Building Engineering   Ofgem   August 2011

UK smart meter programme: preparing for evaluation

Learning from EDRP

Prepared by: Gary Raw
Regional Director

Checked by: David Ross
Regional Director

Approved by: David Ross
Regional Director

Energy Demand Research Project: Final Analysis

Rev No	Comments	Checked by	Approved by	Date
1	First draft to client	David Ross	David Ross	02/08/11

AECOM House, 63-77 Victoria Street, St Albans, Hertfordshire, AL1 3ER
Telephone: 01727 535000 Website: http://www.aecom.com

Job No 60221038 Date Created: August 2011

c:\users\gary\documents\documents\gr people solutions\decc\sm evaluation\lessons from edrp\sm evaluation - lessons from edrp v1.0 - gjr 12 7 11.docx

Table of Contents

Capabilities on project:
Building Engineering

Summary

Introduction

This report is the first deliverable from AECOM's scoping study for energy consumption change measurement. It documents the methodological lessons from the Energy Demand Research Project (EDRP) that are relevant to "the future design of supplier and other trials, benefits tracking and future evaluation". The direct application of the report will be to feed into later stages of the scoping study, culminating in a workshop to take forward planning of the UK smart meter roll-out evaluation. It should also be of value to DECC, its contractors and delivery partners in the design of trials for evaluation of interventions aimed changing householders' behaviour to reduce energy demand.

While most of the material collated in this report is relevant to the design of experimental trials, its applicability to evaluation of the national smart meter roll-out (and benefits tracking) will depend on the general terms agreed for the evaluation. Such issues are the subject of later stages of the scoping project and so all potentially relevant lessons from EDRP are recorded in this report, without prejudice to final decisions over the evaluation.

The report is not a critique of the EDRP trials – just lessons for future evaluation. Neither is it a substitute for expertise in experimental design, knowledge of energy-related behaviour, understanding of energy markets and technology, and skills in project administration. All these will be essential for the design of trials, benefits tracking and evaluating the roll-out; the present report is merely an advisory document.

This summary states which topics are covered in each chapter, so that the reader can go to the detail as required.

Evaluation aims and project design

1. The need to be clear about the aims of the evaluation from the start and – from the aims – to derive a project design that will minimise bias and noise to allow sound deductions to be made from the data collected.
2. The balance of:
    - experimental rigour;
    - the affordable total sample size and study duration;
    - the practical and commercial reality of particular experimental interventions;
    - precision about the impact of specific interventions vs identification of effective combinations.
3. The benefit of central project coordination using experienced professional researchers.
4. The selection of interventions (and combinations of interventions) to modify behaviour related to gas and electricity consumption, considering the range of behaviours and motivations that might be targeted.
5. Cross-sectional vs sequential deployment of interventions.
6. The creation of matched experimental and control groups.[1]
7. Response to new business offerings that become available to customers in general during the trial.
8. Establishing the baseline energy consumption.

---

[1] In this report, the groups (e.g. of households or customers) defined in an evaluation are referred to as experimental groups if one or more or the interventions tested by the evaluation is applied. Groups used for comparison purposes, to determine whether the intervention had any effect, are termed control groups. The evaluation itself is treated as a trial of the interventions and the trial groups are both the experimental and the control groups.

Sampling and recruitment of participants

9. Using sampling and recruitment to ensure that participants are well matched to each other and representative of the population of interest in the trial, considering:
    - the sampling base (the "list" from which candidate participants are selected);
    - the sampling procedure;
    - bias during recruitment, from either refusal or inability to participate or exhausting the supply of some participant types before others;
    - bias during deployment (e.g. as a result of technical problems with installing equipment);
    - selective attrition during the trial.

Sampling

10. Monitoring representativeness (of the sample as a whole and each group) at the start and to track attrition.
11. Deciding on sample size (for the overall sample and for subgroups).
12. Defining the sampling base, taking into account the aims of the trial, the availability of data on potential participants (candidates), regional and local variation, and demographic groups.
13. Criteria for excluding certain households or individuals (and the risk of bias arising from the exclusions).
14. The role of sample stratification and grossing, and the need for supplementary data to understand the strata.
15. The strata used in EDRP:
    - electricity-only vs dual fuel customers;
    - credit vs prepayment customers;
    - Economy 7 customers;
    - indirect demographic classifications (Mosaic, CAMEO, ACORN and Ocean);
    - fuel poverty;
    - pretrial fuel consumption.

Recruitment

16. Deciding between opting in and opting out of the trial, and the level of awareness that participants should have of the trial and/or interventions.
17. Managing and interpreting any resulting sampling bias (overall or differing between groups).
18. The schedule of recruitment in relation to:
    - lags (and variation in lags) due to recruitment, administration or installation issues;
    - assignment of recruits to trial groups;
    - the representativeness and matching of groups in the sample;
    - the options for statistical analysis of the trial data.

19. Managing transitional periods when only some of the participants have been recruited or received an intervention.

20. The impact of the intervention start date on the effectiveness of the intervention.

21. Topping up samples during the trial.

22. The process of approaching candidates in relation to the proportion of candidates who become participants, their understanding of the trial, their response to interventions and the interpretation of findings, taking into account:
    - process effectiveness and cost;
    - recruitment success and selection bias;
    - failure to make contact (initially or as interventions are implemented), to gain access to a home or agreement to participation, or to achieve a successful installation;
    - informed consent to participate;
    - medium of approach (e.g. by letter, telephone or in person);
    - the volume, consistency and accuracy of information given to candidates and recruits.

23. Collecting information from recruits to:
    - ensure they are eligible;
    - achieve a sample that is representative according to chosen criteria;
    - provide data for the analysis of findings, to control for confounders or break down the analysis by subgroups;
    - identify changes during the trial (e.g. in the appliances present or energy-related behaviour).

24. The design of materials used for recruitment to explain the trial and the intervention(s), considering:
    - balance between a large volume of highly detailed information and providing participants with the essential information sufficiently concisely that they are likely to absorb it;
    - balance between technical precision, formality and using clear, attractive language, expressed from the householder's perspective;
    - timing (of what needs to be communicated at some point in the recruitment process or later in the trial);
    - clarity – using simple phrasing and avoiding the jargon of research or the energy industry;
    - the range of approaches to motivating the candidate to participate and the entailed benefits and risks;
    - emphasis on different aspects of the trial, including the compromise between maximising participation, reducing bias and maximising impact.

25. The need for qualitative research to understand the recruitment process better.

The interventions

Introduction

26. The design of interventions for maximum effect.
    - Not just the equipment, feedback or incentive – also the associated user guidance and the media by which it is communicated, to establish the ideal balance of complex functionality and simple, attractive products.
    - Interaction between installation teams and customers.
    - Creating and maintaining participants' awareness of (and interest in) the interventions – the whole household, not just one participant.

Energy efficiency advice

27. Relevance of the form of delivery to the intended audience, considering:
    - the level of advice – from generic information to audits and advice tailored to the individual;
    - the medium – on paper, via the web, email, text, a dedicated TV-based web page, RTDs, social media or verbal;
    - the role of installation visits, reference guides and regular short, simple statements;
    - trust in the source of advice;
    - barriers to reading or following advice;
    - combination with other interventions (e.g. linking advice to energy bills and historic feedback in the bill).
28. The content of the advice, encompassing:
    - choice of topics (e.g. heating and insulation, energy used by lights);
    - structuring advice around means, motive and opportunity;
    - addressing likely barriers to reading or following advice;
    - relevance to the season in which the advice is delivered;
    - risks (e.g. of inadequate ventilation);
    - being clear, correct and unambiguous, and avoiding assumptions about the facilities in the home or the understanding of the householders.

Historic feedback

29. The different roles of historic and real-time feedback.
30. Historic feedback provided to customers vs self-reading of meters.
31. Fitting the feedback to the desired outcome, considering the optimum frequency, duration, immediacy, content (kWh, cost, CO_2, etc.), breakdown (by time, space and appliance), medium (and details such as aesthetics and simplicity of access or use), comparisons (historic or normative), and combination with other interventions.
32. Providing feedback alongside energy bills.

33. The need for feedback to:
    - be easy for the householder to notice and interesting to look at;
    - be clear to read and interpret, with guidance as necessary;
    - limit the detail to what people can easily absorb;
    - not present too much information at once, but show both kWh and cost comparisons.

Benchmarking (comparative or normative feedback)

34. The choice of peer group and benchmark level.
35. Households that are already below the benchmark values.
36. Overcoming people's belief that they are not influenced by normative messages.
37. Targeting responsive households.

Customer engagement using commitment to reduce consumption

38. Giving the commitment a social context.
39. Ensuring the commitment is new.
40. Making the commitment as specific and concrete as possible and linking it to advice.
41. Monitoring the commitment.
42. Reward for making the commitment.

Smart meters

43. Aspects of recruitment and installation that could encourage recipients to save energy.
44. Providing guidance on using the smart meter and saving energy.
45. Installer training.

Real-time displays

46. Combining with other interventions, including external support for response to using an RTD (e.g. to inform purchase choices or to establish norms in relation to comparable households, or to obtain planning permission).
47. Engaging users to have an RTD installed, maintained in working order and actually used (employing the installation process, instructions and guidance, and the design of the RTD itself).
48. The need for clear, simple, well produced guidance including:
    - technical instructions (how to set up, locate and manipulate the RTD to acquire information);
    - motivation (reasons for using the RTD) – at the start and throughout the trial;
    - application (how to use the information acquired to reduce energy consumption).
49. Response to faults in the RTD or communication with the meter.

50. Health and safety warnings.

51. The display content, including:
    - priorities for information to be displayed;
    - variations in preferences for RTD design and display features;
    - default settings that show users the most relevant information without having to operate the device;
    - the role of audible alarms and visible signals;
    - temperature displays.

Heating controller integrated with RTD

52. Potential benefits and risks.

53. The need for clear guidance and well planned display characteristics.

54. Target audience (replacing an existing controller vs installation in homes that have no controller).

55. Location in the home.

Time of use tariff/incentive

56. Emphasis on demand reduction vs load shifting.

57. Possible adverse effects.

58. Tariff levels/ratios and setting incentive levels.

59. Limitations in the flexibility that households are likely to have in when they use energy.

60. Targeting weekends vs weekends, summer vs winter.

61. The effect of household size.

62. Interaction with RTD and online feedback.

63. Keeping it simple and providing clear information about the tariff/incentive and how to respond to it.

64. Clock changes between GMT and BST.

65. Guarantees of customers not paying more.

Incentive to reduce consumption

66. Setting the level of the incentive/reward, its timing and the criteria for receiving it.

67. Enhancing awareness and the perceived value of a reward, through:
    - the form of reward (e.g. cash vs money off the bill vs a prize draw);
    - combination with relevant advice, feedback on progress towards the target or some other intervention;
    - targeting particular households;
    - making targets transparent in how they are calculated; easy to understand; and fair but challenging.

68. Creating a persistent effect on energy consumption.

69. Avoiding detrimental effects on non-financial motives for saving energy.

70. Sequential targets.

71. The use of targets without financial incentives.

72. Using smart meters to make consumers more aware of energy prices and hence more responsive to financial incentives.

Web-based interventions

73. The unrealised potential of web-based services.

74. Possibilities for improving the capacity of web-based services to reduce energy demand, e.g. through:
    - simplified access to web sites;
    - real-time feedback;
    - immediately attractive, clear, accurate and useful content;
    - pitching information at an appropriate level (simple for newcomers but with easy access to detail and tools);
    - user involvement in setting any savings targets.

75. Development of TV-based services.

76. Targeting the users, occasions and contexts for which web-based services are most likely to result in energy savings.

77. The relationship between RTDs, web-based feedback and portable devices for web access.

Community trials

78. Defining the community.

79. Taking into account each community’s baseline of: actions already taken; understanding of what else could be done, and why; motivation to act; and resources to act.

80. Local and central leadership.

81. Support (financial, technical, logistical and networking).

82. Identifying and promoting actions that are relevant to the particular community, the available resources and any targets set.

83. Involving the community in the choice of actions and the use made of any reward offered for saving energy.

84. Effecting promotion and communication.

85. Targets, rewards and monitoring progress.

86. The need to identify the “key ingredients” for more cost-effective versions of the community approaches used to date.

Customer feedback

87. Methods for collecting data on participants' characteristics, experiences, opinions and behaviour during the trial (e.g. face to face, by telephone or by self completion on hard copy or online).

88. Questionnaire survey method and topics covered.

89. The timing of customer feedback, and whether all trial participants should be included, taking into account that:
    - participation in customer feedback may affect subsequent behaviour;
    - requiring participation in feedback may affect participation in the trial;
    - some changes in behaviour may not occur until late in the trial (or after it is finished).

90. Deciding the content of questions, taking into account:
    - the main aims of the trial and the feedback;
    - being realistic about what respondents are able to report (e.g. facts vs opinions, current vs historical information, and precise vs general understanding);
    - clarity and ambiguity;
    - variations in type of energy-related behaviour;
    - the meaning of questions being affected by their position in a survey;
    - the need for supplementary questions in order to interpret the main questions;
    - developing a common structure for questions.

91. Use of cognitive testing, field piloting of questions and development of an analysis plan to define what data will be needed.

92. The analysis of questionnaire data, considering:
    - the level of sophistication that is possible with different levels of data, which in turn depends on the format of questions and response options;
    - sample size;
    - the risks of multiple independent tests;
    - missing data.

93. Interpretation of findings.

94. Investment of customer service time, not just to deal with queries effectively but to gather data on customer understanding, opinion and behaviour.

95. Monitoring the quality and content of customer service responses (treating them as an intervention).

96. Anticipating the types of enquiry that are likely to be received.

97. Minimising queries through attention to the design, testing and installation of interventions, anticipating customer concerns, effective administrative systems and communications.

Analysis of consumption data

98. The secure acquisition, transmission, processing and storage of data, maintaining data quality.
99. Data storage to facilitate analysis.
100. Correcting for differences between seasons and years, and creating annual (or monthly or quarterly) consumption estimates.
101. Data cleaning and outlier exclusions.
102. Designing the analysis based on the trial design and the distribution of data (e.g. whether there is a factorial design and normally distributed data.
103. Including confounding variables in the analysis.
104. Time series analysis.
105. Risks if multivariate modelling is not possible.
106. Treating changes in consumption as differences or as ratios.
107. Reduction in the apparent seasonal variation of consumption by “smoothing” of non-smart meter data because of the need to interpolate infrequent readings.
108. Testing the persistence of the effect of an intervention.
109. Measures of load shifting and accounting for differences between seasons and days of the week.
110. Deciding the criterion for statistical significance.
111. The description of the analysis (statistical rigour vs general clarity).

Technical development and logistics

112. Getting good equipment, in the right place, at the right time, and the equipment continuing to operate effectively.
113. Incorporating new technical solutions.
114. Practical issues in gaining access to properties (contacting and gaining/maintaining agreement from customers; customers cancelling appointments or not being at home; issues with terms and conditions).
115. Installation problems:
    - installation failure/postponement (the implications and reasons for it occurring);
    - the homes of dual fuel customers not accommodating both gas and electricity smart meters;
    - customer dissatisfaction with the installation;
    - homes where the meter enclosure or some aspect of the home has been modified, making installation impossible, and the home owner is legally responsible for re-establishing access;
    - pre-installation site surveys.

116. The need for testing of the equipment, both separately and integrated.

117. The performance of equipment during trials, including:

- Home Area Network (HAN) and Wide Area Network (WAN) data communication;
- fluctuating signal strength;
- signals blocked or disrupted;
- variable internet performance;
- different brands of equipment being used, with different performance, without recording which homes had which brand;
- data security.

118. Suppliers' recommendations on:

- allowing time for product testing, system development and integration, staff training and development of customer services;
- developing suitably sized equipment;
- interoperability standards;
- straightforward installation;
- installation programming;
- properties that are difficult to access;
- reliable system communications;
- data management.

1 Introduction

This report is the first deliverable from AECOM's scoping study for energy consumption change measurement. It documents the methodological lessons from the Energy Demand Research Project (EDRP)[2] that are relevant to "the future design of supplier and other trials, benefits tracking and future evaluation". EDRP was a major research project investigating the impact of a wide range of interventions on domestic energy consumption, in around 60,000 homes in England, Scotland and Wales. The interventions tested included (either individually or in combination with each other):

- energy efficiency advice;
- historic energy consumption information;
- benchmarking of the customer's consumption against the consumption of comparable households;
- customer engagement using self-reading of meters or commitment to reduce consumption;
- smart electricity and gas meters;
- real-time display (RTD) devices;
- control of heating and hot water integrated with an RTD;
- other digital media for delivering information (web, TV);
- time-varying tariffs and incentives to shift load from periods of peak demand;
- financial incentives to reduce consumption.

It may be envisaged that other interventions will be used in the smart meter roll-out but they are not addressed in this report, which draws only on AECOM's final analysis of the findings from EDRP.

The direct application of this report will be to feed into later stages of the scoping study, culminating in a workshop to take forward planning for evaluation of the UK smart meter roll-out. It should also be of value to DECC, its contractors and delivery partners in the design of trials for evaluation of interventions aimed changing householders' behaviour to reduce energy demand. While most of the material collated in this report is relevant to the design of experimental trials, its applicability to evaluation of the national smart meter roll-out (and benefits tracking) will depend on the general terms agreed for the evaluation, such as whether:

- the evaluation uses only routine monitoring data or also experimental trials with subsets of homes;
- any trials are entirely within the remit of the suppliers conducting them or there is central coordination of trials;
- evaluation of the smart meter roll-out is conducted in isolation or coordinated with the design and/or evaluation of other programmes (e.g. Green Deal);
- the evaluation criteria are based only on energy savings or also, for example, customer experience, engagement with interventions, the quality of the interventions, the management and quality of data, and risk management.

This will affect, for example:

- the relevance of the selection and combination of interventions aimed at reducing energy demand through behavioural change;
- the potential for targeting particular types of household or behaviour;
- whether households are knowingly recruited to the evaluation or merely monitored;
- whether control groups are preselected or defined retrospectively using monitoring data.

---

[2] Raw GJ & Ross DI (2011) Energy Demand Research Project: Final Analysis. Office of Gas and Electricity Markets. http://www.ofgem.gov.uk/Pages/MoreInformation.aspx?docid=21&refer=Sustainability/EDRP

Such issues are the subject of later stages of the scoping project and so all potentially relevant lessons from EDRP are recorded in this report, without prejudice to final decisions over the evaluation.

This report is not a critique of the EDRP trials – just lessons for future evaluation. Neither is it a substitute for expertise in experimental design, knowledge of energy-related behaviour, understanding of energy markets and technology, and skills in project administration. All these will be essential for the design of trials, benefits tracking and evaluating the roll-out; the present report is merely an advisory document.

The following chapters summarise the learning, classified according to which aspect of an evaluation is under consideration.

- Chapter 2: the aims of the evaluation and how the project design should reflect those aims.
- Chapter 3: the sampling and recruitment of participants in the evaluation.
- Chapter 4: the design of individual interventions.
- Chapter 5: community trials.
- Chapter 6: collecting and analysing customer feedback data.
- Chapter 7: analysis of energy consumption data.
- Chapter 8 and Annex A: technical development and logistics issues.

There is a glossary of some terms and abbreviations that might not be familiar, in Annex B.

2 Project aims and project design

2.1 Introduction

The aims of a trial or evaluation project and the project design are closely linked. The design includes the high level characteristics of the project that will allow deductions to be made from the data collected, e.g. the population and time period of interest, the interventions used and how they are combined to form an experiment. The details of the method follow from the aims and design and they are addressed in later chapters. In any research, particularly field research on human behaviour, there will be risks of bias or noise in the data, and aspects of the research that reduce confidence in applying the findings more widely to the population. While it may not be feasible to remove all such influences, good design is critical both in achieving usable findings and interpreting those findings in the context of their known limitations.

The aims of EDRP evolved in the course of the project, as the policy, operational and technological questions changed. While this led to some imaginative approaches and a rich source of data, it also made interpretation of some aspects of the project difficult. Three of the four EDRP trials also changed design part way through, which caused problems for intervention quality, definition of control groups and cross-reference between parts of the analysis. The first lesson is therefore to be clear about the project aims and to ensure that those aims are represented faithfully from the start in the design, without distraction by alternative aims. The design is then determined by a balance of:

- experimental rigour;
- the affordable total sample size and study duration;
- the practical and commercial reality of particular experimental interventions;
- precision about the impact of specific interventions vs identification of effective combinations of interventions.

The latter requires a good understanding of both the likely effects of the interventions used and the combinations that are likely to be relevant in practice. As an example, if even five different interventions are being considered, there are 31 possible deployments of one or more of these interventions.[3] This is a large number, before considering control groups or the division of the sample into population segments. Hence some reduction of the combinations is likely to be necessary and the knowledge gained from the project will depend partly on the extent to which the reduction is systematic and governed by the overall aim of the trial or evaluation.

This reduction was partly successful in EDRP. In the EDF Energy trials, for example, some successful combinations of interventions were identified. While the trial design allows some deductions to be made about the “Key ingredients” of the successful combinations, there remains some uncertainty. E.ON’s smart meter trials adopted an incomplete but structured set of combinations, the first group having smart meters only, the second adding more frequent (monthly) bills, the third adding energy efficiency advice and the fourth adding RTDs. This was successful in showing that electricity consumption was achieved only in the group with RTDs but leaves open the question of whether the RTDs would have been effective without the advice and monthly bills. SSE had some fully factorial elements in the design, which enabled stronger deductions to be made about effective combinations, but left some concerns over sample size in each group.

Hence the design must follow from the aim and the main questions that represent the aim but a single evaluation is unlikely to answer all the questions. Nevertheless, a large evaluation project is likely to be more informative with central coordination to ensure that different elements of the evaluation or the work of different suppliers add up to a coherent overall design. This central coordination should include input from experienced professional researchers, to ensure there is a focus on the quality of findings.

---

[3] Based on a factorial design in which every possible combination (including single interventions) is represented.

The design may also be affected by whether the evaluation is purely about the smart meter roll-out or also the interaction with other policy initiatives targeting household energy use. For example, the Green Deal, Feed-in Tariff (FIT), Renewable Heat Incentive (RHI), Energy Company Obligation (ECO) and the Community Energy Saving Programme (CESP) will change the context in which behavioural interventions are played out. Changing the physical form and fabric of the property, the supply of electricity and heat, and the social, legal and financial context will change the means, motive and opportunity of householders to engage in behaviour change. This, in turn, will impact on the efficacy of different packages of behavioural interventions in different ways. The capacity to evaluate the impact of existing behavioural change packages, and to develop and test new packages of behavioural interventions on a continuing basis will be important. In addition, there is a possibility that smart meters themselves, independently of any effects they have on energy demand, will be a useful tool for evaluating the impact of other policy initiatives.

In this report, the groups (e.g. of households or customers) defined in an evaluation are referred to as experimental groups if one or more of the interventions tested by the evaluation is applied. In this sense, merely installing a smart meter may be counted as an intervention if a comparison is made with other groups that do not receive a smart meter. Groups used for comparison purposes, to determine whether the intervention had any effect, are termed control groups. The evaluation itself is treated as a trial of the interventions and the trial groups are both the experimental and the control groups.

The project design is more than just a list of experimental groups and control groups. Other elements of the project method will determine the true design. The timetable for recruitment and scheduling of interventions, for example, affects the experimental relationship between trial and control groups. Awareness of the trial (and whether participants opt in, have the chance to opt out, or are included blind) introduces another factor into the design – SSE showed that this has a significant effect on consumption, fortunately not in interaction with other interventions, but it nevertheless needs to be fully controlled for and balanced across groups in the design.

2.2 Experimental groups

Experimental groups receive one or more interventions as part of a trial. The selection of interventions should be based on:

- which interventions are likely to be effective (individually or in combination), based on experience from past research and an understanding of behaviour change;
- the expected magnitude of benefit in different population segments and the population as a whole;
- any known risks;
- the current state of technical or marketing development and whether a more effective version of the same type of intervention is likely to become available soon;
- the practicality and cost of using the intervention on a large scale.

The effectiveness of various interventions is reviewed in detail as part of EDRP and summarised in Chapter 8 of the final report. There is little merit in repeating that review here and, in any case, the evidence is constantly being added to: the specific combinations of interventions that are most effective in achieving reductions in energy demand, and the details of how and when they should be implemented, are only partly understood. It is becoming clear that electricity and gas consumption are not affected in the same way but the details are only now starting to emerge. The way in which different segments of the population can be engaged, and how they will respond to interventions, also merits further investigation. Prior to any further research, the EDRP literature review should be updated in relation to the specific interventions of interest.

In relation to combinations of interventions, the direct evidence can usefully be supplemented by theoretical consideration of how combinations will deliver the means, motive and opportunity for householders to change their behaviour.

- The *means* is the technology (a characteristic of the building fabric or services) or behaviour that will lead to reduced energy use and/or carbon dioxide (CO_2) emissions. This includes the person making the change having knowledge about current consumption and the technology or behaviour that would reduce consumption.
- The *motive* is the reason why households will want to make the change.
- The *opportunity* is the resource (e.g. time, space or money) to make the change.

In other words, for householders to reduce energy demand, they must know what to do, have a reason for doing it and have the resources to do it. Money appears twice in the framework – as both motive and an essential element in opportunity. But other motives are also important – both those that relate to environmental impact and those that do not (e.g. social influences or achieving a personal sense of control over energy use).

Table 2.1 shows the principal expected contributions of each type of intervention. It is clear that no intervention offers the full package and that combinations therefore need to be designed that will complement each other. This is evidenced in specific cases in the following sections but the table allows other possibilities to be identified, particularly where supplementary interventions need to be identified to support a main intervention. Where they are helpful, more detailed models of behaviour and theories of behaviour change may be applied; while they have limited predictive power, they can be useful in establishing a language for in-depth discussion and hypothesis-building.

Table 2.1 The principal means, motive and opportunity contributions of each intervention

	Means			*Motive*			
	Technical information	*Behavioural information*	*Technology*	*Environmental*	*Financial*	*Other*	*Opportunity*
Energy efficiency advice	✔	✔		✓	✓	✓	✓
Historic feedback	✓	✓		✔	✔		
Benchmarking				✓	✓	✔	
Customer engagement using targets				✓	✓	✔	
Smart meter	✓	✓		✓	✓	✔	
RTD	✓	✓		✓	✔	✓	✓
RTD with heating controller	✓	✓	✔	✓	✔	✓	✓
Time of use tariff/incentive				✓	✔		
Incentive to reduce consumption					✔		✓

Large ticks indicate the likely major elements, small ticks the other likely elements.

Having identified the interventions of interest, the detailed design of those interventions is critical and this is covered in Chapter 4.

The clearest trial designs will compare interventions in different groups of participants. It may sometimes be necessary to apply interventions in sequence in the same group, for example if there are too few participants having targeted characteristics or if one intervention must be established before another is feasible. In such designs, it is not known whether the later interventions would have been equally effective without the earlier ones, even if the earlier ones were ineffective on their own. In such cases, each intervention should be in place for a full year or more so that each intervention:

- can be related back to the full year baseline;
- relates to the same seasons;
- has time to take effect and – if the effect is not persistent – to decline;
- is not confounded by the Hawthorne effect.[4]

It is also worth considering whether particular behaviours or groups of behaviours should be targeted (in the design of interventions, the overall design of a trial or evaluation, or the analysis of data). The EDRP literature review shows that most changes seen in trials of domestic energy demand reduction tend to be – in the short term at least – those that require little investment of time or money.[5] But this is an incomplete characterisation of behaviours, which vary widely in the time and space required and in the cost associated with them. It is also important to understand that these resources may be perceived quite differently according to the spare money, time and space a householder has, and the facilities available in the home. Turning down a thermostat costs nothing if there is already a thermostat; it costs more if there is a central heating system without a (functional) thermostat. It costs even more if there is no central heating system at all. It also matters whether the householders own the space – i.e. whether they are owner occupiers.

Cost and time will also depend on the knowledge and skills of the householder. The cost of insulating a loft, for example, is higher if the householder is not physically able to do it him/herself (but paying someone else to do it will reduce the time demand). The time demand is greater if s/he has to learn how to do it, clear a lot of stored items out of the loft, then (because of inexperience) work at a slow pace to ensure everything is done correctly.

Hence, a simple division of energy efficiency measures into “behaviour” and “installation” does not adequately represent the options. Insulation is not a behaviour, but installing insulation is a behaviour. Turning down a thermostat is a behaviour but it requires the technology to be installed. This needs to be taken into account when selecting and designing interventions.

2.3 Control groups

The purpose of control groups is to know what would have happened in the experimental groups had there been no interventions. In a true “randomised control trial”, participants are first recruited and then allocated randomly to groups. Where this is not possible, the challenge is to get as close as possible to random allocation and to obtain data on the groups to provide evidence on whether or not they are well matched. EDRP showed that, even with strenuous efforts to match trial and control groups, large differences in pretrial consumption can remain.

The control achieved should be not only for deliberately applied interventions but also for extraneous factors that might apply to some or all participants. For example, during EDRP, consumption might have been affected by changes in the economy generally or energy prices, outdoor temperatures, press coverage and national or local energy efficiency campaigns. It was assumed that the effects of such factors are felt equally across all groups; it is

[4] People can change their behaviour because they are being observed rather than because of the intervention itself. This effect tends to be more prominent at the start of a study and then dissipate unless a new intervention is introduced.

[5] Unsurprisingly, changes requiring greater investment are more likely in higher income households and when people move into a new home. Those who do invest tend to save most energy and higher savings are also associated with higher levels of education or income; larger homes and households; strong environmental values; and younger people.

probably necessary to make this assumption but it is potentially flawed (e.g. because these factors create a pressure to reduce energy use in the control group, thus making the baseline change more difficult to exceed in the experimental groups, or because trial participants who would have made investments in energy saving became concerned about spending capital).

It may be necessary to have more than one control group, particularly if the trial involves both smart and non-smart meters, either to control for different elements of an intervention or to match more than one experimental group. For example, if one or more experimental groups is unaware of the trial, the control group should also be unaware and a "business as usual" control group may be appropriate (to control for the influence of extraneous factors). If some experimental group participants are aware of the trial then they should be matched with control group participants who are made aware in the same way (e.g. by participating in a customer survey or simply being told they are being monitored). If the control group is offered an incentive to take part, then the experimental groups should also receive the incentive. Depending on the experimental design, one experimental group may also function as a control group for another experimental group in part of the data analysis. For example, a group receiving a smart meter and no other intervention may function as the control group for a group that is additionally provided with an RTD.

For statistical purposes, ideally the experimental and control groups should be matched at household level so that the groups as a whole are well matched. The disadvantage of this is that a drop-out in one group means elimination of data from the other group too. An alternative is a technique called propensity score matching that defines matches of cases based on their characteristics as measured during the trial. Such approaches in effect allow the control group to be defined after data collection, which may offer a major advantage in the context of a large-scale roll-out.

Where a trial is set in the context of normal business operations, there will probably be decisions to be made as new business offerings become available to customers in general (this happened, for example, in the E.ON trial when historic feedback became part of business as usual). If the trial were the overriding consideration, either the business would have to remain static for the period of the trial or the trial participants would have to be wholly isolated from the changes. Given that neither of these is likely to be plausible, other approaches need to be considered. Depending on the exact circumstances, any or all of the following can be considered:

- anticipate business changes at the start of the trial and build them into the design;
- manage the timing of the business change so that there is a long enough period before and after its introduction to isolate its effect on the trial;
- introduce the changes in a staggered fashion (e.g. regionally) so that any effects on the trial can be identified.

Needless to say, a requirement for any of these to be successful is a strong line of communication from the wider business to the trial managers and a strong line of control from the trial managers to the business as a whole, so that nothing happens by accident, without awareness of the trial.

Some of the control group may have already received one or more interventions, independently of the trial (e.g. they might have bought their own RTD or received one as part of a general distribution, or they might have access to the supplier's website or some parts of it). Excluding such participants from the trial will make a cleaner distinction between the experimental and control groups but also make the sample less representative. The best compromise will need to be decided on a case-by-case basis, depending on the aims and design of the trial.

2.4 Baseline measurements

It is essential to establish the energy consumption of experimental and control groups prior to interventions commencing. Ideally this would be a period of at least a year so that experimental effects can be determined for different seasons and for a period long enough to detect any changes in the magnitude of effect over time. A shorter period may be sufficient if the planned intervention is intended to have an effect for only part of the year (e.g. an intervention targeting heating or use of swimming pools). However, this would entail a gap in data collection and this should be avoided unless the data collection itself is disproportionately expensive.

If the design involves a sequence of interventions over a period of time, each intervention may represent a baseline for the next one but the initial baseline should also be used in each case so that the total effect can be seen.

Groups should be matched on their baseline consumption (and other relevant variable) in order to avoid risks of differential potential to change consumption or regression to the mean.[6] This implies that either (a) baseline monitoring should precede allocation to groups or (b) some cases may need to be removed from the trial in order to match groups. Some differences can be tolerated and controlled for in the analysis but large differences are problematic. Large differences in consumption may not be fully controlled for in the analysis and even non-significant pretrial differences are still relevant to the analysis.

[6] Regression to the mean refers to the statistical phenomenon whereby values above (or below) the mean are likely to decrease (increase) if measurements of the same group are made a second time.

3 Sampling and recruitment of participants

3.1 Introduction

Sampling and recruitment are the means by which a trial is based on participants who are well matched to each other and representative of the population of interest in the trial. The population is not necessarily the national population – it could be a particular segment of the population such as prepayment customers or fuel-poor households.

The representativeness of the final sample depends on:

- the sampling base (the “list” from which candidate participants are selected);
- the sampling procedure;
- bias during recruitment, from either refusal or inability to participate or exhausting the supply of some participant types before others;
- bias during deployment (e.g. as a result of technical problems with installing equipment);
- selective attrition during the trial.[7]

A field trial is never perfect at every stage and often has imperfections at every stage. Hence it is important to track representativeness (of the sample as a whole and each trial group) and, where possible, use the analysis to correct for any biases. Where feasible, it is valuable to conduct some monitoring of households that were approached but did not participate (e.g. by tracking their energy use or inviting them to participate in a customer survey). A low recruitment success rate (i.e. a high level of self-selection) does not necessarily mean that there is experimentally relevant bias, so monitoring can make the difference between credible and non-credible findings. Nevertheless, it is difficult to control for all relevant variables or even to know what the most important variables are: propensity to reduce energy demand may not be well represented by variables that are easy to measure without the cooperation of the participants. Therefore, achieving a high recruitment success rate remains important.

3.2 The sample

3.2.1 Sample size

EDRP to some extent applied statistical methods for determining required sample sizes, based on anticipated variances and mean differences between groups, and the level of statistical significance demanded. Such “power calculations” should be more rigorously applied in the planned smart meter evaluation. More than the overall sample sizes, there is a need to ensure that sample sizes are still sufficient where subgroups are to be analysed separately.

In some EDRP trials, the sample size varied greatly between groups. This was partly because some interventions can more easily and cheaply implemented in a large number of homes. While it is understandable that sample sizes should not be reduced unnecessarily for some groups, purely to match the sample sizes of groups with more challenging interventions, the strategy is not without problems. In particular, larger sample sizes afford greater certainty in findings – this needs to be taken into account when claiming that one intervention has a significant effect but another does not.

---

[7] Retention needs to be monitored to determine whether the characteristics of the sample as a whole or certain groups is changed, especially if the reasons for leaving the trial may be related to the trial or otherwise cause bias. Even where attrition is due to change of supplier or moving house, this might bias the sample in relation to mobility, ease of changing supplier and changing financial circumstances.

Statistical power is greatest when experimental and control groups are the same size. This is often not understood and, consequently, experimental groups tend to be larger (which is also likely to increase project costs).

Determination of sample sizes should allow for the possibility that some cases will be lost in the process of deployment (e.g. because equipment cannot be fitted in all homes) or during the trial because of participants changing supplier or tariff, moving house or being unable or unwilling to continue in the trial. In EDRP, retention in the trial was 65-82%.

3.2.2 Sampling base

The sampling base will be defined by the aims of the trial and the availability of data (including contact details) on potential participants (candidates). EDRP relied on customer databases and the samples would therefore have been broadly representative of each supplier’s customer base, but not the population as a whole. This is an almost inevitable consequence of the trials being supplier-led. Ideally the roll-out evaluation will be centrally coordinated so that a nationally representative sampling base can be used.

In particular, the EDRP trial samples differed markedly by region – partly by design and partly as a consequence of the varying customer bases of the suppliers taking part in EDRP (and proximity to depots handling installations, which would also cause location to differ between trial groups). Fortunately, the evidence from EDRP is that location at macro scale (region of the country) does not have a great effect on response to interventions, suggesting that weather/climatic differences are not the most important to be concerned about.[8]

Nevertheless, other locational variables might be relevant (e.g. urban vs rural location, more and less affluent areas). Data in the sampling base should therefore be sufficient to match groups directly on location or indirectly using demographic classifications that include a geographic element, and the analysis should control for these variables. This would control for local (e.g. sociodemographic) differences and changes (e.g. a local authority or newspaper running a campaign or a major employer making redundancies).

If a trial is intended to target particular groups, the sampling base and/or recruitment process will need to support the targeting. The sampling base may be applied if the targeting is based on past energy use, location or some variable that can be derived from location. In other cases, it is unlikely that targeting can work at the level of the sampling base. For example, it has been argued that targeting an intervention on the households best equipped to use it (in terms of motivation and understanding) could multiply the benefit per targeted household (although the total benefit would necessarily be reduced and a wider range of households could benefit if given some additional support). Such groups would need to be defined in the recruitment process.

3.2.3 Selection criteria

Certain households or individuals may be excluded from sampling for a range of reasons related to the project design, quality of baseline data, commercial conflicts of interest, the practicalities of making contact or implementing interventions, and health and safety considerations. Generally, exclusions should be the same for all trial groups so as to avoid variations in selection bias. Where group-specific criteria are necessary, any customers excluded on these criteria should not be transferred to other groups.

---

[8] Within the range covered in sufficient numbers by EDRP. There may nevertheless be variations in the more extreme highland and island climates.

Looking at the various exclusions applied in EDRP, in some cases there is a clear and logical justification. For example:

- there could be conflict of interest, inside knowledge or perceived obligation to participate if suppliers include their own staff in a trial (although the numbers involved in a random sample are likely to be small);
- there would clearly be confusion if the same participant were included in more than one energy-related trial (which may need cooperation between suppliers to avoid) or had already received one or more of the planned interventions as part of business as usual.

There is also an ethical case for avoiding unnecessary health and safety risks (e.g. not seeking to reduce the energy use of households that are already using so little energy that they may be inadequately heating their homes or placing themselves at risk in other ways). In other cases, the exclusions are merely defining the population of interest (e.g. credit customers or customers without electric heating).

Some exclusions were technical. Some of these were generic (e.g. reflecting the space available for a smart meter or the limitations of HAN and WAN communications). It is to be hoped that such problems will be solved before the smart meter roll-out (for example, the distance limitation disproportionately excluded flats). Other exclusions were specific to the intervention (e.g. the need for a broadband connection or central heating system, or the absence of heating programmer) – while this would make sense for the simplicity and finances of the trial, it could exclude the very people who might benefit most from the intervention.

In other cases the exclusions are logical in the context of the trial but nevertheless risk creating a biased sample and/or making the findings less widely applicable. Examples are exclusion of homes under two years old and customers without two years of pretrial data (with a minimum number of actual readings per year). These requirements were necessary in order to have baseline data for the trial but they are likely to create a bias towards households that are less mobile, more likely to be at home during the day or have a more positive view of their energy supplier, hence more likely to take action on energy efficiency issues. At some point, it will be necessary also to understand how to prompt behavioural change among people who are moving more often or new to a particular home, or in a newly constructed home.

Other exclusions may have been overly cautious or based on convenience rather than good experimental method, such as exclusions on the basis of:

- recent meter replacement;
- non-standard tariffs;
- a history of payment problems or disputed bills;
- the supplier not having a valid telephone number for the customer;
- the customer not speaking English.

Such exclusions could create unnecessary selection bias and/or limitation of applicability of findings. Targeted projects or stratification might be appropriate in these cases.

3.2.4 Stratification

Stratification refers to the adjustment of sampling rates to ensure that certain population groups are sufficiently represented in the sample. It is generally used where one or more groups are important to the research but too few would be present if a random sample were taken. For example, a random sample of 200 households is unlikely to include enough people with medical conditions seriously affecting their finger movement to be able to draw independent statistical conclusions about how that group responds to RTDs. Stratification would allow half the group to be made up of such people, so that comparisons could be made with people having more typical manual dexterity. Stratification serves two purposes – to be able to draw conclusions about different groups and to reduce within-groups variance.

Stratification should be distinguished from three other approaches:

- targeting particular groups of interest, without comparison with other groups (e.g. a study of rural homes with no gas supply);
- providing extra support for groups that might otherwise be less likely to participate (e.g. elderly people, multi-occupancy households and customers who believe they already have low energy usage);
- controlling for differences between participants in the statistical analysis, based on a simple random sample.[9]

The average response of a stratified sample would not represent the average response of the population but the population response may be estimated by the process of "grossing" to adjust statistically for the over-representation of one or more groups in the sample. Grossing requires recording of which participants are in which strata and data on (a) the over-representation of groups in the stratification and (b) the population proportions of those groups. If the latter is not known, then stratification is risky because grossing may not be possible. Stratification on a large number of variables is also risky because grossing may accidentally over-represent some groups if membership of one stratum is correlated with membership of strata based on other variables. In practice, there was stratification but not grossing in EDRP, which places a caveat over the quantitative applicability of the findings.

Aside from the general principles of stratification, some observations can be made on the specific strata employed in EDRP.

A particular stratification used in EDRP was between customers who purchased only electricity from the supplier operating the trial and customers who purchased both electricity and gas (electricity-only vs dual fuel). This is a valid distinction and several findings did differ between the two groups. Nevertheless, the distinction did raise problems with interpretation, because (a) there was little information on why the two groups might have differed in their response to the interventions, (b) it was not always known whether the electricity-only customers had no gas supply or purchased gas from another supplier and (c) no gas-only households were included. These omissions should be avoided in future trials. Ideally, evaluation of smart meters should be coordinated between suppliers so that each participating household is evaluated in relation to all its energy use (including solid/liquid fuel, bottled gas and microgeneration).

Further research could also determine why electricity-only and dual fuel groups differ, the following being possibilities:

- if a customer buys only electricity from the supplier operating the trial, a gas smart meter would not be fitted and the RTD would display only electricity data, hence focusing conservation efforts on electricity rather than gas;
- focusing on electricity might also make the RTD itself simpler to operate;
- dual fuel customers are more likely to see energy efficiency advice as being relevant to them, because it is more likely that at least one specific piece of advice will apply;
- dual fuel customers have made a market choice to take both fuels from one specific supplier and might therefore be more energy-aware (or energy-cost-aware) and willing to look at advice from that supplier;
- the two customer types differ in how easily they could be recruited, and hence in their motivation at the start of the trial.

[9] Or other variations on random sampling that do not create an experimental bias in the sample.

The distinction between dual fuel and electricity-only customers was further complicated by the fact that gas smart meters could not be installed in all dual fuel homes (for reasons related to the space available for the meter and communications with the electric meter).[10] RTDs would also not have shown gas data in these cases. This affected 17-25% of homes across the EDRP trials and was not fully accounted for in the analysis of electricity data. Such failures should be less common in the national roll-out but they would need to be documented where they do occur and the affected homes acknowledged as being different in the analysis of data.

In similar fashion, credit and prepayment customers were treated as separate groups in EDRP. In this case, stratification was appropriate, rather than representative sampling, because the latter did not produce sufficient numbers of prepayment customers for analysis. This is another example of where stratification needs to be backed up by gathering supplementary information to understand differences between strata. Prepayment meters tend to be installed where customers are having difficulty paying energy bills, so they generally have lower consumption and therefore less scope to reduce. However, prepayment customers may be more practiced at monitoring consumption, more aware of costs and have greater (financial) motivation to save energy. With these counterbalancing factors, savings appear to be possible (as shown by SSE) but not always achieved (as shown by Scottish Power). The actual circumstances, motivations and behaviour of prepayment customers therefore need to be better understood if energy demand reduction findings are to be interpreted. Logically, efforts to help prepayment customers reduce consumption might focus on means and opportunity, the motive already being present, but this was not explored in the trials.

Economy 7 customers were generally either excluded from the EDRP trials or treated as a separate stratum but there was some confusion between Economy 7 customers who used storage heaters and those who did not (this needs to be avoided in future).

The EDRP trials also made use of indirect demographic classifications (Mosaic, CAMEO, ACORN and Ocean), in one of two ways: stratification to aim at equal representation of each group or sampling to aim at representation proportionate to population proportions. The former is more appropriate if the aim is to study the effect of demographic group, the latter is more appropriate if the aim is just to control for demographic group in the analysis of other variables. The CSE Fuel Poverty Index for the postcode was also used in one trial to define “fuel poor” and “not fuel poor” strata which did differ in their responses to interventions. This was not an effect of being “fuel poor” as such because households in that category were not directly identified. If this group is of interest in the roll-out then relevant households should be identified, rather than selecting on a postcode-based continuum.

The stratification that proved most troublesome in EDRP was based on pretrial fuel consumption. The issue was not stratification as such but rather the total exclusion of some strata and failures to record the exact basis of stratification. In particular, the findings from a group composed only of high consumers of electricity proved difficult to interpret because the control group was poorly matched and there was no comparable experimental group with medium or low consumption. Stratifying in this way, based on a key outcome variable, is risky because of the difficulty in distinguishing genuine experimental effects from regression to the mean.

Other variables were simply recorded and (where possible) controlled for in the analysis (e.g. dwelling type and age, household size and the age breakdown of the household). This was probably appropriate but, depending on the aims of future trials, stratification may be necessary in order to understand the behaviour of groups that would otherwise be insufficiently represented (e.g. very large households or people aged over 75).

[10] In some cases, it was the electricity meter that could not be replaced but this was much less common.

3.3 Recruitment

3.3.1 Opting in vs opting out

The basis on which participants are recruited is critical to the interpretation of the findings. A distinction is sometimes drawn between opting in (potential recruits are invited to join a trial) and opting out (recruits are told they are in the trial but have the opportunity to withdraw). The usual expectation is that opting out will entail less selection bias than opting in. However, this simple distinction does not cover all the possible options and is therefore often unhelpful. From the perspective of the customers, at least four distinctions are wrapped up in the two categories:

- awareness that there is a trial and they are part of it vs awareness that they are being observed (e.g. their energy consumption is being monitored) vs no awareness of a trial or of being observed;
- awareness explicitly communicated vs awareness by deduction (e.g. because a new meter has been fitted and they are invited to take part in a survey);
- decision to participate (opt-in) vs explicit withdrawal from the trial (active opt-out) vs non-cooperation (passive opt-out, for example by not reading advice or feedback material);
- opting in to the trial as a whole, before being allocated to a group within the trial vs opting in to a particular group (the former may create a difference between the trial sample and the population as a whole, the latter may additionally create a difference between trial groups, so a decision may need to be made between representativeness and experimental control).

Different interventions naturally lend themselves to different approaches. Two examples can illustrate this.

(a) If the intervention is simply mailing energy efficiency advice to the customer, this can easily be done without the customer being aware of being in a trial and no opt-in is required. However, unless engagement with the advice is monitored, the trial will not distinguish between customers who read the material and those who passively opt out, hence the true impact is not known and the likelihood of seeing an overall significant effect is reduced in two ways: the average savings are reduced (because only some customers are benefiting from the intervention) and the variance in savings is increased (because, within the trial group, there is a difference between customers who do and do not read the material). If engagement is monitored, that represents an additional intervention and its effect needs to be evaluated (see Chapter 6 on customer surveys).

(b) Where new equipment is required, it is more likely that the customer’s permission will be required. This means that customers will be asked to opt in to the installation (though not necessarily with any awareness that there is a trial) but can still passively opt out of engagement with the equipment (initially or after an initial period, e.g. when batteries need to be replaced). If asked to install the equipment themselves (sometimes the case with clip-on RTDs), there is a further level of passive opt-out that was prevalent in EDRP: they do not install it. As in the first example, if engagement is monitored, that represents an additional intervention and its effect needs to be evaluated.

Depending on the details of the trial, the opt-in/opt-out strategy may raise issues with both the validity of the findings and the ethics of the trial itself (in relation to whether participants are in any sense deceived).

Opting in has the potential to create a bias for the sample to be composed of customers more interested in energy saving (for financial or other reasons). This bias would be expected to increase with the demands of participation but decrease with the perceived value of the intervention to the customer (e.g. whether they receive a product they value as part of the trial, or some other incentive).

Opt-in bias is often assumed to make the interventions more likely to succeed but it is also possible that those most interested in energy savings will have already taken action and therefore have less potential for further action. While the empirical evidence is not definitive, there does not appear to be a strong overall influence of the take-up rate in opt-in trials. Interventions may have the greatest effect on people who have not yet taken much interest in conserving energy but who could be motivated in the process of recruitment and informed how to use the intervention to fulfil their newfound motivation. Hence, it is not merely the fact of opting in that matters, but the nature of the communications and incentives involved.

While overall sampling bias is important, a difference in bias between groups is likely to be more problematic for the interpretation of findings. In particular, recruitment success rate may be higher for control groups (if they do not have to do anything as part of the trial) or lower (if they have to do something but do not perceive any benefit). In EDRP, the control groups tended to have higher recruitment success rates. If the recruitment to groups receiving different interventions has a different basis, differences in outcomes may be attributed to the recruitment basis rather than the interventions – this in particular should be avoided.

Decisions on the opt-in/opt-out basis of recruitment may not be guided solely by how precisely the findings can be interpreted. There is a reasonable argument that some interventions would only ever be offered on an opt-in basis, so there is little point in evaluating them in any other context. This might apply in particular to the provision equipment or service that incurs a cost to the customer (e.g. an RTD with a higher specification than one offered at no cost to the customer), or might incur a cost (e.g. time of use tariffs).

In making decisions based on realism, it is still important to ensure the project design is able to distinguish between the effect of opting in (to a trial or just to the receipt of the intervention) and the impact of the intervention itself, particularly if it is being compared with other interventions. The very act of opting in (including the information acquired and used in the process) may affect the perception and understanding of the intervention. EDRP, for example, showed that the effect of a smart meter on gas consumption may depend on some aspect of the experience of getting a smart meter, an aspect that is absent if the meter installation is presented as “business as usual” and customers are unaware that they have a smart meter.

Consideration also needs to be given to long-term strategy. For example, an intervention may be effective among those who initially opt in but fail if the evidence from the trial is applied to marketing more widely. Those who initially opt in are unlikely to be self-selected purely on the basis of interest in energy consumption: they may also differ in their capacity to use the intervention (e.g. because of the type of home they occupy, the appliances in the home, flexibility over when activities are carried out in the home or their personal knowledge and skills).

3.3.2 Schedule

The schedule of recruitment (and, if different, the schedule of joining the trial) is an important consideration for the success of recruitment, the representativeness of the final sample and the options for statistical analysis of the trial data.

Matching groups

In most trials, a key aspect of controlling for non-experimental variables will be having one or more control groups, monitored over the same period as the experimental group(s). In this way, if all groups are also well matched by location (see Section 3.2.2) they are exposed to the same variations in weather and other external circumstances. It is important that monitoring and any awareness of the trial or interventions commences at the same time in all groups – it is not sufficient merely to analyse trial data by comparing groups over a period when their participation in the trial overlaps. This is because (a) early recruits may differ from later recruits and (b) behaviour may vary with duration of involvement in the trial and this also needs to be controlled for.

Such matching is more easily stated than achieved because there can be time lags in recruitment and in the implementation of interventions. Lags in recruitment may arise because of varying willingness to join different groups. This can be handled by assigning participants to groups after recruitment (and this has other advantages too, as detailed elsewhere). The penalty for this, however, may be greater difficulty in recruiting and a greater overall selection bias. The choice therefore depends on the size of the population available for sampling and on the relative importance of overall bias in the sample and differences between groups within the sample.

Recruitment or intervention lag may arise because of:

- relative ease of recruitment to different groups (if allocation to groups precedes recruitment);
- variation between strata in the population available for recruitment;
- inherent delays of implementation (e.g. the time taken to install equipment in many homes or because the intervention is tied to the billing cycle);
- technical problems or failures in trial administration.

In some cases, the lag may differ between elements of an intervention (e.g. advice may be mailed simultaneously to all participants but fitting RTDs in a large number of homes may take longer). This is to be avoided (unless it is deliberately part of the design), the obvious method being to adjust everything to the schedule of the slowest element.

Lags can be managed by matching groups at household level, so that groups with less inherent lag have recruitment and commencement slowed to match the slowest group.

If there is not thorough testing of equipment and procedures prior to the trial, losses from the sample may decrease in probability during recruitment, risking bias between groups if recruitment of all groups does not follow the same schedule. It is also worth noting that the period between recruitment and installation or commencement of interventions should be as short as reasonably possible in order to minimise drop-out.

Transitional periods

If lags are not managed, a second problem that can arise is defining the start date of trials. There will be a period between baseline and in-trial periods when some, but not all, participants are recruited and (where applicable) have interventions implemented. This is partly inevitable because it takes time to install equipment in large numbers of homes but there were also cases in EDRP where delays were caused by logistical problems (e.g. changes in the customer database, delays to updating records or meter-reading schedules, cancellation of mailings and problems with the volume of data to be processed).

The impact of transitional periods is greater if data are aggregated over a period (e.g. a transitional period of a month might affect three months of data if aggregated at quarterly level). The impact can be mitigated if all participants are moved to the same billing cycle prior to the trial but this is difficult to manage without exposing participants to the (minor) additional intervention of a shift in billing date, entailing either a small interim bill or a larger than usual bill for one period.

Options for managing transitional periods in the analysis include defining start of the trial period as being the same for all groups (e.g. the first installation, the last installation or when 50% of installations have been completed) or excluding all data from the transitional period. The first option introduces noise into the data; the second makes it less likely that initial transient effects of the intervention will be detected (and may also require the overall trial period to be longer in order to gather sufficient data). The second approach is likely to be preferred unless transient effects are considered important in a particular trial.

Topping up

In the course of a trial, sample attrition may be greater than expected (or some complexity may appear in the emerging findings, implying a need for a larger sample) so that new recruits need to be found. This is a difficult operation and to be avoided if at all possible (e.g. by recruiting larger numbers initially than suggested by assumptions based on knowledge at the time). Where topping up is necessary, some guidelines should be followed:

- the additional sample should be large enough to analyse independently and to compare with the original sample;
- new recruits should be clearly identified in the database;
- all groups that are to be compared with each other should be topped up in the same way, even if numbers are sufficient in some groups;
- the recruitment process should be the same as for the original sample and the same interventions should be applied.

Start date

A final consideration is that the impact of an intervention may vary with the start date. For example, there may be a greater impact on heating if an intervention commences during winter because it is more relevant at that time of year and customer engagement is delayed or lost if the intervention commences in summer. This would be in addition to any seasonal variation in effects on energy demand, once the intervention is under way. This is rarely or never accounted for in published evidence.

3.3.3 Approach to participants

The way in which potential participants are approached is critical to the success of a trial, not just from the perspective of the proportion of candidates who become participants, but their understanding of the trial, their response to interventions and the interpretation of findings. Recruitment may fail altogether because of failure to make contact (initially or as interventions are implemented) or to gain agreement to participation. But there can also be partial failure in terms of the information conveyed to and collected from the participants.

Where recruitment of participants is successful, an intervention may still fail for technical reasons. This would generally have less effect on trial outcome that sampling bias, assuming the failed case is removed from the study (which did not always happen in EDRP). It would, however, render the trial less representative. Issues identified in EDRP, which will need to be guarded against in the roll-out, are described in Chapter 8.

Engaging participants

It is important that participants who are aware of the trial give informed consent to participate. This means that they should understand as much as possible (within the constraints of the project design):

- the overall subject and aim of the trial;
- what the trial organiser is going to do;
- what they as participants are expected to do;
- any risks or benefits to them as participants.

There may be experimental reasons for withholding some or all of this information but this must be considered in the context of both research ethics and the impact on the trial validity.

Depending on the complexity of the trial and the demands on the participants, the approach might be made by letter, telephone or in person (other means, such as advertisements, email or social media, are unlikely to be widely applicable but might be used for focused studies of certain targeted groups). There are clearly also cost implications of the selected approach.

One key point is to make the means and content of the approach consistent across trial groups, between participants and over time, to avoid bias and noise in the data. This does not necessarily mean giving all participants exactly the same information – it could mean a consistent “route map” for the recruiter to identify what would motivate the individual candidate. More complex interventions, for example, may be more difficult to recruit for and have more limitations on eligibility, so it is essential that there is a means of keeping recruitment rates consistent. Nevertheless, unnecessary differences should be avoided. For example, one EDRP trial varied the wording between interventions, saying that smart meters would either "reduce", "avoid” or "remove any need for" estimated bills – all ambiguous but with different nuances of meaning.

Initial contact in person is likely to be excluded in most trials because of cost and rising suspicion of cold calling, unless households need to be visited for some other reason (e.g. to read or exchange a meter). While contact in person may be successful in terms of the number recruited, this depends on what the household is asked to do. Large numbers of households, for example, rejected the RTD offered in cold calls during EDRP. It can also be difficult to maintain consistency in the approach across a large number of callers: each caller should be competent to answer all likely questions consistently and accurately without checking back to a central point. Failure in this area could reduce recruitment success and/or introduce selection bias.

Nevertheless, any visit to implement an intervention should be seen as part of the recruitment process (e.g. EDRP reports numerous cases of installation being refused after making appointment).

Approach by telephone tends to be cheaper and makes it easier to maintain consistency and answer questions. It is nevertheless important to ensure that callers are competent to answer all likely questions. In one EDRP trial, for example, an agency was used to make recruitment telephone calls but the recruitment script was not found to be adequate and staff were given additional training to be able to explain one of the interventions.

The major limitation of recruitment by telephone is simply access to telephone numbers – it emerged from EDRP that even the householder’s energy supplier often does not hold a valid telephone number (this could perhaps be improved before, or as part of, the smart meter roll-out). If a telephone number is available, some householders will not want to take the call (at all or at that particular time). EDF Energy observed that some customers are exceptionally difficult to contact. Even when recruited and "installed on to the trial", gaining access to the home was not possible for around 5% of participants. The principal reason for installation failure was refusal of access to site.

If contact is successfully made, there is a balance to be struck between managing all aspects of recruitment in a single call and keeping the call short to avoid losing the candidate. In any case, the caller should be honest about how long the call is likely to take (e.g. “I only need three minutes of your time but I’m happy to talk for longer if you have more questions”).

Contact by letter has the advantages that it is cheap, includes every household (if there are non-English versions available) and can include more information that the candidate can review at leisure. The problem is that the filtering process (of letters being opened, read, understood and responded to) generally results in a low response rate (e.g. 6% in the Scottish Power trial). The ideal combination may be letters followed by telephone calls but the message and terms used should be consistent between the two approaches so that it is easy for householders to make the connection between them.

Recording information

There are four main purposes for information collected from recruits, i.e. to:

- ensure they are eligible;
- achieve a sample that is representative according to chosen criteria;
- provide data for the analysis of findings, to control for confounders or break down the analysis by subgroups;
- provide a baseline from which to identify changes during the trial (e.g. in the appliances present, household circumstances or energy-related behaviour).

While more lengthy questionnaires might be used later in the trial, recruitment questionnaires should be universal, kept to a minimum length and use simple, clear, unambiguous questions with no technical or industry jargon.

3.3.4 Recruitment materials

Unless there is a good experimental reason, there should be clear advance explanation of the intervention(s), so that participants know what to expect – to enhance the intervention and avoid bias to participants with a better prior understanding of the incentive. Within the specification of information to be communicated, there are issues of:

- balance (between a large volume of highly detailed information and providing participants with the essential information sufficiently concisely that they are likely to absorb it);
- timing (of what needs to be communicated at some point in the recruitment process or later in the trial);
- clarity (using simple phrasing and avoiding the jargon of research or the energy industry);
- emphasis (on different aspects of the trial).

In particular, emphasis is likely to be a compromise between maximising participation (by emphasising the benefits to the participant other than saving energy) and maximising impact (by emphasising how to save energy, or energy costs, through participation). Where one or more of the interventions has an obvious value to the participant other than saving energy (e.g. a Freeview box or a new heating controller), this needs to be handled particularly carefully. Such items may well be a useful incentive to join the trial but they may cause a bias to participants who are interested in the item rather than saving energy. There is also a risk that participants recruited because of an item of apparent value will quickly lose interest in the trial; EDRP participants in the TV group thought the TV intervention seemed attractive at the outset but actual engagement was low and declining. Attractiveness of intervention does not define its effectiveness. There is a clear case here for recruiting participants before allocating them to groups.

Similarly, creating a positive impression of the proposed intervention may aid both recruitment and engagement but could backfire if the participant perceives the description as having been exaggerated or omitted key information (e.g. if real-time feedback is expected but it turns out to be daily feedback, delayed by a day).

Recruitment in EDRP used a number of different assertions (individually or in combination, in addition to the trial interventions themselves) to motivate the customer to take part in the trials:

- saving energy, carbon or money;
- avoiding climate change;
- protecting the environment;
- having the latest technology (smart meter itself or a specific intervention such as a "small portable display unit incorporating a new programmer and controller for your central heating system ... benefit of being able to

programme up to three different heating and hot water sessions per day, which can be different for weekdays and weekends" or a "free, top of the range TV set-top box");[11]

- limited availability of the offer;
- being one of the first UK households to benefit;
- doing what other people are doing;
- online/mobile phone support;
- “interactive equipment” or specific descriptions of the equipment to be provided;
- more information about energy use, including specific services such as annual email summaries, plots of consumption against weather conditions, and forecasts of future energy usage;
- easier or more frequent information on energy use;
- being in control;
- no meter readers (this could backfire if the householder is concerned about employment);
- no need for estimated bills;
- more frequent bills (conceived as a benefit but with a risk of being perceived as a threat);
- advice/audit;
- the project was backed by Government and Ofgem;
- incentive payment/vouchers;
- no catches;
- no additional cost to the customer.

Early feedback was that the “save the environment” message was not effective but there is otherwise little evidence on the effect of different messages. There was also a specific issue with explaining time of use tariffs because customers often did not understand load-shifting or why suppliers would want to save them money.

Arguably the approach to candidates will be more successful if it uses the two or three messages that are most attractive to the candidate but these are likely to vary between candidates. Listing all possible reasons for participating would in theory maximise the chances of at least one message being relevant but could also lose candidates in the complexity (which could, in turn, increase bias towards those who are already energy-aware and interested in participating). It is also possible that some messages could put off as many candidates as they attract. For example, mention of Government could be offputting for some people, especially in the context of providing personal information as part of recruitment).

There is a case for qualitative research to understand the recruitment process better, possibly in the form of follow-up interviews in current research projects. It would also be useful for trials to monitor what motivated each recruit and control for this in the statistical analysis.

Pending further research, there is no definitive list of what information should be offered during recruitment but it should be expressed from the householder’s perspective, taking into account the possible motives to participate, credibility and convenience.

[11] It is worth noting that descriptions intended to “sell” the intervention may make it appear too complicated to some households, or too good to be true (e.g. an expensive piece of equipment offered free of charge). Either could cause bias in the sample.

Letters need to be sufficiently formal that they are taken seriously but not so long or dry that they attract no interest. They should be short, simple and clear, avoiding:

- typographical or grammatical errors (e.g. “data is”, “amount of units”);
- prejudice (e.g. referring to the meter reader as “he”);
- jargon or terms that might have a different meaning to the participant (e.g. “trial”);
- inconsistency (e.g. one EDRP letter said there were a limited number of smart meters available but the supplier wanted to give one “to every customer that wants one”);
- ambiguity.

Where the householder needs to enter into a formal agreement about participation, this should be made as simple as possible and not appear to place excessive demand, risk or liability on the household.

4 The interventions

4.1 Introduction

The selection of types of intervention was covered in Chapter 2; this Chapter addresses the design of the selected interventions for maximum effect. Of course, the reverse order of thinking may also apply in the sense that an intervention may not be chosen if the trial organiser does not believe it can be made sufficiently effective.

Each type of intervention is addressed in a separate section but a few general points are made here.

- The intervention is not just the equipment, feedback or incentive – it includes the associated user guidance and the media by which it is communicated, to establish the ideal balance of complex functionality and simple, attractive products.
- Interaction between installation and customer service teams and customers is itself a potentially effective intervention.
- Awareness of interventions is often poor and this is the first hurdle to overcome.
- Interest in (and information on) interventions is often limited to one member of the household, who may be the person named on the bill but not necessarily the person with most interest in (or time to engage with) saving energy. In fact, interventions can cause conflict in a household, for example when one member wants to make changes in thermostat settings or appliance usage and another does not. Hence there is a case for finding ways to brief the whole household.
- Interventions should not be seen as one-off actions: there is a need to maintain engagement with them (e.g. by designing interventions that are self-reinforcing or by maintaining communications throughout the trial (e.g. through letters, postcards or messages via RTD). These communications should be personalised as far as practicable.

4.2 Energy efficiency advice

4.2.1 Form and medium

Advice can be delivered at a range of levels, through various media, and in combination with one or more other interventions; there is no reason why each approach should be equally successful in the population as a whole or with particular individuals or groups.

(a) *Levels of advice.* At highest level, there are national awareness campaigns, not directed at particular individuals (although they may be oriented towards particular population/market segments). At the most specific level, there is advice worked out for a specific household, according to their social circumstances and the facilities in their home, and expressed in terms that appeal most to them. Strictly speaking, only the second form might be called advice, the first being just information. EDRP used an intermediate approach, with information being delivered to specific households but not tailored to them personally (although they might have perceived it as personal in some cases).

(b) *Medium.* EDRP used written advice in the main but also advice via the web, a dedicated TV-based web page and RTDs. The literature refers mainly to written and verbal advice, with some more recent studies of web-based advice but little use of advice through dedicated TV-based systems. Advice was not delivered verbally in any EDRP trial except in minimal form as part of some of the installations.

(c) *Combination with other interventions.* EDRP delivered advice as a single intervention and in combination with other interventions. With one exception (a heating controller integrated with an RTD) the advice was not accompanied by any technology that would directly save energy (e.g. insulation or energy-efficient appliances) but was sometimes combined with devices to give feedback (RTD or thermometer).

In any trial, the effectiveness of advice may be expected to depend on how appropriate it is to the intended audience on each of these points, in addition to the actual content of the advice (which is discussed in the next section). Ideally the style, content and medium of advice will be tailored to the individual household as far as possible, taking into account the characteristics of the home, appliances and other services used, lifestyle, motives for change and resources of time, space and money. Where this is not feasible, as many elements of tailoring as possible should be provided.

If delivered with energy bills, it makes sense to link advice with the bill to historic feedback in the bill. This is likely to limit the frequency to monthly or quarterly. But providing advice in the same envelope as energy bills is no guarantee that it will be noticed – the bill should explicitly draw attention to the advice and present the motive for reading it. More generally, the delivery of written advice needs to maximise the likelihood of recipients (a) reading it, (b) understanding it, (c) finding it applicable to their household at that particular time, (d) deciding to take action and (e) knowing what action is most likely to save energy.

There is a role for both wide-ranging reference guides and regular short, simple statements on different topics. The former allow householders to find information on a topic that interests them at a particular time but some people may also read them from the beginning so they should get quickly to the substance of the actionable advice without pages of background material. In contrast, regular short, simple statements offer a simple choice – to take a particular action or not – without the complexity of working out what is best. While the action might not be the most cost-effective for the particular person at the particular time, at least something is done.

Some householders are happy to have advice repeated as a reminder or until they feel ready to act on it – not everyone acts on advice first time, or even remembers receiving it. But the focus should be on what is most likely to be missing from readers' knowledge because many householders will "switch off" at the first sentence if they are being told something they already know – reasons for not finding the SSE advice booklet useful included: "Tells you nothing new" (44%), "It's common sense" (21%), "Couldn't be bothered to read it" (11%) and "Too much information to take in" (6%). The general theme is that the information was redundant or too much for the user to work through. The advice should try to offer an aspect of an action that readers might not be familiar with (e.g. most people will now know what insulation is, but may overestimate the cost or difficulty of doing it, or be unaware of local availability of support).

Another reason for not finding the SSE advice booklet useful was "Couldn't use less / already save energy" (6%) and this is where there could usefully be benchmarking or broader statements of how little energy some homes use.

Where communication with the customer is by email, or messages can be sent to an RTD, this would offer greater flexibility over frequency of advice (or reminders to look at the advice already sent). If delivering through electronic media, it is important to make it easy to access the information (e.g. a minimum of passwords and "clicks").

Energy efficiency advice should not be viewed in isolation – the means, motive and opportunity provided by other interventions can be used to complement and draw attention to the advice. For example:

- installation visits can be used to deliver verbal and written advice;
- if historic feedback is likely to motivate the householder to save energy, link the feedback to advice on how to save energy;
- having an RTD makes it more likely that customers will keep and use an advice booklet.

Finally, advice may be more effective if delivered through a trusted source – this might be a community group that has no direct connection with the energy industry, a respected charity or a club or society to which the household belongs.

4.2.2 Content

EDRP delivered a wide range of advice and there is no way of deducing which topics or forms of expression were most effective. Customer feedback suggests that information on the energy used by lights and various electrical appliances may be particularly welcomed (perhaps because there is little “common knowledge” in this area). In relation to other aspects of advice, the most highly rated in EDRP customer surveys were advice on heating and insulation. Advice on saving energy while cooking was least highly rated. Hot water and insulation advice were intermediate. These differences may be attributable to some combination of the advice itself (amount, clarity and quality), the household’s prior awareness of the information provided, and relevance to the household (whether the advice relates to actions the household has not already taken, can take and is prepared to consider taking). Interpretation of the findings is therefore problematic but they may be taken to indicate that:

- advice on heating, lighting and appliances is the easiest place to start;
- more work is needed on delivering advice relating to cooking (perhaps because cooking involves decisions that distract from, override or are unrelated to energy use).

There is generally merit in trying to combine advice on what to do with reasons for doing it (broader reasons – not just saving energy/money) and the resources (time, money, space) that will be necessary and whether help is available to find the resources. For example, advice might explain what loft insulation involves, the likely energy/cost savings, other reasons for doing it (e.g. reducing condensation on ceilings, reducing noise ingress through the roof), likely time and costs if done by householder or by professionals, where to buy materials or services, and whether grants or support from voluntary groups might be available. For additional specific suggestions, see Raw & Varnham (2010).[12]

The advice should also consider what the likely barriers will be and try to address them – means (e.g. don’t know what to do or how to do it), motive (e.g. can’t see why it is worth doing) or opportunity (e.g. cost is too high or it would take too much time).

Perhaps the most obvious point is that the advice should be accurate and well conceived, relevant to the season in which it is delivered and not create risks (e.g. of inadequate ventilation). Otherwise the advice will be misleading to those who do not understand energy saving and appear erroneous or patronising to those who do. The following are examples of problematic advice given to householders during EDRP.

- Use "solar lighting" in the garden – this advice is not inappropriate in itself but it was given in winter and implicitly assumes that the household already has lighting in the garden (otherwise customers who understand energy demand might recognise that getting solar lighting would entail additional energy demand in the form of embodied energy).
- A comfortable temperature range is 20-25°C – this might suggest to customers that they should not reduce their thermostat setting below 20°C.
- Turn down the heating in spring – it is not clear what this actually means (the boiler thermostat, the room/radiator thermostats or the heating period). A properly located and set thermostat should need no adjustment, although the heating period might be shortened (arguably the thermostat should be turned down in *winter*, when people are likely to be more heavily dressed).
- Insulate the floor to reduce drafts (sic) – reducing draughts is not the main point of floor insulation, and misspelling does not give a professional image to the advice.

---

[12] GJ Raw & J Varnham (2010) Focus on behaviour change - reducing energy demand in homes. Department for Communities and Local Government. http://www.communities.gov.uk/publications/planningandbuilding/energybehaviourchange

- Energy cost example of boiling a kettle – this is a very low saving and could give the impression that energy is cheap and reducing use is not worth the effort.

It is also important not to make assumptions about the household facilities.

- Some “no-cost” measures depend on having or acquiring certain equipment (e.g. a thermostat, a shower, a washing machine with a cool wash programme).
- Others have to be interpreted or the logic might not be obvious, e.g. does "the amount of water you need" in a kettle include the water needed in an hour's time, or what does it mean to "overstock" a fridge and why is it a problem?
- Encouragement to hang washing outside to dry may be annoying if there is no secure, clean, dry outside area.
- Low-flow shower heads might not deliver an adequate flow as retrofit and/or householder might take a longer shower to compensate for reduced flow so there might not be actual energy saving for the outlay – a shower head is cheap but taps (and having them fitted) can be expensive.

4.3 Historic feedback

Historic feedback is most useful in offering a motive to change but, with a little guidance, may also help householders to identify what has been contributing to high energy use or high reductions in energy use. If householders create the feedback themselves, by reading their own meters, the impact on energy savings may be enhanced.

Historic feedback may be more relevant to one-off changes that have a persistent impact, such as installing insulation or upgrading a heating system. More fine-grain, real-time feedback is more relevant to routine behaviour and purchases of equipment used intermittently (e.g. washing machines, televisions). By extension, aggregated feedback may be more relevant to the fuel used for heating (most often gas) and real-time feedback to electricity. Historic and real-time feedback also work in different ways – immediate impact vs opportunity to take time to study past consumption.

The feedback should be fitted to the desired outcome, considering frequency, duration, immediacy, content (kWh, cost, CO_2, etc.), breakdown (by time, space and appliance), medium (and details of the medium, such as aesthetics and simplicity of access or use), comparisons (historic or normative), and combination with other interventions.

Monthly or quarterly feedback is the most common emerging format. There is little reason to expect that annual comparisons will stimulate energy saving unless the household has a clear idea of what has caused changes. More frequent feedback (e.g. weekly) appears to be more effective than monthly or less frequent feedback but it needs to be set in a context where householders are expecting it and have a plan for using it. Online feedback may be the most cost-effective way of doing this – perhaps for a limited period for each household.

Providing feedback in the same envelope as energy bills, or even on the same piece of paper, is no guarantee that it will be noticed – the bill and/or other communication should explicitly draw attention to the advice and present the motive for reading it. In the SSE trial for example, only 32% of survey respondents recalled something different about the bill and 27% recalled what it was. Low recall does not necessarily mean that the information was not used when first seen but, asked directly about the historic comparison graphs, 48% said they had looked at the graphs in detail and 38% that it made them think about energy use. Neither should it be assumed that the householder will know what to do about the feedback, even if they notice it.

While the feedback needs to be distinguished from the bill itself, it is unhelpful to use ambiguous phrases such as "this summary is for information only and does not ... require any further action from you". Action is exactly what is needed! In particular, the feedback needs to be presented as an aid to saving energy, otherwise it may just annoy householders who already know they are spending a lot of money on energy.

More generally, the feedback needs to:

- be easy for the householder to notice and interesting to look at (e.g. in EDRP, Scottish Power used a footprint graphic);
- be clear to read (no small print) and interpret (easy phrasing and simple pictures), with guidance as necessary (e.g. keys to any colour codes or abbreviations);
- limit the detail to what people can easily absorb;
- not present too much information at once, but show both kWh and cost comparisons.

4.4 Benchmarking (comparative or normative feedback)

Like historic feedback, benchmarking works on motive rather than means or opportunity but shifts the motive from the financial to the social arena. The choice of peer group is important in making the most of this (e.g. the household might or might not have a personal or geographic affinity with neighbouring households). A comparison group would ideally need to be of households of the same composition, in a similar financial position, in a similar dwelling in a similar climate – this difficult to achieve (especially with the limited information available to energy suppliers). Furthermore, this fact is sufficiently obvious that customers may be suspicious about the validity of their comparison group (perhaps more so for gas because it depends too much on the amount of time spent at home and the characteristics of the home).

Households that are already below the benchmark values may find no reason to reduce consumption and might even increase consumption. This effect is perhaps best evidenced by success in eliminating it – there is evidence that negative impact may be overcome by use of simple normative messages or even minor changes in presentation such as the inclusion of “smiley” icons with low users’ bills.

If the comparison group is not well matched, the challenge of meeting the benchmark value may to too easy or too difficult. Either way, there is likely to be limited impact on consumption.

The more subtle issue is that people tend to place normative information low on their list of what they believe influences their behaviour and yet it can be high on the list of what actually influences their behaviour. Hence, normative messages need to overcome people’s belief that they are not influenced by them, which could otherwise lead to a tendency to ignore them.

Households with stronger left-wing and pro-environmental views may make a greater positive response to benchmarking but there is conflicting evidence on this.

4.5 Customer engagement using commitment to reduce consumption

Given that there is no convincing evidence that this intervention works at all, it is difficult to offer guidance on how to make it more effective. Some possibilities are as follows.

- Give the commitment a clear social context, through groups that have something in common, e.g. because they have volunteered to join an online energy comparison group or local network.
- Ensure the commitment is new: it is easy for people who are already committed to saving energy to sign a commitment to do it.

- Make the commitment as specific and concrete as possible and link it to advice on how the commitment might be fulfilled.
- Monitor the commitment and ensure the monitoring is known to the household.
- If a reward is attached to making the commitment (as distinct from a reward for fulfilling it), people might make a nominal commitment in order to gain the reward.

4.6 Smart meters

The main point of installing a smart meter is that it allows a range of other interventions. But the experience of getting a smart meter may itself influence energy use – in the course of recruitment and installation, customers' subsequent behaviour might be influenced by any or all of:

- being told they are among the first to get the latest technology;
- renewed positive interaction with the supplier;
- reassurance that the meter accommodation is now safer;
- a friendly or unfriendly installer;
- positive or negative attention from friends and neighbours;
- formal or informal energy advice from the installer.

Interaction could focus, for example, on the smart meter being easier to read than the meter it replaces and the opportunity this gives to monitor and adjust consumption (especially for gas, since one-time adjustments to heating or hot water boilers or controls, or new insulation, can have long-lasting effects). It could also remind customers of the billing benefits of having a smart meter but use that as a basis for engagement with saving energy (in EDRP, some householders were happy to benefit from no more meter readers or estimated bills and nothing else).

At the most basic level, the installation process should provide verbal and written guidance on the smart meter – access to a customer service team by telephone or email is useful but no substitute for hands-on demonstration. Guidance should cover not just what the smart meter does and how to read it, but how to use the information it provides in order to save energy and manage budgets (this is expanded in the following section, on RTDs).

In one EDRP trial, considerable efforts were made to prevent customers knowing they had a smart meter. It is debatable whether these are best viewed as smart meter interventions: although the smart meter technology was being tested, the experience of installation and understanding of the meter were not.

In another of the EDRP trials, installers' training did not focus on customer interaction; they gave out spare user guides and answered questions, otherwise customers were referred to a remote customer services team. This minimal customer interaction was intended to minimise "risk of bias". Interviews suggest that customers expected, and could have benefited from, more engagement and instruction at installation stage. Only the most engaged and motivated customers read, absorbed and referenced the booklet. Also, in the absence of training, installers might have invented their own advice (which could increase variance in response to the intervention or even give customers wrong messages).

4.7 Real-time displays

4.7.1 Combining with other interventions

RTDs can support:

- *means* (by identifying what changes could be made) but only if the consumer understands how to do this, and has the knowledge to act upon observations;
- *motive* (by showing how much energy is being used) but only if the consumer already has sufficient motivation to want to check energy use (and possibly also by engaging other motives such as using or showing off new technology);
- *opportunity* (because they save householders time relative to taking a series of meter readings and calculating differences).

The extent to which they actually do this will depend on the other interventions used. The evidence suggests that the uncertainty of impact can be reduced, and the positive impact increased, by combining provision of an RTD with other interventions – including energy efficiency advice but also something else. The “something else” needs to engage the household with using the RTD, for example through a savings target, a time-varying tariff or some other reason to self-monitor consumption. Conversely, providing an RTD can be used to encourage householders to refer to energy efficiency advice.

4.7.2 Engaging users

The first barrier to RTDs being effective in facilitating energy savings is to have them actually installed, maintained in working order (e.g. by replacing batteries where necessary) and used, particularly when they are first installed and there is most to learn. It is essential that installation and set-up is not left to the householder but this is not in itself sufficient, as shown by customer surveys in the EDRP smart meter groups in which RTDs were also supplied and fitted.

- EDF Energy: at the end of the trial, 49-65% were aware they had an RTD. Among those who were aware of the RTD, only 39% were referring to it daily or several times a week and this was mainly limited to one householder (the bill payer).
- E.ON: 48% were aware they had the RTD. Among those who knew they had an RTD, 88% ever looked at it.
- Scottish Power: 77% reported that their RTD was fitted and working at the end of the trial but 11-26% rarely or never checked it.
- SSE: 63% were aware of the RTD and looked at it at least once a week.

Therefore the installation process and user guides are critical to their success and both need to be clear, simple, produced with a high quality finish and written from the householder’s perspective. The written instructions in EDRP may sometimes have been unclear or too complex, especially for people who had been sent the device, rather than making an active decision to buy it. This would tend to reduce the likelihood of the intervention succeeding but also increase the within-group variance, making it yet more difficult to detect changes in energy use. Households should be shown how to use the devices provided, not just given a guidance booklet. Both the demonstration and the guidance should include:

- technical instructions (how to set up and manipulate the RTD to acquire information);
- motivation (reasons for using the RTD);

- application (how to use the information acquired to reduce energy consumption).[13]

The technical instructions need to take account of a general unease that people have about operating electronic equipment – it tends to be perceived as more complicated than it actually is and, even though a sequence of instructions might be easy to follow one step at a time, people tend to be put off if the sequence looks long and complicated. Scottish Power found that 32% of users (across all groups that had an RTD) found it difficult or very difficult to change the settings. Overall, 17% of those who received a clip-on RTD from SSE thought it complicated, or thought the instructions were not clear.

The guidance and the RTD itself should allow for the fact that users may not make the effort to understand the device and therefore establish a default setting of users seeing the most relevant information without having to understand or change anything. This is likely to break down, however, if there is a fault in the RTD or in its communication with the meter. Ideally such faults would be sensed remotely but, failing that, the user would have to report faults and this should therefore be easy and free to do. The RTD should also be checked whenever the supplier's staff or agents visit the property. Similarly, if the unit costs cannot be updated remotely, it should be as easy as possible for users to do it, with instructions included in communications from the supplier every time the price changes.

If health and safety warnings are necessary, their prominence and wording in guidance to householders should be proportional to the risk and not detract from the main purpose of the guidance (which could mean putting warnings in a separate document).[14]

Technical guidance could also include where to place the device. Being able to move around the home with an RTD is most useful when it is first acquired, as householders work out the impact of different appliances. After this initial period, there is less concern about portability and the display tends to be left in a prominent position in the home. In the SSE trial, about half the mains RTDs were located in the kitchen but it is not clear why this was so. It may be because of the number of electrical appliances there, or because that is where there are easily accessible spare sockets, or for aesthetic reasons. Other common locations were the lounge and hall but also "in a drawer/cupboard". Given the choice of a fixed location, customers tend to want the RTD in a cupboard, so this is to be avoided as the device needs to be in a visible position in the living space.

It would be helpful to gather more evidence on how people choose a location so as to offer better guidance in the roll-out. Getting the device in the right location could improve how it is used, or even make the difference as to whether it is accepted into the home at all.

While technical understanding of the RTD is important, it is not enough: householders also need a reason for using it. In fact, motivation may be the first point to make because it needs to be considered when encouraging households to accept an RTD. As with other interventions, the motivation does not necessarily come only from the obvious benefits of saving energy (or energy costs) – appeal should be made to a range of other motives and one EDRP trial found that being in control was a particularly important motive for using RTDs. Reduced "hassle" in

---

[13] While technical instructions, motivation, application may be a logical sequence, it is not necessarily the best order in which to present guidance. The reverse order may be more effective – first telling people what they can do with the RTD, them why they might want to do it, then how to do it.

[14] EDF Energy required smart meter recipients to sign terms and conditions before installation. This originally caused problems as it was presented as a legal document, leading to an initial 90% fallout rate between recruitment and installation. The terms and conditions were revised in line with customers' comments and were subsequently only formally responsible for a fraction of fallout rate. However, EDF Energy expects that many more of cancellations were due to the terms and conditions though customers did not explicitly said they would not sign them. EDF Energy also noted that the terms and conditions were quite manually intensive in terms of managing, answering customer queries and chasing customers to return.

seeing where and when energy is being used, and the immediacy of the feedback (particularly on peaks of consumption), and environmental benefits could also be emphasised.

EDRP records some specific reasons why people found the RTD useful and these could be incorporated into guidance: "It provides reassurance that I'm doing the right things to reduce my energy", "It gives me more control over my energy usage", "It helps me to encourage other members of the household", "It means I have more influence on my bills", "I would miss it if it was taken away". One line that it should ideally be possible to add in future is "it's easy to use".

Motivation should apply not only at the start of a trial (or the national roll-out) but over a period. Of those who had an RTD in the SSE trial but were no longer using it, 32% of credit customers with smart meters had never looked at it whereas 43% had used it for less than "a few months". In the smart meter groups, the dominant reasons for not using the RTD related to the usefulness of the information provided – in general or after a period of use – rather than the functionality of the device itself (except that the greater complexity of the smart meter RTD perhaps lead to a greater problem with knowing "what the buttons do").

4.7.3 Using the RTD

Once householders have started referring to their RTD, they need to know what to do with the information they acquire (or, conversely, they may be more likely to acquire information if they have a plan for what to do with it). It is also logical that finding the RTD useful should in itself encourage further use and sustain its impact on energy consumption. The following points can be made about how RTDs are used.

- A key use of RTDs is to identify specific uses of electricity so that related behaviour can be modified. Users can experience a strong impact of seeing readings jump up and down as appliances are switched on or off. An RTD's effectiveness may depend on users getting to this point and so it would be good to demonstrate it during installation.[15]
- Guidance needs to manage the associated risk that the visual impact of brief use of high wattage appliances may distract users from the potentially large consumption due to extended use of lower wattage appliances.
- The option of appliance-specific monitoring is also attractive to some users.
- While RTDs can be used to identify opportunities for savings, the more powerful use may be in confirming that actions to reduce energy demand have been successful: the RTD may be more effective in maintaining conservation behaviour than initiating it.
- The impact of RTDs is extended where participants go on to discuss the issues raised with their children.
- One specific strategy is to use the RTD to check whether everything is switched off at night. This may be the application that has the most persistent benefit, after all appliances have been checked and usage monitored over a period.
- People will not reduce consumption for a particular application simply because the RTD shows that consumption is high: they have certain non-negotiable uses of energy.
- Most of the energy savings from feedback programmes arise from simple changes in routine behaviours, rather than investments, although the latter can also be encouraged in longer-term programmes or where additional incentives are applied.

[15] In one EDRP trial, installers were trained to do this and also (using a demonstration unit) to show how to view historic information.

- Some EDRP participants questioned the running cost of their RTD and whether it should be switched off at night. It is good practice to switch off all non-essential appliances at night and when away from home so RTDs should have sufficient battery backup to retain data and settings for up to a month.
- External support may need to be readily available in response to using an RTD, e.g. to inform purchase choices or to establish norms in relation to comparable households, or to obtain planning permission for more substantial installations.

4.7.4 Display content

Regarding the RTD itself, for most users it needs to be kept simple; while more sophisticated devices may be imagined, it is more important that users have easy access to the aspects of the display they find most useful and are most likely to use (and that they can easily return to those aspects if they get lost in trying to access other information). The following points can also be noted.

- Cost information tends to be the preferred display, and the most easily understood, but tracking energy use over time – as costs change – probably needs users to refer to kW and kWh, even if they do not fully understand these units.[16]
- Both current and cumulative consumption figures have their different purposes. Monthly figures scaled up to a year or quarter may be more effective than kWh or costs for shorter periods.
- Few people value display of CO_2 emissions and, for any given household, they are of limited additional relevance because they are directly dependent on the energy used.
- Display of gas consumption also scored poorly on awareness and usefulness. This may be partly because extra button pushes were necessary to view gas data, because gas use tends to be constant over a period of hours (e.g. for space or water heating) or because the time resolution of the display was 30 minutes rather than seconds in the case of the electricity display.[17]
- Although accurate numbers are important, a rapidly changing rate is better expressed as an analogue indicator.[18]
- Only the first or second screens in the access sequence tend to be used, so the most useful information should be put on those screens.
- Indications of consumption against a target level can be helpful if someone is trying to reduce consumption but can be a source of stress if money is tight – a constant reminder of money being spent.

Expanding on the final point, some RTDs have a facility to sound an audible alarm (e.g. to indicate that consumption has risen above some target level). In general, EDRP householders did not make much use of an audible alarm where it was available – the reasons are unclear but are probably a combination of being unaware of it, not knowing how to set it, not wanting to bother with setting it and simply not liking it. If such an alarm is used in future, more effort needs to be put into explaining it and it should be under the control of the householder in relation to the basis on which the alarm sounds, how loud it is and turning it off. Audible signals may be particularly useful to blind and partially sighted people but would need to go beyond a simple alarm.

[16] Some EDRP participants were confused by the meter display being in m^3 and the RTD display in kWh. If this will be the normal situation, it needs to be explained and a conversion chart provided.

[17] There is perhaps an assumption emerging that 30 minutes is too long a period to be useful. Equally it may be that it is too short a period. A 30 minute figure could be confusing because it is short enough to be associated with “now”, when the current rate of gas use is in fact different to the past 30 minutes. For example, if someone turns the heating on, then immediately checks the display, it may appear that very little gas is being used. When they turn the heating off, it will appear that a lot of gas is being used. A rolling period of 3 hours or more could be more meaningful.

[18] Although E.ON reported no difference in overall ratings of numeric and graphic display of information.

In contrast, the limited evidence (generally qualitative) indicates that visible signals with clear meaning may be viewed positively and used as part of an energy-saving strategy. Examples in the literature are visual signals of high usage (in relation to a baseline or target, as with the ecoMeter "traffic lights" used in EDRP), the current price band (for households on variable tariffs), or to indicate that the heating was on or that the air conditioning could be turned off. EDRP confirmed the positive user ratings of traffic light signals; it is unclear whether positive ratings mean more energy saved but there is a good chance of this, given that a major element in making RTDs effective is to get householders to engage with them at all. Some householders disliked the traffic lights because they associated the red signal with danger (e.g. a meter error) but these were a minority and this perception may be overcome with better guidance during installation.

Most living rooms do not have a thermometer, many thermostats are not scaled in °C and the UK population is still in transition from thinking in °F. For these reasons, temperature displays (or a separate thermometer) may be useful to households, at least initially, if they are seeking to interpret advice about room temperatures and thermostat settings. There was some limited evidence of this in EDRP. As with visual signals of consumption, temperature displays may be as valuable in encouraging engagement with the RTDs as they are through their specific function.

Aside from these generalities, different market sectors have different preferences for RTD design and display features, giving priority variously to saving money, environmental benefit, acquiring and manipulating data and the 'new gadget' itself with technological and aesthetic appeal. Taking this a step further, it may be critical for installers to have a small set of options for how to present a device to different households.

4.8 Heating controller integrated with RTD

Integrating a controller for gas heating and hot water with an RTD has the potential to provide two benefits:
(a) attracting attention to the RTD main function, making it more likely to be used (and used for a longer period) and (b) providing a direct route to apply the information gained from the RTD in reducing energy demand. At the same time, there is a risk that the combined device will be seen principally as a controller and the RTD screens not used (or used only to monitor gas use).

The EDRP trial of this intervention has yet to be fully reported but there is evidence that the risk outweighed the benefit. There may be changes to the device or instructions that could redress the balance; EDRP does not directly evidence how this might be done but a key factor may be what is on the default screen and how easy it is to change screens.

With the added complexity of integrating a controller with the RTD, clear guidance and display characteristics (as discussed in the previous section) become even more important.

It is also important to be aware that providing any new controller (where there was none previously) will not necessarily reduce energy use and could increase it. Hence the user guidance is particularly important for such interventions. They may be most useful where an existing controller is replaced, which did not happen in EDRP.

On a point of detail, a "clean" intervention would replace an existing thermostat with the new device in as similar as possible a location. This, however, may not be the ideal location for a thermostat and it may not be a good location for communication with other devices, so a decision will need to be made about what is more important.

4.9 Time of use tariff/incentive

In general, the overall savings achieved by time of use (TOU tariffs) are less than for programmes focused on overall energy savings. This is no surprise and the implication is that the aim of an intervention – demand reduction or load shifting – needs to be clear. Even so, while the main consumer motive may be financial, based on shifting load, overall reductions may be achieved if consumers:

- become more aware generally of their energy use and savings options, or
- take more care to eliminate energy wastage when the price is higher (e.g. turn off appliances that are not being used) without compensating by increasing usage at other times.

The opposite effect is also possible, if consumers take less care of energy wastage during cheaper off-peak periods. The design of incentives to shift load needs to be balanced to avoid this and the tariff levels and ratios may be important in this respect but it is difficult to disentangle this from other aspects of tariff scheme and research design. Customers who had been in neither of SSE's incentive groups (incentive to reduce or incentive to shift demand) were asked "How much cheaper than the peak daytime tariff would the night tariff have to be to encourage you to move some of your consumption?" Group average figures ranged from 19-32% (the mean of the group averages was 25%).

Where the incentive is based on a reward for reducing peak time consumption (rather than an explicit time of use tariff), the amount of the reward also needs to be considered carefully: too low and it will have little effect, too high and it may distort behaviour to achieve the reward (e.g. increasing consumption in the baseline period or using excessive amounts of energy in off-peak periods).

Intervention designs also need to take into account the flexibility that households are likely to have in when they use energy. Unless electricity is used for space heating, the opportunities for load shifting are limited. Furthermore, even if electricity is used for heating, this will often have already established load-shifting (using electric storage heating). Response to TOU pricing in the UK depends on ambient temperature, particularly during the daytime – unsurprisingly, colder temperatures prompt people to turn on the heating (with little evidence of shifting consumption into the night). The literature also shows that the impact of TOU tariffs is increased by providing some kind of enabling technology (e.g. thermostats that vary the thermostat set point automatically with tariff or allow remote "direct load control").

The evidence from EDRP is that load shifting is greater at weekends but this does not necessarily mean that interventions should target the weekend. Peak demand nationally and for individual households is typically lower at weekends, so the absolute savings will be lower from a given percentage shift, and there will be less impact on the need to increase peak generating capacity.

While peak consumption was higher in winter, load shifting was not.

The EDF Energy trial found that load shifting was greatest in smaller households (fewer than three people in the age range 16-64) so these households might be targeted. Nevertheless, absolute savings may be lower (as suggested by the Irish smart meter trial)[19] because smaller households tend to use less energy.

The participants in EDF Energy's TOU group had an RTD but it was a basic model that required many button-pushes to access information from a single line display – this may not be ideal in allowing the RTD to support the TOU intervention.

The SSE trial suggested that load shifting is greater where neither an RTD nor online feedback was provided. This is counter-intuitive because better feedback should make it easier for people to manage their consumption. However, there might possibly be some kind of interference effect if too many interventions are in place at the same time. The incentive to reduce consumption was generally perceived as overly complex – while 83% of survey

---

[19] http://www.cer.ie/en/information-centre-reports-and-publications.aspx?article=5dd4bce4-ebd8-475e-b78d-da24e4ff7339&mode=author

respondents were aware of the incentive, only 75% expressed some understanding of how it worked. The perceived complexity might have arisen because of the use of “shadow bills” rather than altering the actual tariff, the shadow bills sometimes being missed (because of smart meter communications failure) and offering other incentives at the same time (to pay on time or by direct debit).

The implication is that the tariff structure needs to be kept as simple as possible. The ratio of charges in different periods should ideally be easy to use in calculations (e.g. 75% and 125% rather than 69% and 131%). A fridge magnet with the tariff periods and rates is a simple option for keeping the information in people’s minds but a visible signal from an RTD may be even better.

The guidance should also be kept simple. One EDRP trial, for example, offers a rather technical explanation of load variation (and refers to "times when usage is abnormally high" to mean "times when usage is usually high") without guidance on how to shift load. The customer information pack refers to three tariffs in four periods; this could be confusing because a separate tariff sheet shows four rates, one of which is the daily standing charge (which clearly does not vary with time of day).

Incentives to shift may be confused by clock changes between GMT and BST – if meters do not automatically adjust for this, the household needs to be reminded each time that the tariff periods have changed.

The tariffs offered in EDRP were set up in such a way that customers would not be charged more than if they had remained on their flat rate tariff. This was probably necessary for recruitment purposes but does not represent a true trial of the tariff. A way needs to be found of getting around this issue. Even with the guarantee of not paying more, some EDRP participants were concerned that the TOU tariff was having a negative financial effect on them – this also needs to be managed.

4.10 Incentive to reduce consumption

The key practical issue is whether it is economically feasible for suppliers to offer a reward that is high enough, and sustained for long enough, to have a lasting effect. A high reward might be economically feasible if the target were difficult to achieve but this may not be sufficient incentive (or could encourage “cheating” by inflating baseline figures or shifting some consumption to a fuel that is not being monitored).

It is possible that the perceived value of a small reward could be enhanced by presenting it as a prize or money off the bill, or by combining it with some other intervention. One EDRP trial combined a reward for reducing consumption with a reward for customers reading their own meter and providing the readings to the supplier; this could increase awareness of energy consumption and create a sustained reduction in energy use but could also negate the reward for reducing consumption by offering an easier way to achieve a similar level of reward. Unfortunately, this intervention did not produce sufficient data to test the effect.

EDRP provides little direct evidence on how to maximise the effectiveness of direct financial incentives to reduce consumption because there is no reliable evidence that any of them worked (or worked for other than a very short period). One important point, however, is that the incentive will have no effect if the household is not aware of it – in the SSE trial, 28% of survey respondents in the “Incentive to reduce” group were not aware of the incentive.

Based on the literature and reasoning from it, the following points may be made.

- Targets should be transparent in how they are calculated, easy to understand, and fair but challenging.
- Rewards have most effect when they follow quickly from the action that leads to the reward.
- When a financial incentive is ceases to be available, people tend to return to their former behaviour.

- A lasting effect might occur if the consumer invests in efficiency measures or more efficient appliances as part of an effort to gain the incentive payment, but such incentives would have to be structured for the long term in order to make the investment worthwhile
- Financial incentives can be counterproductive in the longer term by focusing attention on the financial motive, to the detriment of other motives that the consumer might have had prior to the intervention. Not only would this make it difficult to sustain energy savings, it would render more probable the “rebound effect” whereby energy savings are spent on other goods, services or benefits that might themselves be energy-intensive (e.g. a bigger TV or a holiday flight).
- This erosion of non-financial motives might be avoided by, for example, community or peer group support, or providing the reward in the form of vouchers to spend on energy-saving products. This would need to be tested further.
- It may be safer to target households that are already motivated to save energy but lack the resources to invest in their goal, i.e. to use financial support to provide *opportunity* rather than *motive*.
- Participants should be able to monitor their progress towards the target and it should be easy for them to see whether they are on track.
- Any advice associated with the incentive should be relevant to the incentive (e.g. if the challenge is to reduce electricity use, and most or all of the households use gas for heating, do not advise turning down the room thermostat).
- If targets are set sequentially, it becomes complicated to set new targets to maintain a realistic challenge while taking into account each household’s success in achieving previous targets. Also, householders’ strategies are likely to depend on whether they are told the whole sequence of targets in advance or only one at a time.
- Rewards for reducing consumption are necessarily linked to a target, even if the target is a minimal change and easy to achieve. Hence the financial incentive effect is confounded with a possible target-setting effect: people can be motivated by succeeding in a challenge without any extrinsic reward. Target-setting would be less costly than rewards.
- A possibility that has not been explored in research is that smart meters (and associated feedback) could make consumers generally more aware of energy prices and hence more responsive to price changes or offers, making financial incentives (or the negative equivalent – price rises) more effective in reducing demand.

4.11 Web-based interventions

The web is not an intervention as such but is a rapidly developing medium for delivery of a range of types of intervention. Therefore, it is worth taking an overview of its potential in supporting interventions. The EDRP trials delivered web-based services using household’s own PC or TV. The EDRP findings should not be seen as representing the potential of web-based interventions. The study tested only limited approaches, with no targeting of those customers most likely to use the technology provided.

The important differences between paper- and web-based information are probably that the latter is (a) a more active process for consumers than passively receiving printed material, (b) more able to provide information and advice tailored to the individual user, to support response to feedback and (c) more easily and immediately linked to options for action to reduce consumption and the resources for carrying out those actions. Certainly the web offers opportunities to bring together means, motive and opportunity in a way that would be hugely more resource-intensive through other media. The key point is not that the web has greater impact for a given intervention (and there is insufficient evidence on whether or not this is true) but rather that it creates new possibilities.

Creating the optimum web site for achieving energy savings will almost certainly mean combining various types of intervention – feedback, targeted advice, incentives, etc. While it does seem likely that many customers achieve energy savings as a result of web-based interventions, it is impossible to deduce from the literature what the essential “active ingredients” are or what the optimum web-based feedback would look like. The main points from EDRP are as follows.

- The failure to find significant effects of web-based services reflects the fact that few participants accessed the sites, even though many were regular internet users, and many did not even recall hearing about it – it is not clear whether an effect would have been seen if this initial hurdle had been overcome.
- Access to the website should be easy (e.g. one of the EDRP trials required a username of 14 characters and password of 12 characters) and there may be a benefit in providing a hard copy guide to using the site.
- The lack of real-time feedback may have been a critical weakness in the services offered (feedback was delayed by a day).
- The web site (or TV) is competing with other information – it needs to be immediately attractive, clear and useful (especially if the user does not have fixed price access to the internet).
- Do not offer usage forecast unless it is reliable.
- Any usage targets should be simple and clear, and the user should be able to adjust them.

In relation to the TV-based services in particular, the following suggestions arise from EDRP.

- Fully test all equipment, software and communications before installation in a trial.
- Make the service as compatible as possible with different types of TV and internet connection (including quality of display) so that it can be installed in all selected homes and on the most appropriate TV (if there is more than one).
- Impact may depend on the person with control of the TV also being interested in energy so ensure the right member of the household is briefed.
- Make access to the energy screens as easy as possible, with a dedicated button on the remote control and on a suitable screen on the TV itself (e.g. the programme guide screen).
- Keep the on-screen information simple and clear – avoid it looking cramped with too many words or too much information (decide what the main message is and concentrate on that – no logo or multiple figures).
- Preferences for information to be displayed are similar to those shown in RTD research.
- Assume attention of about 20 seconds and certainly no longer than a typical advertisement break.
- Assume little attention beyond the opening and second screens.

The right question might not be whether web/TV-based interventions work, but rather for whom do they work? The most promising uses of the utility-based websites seem to be with particular subsets of the population and/or specific, focused programmes. Some people would not use web-based service at all, some would make limited use (e.g. to check consumption, but with no intention of reducing it) while others would make effective use of good websites to reduce consumption. Concern about data privacy is likely to be a consideration for some customers; for others, this will seem insignificant in relation to the information that they already exchange online. Alternatively, it may be that most people would use a web-based service but only at certain times (e.g. when moving to a new home or when there is a rise in energy prices) or in certain contexts (e.g. where a more complex energy management demand, such as complex tariffs, benefits from additional information).

In considering targeted use of web-based services, it is also important to pitch information at an appropriate level: making the sites simple and attractive for newcomers, offering an overview of basic information but with easy access to detail and tools for well informed people who are actively researching (and seeking to reduce) their usage.

The relationship between RTDs and web-based feedback also needs to be considered. User feedback suggests that the web is not a good substitute for RTDs in relation to their prime function of real-time feedback: using a PC for this takes extra time and trouble and relies on a good internet connection. RTDs are also more portable and more convenient to leave in rooms such as kitchens and utility rooms.[20] This implies that online data should be seen as a complement to RTDs rather than as a substitute, especially if users can download data from their RTD to a personal web page. The RTD is then used for immediate information, while the web-page can be used to supply graphics, a long view of consumption patterns, and the opportunity for detailed exploration of the data. This distinction is important, both in terms of influencing user behaviour and in relation to the energy used by PCs and TVs in delivering information. The set-top box used in EDRP was estimated to use 200 kWh per year.

The range of ways to access web-based energy information and tools is developing rapidly, with applications now available via mobile phones, personal organisers, etc. An increasing proportion of the British population engages routinely with some form of online material. In this context, there is a case for active research into the various web-based services on offer, to achieve a more robust assessment of their impact and how to optimise them for achieving energy savings and supporting the introduction of smart meters.

[20] The exception may be if web access is via hand-held devices such as web-enabled mobile phones but this was not explored in EDRP.

5 Community trials

5.1 Introduction

This section focuses on the main lessons from EDRP, should community-level interventions be considered as part of the smart meter roll-out strategy. The use of community-based approaches was effective in EDRP but, to be viable for application at national level, further research and analysis is needed to determine whether there are “key ingredients” for more cost-effective versions of the community approaches used to date, and to provide good practice guidance.

The aim here is not to try to relate energy savings to specific energy-saving measures – either installations or behaviours – that would be to miss the main point of community trials. There are three reasons for this.

- The energy savings from specific measures can be estimated without an expensive trial. The estimate will be wrong because the major unknown is whether the measure will be taken.
- Various interventions and energy-saving measures are recorded but the trial design and data do not allow changes in consumption to be attributed reliably to particular interventions or measures.[21]
- The point of community trials is that the community itself decides what actions to take: it is not driven by an energy savings formula. This is not to say that potential savings are irrelevant and, for example, a community should put all its effort into switching off telephone chargers: there has to be a sense of proportion. But this leaves many actions that could sensibly be promoted, for communities to choose from. The sense of engagement comes from making that choice in a way that fits with the particular community.

Therefore, the discussion here is about the process of community trials – how best to engage with a community to deliver energy reduction – not the specific energy-saving measures they should take. In fact, the latter would be difficult to discern from the SSE trials because the contribution of particular measures was not measured.

5.2 Baseline

Each community has a different starting point, in terms of:

- actions already taken;
- understanding of what else could be done, and why;
- motivation to act;
- resources to act – money, time, space and intellectual and social capital.

This is partly dependent on the extent to which the community is truly a community, not just a group of people who happen to live in the same area. Is there an existing sense of belonging, common culture or a community focal point? In this sense, a community does not even have to be geographically defined – it could be members of a group defined by a sports team, church, mosque or school.

All these need to be considered in setting the basic structure for a community-level action. The baselines were quite different in NL, AL and SA but the approaches taken all had the same starting point.

---

[21] Interventions included smart meters, RTDs (clip-on and linked to smart meters), a public community energy display screen, infrared imaging of homes, doorstep energy advice visits, energy advice leaflets, free compact fluorescent lamps and powerdown devices, energy-efficient appliances purchased through a voucher scheme or special offer, home insulation, a community and personalised website, and various community events.

5.3 Leadership

In trials such as those reported by SSE, local leadership groups were established and this is an essential part of community-level action: the leadership has to be seen as part of the community rather than imposed from the outside. It was not surprising that the local leadership teams were bound principally by their motivation to reduce energy consumption; indeed it is difficult to imagine a community succeeding without this.

But a significant weakness exposed in the SSE trials was that a group bound only by energy saving finds it difficult to engage a large proportion of the community, especially if it is newly formed for the trial and therefore has no existing identity in the community. At community meetings, for example, the same people tended to be involved each time. This is to do with the composition of the leadership groups – nothing is implied about their organisational competence.

In one sense it is a significant achievement that substantial energy savings were realised in spite of this limitation. At the same time, one can imagine how much more could have been achieved with more people being actively involved. Three approaches can be considered to address the issue, in addition to the specific points under other headings.

- Think small: the smaller the defined community, the more difficult it will be for free-riders to go unchallenged.
- Start with the community, not the environment. There will be people of local influence and reputation, whether in official positions or not. It is easier to give them an understanding of energy issues than to give energy enthusiasts a local network.
- Launch satellites. Not everything has to be done through a central committee: much can be achieved through a multiplicity of local groups, each with a highly focused aim. NL was particularly successful in this area.

At the same time, wider “community spirit” may be enhanced by an effective energy project and this can be seen as adding to the business case for such projects.

5.4 Support

Saving energy can be a complex business, both in the application of technology and in changing behaviour. While leadership should be local, it is also important to have access to high quality advice and guidance, to make the most of the available resources. External financial support can also be a significant factor in gaining local interest, so long as there does not appear to be strings attached or ulterior motives. This seems to have been well provided in the SSE trials, with expert input, financial support and SSE participation in local meetings.

5.5 Relevance

Many actions can be taken but not all will be relevant to a particular community.

- Some actions will physically make more sense than others (e.g. in a village consisting mainly of 18th century stone cottages, there is little point in talking about cavity way insulation).
- Some will be aesthetically or socially more acceptable than others (e.g. in that same village, talk of external wall insulation could kill the project before it starts).

- The motives that the community can relate to will also be relevant (e.g. one community may be enthused by saving the environment, another by saving money, another by taking pride in displaying the latest technology).[22]
- Actions also need to be tied to the resources available – not just the financial resources but the time that people can put into the action, the space available (e.g. for microgeneration) and the local expertise in designing, procuring or installing measures.

These are potentially complex choices and it is essential to have a fair and open way of people expressing their preferences as well as getting involved in delivery. It is also possible that one or two high profile public actions (such as the replacement of the Christmas lights with more energy-efficient technology in SA) will give a clear reminder of the overall project and engender a wider sense of ownership.

Where a reward is on offer for meeting a target, actions also need to take account of contribution to target. In the SSE trials, the target was based on electricity savings but many of the actions promoted were about heating and insulation, which would be relevant mainly to reducing demand on other energy sources. The impact of this is unknown but it could include reducing the credibility of the leadership, taking resources away from more relevant actions and leaving people feeling aggrieved if they put effort into reducing energy use but without helping the community towards the target.

5.6 Promotion and communication

Not just the actions but also the promotional activities need to be relevant and attractive. In the SSE trials, some of the more successful activities were part of some wider event that people would be attending for other reasons, not events specifically about saving energy (e.g. local fairs).

Websites and newsletters can be useful – if not essential – tools for coordinating and supporting activities but they are passive means of engaging people.

The biggest risk is a promotional activity that backfires because its message or style creates antagonism, embarrassment or confusion. The advice offered in the SA Christmas Poster unfortunately fits into this category, being confusing and sometimes misleading and irrelevant to the target and the motivations of most people.

5.7 Targets and rewards

SSE set up both a target and a financial reward (as did British Gas in the well known Green Streets project). It should not be assumed that either is essential or that each requires the other. A SMARTER target can provide motivation – with or without an associated reward – but so can a competition with judging rules rather than a fixed target (e.g. between similar communities) or just an emerging sense of community pride in what is being achieved. A key factor for maintaining motivation is that any target or competition should be widely viewed as fair – at the start and throughout the project. This was a particular issue for SA, which started from a lower base of organisation and understanding and had to overcome increases in energy use because of local expansion of business and housing. This appears to have been compounded by misunderstanding among participants about the overall aim of the trial in SA.

The nature of any reward should balance relevance to the community, scale and affordability, especially if it is envisaged that a large number of communities will benefit.

---

[22] See Appendix C to the EDRP report for discussion of the range of motives that may apply.

The impression is given that the SSE communities decided how to use the money after it had been won; there is no way of knowing whether this was the best approach for the communities in the SSE trials but there is a case for an alternative. A financial target can be given greater substance by deciding collectively what to do with it, at the start of the project. While £20,000 may appeal as a welcome addition to the community finances, it may also seem small in the context of the whole cost of administering the community. It can seem more real if people can look forward to something specific that the money will achieve. The counter-argument is that some of the community will disagree with the use chosen for the prize money, and therefore be disinclined to support the endeavour. However, if the use is well chosen, it should retain interest from more people than an entirely unspecified use.

Finally, feedback on progress – at household, business and community level – is essential. This should be reliable, easy to access and digest, and regular (but at long enough intervals to create a sense of anticipation and show movement from one set of feedback figures to the next).

6 Customer feedback

The design of a customer survey (or other approach to gathering feedback) and the associated instruments (e.g. questionnaires) is a specialist technical role and each survey needs to be designed according to the specific aims, design and limitations of the study of which it is part. Hence, no definitive or standard approach should be set out. However, this chapter draws on EDRP to offer general guidance (to be applied in the context of basic good practice).

6.1 Approaches used in EDRP

EDRP used several methods to collect data systematically on the participants' characteristics, experiences, opinions and behaviour during the trial:

- questionnaire surveys (face-to-face, by telephone or by self completion on hard copy);
- depth interviews (face-to-face or by telephone);
- focus groups;
- click-through records of the use of feedback devices.

Customer service calls were also analysed in order to gain further insight. Some other possible approaches (e.g. diaries or online surveys) were not used.

Questionnaire surveys were a mixture of recruitment surveys and surveys during or at the end of the trial. Some were sample surveys (covering all trial groups or only selected groups), others approached all participants (though not all agreed to take part). Areas covered by the surveys included:

- household demographic characteristics (e.g. number of occupants, ages);[23]
- dwelling characteristics (e.g. built form, age, thermal insulation);
- energy-using services and appliances (e.g. heating/boiler type, lights);
- reasons for participating, experience and impact of interventions;
- awareness, recall, use and usefulness of the household's RTD, bill information and advice;
- frequency of checking energy use;
- experience of difficulty in keeping warm or cutting back on other items to pay for energy.

This information was sometimes collected as the facts at a point in time and sometime reflected changes during the trial or since the previous survey.

Some EDRP surveys used a mix of approaches (telephone combined with either face-to-face or self-completion surveys). Ideally, the same medium of approach (face to face, by telephone or by self completion on hard copy or online) should be used for all participants in a given survey or linked series of surveys. In this way, any effect of the medium is controlled for.

6.2 The timing of customer feedback

The first point to make is that participation in customer feedback at the initiative of anyone other than the customers themselves is itself an intervention and needs to be controlled for in the experimental design and analysis. In fact,

[23] Indirect demographic classifications were also used, based on postcode – see Section 3.2.4.

EDRP showed that surveys during recruitment can affect willingness to participate in the study at all. This implies a choice for trial designers between obtaining a representative sample and requiring 100% survey participation. A compromise may be to make some incentive (e.g. a participation “thank you” payment) contingent on survey participation.

It could be argued that providing feedback during the trial would affect the control group more (because it is the only “intervention” they receive) or that it would affect the experimental groups more because they have been given more capacity to act.[24] Either way, the survey would be more likely to increase energy-saving efforts although this would not necessarily be true of all households (e.g. if the questions on energy-saving behaviours confused people as to what they should be doing). Pending further research into the effects of feedback participation, it can only be considered as increasing noise in the data.

Including all participants in customer feedback research will eliminate a difference among trial participants but make it impossible to say whether other interventions would have had the same effect without the survey.[25] It may also not be practical (e.g. because of cost or a need to maintain a lack of awareness of the trial). From this perspective, it is better to include only a sample of trial participants in feedback research conducted before the end of the trial but this will mean that the confounders recorded in the research will not be available for analysis of the full sample before the end of the trial. Recruitment surveys that merely collect facts about the household, without questions that would influence energy use, would be excepted from this.

Some changes may take a long time to happen, for example if they are expensive (e.g. a new boiler) and/or are contingent on an existing appliance failing (e.g. an oven may be replaced when it fails but there is then a choice about whether to use energy efficiency as a purchase choice criterion). Hence post-trial follow-up surveys may be useful but this was not done in EDRP.

6.3 Content of questions

Regarding the content of questions, the EDRP surveys highlighted a number of issues with getting usable data from surveys, which can be summarised as follows.

(a) Focus on what is most important to learn. If the main point is understanding changes in consumer behaviour and/or energy use, then avoid demanding too much of the respondent’s time and attention with other information of lesser interest. Depending on the aim of a survey, accurate records of relevant facts (e.g. level of insulation, the services and appliances present in the home) may be more useful, for example, than questions about opinions of the supplier or other sources of advice.

(b) Many issues come down to being realistic about what respondents are able to report. At the most basic level, questions should be reviewed as to whether respondents are likely to know the answer but the following specific points can be made.

- Current facts (e.g. the respondent’s age and how many bedrooms s/he has) are generally easy to report. Nevertheless, even current facts may be omitted or misreported if they are considered sensitive (e.g. income or whether the home is left empty during the day), require documents (e.g. energy bills) to be retrieved and checked, or the question uses technical jargon or other terms that the respondent may not understand. For

---

[24] In the Scottish Power trial, customers interviewed in a survey had lower subsequent electricity consumption but it was not established whether the survey brought about reduced consumption or those with lower or reducing consumption were more likely to agree to participate in the survey.

[25] Arguably a survey should be a normal part of implementation, so lack of a survey should be seen as the potential source of bias, rather than taking part in a survey.

example, one EDRP question was whether the respondent had a smart meter but this required the respondents to know either (a) their meter is called a smart meter and this term does not refer to their RTD or (b) that their meter has been changed, and guess that it is called a smart meter.

- There is a limit to people's precision. For example, they may be able to report (with at least partial validity) whether they found it useful to have an RTD. It is less likely that they will know with any certainty which particular aspects of the display most affected their behaviour. Similarly, they may be able to report that they stopped leaving the TV on standby but less able to say that it was partly or wholly a specific trial intervention that was responsible. Obtaining evidence on such matters may depend on designing the trial to be able to determine precisely the source of effects, rather than assuming that householders can report them.
- Conversely, some questions may not be worth asking at all unless they are composed in a way that delivers sufficient precision. For example, there may be some value in knowing whether households have more often put lids on pans when cooking, but more useful to know how often this is now done, or how many members of the household who ever cook have adopted the habit. Similarly, if the household turned the room thermostat down, did it stay down?
- Questions about opinion or belief tend to be more difficult and it is essential that they are clearly stated, in terms the respondent is likely to understand, and concern something the respondent is actually likely to have an opinion about. It easy to express an opinion in order to avoid the embarrassment of not having an opinion, but such responses have low validity.
- Questions that require memory of the past are more demanding of respondents, even when they are about single concrete facts (e.g. has the loft been insulated in the past two years), more so in the case of multiple actions (e.g. number of bulb or appliance purchases) or routine behaviour (e.g. switching off lights), and much more so when they are leading questions about ill-defined opinion (e.g. has your awareness of energy changed). Difficult memory questions create a risk of misreporting, possibly influenced by what the respondent thinks is the answer the researcher wants to hear.

(c) Questions should not be unduly open to interpretation by respondents because of their wording or because the response choices given for the questions were ambiguous. For example, respondents may vary in their interpretation of questions about whether they "filled the kettle less" – the intended meaning might be that they put less water in the kettle each time they used it but the literal interpretation is arguably that they filled the kettle, then kept reboiling it until they needed to fill it again.

(d) One-off actions that do not need to be repeated (e.g. reducing room thermostat settings) should be considered separately from actions that need to be repeated or become a habit (e.g. putting lids on cooking pans).

(e) Avoid distortion of the meaning of questions by their position in the survey (e.g. because earlier questions create a limited context for the question or because questionnaire routing erroneously excludes some respondents from being asked the question).

(f) Review what supplementary questions will be needed in order to interpret the main questions. For example, questions about actions taken to reduce energy use should take into account whether the action was available (e.g. taking showers rather than baths is an option only if there is a shower; replacing incandescent bulbs is an option only if any were in use at the start of the trial).

(g) To compare the impact of different interventions, it may be useful to build questions around a common structure (e.g. the EDF Energy focus groups referred to usability, engagement and impact on behaviour).

These issues would largely be resolved by a combination of cognitive testing, field piloting of questions and development of an analysis plan to define what data will be needed.

6.4 Questionnaire analysis

Some specific points highlighted by the EDRP analysis are as follows.

- The level of analysis can be more sophisticated if questions and response scales are designed to yield a normal distribution of responses (this is more important than trying to avoid respondents making excessive use of the mid-point of a response scale). Seven-point scales are often optimal for this. For example, with such data it may be possible to include all trial groups in a single multivariate analysis, taking account of confounders, rather than analyse group by group and face the resulting difficulties with low sample numbers and interpreting the analysis.[26]
- Missing data (including responses such as “don’t know” and “refuse to answer”) sometimes have a particular relevant interpretation, depending on the question and its context in the questionnaire.
- Regular checking of an RTD does not mean regular action. It may imply a lack of understanding of what to do with the information acquired.
- Ratings of “useful” may mean “interesting” rather than actually used.
- Positive attitude and intentions were not always translated into less consumption or lower expenditure.
- Awareness and effectiveness are not the same thing. In the EDF Energy trial, awareness of the RTD was greatest among dual fuel customers but this was not associated with greater reductions in consumption.

6.5 Customer service logs

The smart meter roll-out will require an investment of customer service time, with staff capable of dealing with a range of technical and administrative issues. Queries will be partly those that are common to any meter exchange, partly specific to the intervention and partly resulting from customers becoming more aware of their consumption, prompting them to seek further advice or explanations. In order to make full use of the information gathered in customer service calls, all information related to the roll-out will need to be identified, collated and analysed. The quality and content of customer service responses will also need to be monitored as this will itself constitute an intervention.

Based on EDRP, the following types of enquiry are likely to be received (depending on what interventions are used).

- Expected bills, statements or other communications not being sent to customers (e.g. due to communications issues or technical difficulties).
- Gas smart metering service lost due to poor communications, meter not bound, meter not installed, etc.
- Concerns over energy usage or bills, especially if time of use tariffs are used (which might be resolved using smart meter data).
- Explanation of incentive interventions (in the SSE trial, these interventions were the ones most likely to result in customer service calls).
- Discrepancy between forecast/predicted and actual energy use, particularly following sudden alterations in weather.
- Estimated bills and/or manually reading continuing after a smart meter has been installed.

[26] E.g. in one set of analyses in EDRP, only 11 differences were declared significant out of 460 tests. With about 2% of tests being declared significant at the 5% level, no individual test can be regarded as statistically meaningful (one would expect 5% of tests to give a significant result at the 5% level even if there were no differences between any of the groups).

- Meter/RTD faults (or perceived faults).
- Faults with household's own equipment following the installation, which the householder attributes to the installation, or damage done during installation.
- Interventions not starting on time or being rescheduled (e.g. because of delays in logistics or equipment development, or unsuitable weather for outdoor electrical work).
- Delays in billing (e.g. because meter exchange details are not updated on billing system or there are incorrect installation or removal meter readings).
- Energy-saving measures or appliances.
- The aim, function, features or benefit of the smart meter or how to use/read it.
- The aim, function, features or benefit of the RTD or how to use/read it and whether it should be switched off overnight.
- Variations in energy units (kW/kWh, m^3 and possibly older units such as therms or 100 ft^2).
- The energy consumption and associated running cost of the smart meter or other equipment installed in the home.
- Prepayment customers reporting problems topping up or adding credit to their meter, or responding to low credit signals.

Minimising such queries would not only reduce suppliers' costs but also tend to improve the effectiveness of interventions and reduce variance in the trial, thus making it easier to detect effects of interventions. Means to achieve this include the following.

- Make interventions as easy as possible for customers to understand, and provide clear, comprehensive explanations of them (preferably verbally and in writing) – this should cover the intervention itself, what information it will provide, what action is expected of the customer and how the customer can use the intervention to reduce energy use and/or save money.
- Anticipate and address customer concerns, especially in relation to financial issues (e.g. incentive schemes, cost of running equipment/broadband).
- Ensure that administrative systems and procedures are effective in delivering communications to the right people at the right time, that customers know when interventions will take effect and billing/statements are coordinated with the interventions.
- Test equipment for customer acceptance, not just technical performance (e.g. ease of use, tolerance of alarms, positioning of equipment in the home, appearance of equipment).
- Ensure that equipment installations are correct and all communications and functions tested while the installer is still in the home.
- Anticipate technical problems that are independent of the installation but which the customer might blame on the installation (e.g. boiler pilot light not relighting, security system not resetting, DVD recorder malfunction, earth clamp detached from gas meter).
- Give prior notice of reasons why an installation might need to be rescheduled (e.g. heavy rain preventing safe exchange of an electricity meter located outside the home).

7 Analysis of consumption data

7.1 Data management

EDRP had to overcome significant challenges with the acquisition, transmission, processing and storage of large amounts of data (over 800 million data points). The security and quality of the data are critical to the project; while no problems were identified with security, quality was compromised in some cases and resulted in data loss during data cleaning. In addition, data need to be stored in a way that facilitates analysis, ideally without the need for prior data extraction routines or computers being slowed down by applying data corrections every time data are extracted.

The roll-out and associate interventions will similarly create a significant data management challenge for suppliers and the proposed DataCommsCo (DCC), to provide two-way data communication between the meters, other in-home technology and the energy supplier (potentially via one or more third parties) and to store half-hourly data records from both gas and electricity meters across multiple households, all integrated into wider communications and data storage systems. The accuracy and reliability of these systems will be both a likely performance measure in the evaluation and a means of maximising benefits on other measures. The imperatives can be divided into data entry and data processing.

Data entry:

- manual entry of data into the databases during installation must be either avoided or correct;
- all systems involved in the end-to-end process must have all data relevant to a particular customer synchronised so that valid meter readings can be delivered in the billing system;
- there may need to be automated systems and asset tracking (e.g. bar coding and extranets), automated flows to third parties and automated data validation checks.

Data processing:

- managing and processing data will most likely entail development of new data management systems and business processes – these need to be fully tested to avoid ‘teething’ issues;
- the full end-to-end data management process will need to be synchronised (which is best conducted during the installation process);
- systems will need to cope with the large amount of data from smart meters (particularly if the volume of data is higher than normal at any time, owing to the need to reload metered data from previous days because of data communication problems);
- avoiding inconsistencies in smart meter data may require development of new data validation and editing processes;
- *w*hen an error occurs, data volumes are such that manual processes cannot cope with the number of exceptions created by a large number of smart meters – automated processes for fault resolution may overcome this;
- if there are problems with the communication networks (HAN or WAN), the data stored within the meter can be accessed subsequently once the fault is rectified but data retention within the meters needs to be adequate (e.g. at least 7 days).

Rules will have to be devised for gap filling for missing data, especially if the data are to be used for settlement processes.

7.2 Data cleaning and corrections

EDRP highlighted important issues with correcting for differences between seasons and years, and creating annual (or monthly or quarterly) consumption estimates from a few readings of non-smart meters.

- Weather-adjusted energy consumption data takes into account variations in weather conditions for the same day between different years.
- Weather-adjusted and time-corrected energy consumption data takes into account weather variations both between years and for different days within a given year).

Whereas SSE and Scottish Power appear to have been able to apply corrections without difficulty, EDF Energy observed "transformation of the data using time and weather correction factors does not appear to have been successful as seasonal patterns remain". E.ON also reported problems and applied weather corrections were to electricity consumption but no corrections were applied to gas consumption.

The issue may be that the standard corrections are adequate for billing purposes, where any errors would be corrected whenever a meter is actually read, but not sufficiently precise for experimental analysis. Unless and until the issues are resolved, there will be problems with longitudinal comparisons of energy consumption. If annual smart meter data for the same years are compared between well matched groups, then there is little problem. Any deviation from this situation increases the risks. The quality of non-smart meter data could be improved by either (a) selecting only homes with four meter readings in a year or (b) preselecting homes and ensuring a full set of meter readings are taken for those homes. Either approach would risk some bias but more so with option (a).

These corrections are systematic and applied to all data – they do not make the figures more correct, merely make data from different periods more comparable. Errors in the data were managed by a combination of tests run by the suppliers (e.g. exclusion of estimated readings, duplicates and rogue readings) and CSE (not published) and exclusion of outlying data points. The latter differed between suppliers:

- EDF Energy made no outlier exclusions;
- E.ON excluded daily advances that were negative or ≥400 kWh (gas) or ≥100 kWh (electricity);
- Scottish Power excluded all changes between pretrial and in-trial consumption of greater than ±50%;
- SSE excluded the top and bottom 5% of values and any changes of greater than 5,000 kWh (gas) or 10,000 kWh (electricity).

The Scottish Power approach may be the most rational (although the figure of 50% is perhaps arbitrary) because such a large percentage change – even if it is correct – suggests something happening outside the scope of the interventions. Any outlier exclusions could be excluding values that should have been included, thus distorting the analysis, but this is unlikely because of the extreme ranges used. The greater risk is that errors within the allowable range of values go unnoticed.

In the SSE trial, data cleaning removed 16% of households from trials. This was evenly distributed across Mosaic groups but it is not clear whether it was biased by trial group or response to interventions – either way it is clear that attention needs to be given to avoiding errors during the whole period of the trial, rather than trying to remove them at the end.

7.3 The analysis

The appropriate form of analysis will depend on the trial design and the distribution of data. In a factorial design with normally distributed data (or skewed normal data, following logarithmic transformation)[27], it should be possible to create a single statistical model of the effects of interventions, overall and for any predefined subgroups in the sample. Interaction effects should be determined if possible but occasional random findings can be expected and recognised as such when they occur.

Including confounding variables in the analysis serves to reduce the unexplained variance, making it more likely that genuine effects of interventions will be detected. It may also allow the analysis to show interaction effects, i.e. identify variables that modify the effect of interventions (e.g. one population segment being affected more than another). EDRP suggests that the former is more important, only one meaningful interaction with non-intervention variables being found (based on household size). Other confounding variables included in one or more of the EDRP analyses were:

- indirect demographic groups (Mosaic or Acorn/Ocean);
- number of lodgers;
- how financially well off the household describes itself;
- region within England, Scotland or Wales (16 regions in total);
- postcode (first 2 letters of postcode);
- grid supply point or gas zone;
- awareness of the trial;
- fuel type (electric only vs dual fuel);
- number of floors in the home;
- number of bedrooms;
- having a new gas boiler in the last two years;
- type of heating;
- insulation measures;
- number of electrical appliances ;
- whether or not the household had a programmer to control the heating;
- first language of household;
- postcode-specific average daily temperature.

The more variables included in an analysis, the more variance is likely to be explained but the sample size and inter-relationships among confounding variables may limit how many can be included in a single analysis. There may also be decisions to be made about the manner in which variables are to be included and how they are to be coded. One EDRP analysis, for example, analysed the effect of household size on energy use by dividing the sample into “two or fewer” and “three or more”. A boundary lower by one could be more meaningful because one person does not have to negotiate with other household members, whereas more than one person can disagree or, if both are

[27] The statistical technique of log transformation of data simply means taking the logarithm of each value (or each value plus 1, if zero values would otherwise exist). Note that this corrects for skewed distribution, not data processing or measurement errors.

committed to saving energy, can support and remind each other. Hence, the statistical modelling is not a straightforward process and it requires expertise, not just access to a statistical package.

To better understand the effect of confounders, it may be necessary to carry out prior qualitative work to explore the mechanisms of their effect.

One additional variable that should be included wherever possible is whether the household actually engaged with the intervention (e.g. looked at the advice or the RTD provided).

Where possible, there should be time series analysis to determine whether there is any variation in the effectiveness of interventions over time. If different periods need to be analysed separately, the full relevant period should be used, rather than limited “tests points” as used in one EDRP trial.

Where multivariate modelling is not possible, groups may be compared in pairs (e.g. each experimental group compared with its control group). There are two main risks in this:

- the more independent comparisons that are made, the greater is the risk of erroneously declaring an effect statistically significant;
- this problem is made worse if each group is compared with each other group but, without this, there is no basis for saying that one group had a bigger effect than another.

Where normal distributions cannot be obtained, the analysis may need to use non-parametric methods, which are less powerful but more robust.

Most research in the field, including three of the EDRP trials, analysed differences in consumption. The Scottish Power analysis was different, using ratios of consumption (between pretrial and in-trial periods). The reasons given for this were:

(a) it was felt that this better represented differential opportunities to save energy among high and low users;

(b) financial incentives were based on percentage (i.e. ratio, not difference);

(c) the ease of achieving absolute savings levels would vary with season.

However, using ratios means that saving a kWh is given different value for different customers and reductions from a high baseline are less likely to be detected. It is unlikely that customers would have thought about the ratio basis of the targets, especially since none asked how the target was calculated.

Any comparison involving data from non-smart meters will probably need to compare full years of data. The E.ON trial found that smart meter groups show greater seasonal extremes than the non-smart control groups for both electricity and gas, with significantly lower consumption than the control group in almost every spring/summer quarter in every group, sometimes outweighing higher consumption in autumn/winter quarters. This may be entirely due to “smoothing” of the non-smart meter data because of the need to interpolate infrequent readings over each quarter. If so, then analysis at sub-annual level would be valid only when comparing smart meter trials with smart-meter-only groups treated as control groups.

Testing the persistence of the effect of an intervention would ideally use two years of data with the intervention in effect and a longer period of less intensive monitoring of energy consumption. Analysis of changes over time can alternatively use a rolling year rather than comparing each month or quarter. This means that each successive year of data is shifted forward by (for example) one quarter, rather than a full year, and therefore overlaps the previous year by three quarters.

When analysing load shifting, as distinct from changes in overall consumption, a range of variables may be used:

- the consumption in each tariff or incentive period – which includes any overall reduction, not just the reduction specific to the period;
- the percentage of consumption that falls in the peak period (or each period) – which separates out the period-specific reduction but does not show where the "lost" consumption is going;
- calculated shift from one period to another, which is ideal but very difficult in practice.

The evidence is that weekday and weekend load shifting effects need to be separated in the analysis, and season included as a confounder if the analysis is at sub-annual level.

In all the analysis, a criterion for significance needs to be agreed. The default value is $p<0.05$ but this is not fixed. A higher value may be set where multiple independent tests are used or there is a particular risk in acting upon false positive outcomes. A lower value may be set if decisions are to be made on the "balance of probability" rather than "beyond reasonable doubt".

The description of the analysis should be explicit and complete in terms of statistical rigour but also re-expressed in terms understood by intelligent non-statisticians and related back to the aims of the project.

8 Technical development and logistics

The success of the roll-out will depend on having good equipment, in the right place, at the right time, and the equipment continuing to operate effectively.[28] Evaluation of the roll-out should therefore monitor these requirements and seek to minimise adverse impact on energy demand reduction.

8.1 Equipment development

New technical solutions may be needed and the timetable for this may not be predictable. This would have knock-on effects for the trial schedule, reliability of interventions and the customer experience. The roll-out evaluation should therefore be based on technical solutions being developed and proved (individually and in combination with other equipment) on a schedule that does not compromise the evaluation of impact.[29] Even with such prior testing and approval of equipment, actual technical performance and logistics during the roll-out would still be an evaluation issue.

8.2 Access to properties

EDRP identified some common practical issues in gaining access to properties in EDRP, as follows.

- *Contacting customers.* All EDRP trials highlighted a difficulty in initially contacting customers. For example, EDF Energy undertook its recruitment by telephone and it appears that 25% of its calls were to unobtainable numbers. A further 10% of calls were to numbers that did not accept number-withheld calls, customers who had moved address or customers who had insufficient English to take recruitment further.
- *Customers agreeing to a meter exchange.* In all the EDRP trials, the majority of those contacted refused to participate. For the national roll-out, there will be a requirement for the meters to be exchanged but other aspects of the roll-out interventions may be refused.
- *Accessing properties.* EDF Energy notes that 18% of those who had agreed to a date for a visit cancelled in advance (31% of these cancelled as ‘too inconvenient’; 17% due the terms and conditions). Suppliers also had difficulties due to no-one being present when the installer visited the property or the householder being present but deciding not to take part in the trial.
- EDF Energy also notes that even when recruited and installed onto the trial, gaining access to the home (necessary in this instance to implement a technical work-around) was not possible for around 5% of participants, despite significant effort. This finding is consistent with EDF Energy’s experience of customer contact generally.

In general, these issues will be expected to be common to any meter exchange. There is insufficient evidence from EDRP on whether there is a greater resistance to a smart meter replacement.

[28] In addition to the points made here, the recommendations made by suppliers in EDRP are reproduced in Annex A.

[29] As a general rule, the impact of a programme (reduction in energy demand, for example) should be evaluated only once it is clear that implementation of the programme has been successful (e.g. the smart meters are installed and working).

8.3 Installation

Installation failure/postponement is expensive and would tend to have a negative impact on evaluation because of some combination of:

- loss of participants;
- reduced engagement of participants (hence reduced impact and higher variance);
- disrupted evaluation schedule.

Technical reasons for smart meter installation failures in EDRP (having gained access to the property) included:

- inadequate WAN signal;
- health and safety issues;
- inadequate space to install meter;
- inadequate HAN;
- no access to meter;
- premises on an independent gas transporter network;
- other reasons related to the smart meter itself (e.g. access to turn off the electricity or gas, technical problems with meter during installation; problems with pipe work; electrical cable too close to gas meter; unable to remove existing meter; asbestos; floor alteration kit required and customer refused; could not guarantee boiler relight);
- other reasons related to specific interventions (e.g. the TV intervention had to interact with many items in the home including the metering, broadband access point, any existing set top box and the TV set-up; the heating controller had to be connected to the boiler, which in turn needed to be included in the HAN).

In the case of dual fuel customers, this includes cases where there were problems in the replacement of the gas meter and only a smart electricity meter could be installed.

In addition, some EDRP participants thought the installation took longer than anticipated or expressed concern about the amount of equipment that was involved, specifically in the TV intervention (EDF Energy) and 15% said there was minor disruption and 1-2% said that their intervention had been a major disruption (SSE). There is no information on the causes for the differences in responses (in particular, whether they arose from differences in expectation or the actual features of the installation, such as duration, noise, dust or interruption/delay of activities). Such occurrences may have an adverse impact on engagement if repeated in the roll-out.

The specific failures encountered in EDRP are likely to have been overtaken by subsequent improvements in the technology. The key point is therefore to anticipate and prepare for the various difficulties that may arise, prior to roll-out of smart meters or associated interventions. The most difficult cases may be where the customer has modified the meter enclosure or some aspect of the home, and is legally responsible for re-establishing access. While some of these incidences would simply require that customers are given enough forewarning, and that they are made aware of the access requirements so that they have time to comply, customers' resistance may be higher where the remedial work would be destructive and/or costly.

EDF Energy undertook a site survey prior to a separate visit to install equipment but decided that it is very unlikely to install smart meter systems that need surveys first, for reasons including cost, best use of time for skilled resources and also the impact on customer motivation with the delay it causes between recruitment and installation. This may affect the type of smart metering solutions that are suitable for wider roll-out.

8.4 Operation of equipment during trials

The performance of equipment during EDRP created some problems, partly due to insufficient testing of the equipment, both separately and integrated, in advance of the trials. Issue included:

- HAN and WAN data communication drop-out (sometimes resulting in differences in readings between the gas meter and the RTD);
- fluctuating signal strength (so data were stored by the meter and subsequently re-acquired);
- signal blocked by tins of food;
- power line communication (PLC) affected by items like fish tank pumps, anti-surge power sockets or other PLC equipment (e.g. BT home hub set-up) interfering with the internal data communications;
- if broadband used for communications, potential impact of the internet slowing down in periods of peak usage, router switched off by customer or service disconnected by provider;
- two different brands of smart meter and RTD being used in the same trial, with different levels of performance, without recording which homes had which brand;
- data security.

As with the installation issues, the specific issues are likely to have been overcome and new issues will need to be anticipated.

Annex A Supplier recommendations on practical and technical issues (reproduced from the EDRP report)

The energy suppliers highlight a number of learning points but this is a fast changing environment and some of the learning points highlighted below are already being addressed.

1) Sufficient time is required for product testing, system development and integration, staff training and development of customer services.

2) Suitably sized equipment needs to be developed for current installation scenarios. For example, energy suppliers have noted difficulties in installing meters both in newer homes with smaller existing meters and in semi-concealed gas meter enclosures.

3) To deliver a successful end-to-end process there will need to be interoperability standards. This is to ensure that products that will work together can be mixed and matched from different suppliers throughout the process. This includes common communication protocols and data formats and should allow for potential future applications, e.g. to help facilitate the smart grid.

4) The smart metering system needs to be relatively straightforward to install. This will minimise cost, disruption to the consumer and potential need to rectify problems subsequently.

5) Installation needs to be carefully programmed.
   a) EDRP highlighted significant problems in gaining access to some properties.
   b) Action should be taken to identify problem sites in advance where possible, to avoid delays to the installation programme.
   c) The installers need to be given adequate training and tools to carry out reliable installations on site and quality management processes put in place.
   d) As much checking as possible should be undertaken during the installation as to the correct operation of the individual products installed, as well as the HAN and WAN communication networks. This is to prevent the need for post-installation access, which is costly and may be problematic to achieve.
   e) Proactive work is needed to identify and engage with other parties (e.g. the Distribution Network Operator or the Gas Transporter) that may also need to be involved in the installation.

6) As part of the installation process, there needs to be a strategy to deal with properties that are difficult to access and/or where there is remedial work that the customer is legally responsible for and is necessary prior to installing the meters. Suggestions include the following.
   a) Clear guidance (e.g. through a marketing campaign) such that customers understand their responsibilities.
   b) Guidelines as to sufficient time to allow customers to provide access or take remedial action where necessary.
   c) Access to an (independent) body to advise customers on legal responsibilities, help vulnerable groups and possibly mediate in disputed cases.
   d) Processes to address cases of outright customer refusal to provide access or undertake remedial work.

7) The overall smart metering system needs to deliver reliable communication.
   a) Problems were identified with HAN and WAN communications. The majority of trials used low power radio transmission for HAN and mobile (SMS) communications for WAN. Whilst, overall, it appeared that these technologies were successful there were a significant number of problems during installation and operation and greater communication reliability is required if these solutions are used during roll-out. Even a small percentage of problems could result in a large number of homes being affected.

b) Alternative solutions may need to be developed for more difficult installations where the resultant signal strength is reduced because of, for example, meters located in enclosures, large distances between HAN communication devices, intervening building structures or poor mobile reception.

c) Whilst EDRP particularly focussed on the transmission of energy consumption data to the occupants and the energy suppliers, testing is also required of the ability to remotely update the metering technology, e.g. uploading tariff changes to the meter or updating the installed software.

8) A number of data management issues need to be addressed.

a) Manual data collection of fixed asset details (e.g. meter serial numbers) and manual input into back office systems creates a major risk of data errors. Automated systems and asset tracking (e.g. bar coding and extranets), automated flows to third parties and automated data validation checks would support national roll-out. This is noted as being a significant issue as suppliers will have to develop a wide range of new interacting systems to support the roll-out activities. This will include appointment booking systems, mobile platforms (to issue instructions and information to operatives and to enable capture of site-specific data), meter data management and meter asset management systems and new data transmission links with the proposed DCC and other partners.

b) A number of parties are involved in the transfer of data from the meters to the energy suppliers. This creates opportunities for different rules to be applied. There is a need for the development of underlying processes common to all.

c) Smart meter data will have inconsistencies and, therefore, data validation processes need to be agreed to check the data, and edit it as necessary, for correct operation. This is particularly the case where data cleaning needs to happen across several different market participants. Furthermore, rules will have to be devised for gap-filling for the half-hourly data if used for billing purposes.

d) Automated processes for fault resolution need to be developed. When an error occurs, data volumes are such that manual processes cannot cope with the number of exceptions created by a large number of smart meters. Smart metering systems need to have sufficient memory capacity to allow retrieval of meter data once any communication faults are rectified.

e) Testing is also required of data management from the energy supplier to the meter, e.g. uploading tariff changes to the meter.

Annex B. Glossary

ACORN	A demographic classification system. For further details see: http://www.caci.co.uk/acorn-classification.aspx
Benchmarking	Comparison of customer's consumption with that of similar households.
CAMEO	A demographic classification system. For further details see: http://www.callcredit.co.uk/products-and-services/consumer-marketing-data-and-segmentation/cameo-classifications/cameo-uk
Credit customer	Pays for the use of energy after it is consumed.
Grossing	Correcting data from stratified sample to represent the population from which the sample is drawn.
Hawthorne effect	People may change their behaviour merely because they know they are being observed or tested, regardless of any specific attempts to change their behaviour.
HAN	Home Area Network.
Interventions	Measures to reduce energy consumption.
Mosaic	A demographic classification system. For further details see: http://www.experian.co.uk/business-strategies/demographic-classifications.html
Multivariate statistical analysis	Analysis to investigate the effects of more than one variable simultaneously.
National roll-out of smart meters	The UK Government is committed to rolling out smart meters to both domestic customers and smaller businesses. For further details see: http://www.decc.gov.uk/en/content/cms/tackling/smart_meters/smart_meters.aspx
Non-parametric statistical methods	Techniques that do not rely on data belonging to any particular (e.g. normal) distribution or intervals in the data all having the same value.
Ocean	A demographic classification system. For further details see: http://www.caci.co.uk/ocean.aspx
Peak period	The period of the day in which there is the highest energy demand.
Prepayment customer	Pays for the use of energy in advance.
Real-time feedback	A means of providing current (i.e. live) information of the energy used in the home.
Regression to the mean	The statistical phenomenon whereby values above (or below) the mean are likely to decrease (increase) if measurements of the same group are made a second time.
RTD	Real-time display. Also often referred to as "in-home display" (IHD).
Smart meters	In the context of EDRP, these were electricity and gas utility meters that allowed the remote collection of half-hourly energy consumption data.
Stratification	The adjustment of sampling rates to ensure that certain population groups are sufficiently represented in the sample
Time-of-use (TOU) tariffs	Variable tariffs that change with time of day and/or between days. For example, the tariff structure can be set to be highest at periods of peak demand to encourage customers to reduce consumption during that period.
WAN	Wide Area Network.