scale of improvements can be generated, how durable are behavioural changes, and what displays/ functionality are preferred?
- *Longitudinal (time series) analysis for assessing the durability of change.* This analysis explored the extent and durability of changes in observed consumption patterns between and within different groups of trialists. This analysis was performed at several different time scales.
- *Cluster-analysis-based exploration of segmentation of the mains sensors daily profile data to identify clusters of daily energy usage patterns*. Patterns of behaviour were examined in an attempt to identify significant contributing factors that enhanced or provided a barrier to changes in energy consumption.
- *Additional analysis of the time series data at socket/circuit level.* This analysis explored the value of the data in terms of the feasibility of identifying

signatures for types of appliance and patterns of usage for different types of appliance.

3.2 Qualitative analysis:

- *How households react to and engage with the display and how they characterise the importance of feedback in terms of behaviour change.* This analysis applied primarily to the interviews conducted with trialists, but also incorporated some findings from the online surveys

4 Main findings from individual elements of the study

4.1 Analysis of interview data:

For the initial set of qualitative interviews, a stratified random sample of 15 households was selected to include 4 households from each of the SOLO, DUET and TRIO groups, and three from the Non-Display group. This was conducted in October-November 2009 by telephone (9) and face-to-face in the participant's home (6).

For the follow-up interviews, conducted in October 2010, only those households in the SOLO, DUET and TRIO groups were contacted, although one SOLO user declined to be interviewed. All interviews were done by telephone.

4.1.1 Interview themes:

During the interviews, trialists were asked to comment on the following themes:

- How they had got on in the trial and their motivations for doing so
- How they had used and interacted with the device
- If the device had affected their energy awareness or behaviour, and in what ways
- How their usage of the device changed during the trial.
- Any recommendations to help improve the device.

4.1.2 Findings from interim interviews:

The initial set of 15 semi-structured interviews revealed 4 different forms of motivation for participating in the Visible Energy Trial: A) financial – a desire to save money by saving energy was the most dominant motivation; B) environmental – a desire to reduce carbon dioxide emissions; C) technological – a desire to get a new gadget and learn what it could do; D) informational – a desire simply to gather more detailed information on domestic energy use. Whilst in many cases a combination of motivations was expressed, the dominant form of motivation expressed proved crucial to how the GEO monitors were evaluated by householders. For example, those who expressed a financial motivation were often frustrated at being able only to 'save pennies' when they wished to 'save pounds or even tens of pounds' (T1, p2).

Several interviewees were dissatisfied with the installation process for the devices and this had adversely affected their perception and use of the device suggesting this point of first contact is vital. In particular, interviewees called for more assistance in tailoring the devices to their own home and more advice on how they could save energy most effectively. Once installed, interviewees reported that they predominantly used the pounds and pence metric as they did not find carbon dioxide or kWh figures meaningful. Further, they found information on specific appliances most valuable in helping them make sense of their energy use and realise reductions. This was described by one interviewee as a need to make the information 'relational' to everyday life (T4).

Interviewees used the devices to save energy in two distinct ways. First, and most often, the devices were used reactively. Upon noticing a high level of usage interviewees would immediately go around the house to switch lights off and appliances off standby. The devices thus prompted an immediate reaction to investigate 'unnatural' (D4) levels of use and to restore it to 'normal' levels. Second, the devices were used in a more planned and rational manner. Here, interviewees used the devices to a) identify especially wasteful appliances such as old fridge-freezers and dispose of them, b) gradually buy more efficient appliances, and c) make future plans to improve levels of household energy efficiency or, in some cases, to install micro-generation.

Interviewees also reported several forms of limitation in reducing their energy use. First, some aspects of energy consumption were seen as necessary and non-negotiable. Precisely what this was varied enormously between interviewees including, for example, using a kettle, having the heating on, running a fish tank pump, using a bread maker etc. Crucially usage that was considered discretionary was often perceived as extremely small and unlikely to realise any significant savings. Second, interviewees noted that despite their personal willingness to save energy, other household members were not so easily persuaded and hindered their efforts. A key finding from these initial interviews was that these devices are more usefully conceived as being used by whole households than by individuals alone. Third, interviewees regularly cited a lack of meaningful support and information from housing associations, local and central government, appliance manufacturers and retailers as a key reason why they did not make greater efforts to save more energy.

4.1.3 Findings from final (follow-on) interviews:

11 follow-on interviews were conducted with the same householders 12 months after the initial set of interviews. A key finding of these follow-on interviews was that 8 out of the 11 interviewees were still using their GEO monitor and, of the 3 who weren't, only 1 had rejected it and packed it away (in the other 2 cases, one had moved home and the other was experiencing technical difficulties with the device).

The key theme of the follow-on interviews was the way the monitor had ceased to be a novelty and was now simply another item of domestic technology. All of those still using the monitor mentioned that they looked at it less frequently and with no regular routine, but that it had now become part of the background of everyday life; for example, one interviewee described the device as having become 'part of my office furniture' (T1).

Habituation was also expressed in how interviewees used the devices. Here, they showed extremely detailed knowledge of their everyday levels of energy use knowing, for example, exactly how much they spent on energy on an average weekday or exactly how many kWh it took to keep the house 'ticking over'. Interviewees were proud to talk about their expertise in being able to notice even very small discrepancies in usage and, thus, being able to take immediate action to avoid unnecessary energy use.

One, perhaps unintended consequence, of increased expertise in monitoring household energy consumption was a marked unwillingness to reduce energy consumption to below the 'new normal' established by use of the monitors. In almost all cases, following installation of the monitor, efforts had been made to reduce levels of energy consumption. But after this initial burst of energy saving activity, a level was achieved that was deemed satisfactory by the members of the household. Subsequently the GEO monitors were being used to maintain this satisfactory level, rather than to reduce it further. Some interviewees expressed anger and frustration towards any who subsequently suggested they 'should' reduce their energy consumption still further, especially in the perceived absence of more meaningful action to reduce energy consumption from government and big business.

4.2 Quantitative analysis of online questionnnaires:

As mentioned previously, two online surveys were conducted with trialists approximately midway through the trial and one month after the end of the trial (although many of the units were still in place and operating). Only respondents in the SOLO, DUET and TRIO groups were asked to complete the surveys.

The surveys were designed and coordinated by researchers from UEA and representatives from Green Energy Options. Questions covered three themes: Feedback on display functionality and engagement with the display, energy use/energy saving behaviour, and changes in energy awareness and consequent actions taken as a direct result of using the display (e.g. buying low energy light bulbs, appliances etc). Space was provided with every survey question for respondents to add their own comments and ideas and/or to clarify their responses in more detail.

A total of 116 trialists responded to the Interim survey and 127 responded to the final survey. No attempt was made to match the respondents between surveys. The interim survey was conducted in Feb/March, 2010 and the Final Survey in October, 2010.

Interim: n=116		Final: n=127	
OPTION	**NUMBER**	**OPTION**	**NUMBER**
SOLO	33	SOLO	37
DUET	33	DUET	40
TRIO	50	TRIO	50

Responses to selected questions are presented here in order to provide a flavour of the issues explored in the surveys and the range of responses gathered.

4.2.1 Display functionality and engagement:

Satisfaction: For all display types, the largest number of interim survey respondents indicated that they were satisfied with the display equipment. SOLO respondents to the interim survey reported the highest levels of satisfaction (~80% 'satisfied' or 'very satisfied'). DUETs were not far behind at ~70%. TRIO respondents were more likely to indicate neutrality or some level of dissatisfaction (~50). Similar levels of satisfaction were observed in the final survey. Written comments suggest that the greater levels of dissatisfaction for TRIO respondents related to dissatisfaction with the display *unit* (modified laptop), rather than the display functionality. Note that this was corrected later in the trial with provision of a touch screen device, and the majority of recipients indicated this was an improvement over the original kit.

Use of display: Nearly 80% of SOLO respondents reported that they were still looking at the display at least once per day at the midpoint of the trial. DUET respondents were similarly engaged (~75%). TRIO respondents were more likely to report using the display less - over 50% reported that they were using it once or twice a week, or not at all. By the final survey (7 months later) approximately 52% of SOLO respondents indicated that they still used the display at least as much as at the beginning of the trial, compared to 63% for DUETs and 35% for TRIOs. Several of the written comments in the interim survey indicated that levels of interest in the display/ equipment tended to decrease to varying extent as the trial progressed. The interim and final survey data support this conclusion, as do the interview data on this subject.

4.2.2 Energy Saving Behaviours:

Changes in consumption: In all cases, the majority of respondents to the interim survey indicated that they felt that they used less energy than prior to participating in the trial. DUET respondents were the most positive overall – 68% indicated that they used less or significantly less than before the trial, although SOLO respondents are not far behind (63%). TRIO trialists were slightly less likely to report using less in the interim survey (52%), but by the final survey the number had climbed to just over 60%. There is nothing in the written comments to indicate why the TRIO group reported lower levels of energy savings initially, although concerns about display equipment might have an effect here. Interim survey respondents who did indicate decreased energy consumption as a result of participation in the trial report estimated savings in the region of 5-10%.

Switching appliances and lights off: Approximately 65%, 90% and 75% of SOLO, DUET and TRIO respondents to the interim survey reported that they tended to switch off things that they used to leave on (e.g. computers, hi-fis, chargers etc) *at least a little more* than they used to as a result of participating in the trial. These numbers dropped ~10% by the final survey for the SOLO and DUET groups, but all groups are still well above 60%. The majority of interim survey respondents in all three groups (~70%-90%) also indicated that participation in the trial had encouraged them to turn lights off when not in use *at least a little more* than prior to participation in the trial. The TRIO group dropped a little by the final survey (to just over 50%) but overall, the majority of respondents appear to have continued to switch off lights when not in use.

4.2.3 Changes in Energy Awareness and Further Actions:

Other green behaviours: Interim survey respondents in all three groups tended to indicate that they had not engaged in any other non-energy pro-environmental behaviours (e.g. composting, or water saving) as a result of participating in the trial (although they might have engaged in this behaviour). A similar breakdown of responses occurred in the final survey data. However, the interim survey responses also suggest that participation in the trial encouraged/ empowered all three groups of respondents to be more confident to offer energy saving advice to others. The effect seems to be strongest amongst the SOLO respondents, with nearly 80% of respondents reporting a positive change in this area in the interim survey. A similar effect was observed in the final survey, where >60% of respondents in all groups indicated an increased level of confidence in this regard. This is an important, and - as far as we are aware – new, finding. The relationship between use of the real time display and increased 'engagement' with household energy use appears to have had a positive effect on trialists' confidence to talk about greener energy use with others.

Appliance switching: TRIO respondents, who tended to be higher income, were the most likely group to indicate that participation in the trial had prompted them to purchase low energy appliances (23% in the interim survey and 35% in the final survey). DUET respondents were most likely to report that they were considering investing in this area in the interim survey (>65%), and by the final survey, ~35% indicated that they had purchased more efficient appliances as a result of participation in the trial. Roughly 25% of SOLO respondents indicated the same in the final survey. These findings suggest that the trial was successful in *both* encouraging trialists to consider purchasing energy efficient appliance, and in prompting selection of energy efficient appliances when purchases are made. This is an important bridge between awareness/ consideration and actual purchasing behaviour. Written comments for the interim survey suggest that appliance switching was done opportunistically, or as and when the need arose. For most respondents, up-front purchase costs appear to have outweighed considerations about longer-term energy (and cost) savings, until an appliance needs replacing.

4.3 Longitudinal analysis of electricity consumption data:

This section describes the longitudinal analysis of consumption data that was conducted with the aim of assessing the extent and durability of changes in consumption patterns.

4.3.1 The importance of seasonality and normalisation:

Although energy use from central heating is not included in the data, it is reasonable to assume that electricity use other than central heating would increase in winter/ colder months as temperatures cool, daylight hours decrease, and people spend more time indoors. A strong seasonality effect was observed in mean daily data consumption data for all trialists, with the highest mean daily consumption values concentrated near the 1 of January. Further in-depth analysis suggested that ~75% of systems follow this pattern. Because trialists began at different times of the year, discrete time periods could not be compared unless the consumption data were normalised by an appropriate factor, such as degree days, which approximates the general trend observed in trialists' energy use.

After extensive correlation analysis, degree days calculated for Cambridge airport, with a base value of 25.5C were chosen as the seasonal normalisation factor for the trial. The base value of 25.5C was chosen to avoid introducing variance into normalised data where the ratio between consumption and recorded degree days was high (i.e. hot days). Although this base value is higher than the 'standard' value of 15.5C, the correlation values between degree days derived from either base temperature and energy use are approximately equal. Furthermore, because the analysis compared standardised consumption data (which will be explained momentarily) the base value of degree days is inconsequential. Prior to normalisation, consumption data were compared to degree day data using linear regression on a system-by-system basis to determine 'base load' (non-weather dependent) consumption for each participant. Only weather-dependent consumption was normalised.

4.3.2 Further cleansing of data and removal of systems:

Because of the relatively small group sizes, it was important to screen out outlying values that could skew sample means. For the purposes of this analysis, outliers were defined as >+/- 3 standard deviations from mean consumption for all recorded consumption values for each system. Statistically 99.7% of cases in a normal distribution will fall within +/- 3 standard deviations of the mean. This method is more precise than simply specifying percentiles as a cut off point (i.e. middle 95%) because it is more flexible. Outliers were removed from the data set prior to normalisation and after normalisation (in the latter case, the aim was to control for extreme values introduced by the normalisation process). The same mean was used as the base in both cases.

After normalising and removing blank periods, the data were examined to determine which systems could be used for longitudinal analysis. Systems were excluded if they had runs of blank days totalling more than 30% of the base time period under examination. The group of systems compared at the 30 day scale was larger (n=141) than that for the 60 day scale (n=98) because there were a larger number of systems with long runs of missing data in the longer time period.

4.3.3 Standardising observed consumption values:

Normalised consumption data for longitudinal analysis were standardised by dividing the observed consumption value for a particular day by the mean for all observed (all recorded) days for each system up to the cut off point for the dataset (29/07/10). Using standardised values allowed direct comparison of systems with differing volumes of consumption. The mean for all observed (recorded) days for each system is 1. The mean for each system is likely to vary from 1 for discrete time periods within the trial. These periodic variations form the basis of the longitudinal analysis.

It is important to note that because historic baseline data (energy use data prior to installation of the metering/ display equipment) were not recorded, it is impossible to judge whether any of the mean values or gradients examined in this section represent a decrease or increase from the 'true mean' for a participant. All of the differences in consumption values or gradients examined here are *relative* to one time period compared to another and should only be interpreted that way. *These data cannot be used to determine how patterns or trends in the trial data compare to patterns or trends in consumption prior to the start of the trial.*

4.3.4 Longitudinal methodology of analysis:

As mentioned above, two different base scales were used to compare mean standardised consumption in different periods: 30 days and 60 days. Because the standardised mean for each participant is '1,' means tended to move closer to 1 over longer time scales. The 30 and 60 time scales allowed for sufficient variation in mean, but also represented two different time periods of engagement with and adjustment to the display equipment (1 month, 2 months and 4 months respectively). A longer-term comparison of gradients (line of best fit) over 210 days was also performed in order to capture a sense of how energy use trends changed over the course of the trial. However, as the gradient proved particularly susceptible outliers, these results were not considered reliable and are not reported here. All comparisons were performed between or within groups (SOLO, DUET, TRIO and Non-Display). Both appropriate parametric and non-parametric equivalents were run in all cases to examine statistical significance

In terms of statistical comparisons, the null hypotheses tested were:

A) There is no significant difference in mean standardised consumption for all systems between different time periods (this is tested at 30 and 60 day scales using T Tests and Mann Whitney U).
B) There is no significant difference in mean standardised consumption between system groups (i.e. TRIOs vs. SOLOs) for specific time periods (i.e. days 1-30 using ANOVA and Kruskal Wallis tests).
C) There is no significant difference in mean standardised consumption within system groups (i.e. all SOLOs) between different time periods (i.e. days 1-30 compared to days 31-60 using T Tests and Mann Whitney U).

4.3.5 Findings from longitudinal analysis:

Group based comparisons of standardised mean differences for Null Hypothesis A showed no statistically significant differences at different time periods for all systems. Means were approximately similar between time periods and no clear pattern emerged.

Group-based comparisons of standardised mean differences for Null Hypothesis B using ANOVA did show a significant difference between the Non-Display group in the first 30 day period compared to the other groups. Non-Display trialists tended to consume above average levels (about 2% above average) for the first 30 days compared to about 1-2% below average consumption for the other three groups in this period. This difference was found to be significant ($p < .05$) between the control and all other systems. (A significant difference was also found between the Non-Display group and the SOLO group for the first 60 days). However, these trends were not repeated in the second 30/ 60 day period, and subsequent analysis suggested that the early difference stemmed from a few high-consuming systems that had skewed the mean for the Non-Display group upward in the earlier days of the trial. This finding illustrates the effect of the small group sizes on the ability to reliably detect significant differences between groups.

Within group comparisons of standardised mean differences for Null Hypothesis C revealed an interesting pattern for the SOLO group. SOLO trialists appeared to use considerably less electricity on average in the first 30 days compared to the second 30 days (standardised mean of about .95 compared to about 1). Moreover, this pattern holds true for *all* SOLO systems (all of the SOLO systems used comparatively less energy in the first 30 days of the trial than in the second 30 days). Significant differences were not observed for SOLO systems at the 60 day scale, nor for the other groups at any scale.

Given the small group sizes, and the slightly ‘fuzzy’ nature of the data due to normalisation and cleansing procedures, it is perhaps presumptuous to label this finding ‘significant’ in a practical sense, even if the statistical tests indicate that this is the case statistically. However, these results certainly merit further study, particularly because the SOLO group (recruited predominantly from within a housing association) are relatively less affluent overall compared to the DUET and TRIO groups. Their elasticity of demand for electricity might have been more responsive to the real-time price signals from the display equipment – especially when the display was newly installed in the household (see the discussion of longitudinal interest in the display equipment in the survey section).

4.4 Quantitative analysis of demographic data

As discussed in Section 2, VET trialists came from different economic and social backgrounds and live in different circumstances. Some of these differences, such as income levels, or size/ type of property may have had an impact on the relative volume of electricity that trialists used on a day-to-day basis, as well as the range of options available to trialists for household energy reduction. Therefore, it was important to explore similarities and differences between VET trialists with regard to key socio-demographic factors, and to assess the extent to which any observed differences in these areas might have influenced the findings on longitudinal consumption (see section 4.3).

Because the overall number of systems with usable consumption data for longitudinal analysis was relatively small, only socio-demographic factors where a more dispersed range of responses was observed were chosen for grouped statistical analysis. Statistical comparisons of mean standardised consumption by groupings for house type (i.e. detached, terraced, flat etc) income levels, and prior energy saving or environmentally friendly behaviour were performed at the 30 day scales described in the preceding section.

For all three factors, the null hypothesis under examination was: there is no statistically significant difference between socio-demographic or contextual groups and mean standardised consumption groups at 30-day scales. Kruskal Wallis (KW) tests were used to compare mean differences in consumption between groups for the first 30 days, and to compare mean differences in consumption within groups between the first 30 days and second 30 days of the trial.

4.4.1 Findings from demographic longitudinal analysis:

With a few exceptions where group sizes were particularly small, mean differences for all comparisons were observed to be very small (typically <1%) and no clear pattern of differences emerged between or within groups. None of the KW tests of mean differences proved significant. This probably reflects the small group sizes and correspondingly high variance.

4.5 Daily profile analysis of electricity consumption data

The electricity usage data from the mains sensor collected usage for each 15 minute period in the day. These records of 96 epoch values were analysed to search for patterns of daily energy usage behaviour both within and across systems.

Further cleansing had to be performed, removing days where at least one epoch value was 'NULL' or negative, or where all epoch values were '0'. Outlier systems were also removed, i.e. systems whose mains usage were far in excess of the average or almost negligible. Nevertheless, almost 40,000 days of daily profiles covering 157 systems were retained. None of the data were standardised or normalised as we were more interested in the pattern than absolute usage.

4.5.1 Findings from within system clustering:

Two Step clustering was applied to all the daily profiles for a single system in order to identify different daily usage patterns. The Two Step algorithm does not need to be told in advance how many clusters to look for.

The technique was relatively successful, in that it was able to determine clusters showing distinct patterns of behaviour representing weekends (at home), weekdays and also days when more energy was used in the evenings possibly reflecting mode of cooking.

4.5.2 Findings from across systems clustering:

When Two Step was applied to the entire set of daily profile data, it derived just two clusters with same pattern but differing in overall usage. K Means clustering was used with 10 clusters to force the creation of more clusters. Of the 8, 6 had a similar pattern of (weekday) usage, whilst two reflected much more daytime use. Further analysis cold determine the nature of these clusters, but at this stage it can be concluded that the segmentation of daily profiles both within and across systems can shed further light on how energy is used on a daily basis.

4.6 Quantitative analysis of appliance usage data

In addition to gathering mains sensor data by epoch, the TRIO and DUET systems were able to provide appliance usage data for each 15 minute epoch. The users were free to choose which appliances they monitored and subsequently 'labelled'. After further data cleansing, the database held 193,215 days of daily profiles covering over 30 different appliance types of 127 different users.

Analysis showed that the most common choices were: dishwasher, washing machine, fridge, freezer, kettle, tumble drier, computer system, TV & entertainment, oven/cooker, microwave and immersion heater, comprising around 93% of the data set.

4.6.1 Findings from appliance daily profiles:

The daily profiles for each of these 11 appliance types were analysed to identify different user behaviour (e.g. night-time use vs. daytime use for washing machines etc) and also different appliance characteristics (e.g. different cycle times and overall energy consumption for different fridges)

4.6.2 Findings from appliance recognition:

The final analysis carried out was to attempt to assign an appliance type to an unlabelled record of 96 epoch values. This was achieved by splitting the (most commonly used appliances) data set into three sets. One was used to train the classifier, the others to test the performance of the classifier.

Additional features were extracted for each day, such as total energy used, standard deviation, minimum epoch value, maximum epoch value, etc and appended to the record. Two classifiers were used; the tree induction algorithm C5 gave an accuracy of around 67% whilst the K Nearest Neighbour classifier gave an accuracy around 81%.

Interestingly, the confusion matrix (difficulty of classifier to distinguish between 2 given appliance types) showed distinct groupings of appliances with similar patterns (such as computer systems and TV entertainment, or dishwasher, washing machine, immersion and cooker)

5 Summary and conclusions

5.1 Proving the technology:

Whilst there were significant technical challenges, especially for advanced systems, once solved the displays proved generally effective and reliable.

The Powerline was reliable, had excellent range but its performance was limited by low data rates. ZigBee was more complex, less reliable but had more than adequate data rates. Access to IP networks proved problematic with considerable variation in consumer equipment and usage.

A number of initial design and manufacture issues hindered deployment of the SOLOs and DUETs but once the units were established they proved to be reliable. The TRIO, due to its complexity and use of off-the-shelf hardware suffered several set-up and reliability issues.

The low cost option of using SD cards to download data proved to be relatively unsuccessful with less than half the trialists engaging with this strategy. On the other hand, connecting to the web directly via the consumer’s router proved to be successful and very flexible even if the routers are regularly switched off.

5.2 Comparing the displays:

The displays were good, clear and effective. The simpler displays were very well received and the impact of the more complex trio on behaviours and actions was equally effective.

Both the SOLO and the DUET were very well received with their graphics and visibility (back lighting) praised highly. The TRIO was less well received due to the computer format, the complexity and poor reliability.

All three displays proved to be effective in what they were designed to achieve. The simplicity of the SOLO engaged people, the appliance details on the DUET helped people quickly understand their usage and triggered better behavioural responses. The level of detail on the TRIO helped users identify their high usage appliances and to analyse their consumption history resulting in improvements.

Real-time "push" information was central to maintaining interest, alerting users and triggering actions. Historic information was good for helping users really understand their itemised usage and to take more specific action such as identifying and replacing inefficient appliances. The speedometer, appliance details, tick and cross all scored highly whereas control of individual appliances, on-line data and heating information (on the DUET) were considered less effective. The interview data also produced surprising evidence that users were more interested in the 'fuel tank' aspect of the device than absolute current consumption, suggesting that they had set a level of consumption (a budget) and were happy so long as they were meeting this, in general. Indeed there was evidence of resistance to reducing this further. This suggests that there may be a need for (local and national) incentives to provide motivation for continual engagement.

Our experience with the trial also reveals that there is a balance between hand holding to get good results and leaving people to their own devices (which is more representative of what happens in practice). We deliberately ran a 'non-intervention' trial for this reason, and yet feedback from some participants suggested that they wanted and needed more help and assistance to achieve energy savings and understand the feedback from their displays. There is a need for further study into how best to help users of real time displays over the 'learning curve' – both in terms of display design/ functionality and provision of additional information/ learning material.

5.3 Identifying and quantifying the benefits of feedback:

The displays appear to have triggered wide-reaching actions beyond simple behaviour change. Increased understanding and visibility of energy use led to users purchasing energy efficient lights and appliances and exploring other steps such as insulation and micro-generation. Over time, the monitors show a tendency towards habituation. They become 'backgrounded' within everyday life and, as such, gradually lose their ability to encourage reductions in energy use. Indeed, the interview data suggest that after initial energy saving activities, the displays were then seen as helping householders to identify and maintain a 'normal' and satisfactory level of energy use rather than to continually reduce it.

Interview findings reveal that the devices are used by households rather than individuals per-se. This is true in several ways: a) where the device is located in the home has to relate to whole-house routines, b) the devices must match the wider household aesthetic or risk being 'tidied away', c) in shared homes individuals cannot make significant changes without first negotiating them with other household members.

Regrettably, it was not possible to determine actual savings achieved by the trial in kWH or CO2, due to technical problems leading to delayed and staggered start times for each system and also affecting the generation of good quality time series usage data. The lack of sufficient good quality data meant that many of the results found could not be shown to be statistically significant.

Nevertheless, by comparing relative use by each system and system type within the first 30 days of each system's contribution to the trial, we did find a statistically significant drop in usage for the groups with devices with displays compared to the Non-Display group. We are unable to confirm if this reduction was sustained and for how long, once again due to the lack of quality data. However, this would appear to support the evidence that users were interacting with the devices to reduce their energy consumption in this period. This finding is backed up by the results from the online surveys and the interviews.

The more practical benefit of this study has been understanding how people react to these products and what differences they make. Arguably, the displays themselves do not reduce peoples' consumption – it is the actions that they take in response to real time feedback (e.g. buying energy efficient appliances) that impact consumption patterns.

On a broader level, the interview data suggest very strongly that if domestic energy savings are to be encouraged through provision or use of real time displays, there is a need for wider support and action. In the perceived absence of significant action to save energy by government and business, householders felt that their efforts made little difference. These sorts of more qualitative, relational findings between technology, policy and practice, highlight the importance of understanding (and analysing) household energy use patterns in terms of more than just 'the numbers.' Our survey and interview findings provide a richness of data on how energy consumption *happens* that has, until now, been largely ignored in this particular research field. We feel that we have only just tapped into the potential of this level of analysis for contextualising everyday observed consumption patterns/ levels against everyday activities and lifestyles. More study in this area is definitely warranted.

5.4 Exploring the value of the data:

The cluster analysis was relatively successful in that it was able to determine clusters showing distinct patterns of behaviour by periods when houses were occupied, as well as days when more energy was used. By segmenting the mains sensor daily profile data (i.e. 96 epochs), we were able to distinguish different daily usage patterns within a single system, and also different usage patterns across all systems. This shows the potential for this approach to better understand user behaviour in terms of energy consumption and general lifestyle.

Appliance daily profile data highlighted differences in the way users used the particular appliance and also the variability of the appliance characteristics (e.g. daily total consumption of fridges showed a factor of almost 10 in range). Finally, when applying classification techniques to daily profile signals in order to 'classify' an unlabelled signal, we achieved over 80% accuracy (compared to a default 9% baseline accuracy). This is extremely encouraging and work is ongoing to improve the model.

5.5 Final conclusions:

Overall, the VET trial has produced some important new findings - particularly with regard to understanding how households react to, and interact with, advanced real-time energy display equipment. Although the results of the quantitative analysis were inconclusive (see the Appendix for recommendations on how to improve quantitative results in future studies), respondents to the surveys and interviews were generally positive about the devices and about how they had helped them to reduce their energy consumption. Perhaps the most impressive result was the widespread appliance switching behaviour observed across all of the display groups. It seems that the displays were particularly effective at helping trialists to identify energy-hungry appliances, which in turn prompted purchase of more efficient models when the need for replacement arose. Such behaviour creates ‘maintenance-free’ step changes in the energy intensity of everyday routines, and in this manner, real-time displays could have a lasting demand-side impact on domestic energy use that may also boost consumer demand for more efficient products – thereby creating a virtuous cycle that could reinforce existing policies on energy efficiency.

It is more difficult to decipher whether the displays were successful in prompting respondents to perform everyday routines more efficiently, or to engage in more direct conservation behaviour with regard to their ‘discretionary’ energy use. A key theme throughout the interview and survey data was the complexity of the household situations into which these devices enter, and associated differences in the meanings and interpretations that trialists attached to feedback from the displays. Apart from appliance switching, our research shows that the effectiveness of the devices for changing energy-use behaviour is not so much inherent in their technical functionality, as it is reflective of how users engage with, and apply that functionality to everyday routines and the energy used within them. A delicate balance appeared to be required between the devices’ effectiveness in prompting users to reduce consumption, and the frustrations, stresses, and limitations their users may feel in trying to do so when the displays were used in the context of a not-always-willing family, and the accustomed comforts of a modern lifestyle. For most of the interview respondents, there was a clear (but entirely personalised) line between what was considered discretionary and non-discretionary energy use. This finding not only highlights the importance of discretionary energy use from a behavioural and policy-based perspective, but also underscores one of the key conclusions from this research, in that the devices are only as good, or effective, as the broader contexts into which they are inserted and gradually domesticated. There is, therefore, a need for further research into the critical connection between the ways in which real time displays become part of the fabric of normal, everyday life, and the potential of these devices to shape perceptions of discretionary energy use in positive ways over the longer term. Adopting such a user-centred, socio-technical focus, will provide a clearer and more appropriate reference point for understanding how the displays ‘work’ in a household context, and thus the true potential of advanced real time displays as tool for both increasing awareness of energy consumption in everyday life, and reducing the carbon footprint of energy consumers.

Appendix: Considerations for future quantitative evaluations of energy reductions from real-time displays

In this Appendix, we make some recommendations for future quantitative assessments of energy savings from real time displays based on the lessons learned in this trial. This is intended to be used as a methodological guide for research practitioners and as a tool for policy-makers to use in judging the potential for success in further evaluations of real time displays or ‘smarter meters.’

Baseline data:

No electricity use data (in the form of bills or meter readings) were collected from trialists prior to the beginning of the trial. This means that all observed differences or changes in consumption were *relative to the trial itself.* Therefore, we could not determine the extent to which any of the trialists achieved absolute reductions (or increases) in electricity use compared to what was consumed before the trial began. This was a major shortcoming of this trial – both methodologically and in terms of our ability to assess the impact.

Ideally, baseline historical data could be collected over a period of several years to control for the effect of recession, changes in family circumstances and energy prices. However, at a minimum, baseline data should be collected for 1 year prior to a trial in order to create a reasonably robust, and seasonally reflective base for comparison of energy savings. In the absence of this baseline, we were forced to create a ‘virtual’ baseline using the mean for each period, or the gradient for a line of best fit. Both of these measures are highly sensitive to the variability of day-to-day electricity consumption, and neither provided a measurement of absolute savings resulting from the trial.

Control group:

The Non-Display group (without display) recruited for this trial was relatively small (n=~50) and not very representative of the majority of trialists in terms of socio-demographic factors and things like home ownership. It was not a ‘control’ in the true sense of the term, and this is why it was not labelled that way. The lack of consistent start and end dates (see below) added to this problem, as there were periods when there was no, or very little, overlapping consumption data between the Non-Display group and other groups.

Ideally, a fairly homogeneous population of trial participants could be randomly sorted into groups (including the Control) so as to promote consistency of mean volume of electricity consumption between trial groups (thereby decreasing the need to rely on standardisation by means) and to control for the impact of socio-demographic factors on energy use patterns. This was the intention during the recruitment phase of the trial. However, persistent recruitment difficulties necessitated that we accept a more diverse population of trialists in order to achieve greater numbers.

Accounting for disruptions to everyday life:

Our experience working with day-to-day energy use data has changed our understanding of ‘everyday energy use patterns.’ On a day-to-day or even weekly/ monthly scale, there is a high degree of variability or volatility observed in trialists’ energy use. There is some evidence of weekly-level periodicity (for instance, trialists tend to use more electricity on the weekends – presumably because they are not at work) but in general, it is difficult to see a 'pattern' at the weekly or daily scale *as long as everyday routines continue as normal.* Day-to-day energy use tends to look like a random heart rhythm or series of peaks and valleys. In some cases though, we observed cases where trialists' energy use spiked or 'flatlined' for short (i.e. 1 week) periods. These can seriously skew mean consumption levels for a group, particularly where group sizes are small. These deviations from 'natural' consumption levels (which did not meet criteria for cleansing) could be due to a number of factors such as increased occupancy, or going on holiday.

In order to create the best possible chance for identifying and isolating the effect of displays/ feedback on consumption, these sorts of disruptions to everyday life need to be accounted for (presumably they cannot be controlled for) where possible (perhaps by more regular telephone surveys or online questionnaires). Additionally, factors such as recession, job loss/ changes in salary, changes in electricity prices, or appliance switching behaviour should be monitored over the course of the trial and their impact on electricity patterns assessed. We did look for evidence of the appliance switching behaviour reported in the interim and online surveys, but without knowing which specific trialists had purchased these products (and what type), it was impossible to detect the impact of this behaviour.

Consistent start date and longer length:

As mentioned previously, it was originally intended that all systems for all groups would 'go live' on the same day. This would provide a consistent '0 point' for use in making longitudinal comparisons. However, delays in installation created significant differences in start dates (up to 10 months) between systems. This means that we had to rely on seasonal normalisation by degree-days to create comparable 'day 1-day [x]' comparison periods for all levels of analysis. Although seasonal/ weather normalisation will always be necessary in order to make fair comparisons between different time periods, the lack of a consistent start date also meant that we have to trust the normalisation for making comparisons within the same time periods (i.e. comparing standardised mean energy use between groups of trialists in the first 30 days of the trial). This adds an additional level of uncertainty to our analysis/ findings that could have been avoided with a consistent start date. More generally, we wish to note that because degree days offer at best only rough approximation of any individual's actual energy use patterns, it is advisable to avoid relying on them wherever possible.

Per our discussion of baseline data above, the ideal length for a trial of this sort would be at least 365 days (matched to the baseline period if possible). This would provide a measure of control for seasonal variations in energy. Rough comparisons could be made of yearly means (taking into account factors such as recession or loss of income etc as discussed previously) without having to normalise for seasonality. A longer trial would also allow for longitudinal comparison at longer time scales, which could increase the likelihood of finding significant mean differences.