



Solent Acheiving Value from Efficiency (SAVE): household survey data archive report

Tom Rushby (t.w.rushby@soton.ac.uk*[†]@tom_rushby)
Ben Anderson (b.anderson@soton.ac.uk[†]@dataknut)
Patrick James (p.a.james@soton.ac.uk[†]@pab_james)
AbuBakr Bahaj (a.s.bahaj@soton.ac.uk[§])

Last run at: 2019-10-09 10:24:33

Contents

| | | |
|----------|--|-----------|
| 1 | About | 3 |
| 1.1 | Citation | 3 |
| 1.2 | Circulation | 3 |
| 2 | Introduction | 3 |
| 3 | Data available | 4 |
| 3.1 | Household response file | 4 |
| 3.2 | Household recruitment survey | 5 |
| 3.3 | Household survey updates | 6 |
| 3.4 | Time-use diary survey | 7 |
| 4 | Support | 8 |
| 5 | Output Files | 9 |
| 6 | References | 10 |

Document history**Table 1: Document history**

| Version | Date | Author | Notes |
|-------------|------------------|-----------|--|
| Version 0-1 | 25th July 2019 | T. Rushby | Initial preliminary run. |
| Version 0-2 | 29th July 2019 | T. Rushby | Added household response file data. |
| Version 0-3 | 9th October 2019 | T. Rushby | Updated files, cleaning and documentation. |

1 About

This work was carried out by the Sustainable Energy Research Group¹ in the University of Southampton's Faculty of Engineering and Environment².

The work was funded by the Low Carbon Network Fund (LCNF) through the Solent Achieving Value from Efficiency³ project and is (c) 2019 the University of Southampton.

This report forms part of the documentation supporting the deposit with UK Data Archive of data collected under the SAVE project. Further details of the project, including documentation of the interventions tested, can be found on the project website⁴.

1.1 Citation

If you wish to refer to any of the material from this report please cite as:

- Rushby, T., Anderson, B., James, P.A.B. and Bahaj, A.S. (2019) Solent Achieving Value from Efficiency (SAVE): household survey data archive report, University of Southampton: Southampton, UK.

1.2 Circulation

This report is public, to accompany the data release associated with the Low Carbon Network funded (LCNF) Solent Achieving Value from Efficiency (SAVE) project.

2 Introduction

This document describes the household data collected by BMG Research Ltd. during the SAVE project. It then describes the subset of this data that has been deposited at UK Data Archive⁵ and made available for research purposes.

The Solent Achieving Value from Efficiency (SAVE) was funded by the Low Carbon Network (LCN) Fund and sought to establish to what extent energy efficiency measures and behaviour change programmes can be considered a cost effective tool for managing peak demand, and whether they provide an alternative to traditional low-voltage network reinforcement.

The SAVE sample recruited a stratified random address-based sample of households from:

- the County of Hampshire
- the City of Southampton
- the City of Portsmouth
- the Isle of Wight

¹<http://www.energy.soton.ac.uk>

²<http://www.southampton.ac.uk/engineering/index.page>

³<http://www.energy.soton.ac.uk/save-solent-achieving-value-from-efficiency/>

⁴<http://www.save-project.co.uk/>

⁵[data%20link](#)

With some cautions this means that the sample is representative of households across these areas – but not of the UK in general.

The fieldwork initially recruited over 4,000 households in late 2016 but due to attrition this slowly declined. Efforts were taken to refresh and re-recruit new households to replace those that left.

These households were then randomly allocated to one control and three intervention groups. These were then used to test a number of methods designed to reduce demand for electricity during the network peak hours between 4 and 8pm.

3 Data available

The following subsections describe elements of the data collected from households participating in the SAVE project. These are:

- Household response file: a file containing a record for each address that received a SAVE contact letter. Not all households responded;
- Household recruitment survey: a file containing responses to a household survey conducted either at (face to face) or shortly after recruitment;
- Household survey updates: periodic survey updates with recruited households;
- Time-use diary: a file containing responses to a set of time-use diaries conducted during the SAVE trials;

In addition we have created a file containing details of the coding used for responses to the questions in the household surveys as follows:

- Household survey key: contains key to responses for recruitment and update surveys, and time-use diaries.

A list of file names is provided in Section 5.

All household data is linked using the variable `BMG_ID`, a 9-character numeric variable containing the unique household identifier. This identifier can also be matched to the 15-minute electricity consumption data using the **last 9 characters** from the `navetasID` variable.

3.1 Household response file

The outcomes of the recruitment process were recorded and are provided in the file `SAVE_Response_Data.csv`. Note that information used to stratify the sample, Census Output Area (OA), has been removed from the data as this is potentially disclosive and could be used to identify individual households. Variables containing larger geographical units are provided within the household recruitment survey (see Section 3.2).

Table 2 shows the first few rows of this file.

Table 2: First few rows of response file data

| BMG_ID | Intervention | Outcome_Label | Opt Out Date | wgtNonResponse |
|-----------|--------------|---|--------------|----------------|
| 956600002 | 2 | Refused - Not interested/had enough | NA | 0.7598767 |
| 956600005 | 2 | Refused - Not interested/had enough | NA | 0.6554543 |
| 956600007 | 4 | No response (SAVE 400_Install) | NA | 0.6554543 |
| 956600009 | 3 | Refused to do install / Survey (350 new loop) | NA | 0.7162431 |
| 956600012 | 4 | Respondent agrees - new participant contact | NA | 0.7598767 |
| 956600017 | 4 | Respondent agrees - new participant contact | NA | 0.7598767 |

Description of variables included in the response file:

- `BMG_ID` 9 digit unique household identifier;
- `Intervention` single-digit integer describing the trial group (TG) to which the household was randomly allocated;
- `Outcome_Label` provides the outcome for each household contacted during the recruitment process;
- `Opt Out Date` provides, where applicable, the date at which a participant withdrew from the study;
- `wgtNonResponse` non-response weights calculated by BMG research.

3.2 Household recruitment survey

Upon consenting to take part in the research, households were asked to complete a 'recruitment' survey. The first few rows of the data are shown in Table 3. Non-response weights were calculated by BMG research and are provided by the `wgtRakeAgeTenure` variable.

The survey data is provided in the file `SAVE_Household_Survey_Data.csv`. A key to the responses to the recruitment survey is provided within the accompanying file named `SAVE_Household_Survey_Key.csv`.

Table 3: First few observations in household survey data: first 8 variables

| BMG_ID | Intervention | surveyMode | InterviewDate | wgtRakeAgeTenure | Q2 | Q2B | Q2C |
|-----------|--------------|-------------|---------------|------------------|----|-----|-----|
| 956600023 | 2 | CAPI Survey | 2018-08-23 | 0.7337819 | 1 | 3 | 1 |
| 956600025 | 3 | CAPI Survey | 2018-09-06 | 0.9759450 | 2 | 7 | 2 |
| 956600047 | 2 | CAPI Survey | 2018-09-29 | 0.6992979 | 2 | 7 | 1 |
| 956600048 | 1 | CAPI Survey | 2018-09-29 | 1.6202692 | 2 | 3 | 2 |
| 956600076 | 1 | CAPI Survey | 2018-09-29 | 0.6992979 | 3 | 8 | 1 |
| 956610255 | 3 | CAPI Survey | 2018-09-22 | 0.9219489 | 5 | 4 | 2 |

Description of variables not included in the key to survey responses:

- `BMG_ID` 9 digit unique household identifier;
- `Intervention` single-digit integer describing the trial group (TG) to which the household was randomly allocated;
- `surveyMode` the mode through which the participant completed the household survey;
- `interviewDate` date the household survey was completed;
- `wgtRakeAgeTenure` weighting allocated to household (weights modelled using age and housing tenure variables).

Variables relating to geography:

- `la_Name` Local Authority within which the household is located
- `gor_Name` Government Office Region (GOR), now referred to as Regions⁶
- `UrbanRural_Name` urban-rural classification of area
- `UrbanRural_SubName` urban-rural sub-classification of area

Index of Multiple Deprivation (IMD) was used to stratify the sample:

- Index of Multiple Deprivation (IMD) Score measure of deprivation for output area within which the household is located;
- quartile, quintile and decile provide frequency-based groupings based on IMD score.

IMD scores have been removed from the data provided as this can reveal the Census Output Area and potentially be used to identify individual households. IMD quartile is retained for analysis purposes as it is considered a low disclosure risk.

The raw data contains all households contacted during the recruitment (and appearing in the response sheet) but only a subset agreed to participate and completed the recruitment survey first. Table 4 shows the number of recruitment surveys completed by mode.

Table 4: Household surveys completed by mode

| Survey mode | Number completed |
|-------------|------------------|
| CAPI Survey | 871 |
| CATI Survey | 2200 |
| CAWI Survey | 1657 |
| NA | 37742 |

The raw survey data contains all households that were in the response sheet, providing:

- 311 variables
- 42,470 rows of data

Only those households that completed the survey (where `surveyMode` does not equal NA) are retained leaving:

- 4,728 unique household surveys

XXX add Section detailing the recoding of variables to mitigate disclosure risk XXX

3.3 Household survey updates

During the project, update surveys were completed at 12-monthly intervals to ensure that the household data collected using the 'recruitment' survey (and any key household attributes used in any analysis) was as accurate as possible.

Update surveys were provided by the fieldwork contractor in three batches associated with the three trial periods. 61 variables are updated in these surveys including: details of each occupant (age, gender, work status), electricity supply, smart meter installation, and electric vehicle ownership and charging.

⁶See <http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/administrative/england/government-office-regions/index.html>

The update survey included a new question: Q3_3 *How is electricity provided to your home?* with multiple responses for renewable generation (i.e. solar pv, wind turbine, hydro generator, CHP etc).

Again, the raw data for each update contains all households in the response sheet resulting in many empty rows, therefore those not completing an update are filtered out in the UKDA data file leaving just those who were actually updated as shown in Table 5.

Table 5: Update surveys completed with date ranges

| Update period | Number completed | First date | Last date |
|---------------|------------------|------------|------------|
| TP1 | 1825 | 2017-01-19 | 2017-09-22 |
| TP2 | 1592 | 2017-10-04 | 2018-09-13 |
| TP3 | 353 | 2018-11-07 | 2019-01-18 |

A large proportion of households completed more than one update as shown in Table 6.

Table 6: Summary of update surveys by number completed

| Surveys completed | Number of households |
|-------------------|----------------------|
| 1 | 1336 |
| 2 | 1217 |

3.4 Time-use diary survey

Time-use diaries were conducted during each of the trial periods.

- For the first trial period (TP1), the diaries were conducted for a range of dates with a large number targeted for the 'event day' on 15th March 2017;
- For the second trial period (TP2), the diaries were completed for the event on 20th March 2018;
- For the third trial period (TP3), the diaries were completed for the event on 13th December 2018.

The date (ranges) for each trial period are shown in Table 7

The diaries were provided by BMG Research Ltd. in an excel spreadsheet with diaries for each trial period on individual sheets. The data has been combined into one .csv file and a new variable `diaryPeriod` added to specify which trial period the diaries are associated with. The data is provided in the accompanying file: `SAVE_Time_Use_Data.csv`.

The time-use diary key can be found within the household survey key file (`SAVE_Household_Survey_Key.csv`).

Table 7: Update surveys completed with date ranges

| Diary period | No. of households | No. of respondents | First date | Last date |
|--------------|-------------------|--------------------|------------|------------|
| TP1 | 1307 | 1858 | 2017-01-18 | 2017-03-15 |
| TP2 | 882 | 882 | 2018-03-20 | 2018-03-20 |
| TP3 | 738 | 738 | 2018-12-13 | 2018-12-13 |

There is a difference between number of respondents (unique `respid`) and number of households (unique `BMG_ID`). This is because a number of time-use diaries were completed by more than one person in each household.

Table 8: Summary of time-use diaries completed by trial period and number of respondents

| Trial period | No. of respondents | No. of households |
|--------------|--------------------|-------------------|
| TP1 | 1 | 756 |
| TP1 | 2 | 551 |
| TP2 | 1 | 882 |
| TP3 | 1 | 738 |

Example rows from the time-use data are shown in Table 9. The trial period column has been omitted for clarity.

Table 9: Example observations from time-use diary data

| respid | BMG_ID | InterviewDate | diaryDate | Actloop | Q1A | Q2A | Q3 | Q4_Hour | Q4_Min |
|--------|-----------|---------------|------------|---------|-----|-----|----|---------|--------|
| 1 | 956600053 | 2017-01-27 | 2017-01-26 | 1 | 1 | NA | 1 | 5 | 1 |
| 1 | 956600053 | 2017-01-27 | 2017-01-26 | 2 | 6 | 5 | 1 | 5 | 6 |
| 1 | 956600053 | 2017-01-27 | 2017-01-26 | 3 | 4 | NA | 1 | 6 | 2 |
| 1 | 956600053 | 2017-01-27 | 2017-01-26 | 4 | 4 | NA | 6 | 13 | 31 |
| 1 | 956600053 | 2017-01-27 | 2017-01-26 | 5 | 6 | 5 | 1 | 13 | 36 |
| 1 | 956600053 | 2017-01-27 | 2017-01-26 | 6 | 4 | NA | 1 | 14 | 31 |

4 Support

A repository has been set up through the University of Southampton GitLab service to support the SAVE project data deposit at <https://git.soton.ac.uk/SERG/save-public>. Newly identified issues with the data will be logged by the project team and updated through the repository. New issues can be raised at <https://git.soton.ac.uk/SERG/save-public/issues> (registration required) and we will endeavour to respond although note that SAVE project funding has now ended and team members have moved on to other projects.

5 Output Files

The full list of files exported for the archive are presented below:

```
## [1] "SAVE_Household_Survey_Data_v0-3.csv"
## [2] "SAVE_Household_Survey_Key_v0-3.csv"
## [3] "SAVE_Household_Survey_Updates_Data_v0-3.csv"
## [4] "SAVE_Response_Data_v0-3.csv"
## [5] "SAVE_Time_Use_Data_v0-3.csv"
```

Runtime

Report generated using knitr⁷ in RStudio⁸ with R version 3.6.1 (2019-07-05) running on x86_64-apple-darwin15.6.0.

Analysis completed in 26.142 seconds (0.44 minutes).

⁷<https://cran.r-project.org/package=knitr>

⁸<http://www.rstudio.com>

6 References