

TECHNICAL REPORT

Green Deal Segmentation research

Undertaken by GfK NOP and Kantar Media Research

November 2012

Contents

1. Introduction	4
2. Target Audience	4
3. Research method	5
3.1 Segmentation overview	5
4. Sample Selection	6
5. Questionnaire Design	7
5.1. Conjoint section	9
5.2 Segmentation inputs to the questionnaire	9
5.3 Piloting	11
6. Fieldwork	12
7. Initial analysis	14
7.1 Coding	14
7.2 Data tables	14
8. Segmentation analysis	16
8.1 Selection and refinement of input variables	18
8.2 Factor analysis of scale variables	20
8.3 Other variables	22
8.4 Uptake propensity	23
8.5 Segmentation method	24
8.6 Segmentation diagnostics	24
8.7 Segment size	25
8.8 Efficiency	25
8.9 Stability	26
8.10 Recreating the 6 segments in future research	27

8.11 Ability to recreate the segments in the TGI database 28

8.12 Profiling the segments..... 30

1. Introduction

In order to help reduce carbon emissions and improve domestic energy efficiency in Great Britain the Government is launching the Green Deal; an initiative which will enable households and businesses to make energy saving home improvements to their properties. The scheme helps customers pay for some or all of the improvements over time, through their electricity bill. Repayments will be no more than what a typical household should save in energy costs.

GfK NOP was commissioned in early 2012 to conduct research into the Green Deal. The key objective of the research was to understand which key groups exist in relation to potential demand for the Green Deal, what their motivations and barriers are, and to understand which communications channels might best be used to reach them.

This report presents the findings of this study. The aim was to provide clear guidance on which key groups exist in relation to potential demand for the Green Deal, and how they differ in terms of:

- Attitudes to home and to the environment
- Motivations and barriers to uptake of energy saving home improvements
- Communications channel usage
- Demographic, geographic and situational characteristics.

2. Target Audience

The main target audience for the research was households in the following categories :

- **Owner occupiers:** whether owned outright or with a mortgage (including those on shared-ownership schemes)
- **Private renters.**

It is important to note that social housing tenants (Council or Housing Association) were excluded from the research because of complexities related to their tenure.

Newer homes (post-1980) were included in the research since it was reasoned that although new-build properties do tend to have a number of energy efficiency measures installed, not all include all of the potential measures included in the Green Deal, and to exclude them systematically would lead to a biased sample.

Interviews were conducted with the person wholly or jointly responsible for financial decision making in the household (i.e. the person who owned or rented their property, or their partner).

3. Research method

All interviews were conducted face to face in respondents' homes. Interviews were conducted using Computer Assisted Personal Interviewing (CAPI) which means that the interviewers carried a laptop which controlled the questionnaire, the order questions were asked and routing, based on answers previously given. This was particularly important given the complexity of the questionnaire.

A conjoint task was also included in the questionnaire. This element of the research enabled the variable on 'uptake propensity' to be generated. This variable indicates the relative differences in interest in the Green Deal across the segments. The conjoint task was also used to inform a separate piece of research on Incentives, which is covered in a separate report.

3.1 Segmentation overview

The segmentation was required to identify groups which differentiated in terms of potential interest in the Green Deal, and to understand how they differ in terms of:

- Attitudes to home and to the environment
- Attitudes to 'selling points' around energy efficiency and Green Deal
- Motivations and barriers to uptake
- Communications channel preferences
- Demographic, geographic and situational characteristics (e.g. lifestage, life events).

The analytical technique used to identify and profile these target groups is called segmentation.

Segmentation is a statistical exercise which aims to divide a population into distinct groups which have similar response patterns within segments and different patterns between segments. It works by calculating the 'distances' between responses across respondents and identifying groups of respondents whose responses are 'close' to each other. These groups of respondents are known as a cluster and the process aims to find an optimal number of clusters, within which responses are similar but across which they are different and distinct.

Because of limitations on interview length, it was not possible to ask very detailed questions about media use and communications channel preferences. The decision was therefore taken to make use of information contained in the TGI database¹ which is owned by Kantar Media. The segments generated from the segmentation were allocated into the TGI database to enable further analysis of the segments' detailed media preferences, as well as other lifestyle information held on the database. Specifically, this allowed the segments to be analysed by the following:

¹ TGI (Target Group Index) is a market leading single-source consumer survey available in over 65 markets globally. TGI's main strength lies in the breadth of its coverage, enabling in-depth consumer understanding in terms of demographics, attitudes, media consumption and product and brand use across 18 industry sectors. Established in 1969, it delivers large annual sample of c25,000 adults in Great Britain, and enjoys currency status in the media industry: most media agencies and owners use the data to buy and sell advertising. TGI is also widely used in the public sector where clients include DH, HMRC, NEST and COI.

- Key attitudes and circumstances – including the importance of family, career, the environment and technology, personal outlooks (e.g. open-mindedness, enjoyment, stress) and personal circumstances (e.g. convenience/time poor)
- Media consumption – including TV/radio channels and programmes watched, newspaper readership, plus other media consumption including online, magazine, cinema and exposure to outdoor advertising
- Online habits – including types of activities conducted online, frequency of usage and mode of access
- Finance and utilities – including financial holdings, financial engagement, household utilities and upcoming life events
- Trusted sources / sources of advice – whether and where advice is sought (e.g. when making purchases, financial decisions, discussing health issues)
- Community involvement – including charitable giving, volunteering, participation in/membership of organisations and the importance of faith.

In order to enable the segments to be allocated into the TGI database, it was necessary to ensure that there were sufficient common questions across the two datasets to efficiently ‘hook’ them together. During the questionnaire design stage of the project, GfK NOP worked with Kantar to identify appropriate ‘hook variables’ for inclusion in the Green Deal survey questionnaire and these are described later in this report.

4. Sample Selection

The sample was drawn using random location sampling methods: this was considered the most appropriate method for the survey because it offered reasonably robust sampling at a lower cost and more quickly than a random probability method.

A total of 2,050 interviews were conducted, with boost samples conducted in Wales and Scotland to ensure that at least 200 interviews were conducted in each country. This was considered the minimum sample size needed to enable robust conjoint analysis. The boost samples were weighted at the analysis stage to provide a representative sample of the target population in Britain (see section 7.2 for a description of the weighting process).

Fieldwork was completed across 186 sampling points in Britain; 148 in England and 19 in each of Scotland and Wales. Each sampling point took the form of two Output Areas² (OAs) within the same Parliamentary constituency, and interviewers were instructed to work for two days in each paired OA, with the aim of achieving an average of 11 interviews per sampling point.

The sample was drawn in the following stages:

1. **Selection of constituencies:** all constituencies in Great Britain were stratified by urban/rural indicator, social grade and size. A total of 186 constituencies were then selected with probability of selection proportional to the number of residents in each.

² An Output Area (OA) is the smallest area for which detailed 2001 Census results are available. OAs were created specifically for statistical purposes on the basis of data from the 2001 Census. OAs contain an average of 125 households and around 300 residents: the minimum size is 100 residents or 40 households. Where possible, OA boundaries were drawn to contain populations with homogenous characteristics, and around small, free-standing settlements. For more information on Output Area geography, please see http://www.statistics.gov.uk/census2001/geo_methods.asp

2. **Selection of OAs from the chosen constituencies.** OAs were selected at random, following stratification by age, gender, social class, and geodemographic profile (Mosaic classification) within each constituency. OAs with more than 30% social housing were removed from possible selection at this stage for fieldwork efficiency reasons³. Once drawn, the profile of the selected OAs was checked against the national profile to ensure that it was representative by the key variables noted above.

Once the initial selection had taken place, the process was repeated to find a second (or paired) OA in the same ward as the original one so that interviewers had two OAs to work in fairly close to one another. An equal number of substitute points were selected at the same time to be used if any of the original number were ineligible for any reason (e.g. inaccessible gated communities, military housing within closed bases).

3. **Selection of addresses within each sampling point:** Interviewers were provided with lists of addresses which they could approach for interview, and they recruited respondents to quota.

Quotas were set on gender and working status interlocked (Men working full-time, men not working full-time, women working, women not working) and age. Quotas varied from sampling point to sampling point to reflect the profile of the area in which interviewers were working.

5. Questionnaire Design

The questionnaire was designed by GfK NOP in close collaboration with DECC. It used tried and tested questions from surveys such as the 2011 Green Deal survey of consumers, English Housing Survey (EHS), the Low Carbon Communities Challenge (LCCC) survey and the Home Energy Use (HEU) survey. Where necessary, demographic questions were taken from the ONS Harmonised Concepts for Social Surveys.⁴

Because of the need to create a segmentation and to link this to the TGI database, it was necessary to include a number of questions from TGI to act as ‘hooks’ when the survey data was fused. The inclusion of these questions would enable further detailed analysis of media usage, lifestyles and attitudes and behaviours in the TGI database.

The basic questionnaire structure is shown over the page and a copy of the questionnaire is included in Appendix 1.

³ By excluding OAs containing more than 30% social housing, only 1.3% of all private housing stock in Britain was excluded from the survey.

⁴ <http://www.ons.gov.uk/about-statistics/harmonisation/primary-concepts-and-questions/index.html>

Figure 1: Green Deal questionnaire



5.1. Conjoint section

The conjoint section of the questionnaire was used to help determine the relative differences in potential uptake of the Green Deal. Respondents evaluated measures which they 'needed' (i.e. those which were appropriate for their home and were not already installed). Measures which were shown included:

- Solid wall insulation (internal and external)
- Cavity wall insulation
- Loft insulation
- Boiler installation or upgrade.

The structure of the conjoint section is summarised below

Figure 2: Structure of the conjoint section

Establish 'need'	'Needs' were assessed by asking the householder about their property, not by a property surveyor or other professional
Prioritise measures presented	The CAPI script prioritised which measures respondents were presented with, based on a number of rules (as detailed below)
Inform about the measures	Respondents were given information about the measures to enable them to engage with them in the conjoint exercise
Inform about aspects of the package	Respondents were told about the Green Deal package (detailed below)
Conduct trade off exercise	Respondents were presented with a number of packages and asked which they preferred. They completed 10 iterations of this task, each of which showed 2-6 permutations
Conduct calibration exercise	Respondents were presented with 5 packages, and asked how likely they would be to take up each

5.2 Segmentation inputs to the questionnaire

Certain questions needed to be added to the questionnaire as they were identified as either potential segmentation input variables or TGI 'hooks'. The segmentation inputs that were included were as follows:

B11 Thinking about your home in the winter, how easy or difficult is it to keep your home warm when the heating is on? (Very easy – very difficult)

B12 Which one of these best describes how well you and your household are keeping up with your energy bills at the moment? (I/we manage very well - I/we have severe difficulties)

C1 Thinking about your home, how much do you agree or disagree with these things that other people have said? (Agree strongly – disagree strongly)

- I'm the type of person who likes to have the newest gadgets in my home
- I feel very comfortable dealing with financial matters
- It's only worth doing environmentally friendly things if they save you money (TGI hook)
- It's not worth me doing things to help the environment if others don't do the same (TGI hook)
- I'm always looking out for new ideas to improve my home (TGI hook)
- I only take out credit/loans when absolutely necessary (TGI hook)

C2 How often, if at all, do you personally do any of the following? (Always – never)

- Leave your TV or PC on standby for long periods of time (TGI hook)
- Keep a tap running while you brush your teeth (TGI hook)
- When buying new appliances (e.g. fridge, washing machine) choose those which are more energy efficient

D1 Which, if any, of these are you realistically planning to do to your home in the next 12 months?

D3 D3 Which of these reasons would be important to you in making your home more energy efficient?

D4 And which of these things prevent you from doing more to make your home more energy efficient?

In addition, the following questions were used as TGI 'hooks':

- Age
- Sex
- Working status
- Tenure
- Property type
- Newspaper readership
- Frequency of internet use
- Number of adults
- Number of children
- Social grade
- Household income.

Section 8.1 of this report discusses the segmentation input variables in greater detail.

5.3 Piloting

It is worth noting that the questionnaire was not subject to a formal piloting stage. This was primarily a function of timings, but it was decided that a pilot was unnecessary for a number of other reasons:

- The use of validated questions from a variety of sources meant that the questionnaire made best use of 'tried and tested' question wording
- Prior work on the 2011 Green Deal survey of consumers meant that it was already understood how best to communicate technical detail about the scheme
- The questionnaire was subject to internal expert peer review by researchers unconnected to the research
- A small number of interviewers were accompanied at the start of fieldwork to ensure that the questionnaire was working properly and that respondents understood the task they were asked to undertake.

6. Fieldwork

Fieldwork was conducted between 15th February and 23rd March 2012.

To familiarise interviewers with the background to and objectives of the research, written briefing instructions including details of the Green Deal measures were sent to all interviewers working on the survey. Interviewers were instructed to familiarise themselves with the measures and their applicability to properties, and were invited to contact members of the research team if they had any queries specific to the survey or subject matter.

In completing fieldwork, interviewers followed the usual rules to maintain fieldwork quality. Which means that they:

- Only completed one interview per household
- Completed no more than 4 interviews in any one road, etc.
- Completed no interviews with people known to the interviewer
- Registered at a local police station before starting work, to enable them to provide reassurance to respondents if needed.

The survey introduction was carefully worded to encourage as wide a range of potential respondents as possible to take part. The introduction was specifically worded to ensure that it did not put off people who were not interested in 'green' issues, and respondents were not told that the survey was on behalf of DECC until the end of the interview if they asked.

The survey introduction is shown in the box below.

452804	<u>CONTACT SCREENER CARD</u>	Feb 2012
Hello, my name is _____ from GfK NOP, an independent research company.		
We are conducting a survey about people's homes on behalf of a government department. Your views will help to ensure that the government designs better policies that meet people's needs. We would really value your opinions.		
INTERVIEWER: ADD IF NECESSARY		
<ul style="list-style-type: none">◆ The questions will take around 30 minutes to answer.◆ Your name and individual details will remain confidential to the research company and will not be revealed to anyone else without your permission◆ Your answers will be combined with other peoples who complete the survey and will not be linked to your name or address without your consent◆ The interview is being conducted on behalf of a Government department which has responsibility for some aspects of policy related to people's homes.		
INTERVIEWER: SHOW ID		

A Contact Screener Card was provided which included all questions required to enable interviewers to establish eligibility.

In Wales, respondents were given the option of conducting the interview in Welsh, though no respondents took up this option.

Upon concluding the interview all respondents were handed a thank you leaflet, which contained contact details for GfK NOP in case there were any later queries. A large print version of this leaflet was also available.

As previously mentioned, interviewers were set quotas which were specific to the area in which they were working. Quotas were based upon profile information from the 2001 Census.

Table 1 shows the overall quotas that were set along with the number of interviews achieved in each cell. While the working status and country quotas achieved were broadly in line with those set the table shows that younger respondents were under-represented while older respondents were over-represented. These were corrected by weighting the data to the known profile of owner-occupiers and private tenants in Great Britain. The quotas were set to include a 5% 'overage', which would allow for a shortfall in the number of interviews in any area. The target number of interviews to be conducted was 2,000, and a total of 2,050 interviews were delivered.

Table 1: Quotas (set and achieved)

	Quotas set		Interviews achieved	
	N	%	N	%
Aged 18-34	584	29	371	18
Aged 35-54	777	39	743	36
Aged 55+	684	34	936	46
Men Working	624	31	582	28
Men Not Working	375	19	428	21
Women Working	595	30	528	26
Women Not Working	452	23	496	24
England	1,600	80	1,625	79
Scotland	200	10	218	11
Wales	200	10	207	10
TOTAL	2,000	100%	2,050	100%

7. Initial analysis

7.1 Coding

The final questionnaire contained two open-ended questions and 14 questions including 'other' answers which required coding. In order to get the most out of these open responses codeframes were developed by executives working on the project with reference to the objectives of the question. These codeframes were checked at a senior level before being sent to DECC to be signed off.

GfK NOP's team of experienced in-house coders were fully briefed by project executives in advance of starting work and the briefing included the objectives of each individual question, and other relevant material.

7.2 Data tables

Data tabulations were run to a specification agreed with DECC and included information about which key sub-groups to include as crossbreaks (e.g. sex, age, working status, property information, whether measures were already installed etc.)

In addition to standard demographic variables, a number of bespoke variables were created specifically for this research. These included:

- Super Priority Group – For the purposes of this research, this group was defined as those households who fit into at least one of the following three groups:
 - Receive JSA or Income support and have a household income below £16,999 and have someone aged 60+ or receive Disability Living Allowance or have children in the household or someone who has a long standing illness or disability
 - Receive a pension credit
 - Receive Child Tax Credits and have a household income below £16,999.
- Consideration group – A sub-group comprising:
 - “Need measures and have considered” – this group was made up of those who had considered installing at least one energy efficient measure in the last 12 months
 - “Need measures but have no plans to install” – this group was made up of those who ‘needed’ at least one energy efficient measure but had no plans to install any
 - “Need measures but not aware” – this groups was made up of those who ‘needed’ at least one measure but were not aware of measures they could benefit from
 - “Already installed measures” – this group was made up of those who had already installed all of the energy efficient measures which their property could benefit from
- Green Deal segments – This cross-break consists of the 6 groups that were developed as a result of the segmentation exercise

The crossbreaks layout and definitions are shown in Appendix 2 of this report.

Interim tables were run while fieldwork was still ongoing in order to check the data, make any amendments ahead of the final dataset and to make the final delivery of data more efficient upon conclusion of interviewing.

Once data tables had been fully checked by project executives the achieved sample was weighted to the known profile of owner-occupiers and private renters in Great Britain.⁵ The table below shows both the unweighted and weighted sample profiles of the cells which were weighted.

Table 2: Final unweighted and weighted sample profile

	Unweighted profile		Weighted profile	
	N	%	N	%
Aged 18-34	371	18	584	28
Aged 35-54	743	36	768	37
Aged 55+	936	46	698	34
Men Working	582	28	694	34
Men Not Working	428	21	298	15
Women Working	528	26	583	28
Women Not Working	496	24	462	23
White	1,867	91	1,841	90
BME	174	8	200	10
England	1,625	79	1,790	87
North	454	22	491	24
Midlands	380	19	545	27
South	791	39	754	37
Scotland	218	11	151	7
Wales	207	10	109	5
Terraced	570	28	584	28
Semi-detached house	706	34	615	30
Detached house	585	29	553	27
Bungalow	207	10	182	9
Flat	158	8	213	10

The effective sample size was calculated. This describes the effect of the weighting on the accuracy of survey estimates. The effective sample size is dependent upon the size of weights applied to respondents: the more the weights deviate from 1, the smaller the effective sample

⁵ Source: 2001 Census

size and the less accurate estimates will be. As would be expected, the majority of weights ranged between 0.5 and 1.5 and the lowest weight was 0.3 while the highest weight was 4.2.

The effective sample size for this survey was 1,670, or 82% of the interviewed sample size. This is usual for a quota based survey of this type. Much of the impact of the weighting on the effective sample size resulted from weighting by home nation: that is downweighting the boost interviews in Scotland and Wales to their natural proportions.

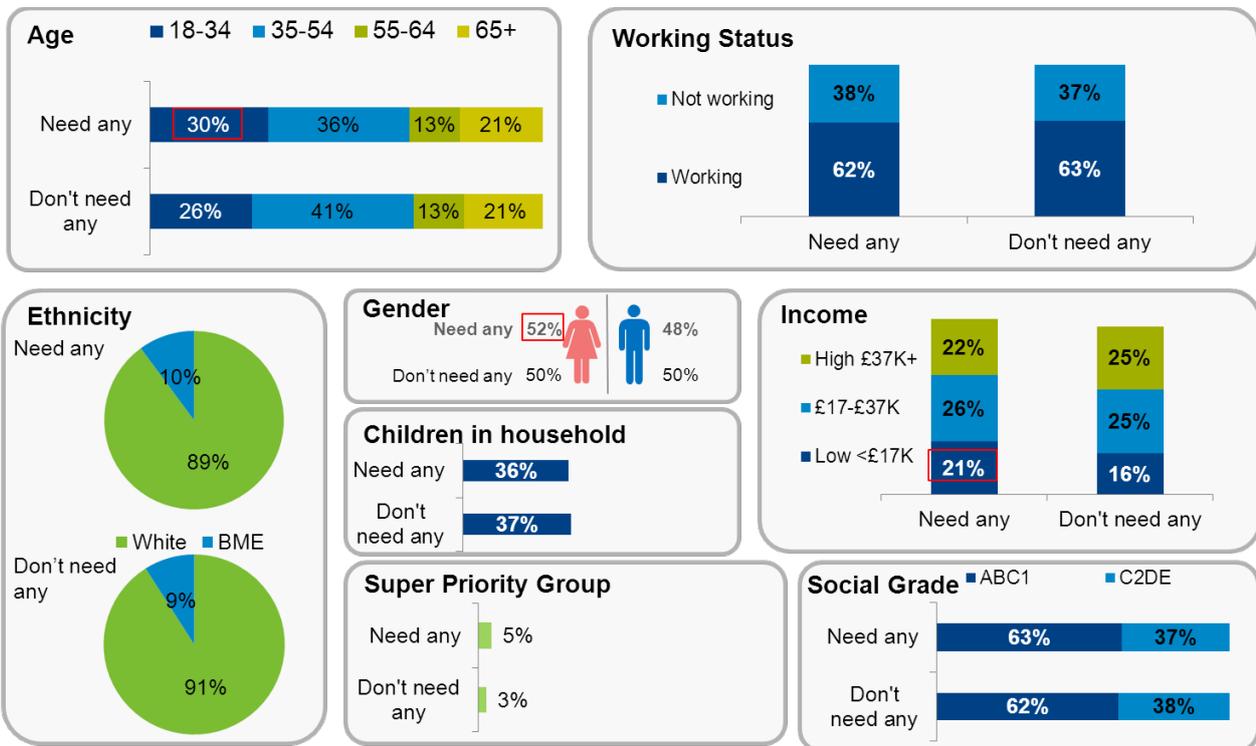
8. Segmentation analysis

The first segmentation solution was created amongst all respondents; however the inclusion of those who did not need any Green Deal measures meant that segments did not differentiate strongly enough on uptake and any genuine differences were diluted.

The need to differentiate between segments in terms of their propensity to take up a Green Deal package meant that the initial segmentation solution was reviewed and it was decided that the segments should be developed only amongst households which 'needed' at least one of the measures included in the research (cavity wall insulation, solid wall insulation, loft insulation, condensing boiler). It is worth noting that 'need' was estimated using responses given by respondents to questions in the survey, rather than being assessed by a building surveyor. As such, it is likely that the proportion of homes needing Green Deal measures is different to published estimates.

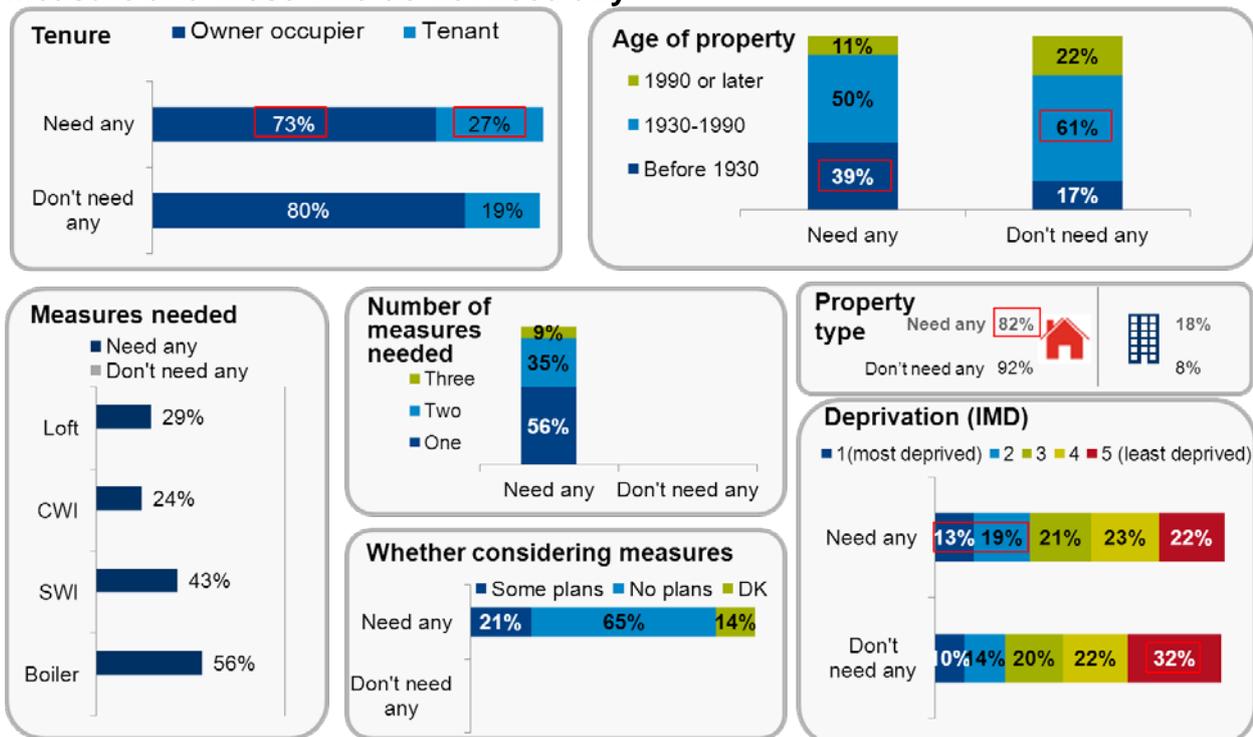
To ensure that this was an appropriate route to take, the profile differences between households 'needing' at least one measure and those not 'needing' any measures were examined, and it was felt that the differences were intuitive and within acceptable bounds. Figures 3 and 4 show that there were very few differences between two groups.

Figure 3: Demographic comparison between those who ‘need’ at least one measure and those who do not need any



Note: Significant differences are indicated by red boxes around estimates.

Figure 4: Property characteristic comparison between those who ‘need’ at least one measure and those who do not need any



Overall, 66% of all households ‘needed’ at least one measure, which meant that the base for the segmentation was 1,352.

In taking this decision, it was noted that it would impact on the allocation of the segments into the TGI database, because there is no way that ‘need’ of any measures could be identified within the database. A further objective of the segmentation process was, therefore, to explore whether it would be feasible and appropriate to allocate segments developed amongst households ‘needing’ at least one measure against those who do not ‘need’ any measures.

8.1 Selection and refinement of input variables

A crucial stage of the segmentation process is identification of the most appropriate input variables, as these determine the nature of the segmentation. Use of demographic variables as inputs to the segmentation was avoided as these tend to dilute any interesting variation obtained from the other inputs. In addition, segmenting in this way does not tend to yield more useful results than cross-tabulations alone.

Potential input variables were discussed and agreed with DECC at the questionnaire design stage and the final selection used to construct the segmentation is shown below. These inputs were based on learnings from previous research which indicated that they might drive decisions around energy efficiency. Other questions which were not used to construct the segmentation were used to profile the resulting segments, and responses to these could still be used in the interpretation and understanding of the segments.

Input variables which also served as ‘hook’ variables to the TGI database are marked in the table below.

Table 3: Segmentation input variables

Q.	Question	Sub-questions	TGI ‘hook’ questions
B11	Thinking about your home in the winter, how easy or difficult is it to keep your home warm when the heating on? (Very easy – very difficult)		
B12	Which one of these best describes how well you and your household are keeping up with your energy bills at the moment? (I/we manage very well - I/we have severe difficulties)		

C1	Thinking about your home, how much do you agree or disagree with these things that other people have said? (Agree strongly – disagree strongly)	<ul style="list-style-type: none"> • I'm the type of person who likes to have the newest gadgets in my home • I feel very comfortable dealing with financial matters • It's only worth doing environmentally friendly things if they save you money • It's not worth me doing things to help the environment if others don't do the same • I'm always looking out for new ideas to improve my home • I only take out credit/loans when absolutely necessary 	<ul style="list-style-type: none"> • It's only worth doing environmentally friendly things if they save you money • It's not worth me doing things to help the environment if others don't do the same • I'm always looking out for new ideas to improve my home • I only take out credit/loans when absolutely necessary
C2	How often, if at all, do you personally do any of the following? (Always – never)	<ul style="list-style-type: none"> • Leave your TV or PC on standby for long periods of time • Keep a tap running while you brush your teeth • When buying new appliances (e.g. fridge, washing machine) choose those which are more energy efficient 	<ul style="list-style-type: none"> • Leave your TV or PC on standby for long periods of time • Keep a tap running while you brush your teeth
D1	Which, if any, of these are you realistically planning to do to your home in the next 12 months?		
D3	Which of these reasons would be important to you in making your home more energy efficient?		
D4	And which of these things prevent you from doing more to make your home more energy efficient?		
	Uptake propensity	Derived from conjoint task	

The full questions can be found in Appendix 1.

In addition to the choice of input to the segmentation, the way in which they are input also impacts on the nature of the resulting segmentation. Simply inserting the questions as they stand would be likely to significantly skew the resulting segments: this is because some may be significantly inter-correlated and their inclusion may lead to double counting the impact of a particular view or action on the resulting segment. In addition, differences in the way in which questions are asked (e.g. some scales, some chosen from lists) may also bias the resulting segments. It was therefore important to carefully construct the segmentation inputs to be used in driving the segmentation. Different methods were employed for different types of variables, as shown in the following section.

It is also worth noting at this stage that before they were entered into the segmentation programme, all input variables were 'standardised'. This means that they were transformed using a linear transformation that set their mean to zero and their standard deviation to one. This is a standard practice in segmentations which aim to ensure that the variables with differing scales do not have disproportionate influence on the segmentation simply by virtue of their scale.

8.2 Factor analysis of scale variables

A factor analysis is a statistical procedure that attempts to generate a number of underlying concepts (or factors) which are functions of the questions asked. These factors are linear combinations of the questions and are typically designed to be orthogonal, i.e. completely uncorrelated with each other, which means that they will all have equal weight in the segmentation. Only scale questions lend themselves to factor analysis, so these methods were employed to look only at questions B11, B12, C1 and C2.

The first stage is to examine the correlations between these variables and table 4 below shows the correlation matrix. Correlation coefficients run from 0 to 1 (or -1):

- A correlation coefficient of 0 means that the two variables are entirely unrelated to each other
- A coefficient of 1 means that the two variables are entirely related to each other
- A coefficient of -1 means that the two variables are entirely related to each other, but in a negative direction.

The shading in the table below indicates the magnitude of correlations: those in the darkest red show the strongest positive correlations, and the darkest blue show the strongest negative correlations. Coefficients shown in the table are not particularly high which indicates that there are relatively low levels of correlations between all the variables: the exceptions are the two questions related to environmentally friendly behaviours ("It's only worth doing environmentally friendly things if they save you money" and "It's not worth me doing things to help the environment if others don't do the same") and the two questions related to ease of heating the property ("... how easy or difficult is it to keep your home warm when the heating is on?" and "... how well are you and your household keeping up with your energy bills?").

Table 4: Correlations between questions C1, C2, B11 and B12

	c01_01	c01_02	c01_03	c01_04	c01_05	c01_06
c01_01 C1. I'm the type of person who likes to have the newest gadgets in my home		0.08	0.09	0.06	0.30	0.07
c01_02 C1. I feel very comfortable dealing with financial matters	0.08		0.00	-0.05	0.10	0.09
c01_03 C1. It's only worth doing environmentally friendly things if they save you money	0.09	0.00		0.44	0.07	-0.06
c01_04 C1. It's not worth me doing things to help the environment if others don't do the same	0.06	-0.05	0.44		-0.03	-0.06
c01_05 C1. I'm always looking out for new ideas to improve my home	0.30	0.10	0.07	-0.03		0.10
c01_06 C1. I only take out credit/loans when absolutely necessary	0.07	0.09	-0.06	-0.06	0.10	
c02_01 C2. Leave your TV or PC on standby for long periods of time	0.17	-0.01	0.02	0.07	0.05	0.01
c02_02 C2. Keep a tap running while you brush your teeth	0.05	-0.05	0.05	0.07	-0.01	0.01
c02_03 C2. When buying new appliances (e.g. fridge, washing machine) choose those which are more energy efficient	0.00	0.07	-0.03	-0.09	0.14	0.02
B11 B11. Thinking about your home in the winter, how easy or difficult is it to keep your home warm when the heating on? Is it...	-0.01	0.11	0.03	0.02	-0.05	0.01
B12 B12. Which one of these best describes how well you and your HH are keeping up with your energy bills?	0.09	0.23	-0.06	-0.04	-0.04	0.05

Overall, it was felt that a factor analysis was useful because it reduced the number of input variables to a more manageable number. A number of factor solutions were examined, but the solution shown below was considered the most useful and easy to interpret. This 6-factor solution explained 70% of the variance within the statements.

The table shows factor loadings of each statement with the 6 factors: factor loadings are the correlation coefficients between the statements (rows) and the factors (columns). The higher the factor loading, the more the statements are represented by the factor. By looking across each row, the factor with which each statement is most closely correlated can be found. Statements are grouped under each factor for the purposes of interpretation.

The table has been sorted so that statements which are most strongly correlated with a particular factor are shown together, and these are separated from the next factor by a thicker horizontal line. Note that the third statement (C1: comfort dealing with financial matters) does not correlate particularly strongly with any factor, but has been grouped into the factor with which it most strongly correlates (factor 1).

Table 5: Factor loadings

	1	2	3	4	5	6
rb12 B12. Which one of these best describes how well you and your HH are keeping up with your energy bills?	0.82	-0.09	0.06	0.01	0.01	0.00
rb11 B11. Thinking about your home in the winter, how easy or difficult is it to keep your home warm when the heating on? Is it...	0.74	0.11	-0.16	0.05	0.12	0.00
rc01_02 C1. I feel very comfortable dealing with financial matters	0.48	-0.05	0.35	-0.29	-0.07	0.15
rc01_03 C1. It's only worth doing environmentally friendly things if they save you money	-0.02	0.85	0.12	-0.03	0.02	-0.04
rc01_04 C1. It's not worth me doing things to help the environment if others don't do the same	0.01	0.83	-0.04	0.11	-0.09	-0.03
rc01_01 C1. I'm the type of person who likes to have the newest gadgets in my home	0.10	0.05	0.76	0.21	-0.08	-0.04
rc01_05 C1. I'm always looking out for new ideas to improve my home	-0.15	0.05	0.74	-0.08	0.24	0.13
rc02_02 C2. Keep a tap running while you brush your teeth	-0.08	0.10	-0.10	0.76	0.05	0.14
rc02_01 C2. Leave your TV or PC on standby for long periods of time	0.08	-0.03	0.29	0.68	-0.12	-0.16
rc02_03 C2. When buying new appliances (e.g. fridge, washing machine) choose those which are more energy efficient	0.09	-0.07	0.08	-0.04	0.95	-0.02
rc01_06 C1. I only take out credit/loans when absolutely necessary	0.05	-0.07	0.08	0.02	-0.02	0.96

Factors were given names based on the statements with which they were most strongly correlated, and these working names were used in the interpretation of the resulting segments. However, it should be noted that the individual questions/statements were also profiled against the segments and used in the interpretation of the segments:

- Factor 1: Ease of paying energy bills / comfort with finance
- Factor 2: Green if convenient
- Factor 3: Home improvement/ gadget conscious
- Factor 4: Wasteful of resources
- Factor 5: Choose energy efficient appliances
- Factor 6: Only take out loans when necessary.

8.3 Other variables

The other input variables did not suit factor analysis because they were not scale questions: instead they comprised a number of binary items (i.e. answers which were either chosen or not chosen by the respondent from lists shown).

Question D1 listed types of works which the respondent was planning to make to their property in the next 12 months. Because it is known that undertaking other works is a very important trigger to installing energy efficient measures, it was important that this variable identified households who were most likely to undertake works which could be triggers. Rather than making a simple count and following guidance from DECC, a decision was made to group the types of works as shown below. The variables were constructed as binary variables, so that if a respondent said they were planning to undertake any works in a category, they were coded as 'yes' for that category (code 1). Otherwise, the code for that category was zero.

Table 6: Grouping works to be undertaken in the next 12 months (D1)

Category	Works included
Energy efficiency works	<ul style="list-style-type: none"> • Install double glazing/new windows • Updating or installing central heating/new boiler • Installing other energy efficiency measures (e.g. loft insulation, cavity wall insulation, solid wall insulation) • Install solar panels/solar heating
Other major works	<ul style="list-style-type: none"> • Whole house refurbishment • Extension or loft conversion • Substantial refurbishment to a single room, such as fitting a new kitchen or bathroom • Other work to the outside of your property
Other works	<ul style="list-style-type: none"> • Other refurbishment to a single room, such as decorating, fitting new carpets, etc • Garden make-over
None	

In addition, questions on benefits and barriers to undertaking energy efficient improvements were carefully examined to ensure their suitability for inclusion. Because some responses were given by very few people, their inclusion as inputs would unduly skew the segmentation solutions, so very low responses were grouped or removed to ensure that no input variables were used which fell outside of the range 20%-80% (i.e. no fewer than 20% of people giving an answer, no more than 80% of people giving an answer). Once again, it should be noted that the full questions were also profiled against the segments and used in the interpretation of the segments.

8.4 Uptake propensity

A variable which gave an indication of potential to take up the Green Deal was included as an input into the segmentation. This was generated from a conjoint exercise included in the questionnaire.

This variable was included to make it more likely that the segments would differentiate by propensity of uptake and was used to provide starting centres for the segments. It was effectively downweighted to produce a segmentation which could be described by the other inputs and so that it did not have an undue influence on the segments.

8.5 Segmentation method

The segmentation approach employed was called K-means: this attempts to minimise the differences within each of the segments and maximise the differences between them. The process can be summarised as follows:

- It is necessary to specify in advance the number of clusters to be modelled
- It is then necessary either to form initial clusters at random, or predefine some cluster centres
- The algorithm 'shifts' the cluster centres around by re-assigning respondents to different clusters, each time reducing the overall aggregate distance between the cluster centres and the respondents in them across all clusters. This continues until convergence is achieved, i.e. it is no longer possible to reduce the distance any further, given the existing centres.

With regard to the final point, it is worth noting that this does not mean that there is no other possible placement of the cluster centres that has a lower total distance, only that it is not possible to change the existing combination without increasing the distance.

This is not an entirely mechanical process as the segments are a function of the 'starting centres' but in this case the uptake propensity range was used to provide an indication of optimal cluster centres to maximise the efficiency of the process.

A number of segmentation models were developed with different variations on the input variables, and each containing different numbers of segments. For the sake of brevity, the remainder of this section describes only the core characteristics of the segmentation selected.

8.6 Segmentation diagnostics

The most important indicator of the quality of a segmentation is how well the resulting segments answer the key research questions: it is always important to construct segments which:

- Are recognisable
- Are easy to interpret and communicate to others
- And can be used for future marketing.

A number of segmentation solutions were discussed with the team at DECC before the optimal solution was agreed.

In addition, a number of other outputs and diagnostics were checked to ensure the statistical robustness, reliability and replicability of the segmentation solution. These key metrics are described below.

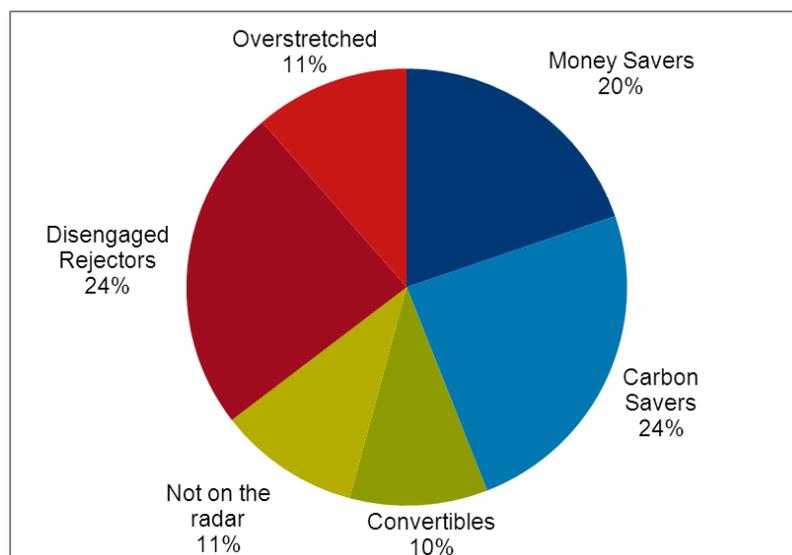
8.7 Segment size

One key metric is the relative sizes of the segments:

- If very large segments are discovered, this could indicate that the segmentation has not been effective as there is a large group of respondents for whom no differentiation was detected. This could, of course, be a reflection of reality, but it is more likely to mean that the distribution of the input variables was not optimal
- If very small segments are discovered, the sample size becomes too small to say anything meaningful about them or to be able to profile them reliably. In addition, they would be too difficult to target through marketing or communications and would therefore be less usable.

In this instance there was a reasonable distribution of sizes amongst the segments with none making up less than 10% of those 'needing' at least one measure, and no segment larger than 25%.

Figure 5: Segment sizes

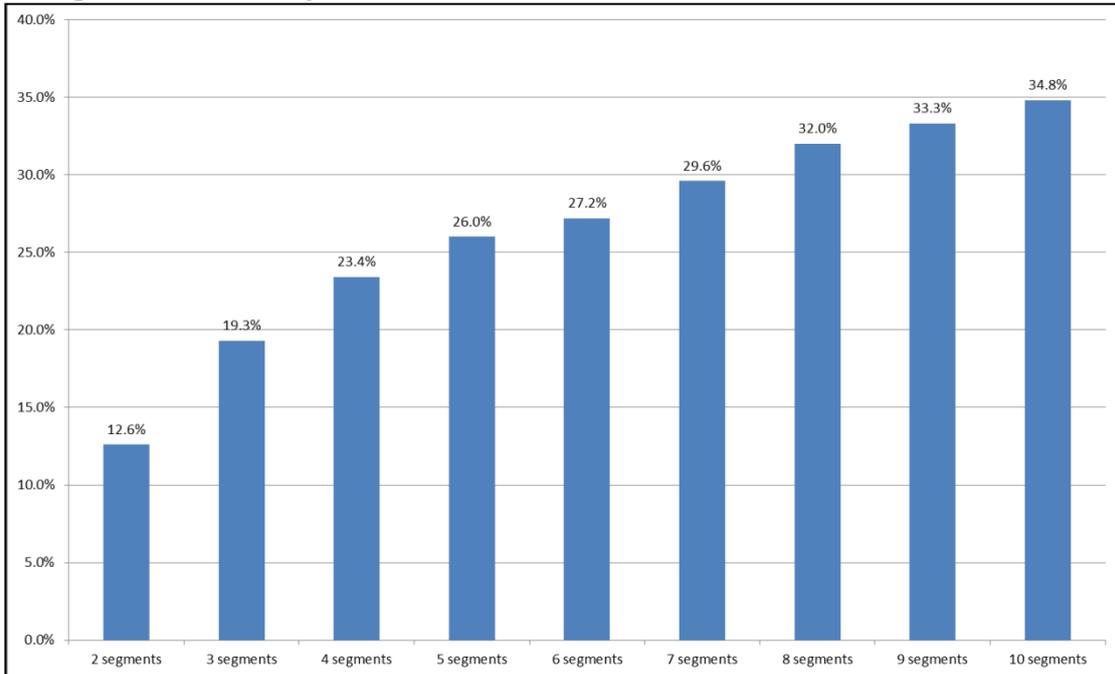


8.8 Efficiency

A further measure that was used to assess the segmentation is its efficiency: this refers to the percentage of the variation in all the input variables that is explained by grouping them into the respective number of segments. For example, one segment only will, by definition, explain none of the variance and having the same number of segments as there are respondents, will explain 100% of the variance.

For segment solutions in-between, the efficiency level is a function of the number of inputs – so fewer inputs tends to lead to a greater level of efficiency (and it is notable that this segmentation employed a fairly large number of input variables). The relative increase in efficiency by each segmentation solution is shown over the page.

Figure 6: Segment efficiency



The segment efficiency plot above suggested that segment solutions between five and seven should be considered. However, as already mentioned, a crucial part of the decision process is the interpretation of the clusters, and the usefulness of the 6 cluster solution ultimately led to its selection.

8.9 Stability

A further important measure of the segmentation is how stable it is. If the segmentation were to be run again on slightly different data, would this have a large impact on the segments? Would they remain the same or change?

To test stability, the same clustering procedure was applied to subsamples of 80% of the original data. On the basis of this, the allocation of clusters against the original cluster solution (based on everyone) was checked to see how closely they matched. This was run four times and the average similarities between the original cluster solutions and the sub-samples were 69%, 84%, 72% and 72%. In our experience, this indicates a good level of stability, indicating that typically three quarters of all respondents remain in the same segment, given that 20% of the respondents were omitted.

8.10 Recreating the segments in future research

The segmentation can be recreated in other relevant studies for analysis and/or tracking purposes. The most perfect way to recreate the segmentation is to add all of the questions which were used into the new questionnaire, but this is not always feasible (e.g. in terms of space within the questionnaire).

As a result, it is necessary to develop an algorithm which determines the best balance between the *number* of questions used by the algorithm and the level of allocation *accuracy*.

Table 7 shows the allocation efficiency across a number of potential question combinations. It became clear at a very early stage that questions D3 and D4 were very important in allocating the segments. For each of these combinations, the allocation accuracy was found to be high (it is desirable to aim for an allocation efficiency of 80%+). However, adding further questions did not increase the accuracy greatly, so from a statistical point of view adding in D1 and the statement from C2 ('Leave your TV or PC on standby for long periods of time') only increased accuracy by 0.3% each. Given this, the final algorithm was based on questions D3 + D4 + 5 questions as per the table below.

Table 7: The allocation efficiency across potential question combinations

Number of Questions	D3 + D4 + 5 questions	D3 + D4 + 6 questions	D3 + D4 + 7 questions
Allocation Efficiency	80.4%	80.7%	81.0%
D3. Which of these reasons would be important to you in making your home more energy efficient? - All items	1	1	1
D4. Which of these things prevent you from doing more to make your home more energy efficient? -All items	1	1	1
B12. Which one of these best describes how well you and your HH are keeping up with your energy bills?	1	1	1
C1. It's not worth me doing things to help the environment if others don't do the same	1	1	1
C1. I'm the type of person who likes to have the newest gadgets in my home	1	1	1
C1. I'm always looking out for new ideas to improve my home	1	1	1
C2. When buying new appliances (e.g. fridge, washing machine) choose those which are more energy efficient	1	1	1
D1. Which of these are you realistically planning to do to your home in the next 12 months? - Other refurbishment to a single room, such as decorating, fitting new carpets, etc		1	1
C2. Leave your TV or PC on standby for long periods of time			1

Although this algorithm predicted segment membership in 80.4% of cases overall, this does vary from 57% to 91% between segments, as shown in Table 8.

Table 8: Predicted segment membership for final algorithm

		Predicted Segment Membership					
		Carbon Savers	Not on the Radar	Disengaged Rejectors	Overstretched	Convertibles	Cost Savers
Actual	Carbon Savers	88%	1%	3%	5%	0%	3%
	Not on the Radar	4%	75%	13%	1%	6%	1%
	Disengaged Rejectors	4%	1%	91%	0%	0%	3%
	Overstretched	8%	2%	2%	57%	8%	22%
	Convertibles	2%	2%	0%	5%	85%	5%
	Cost Savers	6%	3%	8%	12%	1%	70%

The segmentation was developed based only on households ‘needing’ at least one of the energy efficiency measures included in the Green Deal survey. Future surveys may not necessarily ask about ‘need’ for measures (nor for the same measures that were included in the Green Deal Survey, nor in the same way); to do so would greatly increase the number of questions that were needed as it would be necessary to work out whether measures are appropriate to the property first, and then whether they are already installed.

It was therefore important to ensure that the algorithm could be used with a general population survey. This is discussed further in the following section.

8.11 Ability to recreate the segments in the TGI database

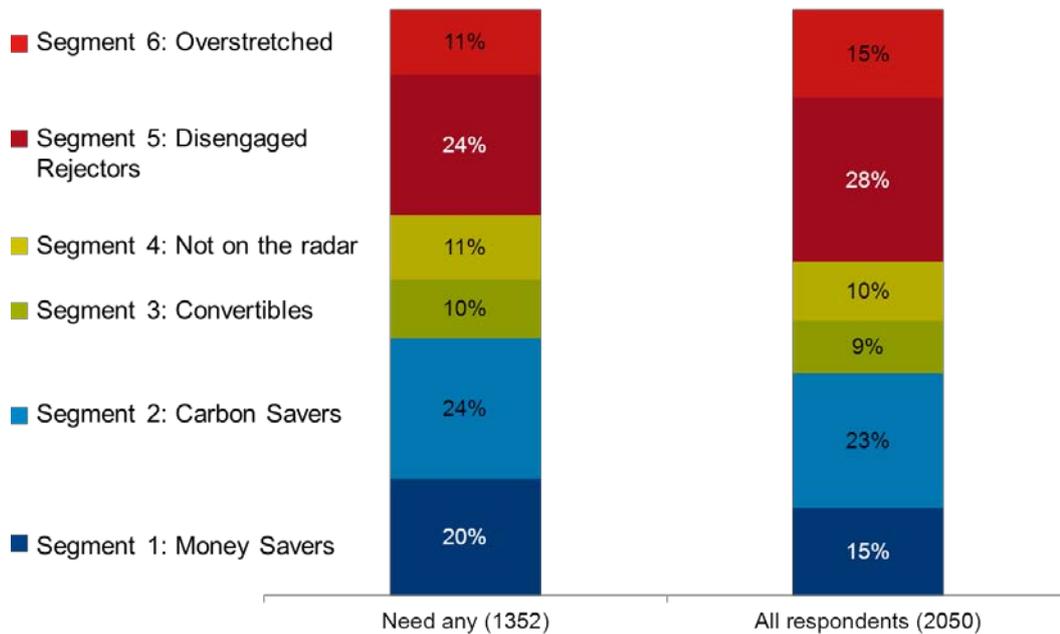
This is clearly not a standard measure of the quality of a cluster analysis, but given that the clusters needed to be allocated into the TGI database, it was important that the cluster solution chosen could be recreated at a reasonable level of confidence using only the questions which are common to both the Green Deal survey and the TGI database.

It should be remembered that the segmentation was developed based only on households ‘needing’ at least one of the energy efficiency measures included in the Green Deal survey, but the TGI database does not include any information about energy efficiency needs. The next stage in the process was therefore to test the appropriateness of the segmentation amongst households with no energy efficiency ‘needs’ to assess whether it could feasibly be used on a general population sample (such as the TGI database).

An allocation algorithm (using the full range of questions included in the questionnaire) was developed and, using this, respondents who did not ‘need’ any energy efficiency measures were allocated to segments. The efficiency of this exercise was 98% (e.g. in 98% of cases respondents were allocated to the same segment).

The resulting profiles were examined to ensure that the segment sizes were not too dissimilar and, where they varied, the differences could be explained. The demographic profiles of the segments within each ‘need’ and ‘no need’ group were also examined, and shown to be a good fit.

The chart below shows the segment profiles and how they differed, and again it was felt that the differences were within acceptable bounds and could be explained by profile differences.

Figure 7: Segmentation profile comparison

The next stage was to conduct a further discriminant analysis to develop an allocation algorithm using only the hook questions included on the Green Deal questionnaire. The resulting algorithm predicted segment membership in 61% of cases: though this varied from segment to segment, as shown below:

- Segment 1 (Money Savers): 73%
- Segment 2 (Carbon Savers): 64%
- Segment 3 (Convertibles): 51%
- Segment 4 (Not on the Radar): 49%
- Segment 5 (Disengaged Rejectors): 64%
- Segment 6 (Overstretched): 47%

These were felt to be within acceptable bounds for an allocation exercise such as this.

It is notable that allocation efficiency was higher for segments which were demographically homogeneous: reflecting the demographic nature of many of the TGI hook questions.

The segments were allocated into the TGI database amongst only those who were owner occupiers or private rented tenants (i.e. social rented tenants included in the TGI database were excluded from the allocation process). It was not possible to exclude adults who were not householders or partners from the TGI database because they could not be identified using the questions included in the database.

As a result the full TGI database can be analysed by segment. The final sample size for this analysis is 21,035.

8.12 Profiling the segments

Once the final segmentation solution had been agreed, the segments were profiled against all other key variables in the dataset for the purposes of interpretation and reporting. The profiles of the 6 final segments are shown in Appendix 3 of this report.

9. Appendices

Appendix 1: Questionnaire

Please click on the icon below to open the final questionnaire.



DECC Green Deal
incentives and segme

Appendix 2: Cross breaks and derived variables

Please click on the icon below to open the cross breaks and derived variables.



Cross break Layout
and Definition.docx

Appendix 3: Profiles of the 6 segments

Please click on the icon below to open the cross breaks and derived variables.



Green Deal
segmentation profiles

© Crown copyright 2012
Department of Energy & Climate Change
3 Whitehall Place
London SW1A 2AW
www.decc.gov.uk

URN 12D/415