ofgem

Making a positive difference
for energy consumers

Ofgem:
Guidance on Running Trials

Ofgem: Guidance on Running Trials

This guidance is intended for any energy supplier who plans to run randomised controlled trials. It describes the purpose of trials, why they are an invaluable method, both to gain insight into consumer behaviour and to inform policy decisions, and how to run them. Here you can also find out about Ofgem's support in developing and running trials, and there is a checklist with the minimum standards for conducting trials.

Contents

Section one: What is a trial and why are they useful? ..... page 3
Section two: Background to energy trials ..... page 3
Section three: Designing your trial ..... page 6
Section four: Running a trial ..... page 13
Section five: Trial checklist ..... page 16
Section six: How will Ofgem be involved in the process? ..... page 17
Annex A: Further resources ..... page 18
Annex B: Expression of interest for trials ..... page 18

Section one: what is a trial and why are they useful?

A randomised controlled trial (RCT), or simply, a trial, allows the researcher to find out if an intervention works or not, and determine how effective it is. Trials let you measure actual consumer behaviour, as opposed to reported behaviour.

For example, you can ask participants to report using a survey how much energy they use at home by recording the number of times they run various appliances. But this is likely to be very different to what they actually use, which can be measured accurately with energy meter readings. It is partly because people cannot accurately recall, or misremember their behaviour or because they feel influenced by surveys to report an answer they feel is right, rather than true.

Government and regulators are using RCTs more and more. In an RCT, participants are randomly allocated to either a control group or one or more treatment groups. The control group is treated as though there was no trial, and the treatment group(s) receives whatever new intervention or policy is being trialled.

By comparing outcomes at the end of the trial, researchers can see if the treatment was similar to, more, or less effective than the control group. So, crucially, the control group provides a counterfactual – what would have happened without the intervention. This is what makes evidence from (properly constructed) trials robust and convincing. We encourage suppliers to use them to inform their understanding of their customers, to improve products or services, or provide robust and credible evidence to influence external stakeholders, policymakers and regulators.

Section two: Background to energy trials

In 2016, the Competition and Markets Authority (CMA) published their final report on the energy market investigation. They recommended that Ofgem establish a programme to identify, test (through RCTs where appropriate) and implement measures to give domestic customers and micro-businesses different or additional information, and get them more engaged in the retail energy markets.

The CMA further recommended that Ofgem introduce a licence condition to require suppliers to participate in this programme, to help ensure the programme's effective implementation. We have now introduced this licence condition (SLC 32A) and published selection criteria, setting out transparent criteria for how we will select suppliers to partner with us on trials relating to domestic customers.[1]

We believe trials can help us gain better insight into what works best for different consumers and make customer communications more effective. These trials will be developed and run in partnership with energy suppliers. In some cases, Ofgem will lead the trials, in others, we will expect suppliers to take the lead and to provide the majority of the resource for the planning and delivery of the trial.[2]

[1] https://www.ofgem.gov.uk/system/files/docs/2017/01/slc_32a_decision_final_website.pdf

[2] Unless otherwise specified in a direction or a derogation, suppliers will be expected to comply with usual licence conditions.

When should an RCT be used?

Not all research questions can or should be answered by using an RCT. Some will require qualitative research, some can be tested by surveys, some by user testing.

An RCT should be used when:

- You want to conclusively say whether an intervention works or not[3]
- There is a clear, measurable behavioural outcome
- The likely benefit of the intervention outweighs the cost of trialling

RCTs are experiments which are conducted in the 'field', which means they take place in real world settings. Lab tests use a similar methodology, but take place in a highly controlled environment (not necessarily a lab). Many lab tests are now done online with people who have chosen to take part. For example, an online experiment may randomly show a group of participants one of two versions of a letter, and then participants are asked how well they understood the letter and how they think they would have acted on receiving it. The disadvantage of lab testing is that they don't necessarily provide a real world understanding of behaviour. But they can often be done cheaply and quickly, and are a good way of getting an indication of whether a field trial would be worthwhile.

Ways to design trials

There are many different trial designs. Here is an outline of each one.

- Parallel: Each participant has an equal chance of being allocated into a control or treatment group, and each group receives a particular intervention or business as usual treatment, depending on what intervention is being trialled.
- Cluster: groups of participants (for example, postcode groups of customers) are randomly allocated to either a control or treatment group.
- Factorial: Each participant is allocated to a group which receives a particular combination of interventions. For example, participant one receives letter one followed by email one, but, participant two receives letter two followed by email one. This design allows you to isolate the most effective part of your intervention. For example, the table on page 5 is a 2x2 factorial design that evaluates the effectiveness of two letter types for a first contact letter followed by a second follow-up letter. This design creates four experimental conditions.

[3] It is very likely that price increase notifications, press coverage, or other high profile activities will affect customer behaviour and create 'noise' in your trial, which makes it harder to see the specific effect of an intervention. As far as is possible, try to conduct trials in periods of time which avoid these activities.

Experimental Condition	First Letter Type	Second Letter Type
1	1	1
2	2	1
3	1	2
4	2	2

Alternative research methods and designs

Doing a trial is a serious decision, so take the time to work out the most appropriate research method to answer your research questions. Sometimes alternative methods may be more appropriate than an RCT. This section introduces other research methods, but for a fuller description, see the Magenta Book.[4]

Quasi-experimental methods

If an RCT is not possible because, for example, randomisation is not an option, then a quasi-experimental method can be used instead. Examples include difference-in-difference designs[5] and using a matched comparison as a control group.

Surveys

Surveys are used to get quantitatively robust data on a population. They are useful for capturing information on attitudes, beliefs, opinions, characteristics and behaviours. However, surveys are generally weaker at understanding past or hypothetical decision-making or behavioural drivers. In these cases, techniques such as ethnographic observation are more appropriate.

Just as survey design can be varied to meet different needs, the actual data collection design can also take a number of forms. Broadly speaking there are two types of survey design: self-completion (where the respondent fills out a survey on their own) and interviewer-administered (where an interviewer takes a respondent through a survey). Surveys can be completed face-to-face, over the phone, online or via the post.

[4] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf

[5] Difference in difference is a statistical technique that attempts to mimic an experimental research design by using observational data and comparing the average change over time in the outcome variable for the treatment group, compared to the average change over time for the control group.

Qualitative methods

Qualitative methods include interviews, focus groups, observational research, and case studies, amongst other methods. They are normally more in-depth than a survey, and are often used to understand why people act in certain ways or hold certain attitudes. The more open and unstructured element of qualitative research (relative to quantitative research) allows you to understand in greater detail the nuance and breadth of response to issues. However qualitative research is not suitable for gathering any quantitatively robust data, and findings from qualitative research cannot be assumed to be true of a wider population.

It is often advisable to use some of these other research methods alongside a trial to inform its design, or to help understand why it worked or not. For example, in a trial to see if a new leaflet was effective at prompting switching, you may want to do some testing with users, such as focus groups, before your trial to make sure your audience understands the leaflet. You may also want to do a survey after your trial to understand how many participants read it and conduct qualitative interviews to understand how they felt about the information in the leaflet.

Section three: Designing your trial

Ofgem is interested in generating robust evidence on customers in the retail energy market. We expect that proposals for trials will be primarily quantitative, although they may be accompanied by qualitative research.

This is because although an RCT will tell you what works, it will not tell you why. We are interested in customers' actual, rather than intended, behaviour. So only interventions which result in measurable behavioural outcomes (such as rate of switching, as opposed to a customer's willingness to switch) will be considered for Ofgem or supplier-led trials.

You might find there are many experimental and quasi-experimental methodologies appropriate for you[6], but Ofgem prefers suppliers aim for an RCT where possible.

This will result in trials that involve hundreds, and possibly thousands of customers, and may take several months. Some trials may be done with a wide group of customers (for example, all customers on a standard variable tariff for over three years), or with a particular, targeted group (for example, only customers with restricted meters).

There are many steps to complete before a trial goes live. This will make it more likely that the trial can proceed as planned and will yield robust evidence. Bring these steps together to form a trial protocol, a document which records all the details of why the trial is being run, how and by whom. This also includes any instructions about how the logistics of the trial will run in practice. The trial protocol should be agreed before the trial goes live.

[6] It is not always possible to meet all the conditions needed for an RCT, and in these cases, quasi-experimental methods can be used. These methods aim to mimic randomisation as far as possible to try to ensure any impact can be attributed to the treatment being tested. See here for more information: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/190984/Magenta_Book_quality_in_policy_impact_evaluation__QPIE_.pdf

The next section describes the steps in designing your trial.

1. Define a SMART research question

Before designing and developing a trial it is essential to have a clearly articulated research question that you are trying to answer. Try and make the question as SMART as possible (Specific, Measurable, Assignable, Realistic, Timebound).

Your research question might be based on previous research or testing, or knowledge of the market based on past experience. It is also important that the question you are seeking to answer is based on established behavioural theory. It is only by being clear on the question you are trying to answer that an appropriate methodology can be developed and the treatments that you are applying evaluated robustly.

RCTs are designed around one primary research question, but can also be designed to answer other, secondary, research questions. However, attempting to design a trial to answer too many questions can make your trial complicated to design or interpret. It is important to recognise that the answer to your primary research question may well be that your intervention has no impact, or makes outcomes worse. This can occur despite your intervention being well-designed and founded on good evidence.

2. Identify the trial population

First you need to determine which group of people you want to target with your intervention, or your 'population'. The 'population' used for the trial are then all the individuals who meet your inclusion criteria. For example, it may be all domestic energy customers on fixed tariffs. Choose inclusion/exclusion criteria to ensure that the population represents the group of individuals who would be targeted by any treatment if it were to be rolled out more widely. These criteria need to be things that can be identified from new or existing data.

3. Define your outcome measures

Next, you need to define the primary outcome you are interested in measuring. This will partially be decided by your primary research question, and partially by the data available. In the context of increasing customer engagement, the outcome measure will often be proportions of customers switching tariffs or suppliers.

Your secondary research questions may require different outcome measures to be defined, for example call volumes or emails to suppliers. Sometimes, ideal outcome measures may not be available, and you may need a proxy measure instead.

4. Specify your intervention

Be clear at the outset about what it is you want to trial. This is known as your intervention. This will increase confidence in the validity of your results, help anyone who wants to replicate the trial[7], and will also be invaluable if the intervention works and we want to roll it out more widely.

It is possible to use an RCT to test a number of different interventions at once, or a number of versions of the same intervention. In these cases, it is very important that the treatment groups only differ in the essential features that you are hoping to test. For example, you may want to know the most effective of four messages used in marketing material. In this case, only the messages themselves should be changed, and the rest of the leaflet should remain the same. Otherwise, you will never be sure if it was the message itself that was the effective part.

[7] We encourage suppliers to be transparent with Ofgem regarding the methodology used for any trial, so that it could theoretically be replicated. However, we do not expect suppliers to share all such details with their competitors.

Align your intervention design with any planning of the design of the trial and how it will be analysed. Consider at this stage how practical and cost-effective the intervention would be if rolled out on a wider scale.

5. Identify your trial arms

A trial will compare outcomes of participants in a control group to those of participants in one or more treatment groups. These are known as trial arms.

In a trial, a control arm should include individuals who are part of the trial but are not given the treatment. Data about this group is still collected and analysed, but they receive 'business as usual' treatment.

This group provides the comparison (or counterfactual) to the effectiveness of your treatment. For example, if you wanted to trial a simplified version of a customer bill, the control group would continue to receive the standard bill as normal, and the treatment group would receive the simplified version. Any difference in outcomes between the two groups is compared at the end of the trial, with the control group providing the baseline.

The number of treatment arms that you can have will depend on the research question you are trying to answer, the potential number of participants in your study trial, and the available budget and resources.

6. Choose your unit of randomisation

Most trials are randomised at the individual level, so each customer has an equal chance of being allocated in to each one of your trial arms. However, in some cases, the randomisation will be done with groups of individuals, this is known as cluster randomisation. For example, this might be done by local authority area, postcode, or distribution area. This can make a trial practical, and help avoid contamination, but can increase the analytical complexity, and require more participants than individually randomised trials. We would recommend seeking specialist advice for the design of cluster trials.

7. Define your effect size

Trials are designed to measure a particular size of effect (called the minimum detectable effect size) between the control group and the treatment group. The effect size will change from trial to trial, and will depend on the research question, the potential cost of the trial, and rollout of a successful intervention. For example, if you are trialling a new version of an existing letter, which costs no extra to produce, then a small effect, like a 1 per cent increase in your primary outcome measure, may be enough for you to consider the new letter successful and worth your while rolling out to all customers. It is important to specify early in the trial design process what the effect size you are interested in detecting so you can calculate your sample size accurately.

8. Calculate your sample size

The sample size is the number of participants in your trial broken down by trial arm. It is important to have a sufficiently large sample so you can be confident in the findings from your trial. Unless suppliers have in-house statisticians, we recommend you seek external advice on how to carry out statistical calculations to establish what a suitable sample size would be. How the sample size has been determined, the sample size used and the rationale for these choices should be included in your trial protocol. You can use these calculations to determine whether a study has enough "power" – the probability of detecting an effect statistically, given that one exists. This would let you design the trial in a way that would be powered sufficiently to detect an effect size of interest on your primary outcome measure (for example, switching rates). Control and treatment arms are normally expected to be equivalent sizes.

9. Work out your trial duration

In many instances, the duration of the trial will be determined by the necessary sample size. For example, if sample size calculations show that 1,000 participants are needed, but only 100 participants are eligible for inclusion in the trial each month, the trial will need to be run for 10 months. For some interventions, you may be interested in their durability (whether the impact diminishes over time). In these cases, you would want to measure outcomes over a longer period (for example, if a new prompt had been developed for a bill sent out every three months, you might want to measure outcomes over multiple billing periods, to see if the impact of the prompt increases or decreases over time, or if there are any seasonal effects). For others, such as a new customer communication you might want to measure for a couple of months to avoid just measuring the 'novelty effect'. Establish the likely duration of the trial before the start of the trial, and record this in the trial protocol.

10. Identify your sample and randomise

An RCT involves comparing the impact of an intervention against the baseline level of a control arm. It is important that a representative sample[8] is taken from all of the individuals who meet the criteria for being included in the trial (known as the trial population) and then randomised into the various trial arms. So, for example, for a three-armed trial, you would randomly allocate your sample to each of the three arms in equal amounts. For example, if you have 500,000 in your eligible population, you randomly allocate each participant to one of 4 conditions: (1) Out of trial (n=425,000), (2) Treatment Arm 1 (n=25,000), (3) Treatment Arm 2 (n=25,000), and (4) Control Arm 3 (n=25,000). This means that each customer in your sample has an equal chance of being allocated to each of the three arms, including the control arm, to reduce the chance of introducing bias into your trial results.

Randomly allocate individuals who are part of the trial to the control arm or treatment arm, using appropriate software and methodology. Guidance on this should be sought from trial statisticians. Do not use approximations for randomisation because it introduces the risk of bias into your sample. For example, alphabetising a list of participants' surnames and dividing into groups could be used as a proxy for randomisation, but since different ethnic groups have higher proportions of surnames beginning with different letters, you risk that the groups will not be representative of the overall population. The participants should be 'blind' (i.e. unaware of whether they have been allocated to a treatment or control group)[9].

[8] A representative sample is one which accurately reflects the make-up of the general population.

[9] To ensure participants are 'blind' to an intervention, you should also minimise any communications that might increase the visibility of your trial and jeopardise your results.

Randomisation can, by chance, produce skewed outcomes. After random allocation, you should conduct balance checks to ensure that each trial arm is balanced - i.e. the people in each arm are roughly similar. Carrying out balance checks will depend on the information you have available about your sample. For example, you might want to check that the average energy consumption, or average length of tenure are similar in each trial arm.

11. Avoid contamination

Contamination across trial groups should be avoided as far as possible so no bias is introduced. An important factor in avoiding contamination is ensuring that participants do not become aware of the treatment they are receiving and the existence of other treatment variants e.g. by communicating details of the trial publically.

To ensure any impact observed is as a result of the intervention, do not engage in any additional communication (other than business as usual) or other trials with the trial sample during the duration of the trial. Ideally, you should protect the whole trial sample from any such communication while the trial is running, but if this is not possible, then aim for consistency across the trial arms.

You also need to consider whether there is any potential for the treatment arms to be mixed up, and ensure that customers are exposed to the correct treatment arm as planned.

In most trials, there will be some customers who become ineligible for the trial while it is running. This is known as attrition. For example, a customer may be chosen to participate in a trial of a new leaflet, and then before receiving it, chooses to switch supplier. Your trial should have enough participants to be sufficiently powered to cope with your expected level of attrition. Any customers that do need to be removed from the sample, for whatever reason, should not be replaced. The randomised design means that this should be no more likely in one trial arm than another.

12. Consider the ethics of your trial

Many interventions will be variations of existing service delivery, and therefore there will be little risk that customers come to harm as a result of these interventions. However, there will be exceptions to this, and we encourage suppliers to discuss and clarify any potential ethical issues[10] before a trial begins. In the majority of cases, it is will not be necessary to tell customers that they are part of a trial. Revealing this information may influence their behaviour, and introduce bias to the trial. Only if there is a risk of any harm coming to customers, would they need to be informed of the trial and of these risks. Sometimes, it may be necessary to tell customers about a trial for practical reasons, for example, to allow their data to be shared between trial partners, or to gain consent to be contacted as part of any follow-up research. Trial protocols should consider ethical issues. They should also include when a trial would be stopped – if, for example, it became clear the trial was having a strong negative impact on customers. This might require monitoring while a trial is live, for example by recording the volume of extra customer calls or complaints that the trial might generate.

[10] See the Government Social Research guidance on ethical issues in social research for a discussion of some of the ethical issues you may face: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/515296/ethics_guidance_tcm6-5782.pdf

13. Address any logistical issues

Trials often involve making changes to existing systems or processes. Even simple sounding interventions, such as changing the text on a current letter to customers, will involve changing systems to allow two versions of a letter to be sent out concurrently, changing the text on the letter template, and getting sign-off for the new wording. These issues need to be ironed out before the trial begins.

Example: End of fixed-term tariff trial

One example of a new prompt to encourage consumer engagement that the CMA recommended was changing the name of the default tariff. In this example we are interested in finding out whether changing the name of the standard variable tariff affects customer behaviour.

A SMART way of setting out this research question would be: 'Does changing the name of the standard variable tariff on the end of fixed-term notice have an impact on the proportion of customers who actively choose another tariff, rather than defaulting onto the standard variable tariff?'

All customers due an end of fixed-term notice (EFTN) in a specified time period would be eligible for inclusion in the trial. Customers who switch suppliers before receiving their notice letter would be excluded from the trial. The outcome measure would be the proportion of participating customers who actively switch tariffs during the trial period. The randomisation will be carried out at the individual level, with customers having an equal chance of being placed in each of the two arms. The customers in both arms would be recruited to the trial at the same point in time.

The trial arms would be:

- Arm one (control arm): Customer receives a standard EFTN with the current wording
- Arm two (treatment arm): Customer receives a standard EFTN with the standard variable tariff described differently e.g. as the 'out of contract'.

You will need to understand the baseline rate of switching among this customer base to calculate the sample size. This should be available from existing data held by suppliers.

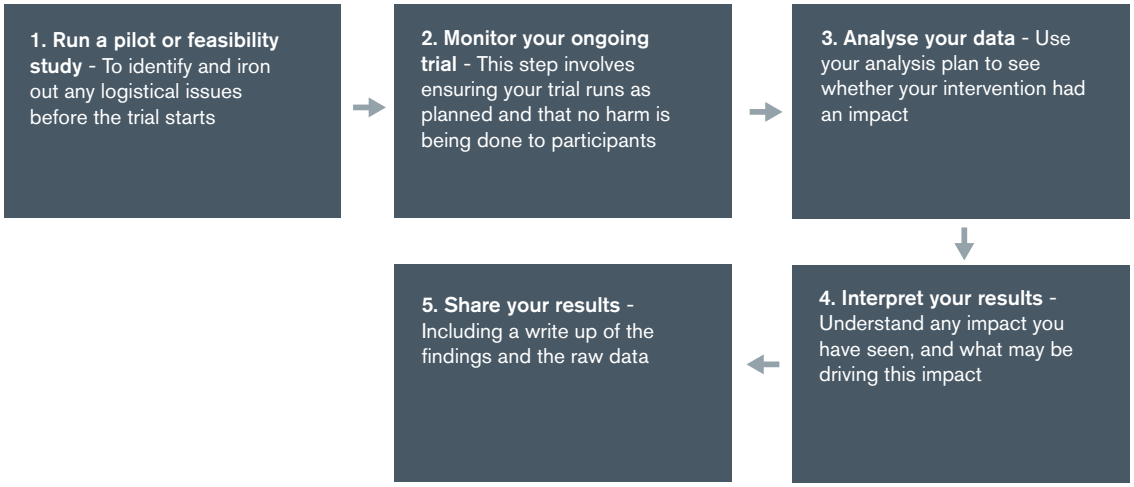
The trial will need to give customers a chance to act after receiving the letter. The same amount of time should be given to those in the control and treatment arms. Around four to six weeks is probably sufficient in this case, but this should be informed by suppliers' knowledge of their customers' behaviour.

To avoid contamination, remove from the sample any customer who decides to take action before the EFTN is sent. For example, a customer has been randomly selected for the trial, and is due an EFTN letter in September. They decide to switch supplier in August and are therefore no longer eligible for the trial. They should be removed from the sample and not replaced. Suppliers should look at historical data to understand how many customers due an EFTN would normally be expected to behave in this way[11].

[11] Note that this may not always be the case. In some trial designs an intention to treat (ITT) analysis may be more appropriate. This means that all participants are included in the final data analysis, regardless of whether they received the intervention as planned or not. This is particularly important if the trial is designed to investigate how an intervention would work in a real life environment.

Section four: Running your trial

Once you have completed the steps above, you should have a trial protocol, which includes any instructions for those running or delivering the trial. You are now ready to proceed with your trial.



1. Run a pilot or feasibility study

RCTs can be complex, costly and time-consuming. It is useful if potential logistical and practical issues are identified and ironed out before a full RCT starts. We recommend that a smaller-scale pilot is run before a full trial, with the aim not to measure outcomes, but instead to learn about the process of running the trial. For example, does the randomisation strategy result in equally-sized trial arms, are there any problems with data flows etc?

2. Monitor your ongoing trial

Monitoring implementation is a key part of trialling; for instance, to let you check that randomisation is correctly being carried out, and if the intervention is being delivered in the way that it was designed to be. Aspects of delivery that you should monitor include:

- Have inclusion and exclusion criteria been followed correctly?
- Have customers been allocated to the treatment or control group?
- Have sufficient numbers been recruited to the trial, and at the right levels for each treatment group?
- Have trial customers received the intervention as set out in the protocol?
- Have customers allocated to the control group experienced ‘business as usual’?

These steps should be audited, ideally by separate staff, to ensure they have been carried out correctly. In particular, the application of inclusion/exclusion criteria, and the random allocation of customers to trial arms, are crucial aspects of delivery and should be thoroughly quality assured. This will save you time later, and increase the likelihood of a successful and robust trial.

3. Analyse your data

There are many ways to evaluate results of trials, and the method used will depend strongly on the trial design. It is important to plan how you will do your analysis before the trial starts, as this will make your findings more robust. Multiple regression is the most commonly used method for analysing RCT data. Regression involves testing the ability of your intervention to predict an outcome variable, while controlling for (or holding constant) a range of other variables. Regression can be used with RCTs to get more precise confidence intervals[12] than when using ANOVA[13] or t-tests[14], and for this reason is recommended over such methods.

We suggest discussing your plans for analysis with trialling experts if needed.

4. Interpret your results

Generally, the analysis of trial data tests whether there is no difference between the treatment and control groups on the primary outcome measure that you specified in your analysis plan, i.e. that the intervention being trialled has no impact. This is known as the null hypothesis. If you find a statistically significant difference between the groups, this will allow you to conclude that any difference in impact between the treatment and control groups is down to the intervention, rather than chance. In statistical language this is called “rejecting the null hypothesis”, and means that the intervention did have an impact.

The analysis methods used will give information on the:

- statistical significance of the observed difference
- uncertainty associated with the impact estimate
- importance of the impact for policy development (but this depends on the size of the impact, along with the policy context and the cost of the intervention)

Statistical significance is shown as the p-value, and represents the strength of the evidence against the intervention having no impact. The smaller the p-value, the stronger the evidence that your intervention has worked. The p-value should be compared to your pre-specified rules or cut-off value. This value is normally 0.05. If the p-value is above your cut-off, this means that any effect is down to chance, if it is below, this means you can be confident that any impact is down to your intervention.

[12] A confidence interval is a range around a measurement that conveys how precise the measurement is.

[13] Analysis of variance (ANOVA) is a collection of statistical models used to analyse the differences among group means and their associated procedures

[14] T-tests assesses whether the means of two groups are statistically different from each other

It is important to note that trials can generate a lot of data allowing a range of statistical tests to be performed on different variables within your dataset, and it is likely that one or more of these tests will result in a statistically significant finding. However, this may still be down to chance rather than the impact of your intervention. This is why it is important to specify your analysis in advance of receiving your data to ensure transparency, and so that your data is not misinterpreted.

Even if your analysis shows that your intervention has caused a statistically significant change, the actual importance of that change will have to be assessed in a wider context. For example, is an intervention that creates a (statistically) significant yet small behaviour change worth the cost of rolling it out more widely? This is a judgement to be made, based on wider aims and resources.

5. Share your results

We encourage you to share with Ofgem your trial protocol, a write up of your trial results, as well as anonymised trial data. We will treat all results in the strictest of confidence.

Section five: Trial checklist

To design a robust trial, the following four key questions should be addressed:

- Is there a clearly defined, SMART, primary research question?
- Is the outcome measure a measure of actual, rather than intended, behaviour?
- Do you have a way of identifying the population of interest in order to define the sample for your trial?
- Is the trial sufficiently powered to detect a specified increase/decrease in the primary outcome measure (e.g.switching)?

The checklist below provides a fuller list of questions that we encourage suppliers to ask to determine whether the trial proposal is sufficiently robust to provide useful evidence that could be used to inform policy.

- ✔ Is there a clearly defined, SMART primary research question?
- ✔ Is the outcome measure a measure of actual, rather than intended, behaviour?
- ✔ Do you have a way of identifying the population of interest in order to define the sample for your trial?
- ✔ Is the trial sufficiently powered to detect a specified increase/decrease in switching?
- ✔ Are you able to randomise your population?
- ✔ Do you have a way of tracking the outcomes for your sample?
- ✔ What are your criteria for inclusion and exclusion from the trial?
- ✔ How many arms will your trial have?
- ✔ Which interventions will be tested as part of your trial?
- ✔ What is the difference between the interventions? Are they suitable to address your primary research question?
- ✔ Do you have a plan for analysing the data from your trial? Has this been reviewed by a statistician?
- ✔ Will you be able to share full (anonymised) data with Ofgem?
- ✔ Will you be able to report full findings (including any learnings about the process of running the trial) to Ofgem?

Section six: How will Ofgem be involved in the process?

Ofgem is keen that all trials (both those led by us and those led by suppliers) are robust enough that we can use the evidence generated to improve outcomes for consumers. For Ofgem-led trials, we will lead on drafting the trial protocol, issue directions where necessary, and collaborate closely with suppliers. While we cannot not pay for suppliers' costs incurred in designing or running a supplier-led trial, we can work as partners by reviewing trial plans and providing advice.

Alongside this guidance we have published an open letter which sets out our approach to supplier-led trials, including areas of interest for trialling. If you are considering running a trial in one of these areas, we encourage you to get in touch with us early in the process. To help you, we have developed a template (in Annex B) which you can use, if helpful, to outline your proposed trial idea and share key information with us.

Sharing results with us

For us to consider the evidence from your trials as part of our policy development work, you must share your results openly and transparently with us, even if a trial did not proceed as planned or did not have its intended impact.

How we decide to monitor ongoing trials will depend on the complexity and scale of the trial. As a minimum, we will need to be assured that there are plans for ongoing monitoring before the trial starts.

Annex A: Further resources for conducting good trialling and developing prompts

The following resources may be useful further reading in developing your trials:

Test Learn Adapt[15]

Magenta book[16]

Quality in impact evaluation[17]

EAST behavioural framework[18]

Annex B: Expression of interest for trials

If you are interested in conducting a supplier-led trial, please provide us with relevant information (as outlined in the example template below), and send to promptsenquiry@ofgem.gov.uk.

	Example
Trial name	Framing the name of the standard variable tariff in End of Fixed Term Notices (EFTNs)
Submitting organisation	Supplier A
Lead contact and contact details	Jen Brown, Project Manager, Supplier A Jen.Brown@supplierA.com; 0208 123 4567
Relevant internal experience	We have completed other trials testing variations to EFTNs
Description of intervention	All customers on fixed tariffs receive a notice telling them they will shortly be moved to the standard variable tariff if they do not take action. We want to test whether changing the wording of this notice, and particularly, reframing the name of the SVT to the 'out of contract' tariff, results in a change in behaviour.
Overview of project	We want to work with Ofgem to run a trial to understand if making a simple change to the EFTN sent to customers six weeks before the end of their fixed tariff increases levels of engagement.
Rationale for project (including any existing evidence)	We know that customers are confused by tariff names. We suggest that by reframing the name of the SVT as the 'out of contract' tariff, it may prompt customers to switch or re-fix after receiving their EFTN
Population of interest (including size of population)	Customers due to receive an EFTN in September and October 2017.

[15] https://www.gov.uk/government/publications/test-learn-adapt-developing-public-policy-with-randomised-controlled-trials

[16] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf

[17] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/190984/Magenta_Book_quality_in_policy_impact_evaluation__QPIE_.pdf

[18] http://www.behaviouralinsights.co.uk/publications/east-four-simple-ways-to-apply-behavioural-insights/

Proposed method (including any other complementary research alongside the trial)	We propose to run a RCT, randomising customers due to receive an EFTN into two groups, one who would receive an existing EFTN, one a new one with a different framing for the name of the standard variable tariff.
Research questions	Does changing the framing of the name of the SVT on an EFTN impact on the proportion of customers switching or re-fixing?
Trial arms	Two: Control - receive an EFTN as normal Intervention - receive an EFTN with new wording to describe the SVT
Estimated sample size	10,000 (5,000 in each arm)
Outcome measures	The proportion of customers switching to a new tariff or re-fixing. This data is already routinely recorded by Supplier A.
High level analysis and reporting plan	Our internal statisticians will undertake analysis comparing the number of customers who switched tariff or re-fixed after receiving their EFTN in the two trial arms. We will use logistic regression to isolate the main effects of the letters. Results will be shared with Ofgem in a Word report. Anonymised data will also be shared at the end of the trial (expected to be in early Winter 2017)