

## **When and how we use field trials**

July 2018

# Contents

1	Executive Summary	2
2	Introduction	4
3	When are we likely to use field trials?	8
4	How we run field trials	14
	Annex 1: References	22
	Annex 2: Trial Analysis Plan	24
	Annex 3: Ethics policy and procedure	35

# 1 Executive Summary

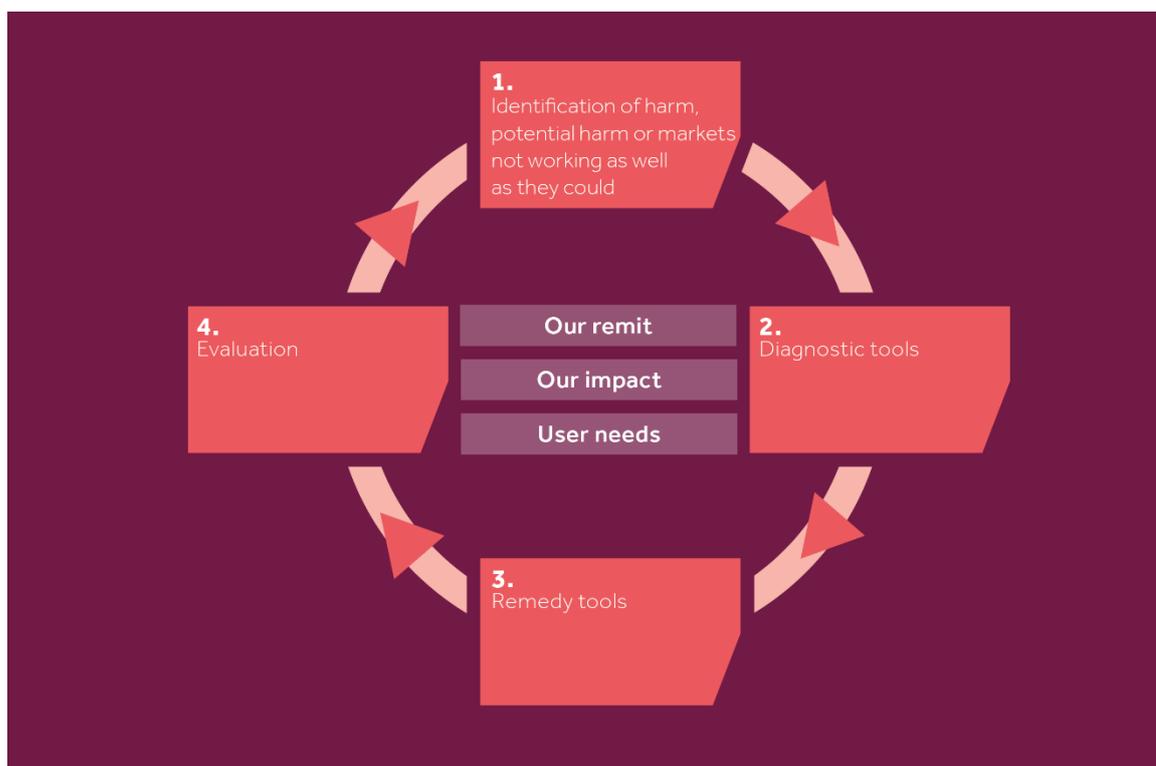
- 1.1 Field trials are real-life experiments which test directly whether proposed interventions actually work. This makes them powerful tools for gathering evidence for making policy. But, as with all research methods, they come with costs, such as time and resource. This paper explores the circumstances in which we would be more, or less, likely to consider the use of a field trial and gives an overview of how we would normally carry out a trial.
- 1.2 Field trials (a type of randomised controlled trial or RCT) help us to understand the causal impact of our policies. In an RCT, we divide the population to be tested randomly into two or more groups. One group receives service as usual (the control group) and the other receives the new intervention (the treatment group). Because the groups contain people who should be roughly similar to each other on average, we can then compare the difference in outcomes between the treatment and control group. This difference is a precise measure of causation, telling us exactly what effect our intervention has had.
- 1.3 As part of our decision making framework, which is set out in the FCA Mission, we explore remedy tools after identifying harm and diagnosing a problem. Field trials help us to assess what remedies are likely to be effective, because they allow us to test them on a subset of real consumers making real decisions. They are also the best way to estimate causation, as they control for other factors which might affect consumer decisions; many other research methods simply identify correlations. However, field trials can take time – usually months, if not years – and require the cooperation of financial service providers. Costs are often lower than other research activities, depending on the amount of in-house expertise. In many cases, other methods may be more suitable and proportionate to inform policymaking.
- 1.4 We ask three main questions when deciding whether to carry out a field trial:
  1. Is a field trial possible and appropriate? For example, does it test consumer behaviour in a representative setting and a timely fashion?
  2. Is evidence from a field trial important for the policy decision? We are more likely to consider a field trial an important part of the evidence base when there is little existing evidence, when other methods are less suitable and when the proposed intervention would pose high costs.
  3. Is a field trial proportionate? We will only conduct field trials when it means we are using our resources in an efficient and economical way and we are imposing burdens which are proportionate to the expected benefits.
- 1.5 We undertake a number of steps when carrying out a field trial, which are set out in Chapter 4. First, we aim to diagnose the problem and design remedies which address this as closely as possible. We then call for partners, with whom we work closely to discuss trial parameters and get feedback on treatment designs. We document our plans through the trial protocol, which includes undertaking an ethics review and developing a pre-commitment to analysis. After carrying out the trial, we analyse the results, conduct quality assurance, and communicate them to the firm, before publication. We then collect feedback and review the lessons we have learned.

- 1.6 By setting out the factors we consider when deciding whether to conduct a field trial and explaining the steps that we go through, we hope to provide industry with clarity about our approach.

## 2 Introduction

2.1 'Our Mission 2017' commits the FCA to using its regulatory tools to prevent and reduce harm in a cost-effective way and thereby create public value. We have adopted a decision-making framework to focus decisions on public value, as shown in Figure 1.

**Figure 1: Our decision-making framework**



2.2 We identify and diagnose harm, and decide whether it is cost effective to use remedial tools (e.g. new policies, disclosure, price caps) to address that harm. Deciding when these will be cost effective, or effective at all, is one of the toughest challenges we face as a regulator. There are a number of ways we can assess remedies to decide if they will be effective. Field trials - randomised controlled trials conducted in the real world - are one part of a broader set of evidence-gathering techniques that we can use to help develop policy. We have used these to provide evidence for policymaking in the savings, general insurance and pay-day loans markets, among others.

2.3 The FCA makes rules based on all the evidence available from a variety of sources. Field trials will not always be a suitable or necessary part of that evidence base. Alongside this paper we are publishing a document on 'How we analyse the costs and benefits of our policies' and an Occasional Paper on economic welfare estimates for regulation, both of which are relevant to this paper and to how evidence informs our decision making.<sup>1</sup>

2.4 As well as thinking about rules before we make them, the FCA is committed to evaluating our interventions after we have implemented them, where possible and

<sup>1</sup> See <https://www.fca.org.uk/publications/corporate-documents/how-analyse-costs-benefits-policies> and Lee (2018)

relevant. We recently published a Discussion Paper on our ex-post impact evaluation framework.<sup>2</sup> Where field trials are not feasible, it may be appropriate to stagger roll-out of the intervention using a stepped wedge to assess any changes in behaviour and outcomes.<sup>3</sup> A further approach to minimise ineffective policy, is to implement sunset clauses that are triggered in certain circumstances (for example unless an evaluation has found the policy to be effective in some predefined way).

- 2.5 Field trials provide the FCA with robust causal estimates of the effectiveness of our interventions in the real world. Importantly, they are the only way we can get such estimates before implementing a policy, unless, by chance, one or more firms have changed their behaviour and we are able to examine the outcomes in the data we collect. Field trials are therefore regarded as the one of the strongest research methods available to practitioners and academics.<sup>4</sup>

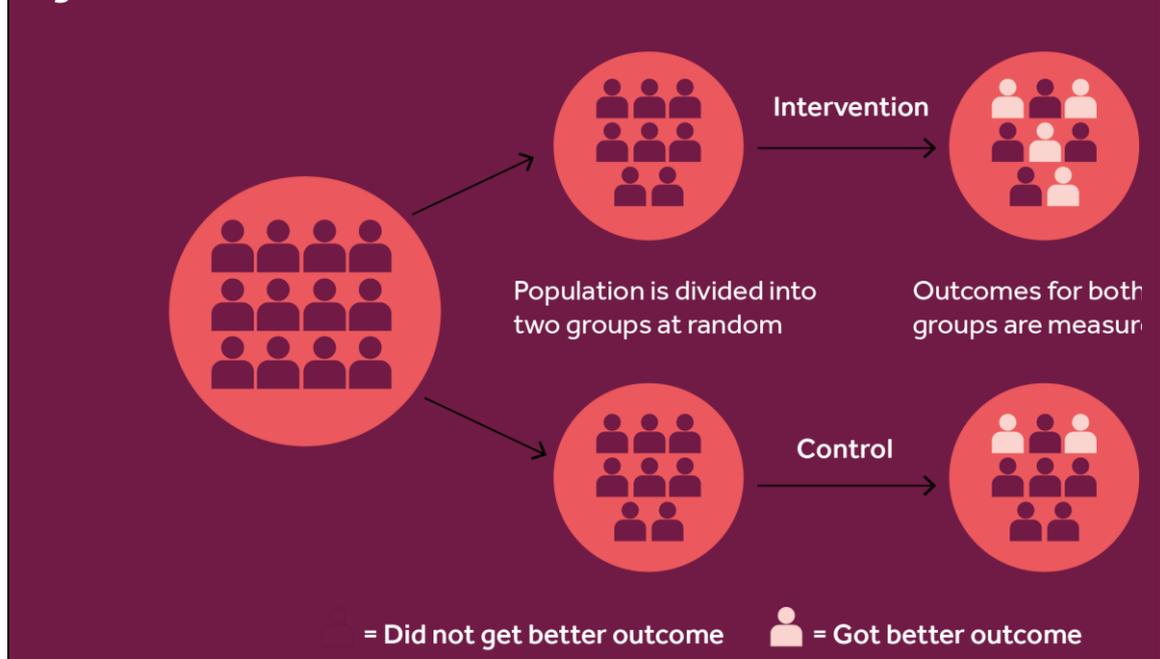
**Box 2.1: Field trials, causality and the external validity of research**

Field trials and causality

One of the most difficult challenges for a researcher is to estimate causality. Many research methods can show us correlations, such as as spending increasing when consumer confidence increases. These cannot usually tell us what caused what however. This is important in policy as we want to know the causal impact of our policies.

Randomised controlled trials (RCT), of which field trials are one type, help us to answer this. In an RCT, we divide the population to be tested randomly into two or more groups. One group receives service as usual (the control group) and the other receives the new intervention (the treatment group). Because the groups contain people who should be roughly similar to each other on average, we can then compare the difference in outcomes between the treatment and control group. This difference is a precise measure of causation, telling us exactly what effect our intervention has had.

**Figure 2: A basic randomised controlled trial**



<sup>2</sup> Discussion Paper on ex post impact evaluation framework - <https://www.fca.org.uk/publication/discussion/dp18-03.pdf>

<sup>3</sup> See for example Hussey and Hughes (2006).

<sup>4</sup> RCTs are the only policy evaluation type to achieve the highest level of 5 on the Maryland scale.

Field trials are RCTs that are carried out on people making real-life decisions who are usually unaware they are part of an experiment. This is important, because if they were aware, they might change their behaviour, for example to look better in the eyes of others (social-desirability bias). It also helps to avoid the problem of self-selection; people who volunteer to take part in experiments are often unrepresentative of the wider population.

#### Internal and external validity

Research is more useful to the FCA if it maximises internal and external validity.

**Internal validity** describes whether the effects seen in an experiment are truly caused by the intervention that is tested, or whether they are due to something else. As discussed above, RCTs tend to be internally valid as they standardise the conditions seen in the different intervention groups (eg using the same envelopes for all letters in a mail-out) and therefore control for other factors that might affect the outcome.

**External validity** describes how far the effects observed can be generalised to other contexts. Field trials tend to have high external validity because they observe the behaviour of people making real decisions in their natural environments. This helps avoid:

- consumers making hypothetical decisions (as in laboratory or online experiments)
- having to rationalise how they are thinking about particular decisions (as in focus groups or other survey tools) or

2.6 Field trials have been used to inform many of the world's most difficult policy challenges over the last few decades. They provide robust evidence on policies without wide-scale implementation and avoid many of the pitfalls of using historical data to evaluate policies. For example, the use of field trials was instrumental in identifying the effect of job seeker training in the 1970s and 1980s where self-selection into training programmes had previously prevented an unbiased evaluation.<sup>5</sup> This can be a powerful research tool to ensure that interventions are effective in the real world, before being rolled out to the market.

2.7 As well as helping policy makers understand what does work, field trials can also deter us from interventions that don't work, or that are even harmful. For example, in the medical domain for many years steroid injections were a standard and common treatment for head injuries. Only many years later, through the use of a randomised controlled trial (this time using a placebo), did it emerge that the use of steroids had actually been harming patients.<sup>6</sup> And more recently field trials have contributed to advances in understanding what is effective and what is not in education and health policy, among others.

2.8 Field trials are now commonplace in assessing changes to central government policy. The Behavioural Insights Team (part owned by the Cabinet Office) started 163 trials

<sup>5</sup> LaLonde, 1986; and Wandner, 2010.

<sup>6</sup> CRASH Trial Collaborators. (2005).

in the 12 months covered in their 2016/17 report.<sup>7</sup> Increasingly other regulators in the UK and overseas have also begun to apply these methods.<sup>8</sup>

- 2.9 In addition, developments in our understanding of consumer decision-making from fields such as behavioural economics and psychology show that rational models of consumer decision making can no longer be relied on to accurately predict consumer decisions in the real world.<sup>9</sup> This more nuanced understanding of human behaviour emphasises the importance of context, calls for policy makers to test ideas in the context in which they are going to be implemented. The combination of behavioural economic theory and experimentation has been crucial in the last five years in helping to 'consider how real consumers and firms make decisions in practice'.<sup>10</sup>
- 2.10 Since 2013 we have carried out field trials in areas such as redress<sup>11</sup>, general insurance<sup>12</sup>, cash savings<sup>13</sup>, pensions<sup>14</sup>, mortgages<sup>15</sup>, credit cards<sup>16</sup> and current accounts<sup>17</sup>. During this time the FCA has developed and refined its approach to field trials, how they can be integrated into regulatory policymaking and how they can be supplemented by other research methods. From these experiences we have a deeper understanding of when field trials may be appropriate, as well as learning how best to combine them with econometric analysis of historical data and online experiments.
- 2.11 Despite these clear benefits, field trials, like all research methods, require time and resource and this can present challenges.<sup>18</sup> We need to prioritise those policy decisions where field trials will be most helpful in allowing us to fulfil our duties and generate public value. In Occasional Paper 23 the authors discussed some of the logistical and methodological challenges faced when running field trials.<sup>19</sup> This paper looks beyond the simple logistical challenges, and sets out a range of factors that help us decide when we might use field trials to support our decision making. In addition, we also provide a simple step-by-step guide to a typical field trial, so that firms, consumer groups and other stakeholders understand how we conduct this type of research.
- 2.12 This paper focusses on field trials because they are a relatively new technique for developing regulatory policy and the FCA's use of them often requires some interaction or collaboration with the financial services industry. We hope being clear about when and how we use them will give firms a sense of our expectations and procedures, and the situations and circumstances when trials are more or less likely to form part of our evidence base.
- 2.13 The paper is laid out as follows: Chapter 3 covers some broad principles on the use of trials in the FCA, specifically around feasibility, importance and proportionality ("when to trial"). Chapter 4 provides a blueprint for successful deployment of field trials as part of regulatory policy development ("how to trial").

<sup>7</sup> <http://www.behaviouralinsights.co.uk/publications/the-behavioural-insights-team-update-report-2016-17/>

<sup>8</sup> For example the UK Competition Network's initiative to develop a knowledge base on demand side issues; CMA's recommendation to use field trials in both energy and banking markets and OECD's report into behavioural insights for policy. OECD estimate there are now 130 distinct teams working across the globe.

<sup>9</sup> See Erta, Hunt, Iscenko and Brambley (2013)

<sup>10</sup> FCA Mission, 2017

<sup>11</sup> Adams and Hunt (2013)

<sup>12</sup> Adams, Baker, Hunt, Kelly and Nava (2015)

<sup>13</sup> Adams, Hunt, Vale and Zaliauskas (2015) and Adams, Hunt, Palmer and Zaliauskas (2016)

<sup>14</sup> Adams and Ernstson (2018)

<sup>15</sup> Smart (2016)

<sup>16</sup> Adams, Guttman-Kenney, Hayes and Hunt (forthcoming)

<sup>17</sup> Adams, Grubb, Kelly, Nieboer and Osborne (2018)

<sup>18</sup> There is ongoing debate and critique of experimental methods. For example Deaton and Cartwright (2017).

<sup>19</sup> Smart (2017)

### 3 When are we likely to use field trials?

- 3.1 The FCA Mission notes, 'when we assess harm and regulatory benefit, we consider how real consumers and firms make decisions in practice, using a range of empirical techniques, from focus groups to randomised controlled trials.'
- 3.2 We published 44 consultation papers in 2017, many of which contain multiple policy proposals. While field trials are a powerful research tool to understand the effects of our interventions, it would not be an efficient use of our resources nor a proportionate burden to expect collaborating firms to carry, to run large scale, multi-firm field trials for every new policy proposal. We should prioritise our and the industry's resources to focus activity where it is most needed. How should the FCA trade-off the costs and benefits of field trials compared with those of other activities it could undertake? When do the benefits of using a field trial outweigh the cost?
- 3.3 This chapter sets out some high-level factors that we take into account when considering whether to use field trials in policy development. These are not hard-and-fast rules, and the type and extent of evidence needed for each policy is judged on a case-by-case basis. Sometimes field trials may not be suitable. There may be times when we need to implement policy quickly and choose not to use field trials even if they might be possible. Sometimes we would prefer to run a trial, but firms are unwilling or unable to test the interventions we want in the timescales we have set out. At other times, other analytical tools may be sufficient to allow us to reach well-informed policy decisions.
- 3.4 We ask three questions when deciding whether to run a field trial:
- **Is a field trial possible and appropriate?** This means:
    - i. that the causal chain should involve **changing consumer behaviour**
    - ii. that the field trial should **sufficiently recreate the policy intervention** on the relevant population
    - iii. that we should be able to observe **relevant outcomes** from the trial in a **timely fashion** (for example, balanced against ongoing harm)
  - **Is evidence from a field trial important for the policy decision?** We are more likely to consider a field trial an important part of the evidence base:
    - i. the less **existing evidence** tells us about the effectiveness of the intervention
    - ii. the less suitable **alternative research methods** are
    - iii. the higher the estimated costs of **our intervention**
  - **Is a field trial proportionate?** We will only conduct field trials when it means we are using our resources in an efficient and economical way. Burdens imposed by a field trial should be proportionate to the anticipated benefits.
- 3.5 We discuss each of these questions in turn, providing details on the factors which we consider important and giving examples where possible.

## Is a field trial possible and appropriate?

### (i) Does the causal chain involve changing consumer behaviour?

- 3.6 Field trials can be helpful and relevant when assessing policy, depending on the precise policy in question. In general, field trials are most useful to the FCA in testing policies intended to influence consumer behaviour.
- 3.7 Some policies are intended to change firm behaviour directly to protect consumers indirectly. For example, setting prudential capital requirements or setting the standards by which firms must look after client money requires firms to change but does not directly affect consumer behaviour. In most of these cases, firms must change their behaviour to match the rules, or risk facing legal action from the FCA. Given the lack of discretion, we would not expect much variation in how firms respond to these rules, so running field trials would not teach us very much.
- 3.8 If the causal chain for a proposed policy shows that it will only be effective if all the firms in a market were to implement it, a field trial is unlikely to test whether it works. For example, consider the market-wide standardisation of cost information for a particular product. The intention of this policy is to help consumers search and shop around across the market. Running a field trial with a subset of firms in the market does not sufficiently recreate the policy of market-wide standardisation, and running a field trial across the whole market could be prohibitively complex and expensive to coordinate. If so, an alternative analytical option could be to run an online hypothetical experiment, where the online environment recreates the search and shopping environment and tests different standardised cost measures within that.<sup>20</sup> Whether or not field trials or alternatives such as online experiments are appropriate can be informed by the causal chains we have in mind when designing our policy.

### (ii) Can a field trial recreate the proposed policy?

- 3.9 At a very practical level, we should only seek to test options that can be codified into rules, guidance or voluntary agreements. Options tested should also comply with current and proposed legislation, be aligned with wider FCA policy intentions, and not be at odds with the actions of other policy-making bodies.<sup>21</sup>
- 3.10 We are more likely to consider a field trial where the context of the trial can be designed so that it is similar to the actual market for which we plan to make rules.<sup>22</sup>
- 3.11 We look for cases where:
- The sample for the trial is similar to the **target population**. For example, testing retirement planning aids on people approaching retirement is more helpful than the same trial run with university students.
  - The **decision** in the trial is similar to the targeted decision we wish to influence through policy. For example, testing how we can encourage shopping around in the home insurance market is generally a similar process of decision making to shopping around in the general insurance market. But testing how we can encourage shopping around at renewal might be very different to encouraging good purchases at point of sale. The decisions are contextually different.

<sup>20</sup> The FCA have commissioned several of these types of studies including to inform the development of income drawdown cost metrics, pay day lending price comparison sites, annuity quote comparisons and fund cost disclosure.

<sup>21</sup> The FCA Sandbox is an exception to this, as we may choose to waive rules in order to conduct a Sandbox test using a field trial

<sup>22</sup> In technical terms the extent to which research findings generalise to other settings is known as external validity.

- The **environment, context and implementation** in the trial is similar to the market-wide environment and context. For example, a trial run during the Christmas holidays may not give a true representation of consumer behaviour all year round.

**(iii) Are relevant outcomes observable (in the timeframe available)?**

- 3.12 When designing field trials we always need to keep in mind the desired outcomes of our policy and how we would measure these as part of a field trial. When we assess our remedies or evaluate our rules, we rarely observe the final outcomes that we are most interested in, such as consumer welfare or social value directly. Instead, we have to infer these from a range of objective and subjective outcomes that we can observe. The same is true for field trials. When designing field trials we need to be sure that the outcomes we seek to measure are relevant and available.
- 3.13 The closer a field trial outcome is to our policy outcome of interest, the more likely we are to consider a field trial. A very direct example is the field trial conducted to investigate consumer responses to redress letters. Here the intended outcome is to maximise redress for consumers who were mis-sold financial services. A field trial was used to improve engagement with the letters and increase response rates.<sup>23</sup> A more nuanced example might be to encourage increased competition in an effort to drive down prices and push up quality. One way to do this might be to encourage consumers to shop around or switch their provider. In that case, an RCT using switching as an outcome measure would be relevant. Although this is an intermediate outcome towards lower prices and improved quality, it is a crucial part of the causal chain of the intervention.
- 3.14 There might be situations where outcomes at the individual level and over a relatively short time period do not capture all of the potential outcomes. For example, a simple message sent to consumers to save a certain proportion of their income might help individuals to save and we can measure this in an RCT. But an RCT would not be able to tell us the impact of the policy being rolled out nationally and the message becoming a widely accepted social norm. Or, in another situation, we might think that the effect of information on consumers gets stronger the more often they see it, so seeing the same disclosure year after year will mean an individual gets better at finding a good deal. In both cases, an RCT will underestimate the total impact of the policy. We need to take these into account when thinking about the outcomes we measure, the time period we measure them over, and the causal chains we have in mind for the policy.
- 3.15 Importantly, outcomes should also be measurable within the time available to make policy. If we only observe outcomes from a field trial several years after an intervention, then this might not be an appropriate area for a field trial. This depends on the extent of ongoing harm in the market and therefore the urgency of trying to reduce that harm. There may be some circumstances where we judge the ongoing harm to consumers to be sufficiently high that waiting for the results from a field trial would not be beneficial. In other words, the added benefits of certainty we may have in our policy through use of trials are outweighed by the losses consumers suffer while the policy is being tested.<sup>24</sup>
- 3.16 However we need to be careful when implementing policy quickly without testing. If harm is high, then quickly implementing an untested intervention may be worse than waiting for the results of further research. For example, an untested, but well-intentioned, disclosure may turn out to be ineffective but we might only find out

<sup>23</sup> See Adams and Hunt (2013) for more details.

<sup>24</sup> In extreme circumstances if we judge the potential harm to consumers to be very high then we may choose to use our early intervention powers

much later by conducting an evaluation after the fact (ex-post), during which the harm has persisted. It could even turn out to be harmful, and result in a decrease in of the desired behaviour. In the cash savings market study we used a field trial to test several different disclosures and found that they were broadly ineffective at encouraging switching between providers in the easy access cash savings market, despite our intuition and expectation that they would help.<sup>25</sup> This has led us to rethink the policy to ensure that we sufficiently address the identified harm. Without the trial we could have made disclosure rules and then assumed that they would fix the problem. This adds weight to the importance of our ex-post impact evaluation framework.

- 3.17 It is important to consider the extent to which there is sufficient sample to test interventions, and that the different groups receiving the control and treatment interventions can be effectively separated. If only a very small number of people would be affected or there is a possible crossover between those who receive the treatment and those that don't, then a field trial might not be the best approach.

### **Is evidence from a field trial important for the policy decision?**

#### **(i) What does the existing evidence tell us?**

- 3.18 Efficient use of our resources means we should only dedicate additional resource to understand the effects of policy where that will help to determine our course of action. If the existing evidence we have demonstrates that the costs of the policy are known to be small and the benefits are large, then conducting further research would not be a proportionate use of resources. This concerns the proportionality of additional research to understand the net effects of policy, rather than the proportionality of the policy itself.
- 3.19 When assessing the existing evidence it is important to consider how confident we are in the research that generated the evidence we rely on. As discussed in Box 2.1 in Chapter 2, the internal and external validity of that research for a particular policy proposal allows us to assess the strength of the available evidence.<sup>26</sup> The strength of the existing evidence will determine the extent to which running further research, including field trials, would provide important additional evidence for our decision making. In particular, and as discussed, existing evidence should be compared to the evidence that field trials can provide, namely precise estimates of the consumer response to an intervention. Theoretical arguments need to be carefully applied to make sure that we do not over- or under-weight existing evidence, and that the conclusions are appropriately applied to current policy.
- 3.20 We also need to consider prior field-trial research that is directly relevant. If we have direct, robust, field-trial evidence on a proposed policy, further trials to understand the impact of minor additions or changes to the policy are unlikely to provide substantive new evidence that is worth the additional effort. For example, we undertook extensive analysis to assess the effectiveness of disclosure in the cash savings market. We found that disclosure had limited effects on internal switching and no effects on external switching. We considered making small changes to the disclosure but did not think the additional information we would gain from running a field trial using those changes would be sufficient to justify the additional resources. In that case, in light of the existing field trial evidence we concluded that disclosure

---

<sup>25</sup> See Adams, Hunt, Palmer and Zaliauskas (2016) and CSMS Final Report

<sup>26</sup> An important complication for reviewing existing research is the distorting effect of publication bias. Publication bias refers to the phenomenon of academic journals being biased towards publishing research which display positive (and therefore interesting) tests of hypotheses. See Smart (2016) for more.

would not be sufficient to tackle the identified harm to inert customers and therefore began to explore alternative remedies.

### **(ii) What other types of research might provide better evidence?**

- 3.21 We undertake research to get evidence about how well a proposal is going to work. The value of any piece of research is the marginal increase in evidence over the next best form of research. So when we think about the value of field trials, we have to think about alternative options that could provide evidence that is good enough for that purpose. A number of other research options allow us to diagnose and evaluate remedies, and there are many ways the FCA can diagnose the causes of harm, develop potential remedies and evaluate their effectiveness. Occasional Paper 13 considers a wide range of research tools and their relative strengths and weaknesses.<sup>27</sup> Two of the best alternatives to field trials are natural experiments and online/laboratory experiments, which we discuss briefly here.
- 3.22 In some circumstances we may be able to use **natural experiments** from past policy changes to evaluate the market-wide impact of these changes. While this might not perfectly match the new policy options being considered, we can use theory to extrapolate carefully where possible and to define the boundaries of how to use the results. If firms have tried something sufficiently similar to the proposed policy, and it has been rolled out in stages to random parts of their customer base, then it may be possible to use that firm's granular data of consumer outcomes to estimate the causal impact of the change on those customers under those circumstances. We could then try to use existing knowledge to understand whether to expect a similar impact on the wider market population. Careful application of theory can help us understand the extent to which we can extrapolate findings from one firm to the wider market.
- 3.23 We can also use **laboratory/online** experiments to understand consumer decision making. In some circumstances, lab experiments may be preferable to field experiments. For example and as noted previously, if we want to understand the effect of standardisation on search and shopping behaviour, then a lab or online experiment is a suitable place to test different ways to standardise information. They may also be more helpful than field trials when trying to diagnose problems, as they allow a far greater degree of control and observation. However in some circumstances they may not be appropriate. For example if the diagnosis of harm has indicated that consumers do not normally pay attention or that they often fail to take action, then we might be concerned that evidence developed in a lab would find large positive effects which would be unrealistic in a real-world environment. A lab experiment would therefore not have sufficient external validity.<sup>28</sup>

### **(iii) Costs of implementing the policy**

- 3.24 If an intervention is highly costly to implement for firms then this is likely to improve the case for undertaking a field trial. This is because the costs of implementation should be considered as a reduction in public value, and for more expensive policies we should be more confident in the benefits of our interventions. Put another way, we can compare the cost of doing the trial against an estimate of the resources wasted if the policy were implemented but had no effect.<sup>29</sup>
- 3.25 Another cost of implementing policy is the potential for it to have unintended consequences. Such unintended consequences can be important and relevant in

<sup>27</sup> Iscenko, Andrews, Dambe, and Edmonds (2016)

<sup>28</sup> Lunn and Choisdealbha (2017) and Iscenko, Duke, Huck and Wallace (2014) discuss in more detail when lab experiments can be usefully deployed

<sup>29</sup> This type of reasoning has been adopted by others, for example see Haynes, Service, Goldacre, & Torgerson, (2012)

determining the overall benefits of a policy. If the intervention is quite radical, then there may be a risk that consumers respond in unexpected and unpredictable ways that are detrimental. Using field trials enables us to identify any unexpected behaviour before the intervention is rolled out to the entire market, avoiding those unintended consequences.

- 3.26 The proportionality of each field trial will be considered on its own merits, in light of the facts and circumstances of that particular field trial and the specific proposed policy being considered.
- 3.27 In discharging our general functions (which include activities relating to rule-making, such as field trials) under the Financial Services and Markets Act 2000, we must act, as far as reasonably possible, in a way that is compatible with our statutory objectives, and have regard to the regulatory principles in section 3B FSMA including:
- a) the need to use our resources in the most efficient and economical way
  - b) the principle that a burden or restriction which is imposed on a person should be proportionate to the benefits, in general terms, which are expected to result from the imposition of that burden or restriction.
- 3.28 These principles relate to the general question of the type and extent of the research we undertake. They provide us with a framework for thinking about the marginal costs of further research. The most efficient and economical use of resources relates closely to the feasibility and impact considerations mentioned previously.<sup>30</sup> Where there are alternative research methods that would enable us to assess the effectiveness of our proposals sufficiently, then it may not be an efficient or economical use of FCA resources to conduct additional research (of any kind) to further increase certainty of our assessment.
- 3.29 Given that firms usually agree to take part in field trials voluntarily, section 3B(b) is unlikely to be relevant, but we would take the principle into account in respect of expectations of firms and the exercise of other compulsory evidence-gathering powers we may use to complement field trials, such as information requests to firms.
- 3.30 As a public authority, the FCA must act rationally, follow proper procedures and satisfy itself of appropriate matters in exercising its functions. As field trials involve working with firms it may not be proportionate to expect firms to carry out field trials if the benefits from running a field trial (in terms of the value of additional information in forming our assessment of the policy) are unlikely to outweigh the costs.

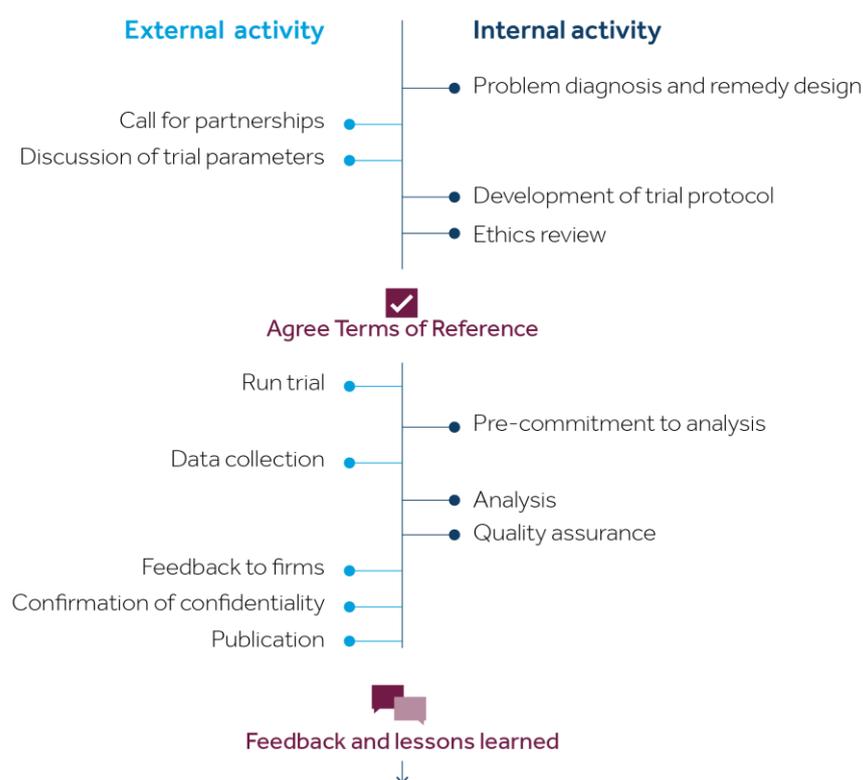
---

<sup>30</sup> Iscenko, Andrews, Danbe and Edmonds (2016) set out that “the resources that regulators spend on impact assessments must be proportionate to the problem at stake, the costs of additional analysis and the incremental improvement in certainty about the outcome of the intervention.”

## 4 How we run field trials

4.1 To help regulated firms understand the steps we go through when conducting this type of research, this chapter provides a high-level overview of the process. It will be of interest to the regulated firms we may collaborate with when conducting field experiments, but also to other regulators and practitioners who want to conduct field trials for policy development, especially in regulated markets. We have had to simplify the steps for the purposes of this paper, and individual field trials may have different or additional steps in the process. Some of the early steps concerning the diagnosis of the problem are also covered in Occasional Paper 13.<sup>31</sup>

**Figure 3: Typical timeline of internal and external processes**



### 1. Problem diagnosis

4.2 Even before considering which research tools to use to assess policy, it is important to ensure that policy proposals are designed to tackle the root causes of problems. This is true for all policy proposals, not only those that are to be tested in field experiments. OP13 has a full list of research tools that can help diagnose problems within markets, including:

- detailed data analysis (particularly at the individual level and using natural experiments where possible)
- qualitative and quantitative survey techniques
- online or lab experiments as well as existing theoretical and empirical literature.

<sup>31</sup> Iscenko, Andrews, Dambe and Edmonds (2016)

- 4.3 For example, the FCA used a field experiment in the early stages of the cash savings market study to understand whether consumer attention was a problem (Adams, Hunt, Vale & Zaliauskas, 2015).

## **2. Remedy design**

- 4.4 A clear diagnosis of the problem helps us to develop potential remedies. The process for developing and refining these into feasible policy options to test in a field trial (or through other research methods) will vary depending on the specific problem, the market and the available evidence. We also consider whether field trials are a feasible and appropriate research tool to investigate the efficacy of the proposed policy. To do so, we will look at the factors discussed in Chapter 3 of this paper.
- 4.5 If we decide a field trial is appropriate, then we will undertake preparatory research to design and refine interventions to make them as effective as possible in the field. This includes:
- looking at existing behaviour from micro-level datasets
  - qualitative and ethnographic research
  - quantitative surveys
  - lab experiments
  - design and user experience testing and
  - pilots or other forms of prototyping
- 4.6 As a field trial is a relatively complex process, upfront work ensures we get the most out of the trial. For example, by collecting data in advance, we can understand the existing consumer outcomes in the market and use this to inform our sample size calculations. Or we might use online testing and experiments to improve and refine our treatments to ensure consumers comprehend them. Finally, we might use an iterative policy design process to ensure that we design the interventions with end users in mind. In some cases, this research might lead us to change our policy plans and/or choose not to undertake a field trial.

## **3. Call for partnerships**

- 4.7 For most retail consumer products, the FCA does not have a direct link or communication channel with consumers. Our rules dictate how firms behave and we protect consumers by the boundaries and requirements we set in those rules. Given this indirect influence on consumers, if we decide that field trials would provide useful evidence we would seek partners in the appropriate industry. So far, we have worked with over a dozen different firms to deliver trials and this approach to collaboration has worked well.
- 4.8 When we seek collaboration from industry to conduct a field trial, we will issue a general call for partnerships to firms within the relevant market. This will either be published alongside any related market study or policy papers, as with the Credit Card Market Study, or we will approach firms in the particular market directly, as we did for our research into general insurance automatic renewal. Where we publish the request, we often follow up directly with firms in the market.

- 4.9 We cannot always conduct trials with all the firms who express interest. Working with additional firms requires additional resource and so we tend to work with the largest and most representative firms where possible. Some firms will not have a sufficiently large or representative customer base to allow for a rigorous test with acceptable sample sizes. If a firm has too few customers, this would prevent us proceeding with a field trial. If they serve only a particular subset of consumers, we could consider trialling with the firm, although we would have to be particularly careful about drawing conclusions from the results about the wider market.
- 4.10 Where we are unable to find appropriate partners to run field trials with us, this does not mean that we will abandon or delay the development of policy. In this case we use other research methods as far as possible to understand the impact of our policy options. We move ahead when we feel we have undertaken sufficient inquiry and have a good case for the policy.
- 4.11 Likewise, we will not work with firms to test policy options which are not close enough in design or intent to the intended policy. Firms can choose to run their own experiments (either lab or field) on approaches which they think might improve customer and market outcomes, and we welcome them sharing the results with us. However we do not think it is an efficient use of FCA resources to help run field trials with firms on options that are tangential to policy options being proposed or developed by the FCA.
- 4.12 So far, collaboration with industry has worked effectively and firms have reported a positive experience in working with us. We wish to maintain this positive working relationship and will continue to invite participation from industry on projects.

#### **4. Discussion of trial parameters**

- 4.13 Once we have initial expressions of interest from firms we will enter into a period of discussion to ensure that firms understand what the field trial requires in terms of systems changes, timescales and data sharing. We will draw up a detailed Terms of Reference for the field trial. It will be unique to each individual firm in the trial and will detail what we will expect from the firm and what the firm can expect from us. This will set out relevant details such as outcomes we are seeking, how we will randomise, what sample sizes are required, treatments to be tested, how to ensure external validity, data to be collected and data sharing arrangements.

##### **Outcomes**

- 4.14 We define the outcomes we seek through our treatment both in terms of overall changes in consumer behaviour as well as detailed changes in specific administrative data. As noted, it is important to ensure that the outcomes we measure as part of the trial are related to the overall objective of the policy and to our mission to generate public value. Where appropriate we define and rank multiple outcomes in terms of primary and secondary outcomes of interest.

##### **Randomisation**

- 4.15 A crucial part of field trials is to ensure that treatment and control conditions are delivered randomly to participants. In most projects we agree with the partner firm that the FCA research team will undertake the randomisation to ensure that it is done correctly. This requires participating firms to securely share non-personal details of customers with us. We then stratify (where appropriate) and randomise

before sending back to firms unique consumer identifiers with an indicator of which treatment each customer receives. In some cases we do allow firms to randomise based on agreed systems and processes. If this is the case then we document these thoroughly in the Trial Analysis Plan to ensure we have a record of the approach. In both cases the randomisation procedure and how it is implemented is checked thoroughly to ensure there are no mistakes which would undermine the validity of the trial.

### **Sample size**

- 4.16 We will agree with the firm upfront the number of individuals to be included in the trial and in the different intervention groups. This is so that we have sufficient information (power) to be confident in the results of the trial. We need to balance the need to have a sufficient sample size for our research with the impact on the partner firms, so we do not ask for more sample size than is strictly needed for our analysis. The sample size we require will depend on the number of different intervention groups we are testing, the expected outcomes in the control group, the power we wish to have (typically we target 80% power) and the size of the effect that we believe is relevant for the policy. Where we already have an estimate the costs of the intervention, then we can use this to determine the size of the effect that we would need to make the benefits of the policy outweigh the costs. In other cases we may have to use reasonable estimates of what would be an economically meaningful change in outcomes.<sup>32</sup>

### **Treatments**

- 4.17 We will share and discuss treatment designs with collaborating firms. This will include taking on feedback from designers and potentially conducting qualitative reviews with the firms customers to refine treatment design. We will also work with firms to help implement treatment designs within the brand guidelines of those firms where appropriate. As noted in Section 3, we will not compromise the policy intent of our designs.

### **Maximising external validity**

- 4.18 As discussed in Box 2.1 in Chapter 2, external validity is a general term meaning the extent to which the results of a study can be generalised to other people in other situations. This is particularly important for using field trials in the development of policy because we need to be relatively certain of the impact of our intervention once a policy is rolled out. External validity is influenced by a number of factors:

#### **External validity > The sample of customers in the experiment**

- 4.19 Picking a specific subset of consumers will give you detailed information on the responses of that subset but might not be able to be generalised to the wider population. This may be appropriate, and more cost efficient, if we are developing a policy aimed at a specific target group, however in most cases we are most interested in market-wide effects.
- 4.20 Therefore, when choosing the firms to work with for trials, and the samples of consumers within those firms, we are careful to choose samples which are as representative as possible of the wider market at which the policy is aimed. Or, where we deviate from this for a specific purpose, we are clear about what we are doing, why we believe the benefits outweigh the loss of representativeness, and

---

<sup>32</sup> See Bloom (2008) for more detail.

what it might mean for the target population. In addition, where we can, we make direct comparisons of the test sample with the market population.

### **External validity > Firm specific factors**

- 4.21 Specific characteristics of a firm we are working with on a trial can influence consumer behaviour, and therefore mean that the results from a trial with one firm can be hard to apply to other firms, or to the market as a whole. For example, consider a trial to promote switching between car insurance providers. Here, trialling the intervention with a provider that actively targets older consumers might not provide results which we can use for the wider market.
- 4.22 As we currently rely on voluntary participation from firms, it is possible that there might be a correlation between the type of intervention we're trialling and the willingness to participate. If so, then the effects found with a specific firm may not generalise to the wider population. For this reason, where possible we try to test identical or substantively similar interventions with multiple firms. Where this is not possible, trials may still prove beneficial, but care should be taken to identify and highlight in advance any idiosyncrasies of the firms that we partner with.

### **External validity > The environmental context**

- 4.23 Where possible, trials should be designed to limit the extent to which environmental factors might make results less valid. This means taking the reality of the wider environmental context into account. A trial in which customers are contacted during the Christmas or summer holidays is likely to underestimate how they respond during other times of the year. Therefore it might not represent the average or expected response if we were to apply the policy across a whole year. Researchers should think carefully about the environmental context and try to collect information that might account for it, or at least help to interpret the results in a different context.

It is not always possible to avoid environmental factors completely when undertaking research. For example the cash savings market study was conducted during an unusual period of historically low interest rates which depressed the monetary incentive to switch accounts. Arguably this affects our ability to extrapolate the results of our trial to the future when interest rates might rise. To overcome this we looked at whether those individuals with high balances, and therefore more to gain, were more likely to switch. We used this to help decide whether the intervention would be more effective when interest rates rose.

## **5. Ethics review**

- 4.24 We have developed an ethics policy which all researchers conducting human subject research need to read, understand and apply. This is used alongside an ethics process that researchers must complete. The process is designed to prevent unethical research being conducted, for example that might involve deception or misuse of participants' data, and to ensure that potentially high-risk cases are judged by external parties, rather than the research team. The policy and the ethical process form are included as Annexes 1 and 2 to this paper for reference and in case others wish to use them.
- 4.25 We are aware that the process outlined does not have the same level of scrutiny as the independent review boards or other ethical procedures used by academics. This

is partly because the FCA already has a legal obligation to protect consumers as well as to handle data correctly under the GDPR. We believe this obligation and our public service function more broadly means the risk of unethical research is lower. Moreover, we do ensure that there is senior staff oversight and sign-off of our field trials to ensure that the research team is acting ethically.

## **6. Agreement of Terms of Reference**

- 4.26 Once discussions have concluded between the FCA and the collaborating firm, we will finalise and agree a Terms of Reference. This agreement will be prepared in parallel to other work happening internally to the FCA including the Trial Analysis Plan. Once the Terms of Reference and the Trial Analysis Plan have been agreed, we will sign both documents off at an appropriate level of seniority. We encourage the firm to do the same. It is important that both the FCA and the firm are clear on what is required and when to ensure that expectations are set and agreed from the outset to avoid any subsequent misunderstanding.

## **7. Pre-registration**

- 4.27 Where possible we consider pre-registering or pre-committing the trial protocol or analysis plans before we begin collecting data for a trial. Pre-registration restricts researchers in what they can analyse, reducing the scope for selectively presenting the results. The purpose is to make clear our aims and objectives and the main outcomes we will be assessing. This reduces the risk of changing the outcomes we're interested in after the fact to fit any positive results we obtain. It is not always possible to pre-register our research with external bodies, especially where it relates to confidential policy development. However, where this is the case, we can use other internal and external mechanisms to commit ourselves to the consumer outcomes we will measure in the trial.

## **8. Run trial**

- 4.28 Once the terms of reference are agreed (and in parallel to internal FCA work to pre-commit to trial outcomes), the firm will implement the trial according to the details specified in the Terms of Reference. This can involve close coordination with the FCA team to ensure correct implementation and to check in regularly to ensure any problems are dealt with promptly.

## **9. Data collection process**

- 4.29 We collaborate with firms to agree a secure process for transferring data to the FCA to be analysed. This will normally include sharing dummy datasets so that both parties are certain what is being shared and so that FCA can prepare the relevant analysis. Real versions are checked to ensure the data is not corrupt and that the groups are balanced across the different intervention groups.
- 4.30 Security of proprietary and personal data is of the utmost importance to the FCA and to the partners we work with. We only collect data where absolutely necessary and minimise the extent of the personally identifiable information we hold. Data is encrypted and there are a number of different approaches we can use including

encrypted hard drives or Secure File Transfer Protocol (SFTP) to ensure this works for our partners.

## **10. Quality Assurance**

- 4.31 We frequently collaborate with academics on our projects to ensure our methods are state of the art. Where we do not work with them directly we might still seek the advice of senior academics in a particular field. Regardless of any academic involvement during the trial, when analysing and writing up the results we always ensure an independent academic reviews our work, to ensure we have analysed and interpreted the results correctly. Academics working on FCA projects are under strict non-disclosure and confidentiality agreements.
- 4.32 There are other ways in which FCA researchers monitor the quality of our work, such as checking that randomisation has been implemented properly and the interventions are being carried out as expected. Monitoring as things go into the field is important to ensure that the trial is proceeding as expected. Throughout the project, the project team face theoretical and analytical challenge from senior staff, as well as conducting cross-team reviews of analytical code.

## **11. Feedback to firms**

- 4.33 As we analyse a particular trial we discuss any emerging findings from a specific firm with each participating firm. This is a chance for the firm to understand the impact of the interventions they have tested and to question the FCA researchers. It also allows us to ensure that our understanding of the firm's data is correct and to hear if there are specific nuances in the firm that we need to take into account.

## **12. Confirmation of confidentiality**

- 4.34 Before publication we share the relevant, firm-specific, sections with each firm involved in the trial. This is to ensure that the collaborating firm is happy that the material does not disclose anything confidential about it. We ask for email confirmation of this. Results from one firm's trial are not shared with other firms also involved in the research. Nor do we share the broader conclusions or regulatory implications that we make as a result of the range of research in any one project.
- 4.35 Note that we never publish or disclose the names of collaborating firms unless specifically requested, and then only after publication. Firms should not volunteer their own names in advance of publication as this may restrict what information we can later include as part of the analysis of the trial, which reduces its public value.

## **13. Publication**

- 4.36 As noted in OP23, we strive to publish all of our research, whether experimental or not, and regardless of what we find. This fits with our FSMA Principles of openness and disclosure, and of transparency. And it is good research practice, ensuring we're adding to the body of available knowledge, even when we do not find interesting results or find results that do not support our expectations.

## **14. Feedback and lessons learned**

- 4.37 We always seek feedback from firms that we partner with to understand how the research process was for them and how we can improve our ways of working. This is important to ensure we learn from firms' experiences and for us to adapt accordingly.

## Annex 1: References

- Adams, P., & Hunt, S. (2013). Encouraging consumers to claim redress: evidence from a field trial. *Financial Conduct Authority*.
- Adams, P., Hunt, S., Vale, L. & Zaliauskas, R. (2015). Stimulating interest: Reminding savers to act when rates decrease. *Financial Conduct Authority*
- Adams, P., Baker, R., Hunt, S., Kelly, D. and Nava, A. (2015) Encouraging consumers to act at renewal. Evidence from field trials in the home and motor insurance markets. *Financial Conduct Authority*
- Adams, P., Hunt, S., Palmer, C., & Zaliauskas, R. (2016). Attention, search and switching: Evidence on mandated disclosure from the savings market. *Financial Conduct Authority*
- Adams, P. & Ernstson, E. (2018) Testing retirement communications: Waking up to get wise. *Financial Conduct Authority*
- Adams, P. Grubb, M.D., Kelly, D., Nieboer, J. & Osborne, M. (2018) Time to act: A field experiment on overdraft alerts. *Financial Conduct Authority*
- Bloom, H. S. (2008). The core analytics of randomized experiments for social research. *The Sage handbook of social research methods*, 115-133.
- CRASH Trial Collaborators. (2005). Final results of MRC CRASH, a randomised placebo-controlled trial of intravenous corticosteroid in adults with head injury—outcomes at 6 months. *The Lancet*, 365(9475), 1957-1959.
- Deaton, A., & Cartwright, N. (2017). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*.
- Duke, C., Huck, S., Iscenko, Z., & Wallace, B. (2014). How Does Selling Insurance As an Add-On Affect Consumer Decisions? A Practical Application of Behavioural Experiments in Financial Regulation.
- Erta, K., Hunt, S., Iscenko, Z., & Brambley, W. (2013). Applying Behavioural Economics at the Financial Conduct Authority. *Financial Conduct Authority*.
- Haynes, L., Service, O., Goldacre, B., & Torgerson, D. Test, learn, adapt: developing public policy with randomised controlled trials. *Cabinet Office*.
- Hussey, M. A., & Hughes, J. P. (2007). Design and analysis of stepped wedge cluster randomized trials. *Contemporary clinical trials*, 28(2), 182-191.
- Iscenko, Z., Andrew, P., Dambe, K., & Edmonds, P. (2016). Economics for effective regulation. *Financial Conduct Authority*.
- LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *The American Economic Review*, 604-620.
- Lee, W. (2018) Estimating the benefits of interventions that affect consumer behaviour. *Financial Conduct Authority*

Lunn, P.D., & Choisdealbha, Á.N. "The case for laboratory experiments in behavioural public policy." *Behavioural Public Policy* (2018): 1-19.

Smart, L. (2016). Full disclosure: experiments into giving information. *Financial Conduct Authority*.

Wandner, S. A. (Ed.). (2017). Lessons Learned from Public Workforce Program Experiments. WE Upjohn Institute.

## Annex 2: Trial Analysis Plan

**How to use this document:** This template is designed to help plan and record the key steps in a randomised controlled trial. You should fill it out once a decision has been made to run a trial. It ensures that you have a thorough audit trail which can be shared with others working on the project, giving them all relevant information in one document. The project lead should ensure all sections are completed and kept up to date, including links to relevant documents. This template also contains the ethics review template.

It is also intended to **help you design your trial and plan your analysis**. It ensures that you clarify your hypotheses, run your power calculations accordingly, record your statistical models and to distinguish between confirmatory analyses (ie testing a presupposed hypothesis) and exploratory analyses (ie suggested by the data) upfront.

**Title of trial:**

**Stage**

**Describe/provide  
rationale**

**Comments /  
attachments**

### 1. Trial description

*Introduction/background*

- What problem is it setting out to solve?

<b>Title of trial:</b>		
<b>Stage</b>	<b>Describe/provide rationale</b>	<b>Comments / attachments</b>
<p><i>Aims</i></p> <ul style="list-style-type: none"> <li>• Primary aim: Short statement on the desired outcomes of the research. i.e. what is to be accomplished. This should focus on the key hypothesis/es you are testing with the trial and what the causal chains are.</li> </ul> <p><i>Secondary aims</i></p> <ul style="list-style-type: none"> <li>• Any secondary/ exploratory analysis you intend to undertake?</li> </ul>		
<b>2. Sampling</b>		
<p><i>Population/sampling frame</i></p> <ul style="list-style-type: none"> <li>• Demographic information for population of interest</li> <li>• Rationale for choosing population</li> </ul>		<i>Link to data / relevant planning documentation</i>
<p><i>Sample</i></p> <ul style="list-style-type: none"> <li>• Which groups will be included and excluded, and why?</li> <li>• How will the sample be recruited?</li> </ul>		

<b>Title of trial:</b>		
<b>Stage</b>	<b>Describe/provide rationale</b>	<b>Comments / attachments</b>
<p><i>Power calculations</i></p> <ul style="list-style-type: none"> <li>• Which outcome is being powered for?</li> <li>• What effect size is the study powered for and what is this based on?</li> <li>• Sample size?</li> <li>• What is the estimated power?</li> <li>• Equal/unequal allocation ratios</li> <li>• Potential for attrition and plans for addressing it? What checks will be done and what adjustments will be made for non-random attrition? Will sample size be increased for potential attrition? Can you adjust the experimental design to reduce the risk of attrition?</li> <li>• Do we expect heterogeneous effects? Why and in which sub-groups? Do we want to power the study for these sub-groups?</li> </ul>		<p><i>Link to calculations</i></p>
<b>3. Data sources</b>		
<p><i>Data collection methods and procedures</i></p> <ul style="list-style-type: none"> <li>• What data will be collected?</li> <li>• How will you receive the data and when will you receive interim and final batches?</li> <li>• Has the partner confirmed they are able to provide the data in the agreed timetable?</li> <li>• If planning to run a survey, then include relevant plans (including the list of questions, how the survey will be delivered and how the sample will be selected).</li> </ul>		<p><i>Provide full data tables that have been agreed with the partner, along with a data dictionary with a description of fields.</i></p>

Title of trial:		
Stage	Describe/provide rationale	Comments / attachments
<b>4. Trial design</b>		

<b>Title of trial:</b>		
<b>Stage</b>	<b>Describe/provide rationale</b>	<b>Comments / attachments</b>
<p><i>Randomisation</i></p> <ul style="list-style-type: none"> <li>• What is the unit of randomisation?</li> <li>• Will the randomisation be clustered? If so, how many clusters?</li> <li>• Who will carry out the randomisation and how? If carried out by the partner, state the process and how it will be quality assured.</li> <li>• Will the randomisation be stratified? On which variables?</li> </ul>		

<b>Title of trial:</b>		
<b>Stage</b>	<b>Describe/provide rationale</b>	<b>Comments / attachments</b>
<p><i>Interventions/control</i></p> <ul style="list-style-type: none"> <li>Describe the treatment(s) and how they will be delivered</li> </ul>		<p><i>Link to copies of signed-off versions</i></p>

<b>Title of trial:</b>		
<b>Stage</b>	<b>Describe/provide rationale</b>	<b>Comments / attachments</b>
<i>Timing of the trial</i> <ul style="list-style-type: none"> <li>• Start/end dates</li> </ul>		
<b>5. Ethics</b>		

<b>Title of trial:</b>		
<b>Stage</b>	<b>Describe/provide rationale</b>	<b>Comments / attachments</b>
<p><i>Protecting rights, dignity and welfare of individual</i></p> <ul style="list-style-type: none"> <li>• What are the potential risks (if any) to the rights, dignity and welfare of any individuals involved in the research?</li> <li>• How will you ensure that you mitigate any risks? Eg Will you seek informed consent? If not, why?</li> </ul>		
<p><i>Balance between individual and society</i></p> <ul style="list-style-type: none"> <li>• What are the benefits and risks to individuals taking part in the research compared with society/science?</li> <li>• How will you balance these? Eg. What is the counterfactual if the research is not carried out?</li> <li>• Is the trial a pilot?</li> <li>• Are the groups involved in the research similar in composition to the groups with the potential to benefit from the output of the research? If not, why not?]</li> </ul>		
<p><i>Balance between risks and benefits</i></p> <ul style="list-style-type: none"> <li>• How will you balance the potential benefits against any potential risks? Eg What existing evidence is there for the treatments?</li> </ul>		

<b>Title of trial:</b>		
<b>Stage</b>	<b>Describe/provide rationale</b>	<b>Comments / attachments</b>
<p><i>Responsible and robust research</i></p> <ul style="list-style-type: none"> <li>How will you ensure that your research is good quality, scientifically rigorous and transparent? Eg Will the research be reviewed by an academic?</li> </ul>		
<p><i>Additional risk factors</i></p> <p>Does your research involve:</p> <ul style="list-style-type: none"> <li>Vulnerable individuals?</li> <li>Financial inducements?</li> <li>Sensitive data?</li> <li>Secondary use of data?</li> <li>Psychological manipulation?</li> </ul> <p>If yes, please give details and explain the rationale and any risk mitigation strategies.</p>		
<b>6. Analytical strategy:</b>		
<p><i>Which variables will you include in balance tests?</i></p>		<i>Link to code / outputs</i>
<p><i>Analysis plans (to be set out before receiving the data and ideally at the design stage):</i></p> <p><i>Specify:</i></p>		

Title of trial:		
Stage	Describe/provide rationale	Comments / attachments
<ul style="list-style-type: none"> <li>• <i>Hypothesis/es being tested</i></li> <li>• <i>Primary (or confirmatory) analysis</i>: this follows from the pre-defined hypotheses that the trial was designed to answer.<sup>33</sup> State exactly what the trial will test, including specification of outcomes.</li> <li>• <i>Secondary (or exploratory) analyses</i>: any other tests you will undertake to understand variation in findings, but which you are not testing against a pre-defined hypothesis.</li> </ul>		
<p><i>Analysis plans continued</i></p> <ul style="list-style-type: none"> <li>• <i>Approach to statistical analysis</i>: Explanation of model approach, including treatment effect equation for your primary and secondary analyses</li> <li>• <i>Heterogeneous treatment effects (subgroups)</i>: <i>Linked to the choices about sample, which sub-groups are you interested in? Will you test these as part of your primary analyses? Why?</i></li> <li>• Whether the statistical tests are <i>one or two-sided?</i></li> <li>• <i>Treatment compliance</i>: will an intention to treat or treatment on the treated be estimated? If the latter, how will potential bias be accounted for?</li> <li>• How will testing of <i>multiple hypotheses</i> will be accounted for (if applicable)?</li> </ul>		<i>Link to code for analysis</i>
<i>Checking and cleaning the data</i>		<i>Code for data</i>

<sup>33</sup> It is important to separate confirmatory and explanatory analysis and state them before conducting analysis as your confirmatory analysis will inform your trial design. Doing so also guards against 'p-hacking'. For further reading: <http://ordination.okstate.edu/motivate.htm> and [http://www.koestler-parapsychology.psy.ed.ac.uk/Documents/explore\\_confirm.pdf](http://www.koestler-parapsychology.psy.ed.ac.uk/Documents/explore_confirm.pdf)

<b>Title of trial:</b>		
<b>Stage</b>	<b>Describe/provide rationale</b>	<b>Comments / attachments</b>
<ul style="list-style-type: none"> <li>• Data cleaning. How have outliers, unusual data and gaps been dealt with?</li> <li>• How have any new variables been constructed?</li> </ul>		<i>cleaning / exploration / construction of new variables</i>
<p><i>Code and analysis review:</i></p> <ul style="list-style-type: none"> <li>• Code QA: who will do it and at what stage?</li> </ul>		<i>Comments / amendments from reviewers</i>
<b>7. Use of trial results</b>		
<p><i>Application of trial results</i></p> <p>(optional)</p> <ul style="list-style-type: none"> <li>• Provide some commentary on the degree of external validity for the trial. Could it be used to help answer other related policy questions?</li> <li>• Would it be possible to replicate the trial in other contexts, based on the information provided here?</li> </ul>		

# Annex 3: FCA human research subjects ethics policy and procedure

## **1. Introduction**

This paper describes the ethics policy and procedure for research involving human subjects carried out at the Financial Conduct Authority.

This policy has been created and designed by the Behavioural Economics and Data Science Unit in consultation with Matt Levy at the London School of Economics, John Gathergood of the University of Nottingham.

The policy and procedure is informed by the best practice of thirteen leading universities and other organisations who carry out experimental research, including Harvard, University of Cambridge and Abdul Latif Jameel Poverty Action Lab. It is also informed by a number of academic articles and other relevant sources.

Chapter 2 outlines the FCA's ethics policy for carrying out human subjects research and Chapter 3 describes the procedure that researchers and reviewers should follow. All FCA researchers should read and familiarise themselves with the policy and procedure in this document. From commencement, relevant staff will be required to sign a declaration that they have read and understand the policy and procedure before planning or carrying out experimental research.

The policy and procedure has been tested on a number of live projects and is reviewed at regular intervals to make sure that it is fit for purpose.

## **2. Ethics policy**

### **2.1 Rationale for policy**

As a regulator carrying out research involving human subjects, it is important that we acknowledge ethical considerations in our work and attempt to mitigate any ethical risks which may arise. There are two fundamental reasons for this.

- **As a researcher**, we investigate human behaviour in the pursuit of knowledge, which is generally held to be a good thing in most societies. However, various ethical considerations may arise when becoming involved in others' decision making or behaviour with an objective other than to help the person in question and with the possibility that the person could be worse off following intervention. The conflict between responsibility to each individual and responsibility to society has long been recognised in medical research and numerous rules, guidelines and authorisation procedures have been established to control the parameters in which research can take place. This is now also the case in most university faculties and many other organisations which carry out research.
- **As a regulator**, we hold a position of responsibility with regard to the firms we regulate and other direct or indirect beneficiaries of our work. It is important that we formally recognise that our responsibility and remit brings with it the duty to protect those who may be involved in and affected by any research we carry out. While a general duty to protect consumers is written into our operational objectives, it is also important to consider the specific ethical implications of human subject research, as this does not specifically feature in any legislation or objectives relating to the FCA, and because research may necessarily involve potential benefits to some parties and potential negative outcomes for others.

Given the increasing popularity and ease of carrying out RCTs (particularly A/B testing), having an agreed framework and process for considering the ethical implications of our research work will allow us to make sound judgements on projects.

### **2.2 Scope**

The following document takes in all research projects which include human subjects.

The scope of the proposed review process may exclude a number of situations, which are outlined in the draft flowchart in Section 2.2. For example, any project which has already undergone a robust ethics review by another institution should not be subject to this process. Similarly, research involving observing historical, anonymised data is likely to have no ethical implications relevant to our review (ethical considerations with regard to privacy of information or data are legislated for under the Data Protection Act 1998) and so will not be subject to review.

### **2.3 Principles**

Many organisations set out overarching principles on research ethics which detail the main objectives for carrying out research and the theoretical underpinnings of the

conflicts that an ethics process seeks to control. These often relate explicitly to the code of conduct for these institutions.

The FCA already has an operational objective to "secure an appropriate degree of protection for consumers". Indeed, it is often this, among the other objectives, which forms the basis of the rationale to test. This objective remains an important and overarching principle as an ethical guide, but more detail is required for the specific circumstances of human subjects research.

Based on a review of similar organisations and the literature regarding ethical principles (see references) our four proposed principles are:

#### **1. Protecting the rights, dignity and welfare of the individual**

Researchers must endeavour to protect individuals from harm resulting from their participation in research. In cases where research includes vulnerable individuals (for example by virtue of their financial position or health situation) or sensitive data (for example data pertaining to a specific individual's financial position), additional measures to protect individuals should be considered where appropriate. However, research should be inclusive, and should not target or exclude particular groups without good reason. In general, the group which takes part in research should be similar in composition to those who would be likely to benefit from the gains of research as far as possible.

#### **2. Balance between individual and society**

Researchers must endeavour to balance the potential advantages to future beneficiaries of the results of research with the potential advantages and disadvantages to current participants.

Researchers should endeavour to take into account the counterfactual to carrying out research. In cases where policies will otherwise be implemented on the basis of no research or inadequate evidence, a trial may be the ethically preferable option.

#### **3. Balance between risks and benefits**

Researchers should endeavour to ensure that potential risks (such as financial losses or emotional distress) arising from participation in a trial or experiment are proportionate to potential gains.

#### **4. Responsible and robust research**

Researchers should endeavour to produce good quality, scientifically rigorous research, which is transparent and adheres to the FCA's Records Management policy, such that methods can be replicated.

Researchers should endeavour to remain aware of any conflicts of interest (for example, between clients' desire for or expectations of particular results and scientific rigour) and manage them appropriately. Researchers should endeavour to publish the results.

#### 2.4 Guidelines for assessing ethicality

Each case should be judged on its own merits, but the following factors may be taken into account when judging the ethicality of particular research projects:

- Whether participants give informed consent

- Existing evidence on the treatments to be tested
- The proposed sample size and whether the research is a pilot ahead of larger scale implementation
- Whether the research includes additional risk factors

These factors are discussed in more detail below.

#### **2.4.1 Informed consent**

The FCA carries out a number of randomised controlled trials, which as is customary, do not seek informed consent from participants. While allowing individuals to make an active choice with regard to participation in research mitigates the risk of subjects being involved in research against their wishes, it may also mean that the research may suffer from selection bias, as well as attrition of the sample. Awareness of being part of a trial is also highly likely to affect individuals' behaviour, which may distort results. In many cases, requiring informed consent for a field trial is logistically unfeasible, particularly in cases where data at the individual level is unavailable.

For these reasons, informed consent will not usually be required for field trials (although it is likely to be appropriate for other types of randomised controlled trials, such as laboratory experiments). However, there is a level of experimentation which most reasonably informed individuals are comfortable with in everyday life (for example, A/B testing on shopping websites). With this in mind, researchers should be mindful of Principle 3 (balancing risks and benefits) when designing research and should consider that the potential risks arising from participation in the trial would be acceptable to most reasonably informed individuals, given the potential benefits.

#### **2.4.2 Existing evidence**

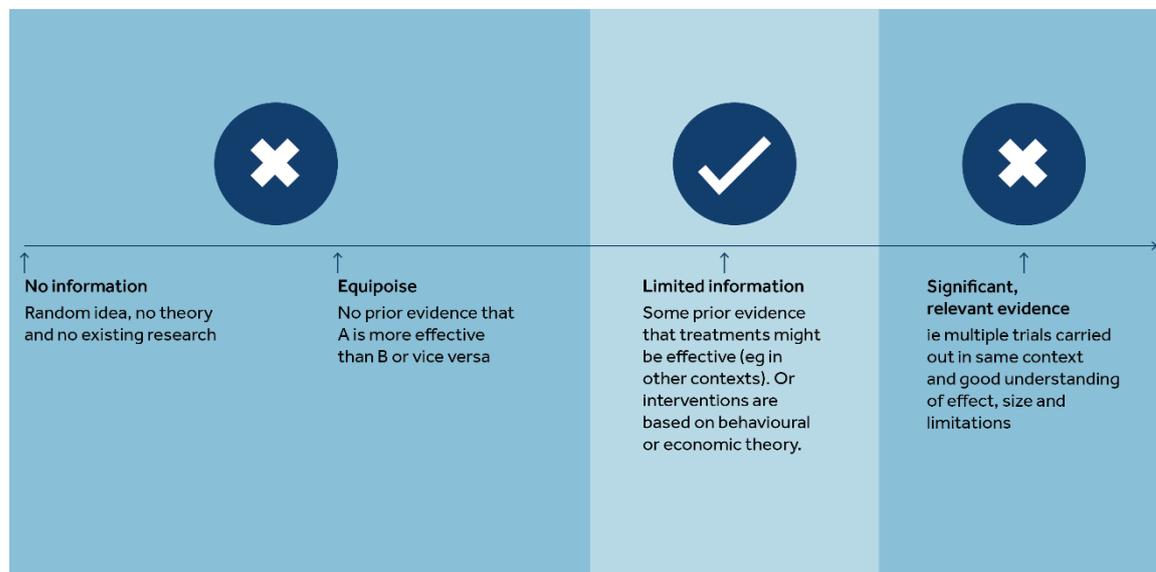
Medical research often relies on the principle of equipoise, that there is genuine uncertainty over whether a treatment will be beneficial (Edwards, Lilford, Braunholtz, Jackson, Hewison, & Thornton, 1998). Observing equipoise means that trials should only be conducted in cases where there is no prior evidence that the treatment is more effective than the control. While there is some debate over the threshold at which indications that a treatment is more effective becomes "evidence", this addresses to some extent medical researchers' conflict between their responsibilities to their current patient and their responsibilities to the health of future patients. Since this conflict is less acute at the FCA (the FCA's responsibilities are towards all consumers rather than individual consumers), it is likely that striving for equipoise may not be suitable in the context of trials conducted by the FCA.

In fact, we would usually regard ethical practice to involve trialling treatments based on theory and/or evidence of efficacy in other situations (see figure below). This prevents the trialling of treatments for which the potential risks are completely unknown (and therefore could be significant), as well as avoiding wasting time and the goodwill of participants.

There must, of course, be a balance. We do not advocate trialling interventions for which there is already strong evidence, although we would consider interventions with evidence of efficacy in other areas when applied to new contexts. We would also seek to trial when the intervention in question would otherwise be rolled out to everyone in any case and there is no prior evidence of its efficacy. Finally, we would

be unlikely to approve trials which test an intervention for which there is negative evidence, unless the intervention was current practice or agreed future practice and was to be tested as the control group against a new idea. However, we may consider trials which involve removing services rather than adding them if there is sufficient evidence that this would be unlikely to be harmful.

**Figure 4: Spectrum of existing evidence**



### 2.4.3 Sample size and pilots

Researchers should choose a sample size which is sufficient to detect effects of the treatments (judgement and knowledge of similar experiments should be applied when deciding which effect size is most appropriate). However, researchers should avoid selecting an unnecessarily large sample when more overall benefit could be gained from testing a smaller sample and then applying the most effective intervention to the rest of the population.

In cases where field trials are to be carried out ahead of large scale implementation of e.g. a new policy, carrying out a pilot on a smaller sample of the whole population could help to mitigate the level of detriment potentially arising from any risks of participating in the trial. Where possible, pilots should be carried out and should be discontinued when there is sufficient evidence in favour of and against different interventions. The rest of the eligible population should then receive the best performing treatment(s).

Pilots may be seen as distinct from general testing to inform policymaking, where a "control" does not currently exist, or where there is no clear plan to introduce (a) specific new intervention(s).

### 2.4.4 Additional risk factors

Many university and institutional ethics policies make special provision for groups and factors which may present additional risks. These may include: vulnerable individuals, intrusive methods, sensitive topics or use of sensitive data and others. Factors relevant to the FCA are discussed below.

#### a) Vulnerable individuals

It is likely that field trials carried out by the FCA may include vulnerable consumers, since many individuals' vulnerable status will not be observed. In some cases, it is possible that a large proportion of the population in a research project may be classed as "vulnerable", for example, pensioners or those taking out high cost, short term loans. If this is likely to be the case, researchers should explicitly state how they intend to ensure that vulnerable consumers are not taking on more risk than non-vulnerable individuals by participating in the trial, with regard to the specific circumstances of the research. In cases where vulnerable consumers take on the same amount of risk as others but the impact of this risk may be higher, researchers should consider whether the extent of the impact and whether the person taking on the risk would be likely to agree to take part in the trial if asked.

In cases where informed consent is desirable and achievable, researchers should take into account the individuals' capacity to consent. In general, it is not likely that research at the FCA should require participation from those deemed unable to consent (e.g. children).

#### *b) Financial inducements*

Some university and institutional ethics procedures have special provision for financial inducements, to manage cases in which an individual participates in or continues with a trial against their will because they need the money. Many market research companies do not provide financial inducements at all. However, a large proportion of Psychology and Economics research relies on (usually small) financial rewards to incentivise effort in a task, whether linked to performance or not.

It may be appropriate for the FCA to provide financial inducements to participants, particularly in laboratory trials which aim to mimic real life stakes in an experimental setting. If financial inducements are to be used, this should be made clear in the trial protocol and attention should be given to any adverse effects which could result from them. Researchers should mitigate the risk of adverse effects by ensuring that any compensation is commensurate with the participant's contribution and is paid afterwards in order to avoid participants feeling obliged to continue with the experiment if they would like to drop out, because they have already received payment. However, using financial inducements should not preclude the trial from being approved, if properly managed.

#### *c) Use of sensitive data*

The sensitive data most likely to be used in FCA research concerns individuals' financial data (e.g. spending and saving, balance, investments). For field trials, customer data may be used by both the firm and the FCA for the purposes of FCA-instigated research.

In many cases, the customer may have already agreed for their data to be used for the purposes of research through their contract with the firm. In other cases, the legal requirements on data protection may be covered by the Data Protection Act 1998 and/or the Financial Services and Markets Act 2000.

Sensitive data should be anonymised at the first possible point and ideally before it reaches the FCA. Researchers should ensure that they adhere both to the Data Protection Act and Financial Services and Market Act 2000. In addition, they should consider whether the use of sensitive data for research purposes complies with the four principles above and make a record of this in the trial protocol.

*d) Secondary use of data*

There may be cases where researchers may wish to reuse data collected for a specific purpose for other research purposes. Again, any reuse should comply with the Data Protection Act and Financial Services and Markets Act 2000. In addition, researchers should consider whether the reuse of data for others purposes complies with the principles above and make a record of this in the trial protocol.

*e) Psychological manipulation (deception, emotional manipulation)*

In some laboratory experiments, it may be necessary to obscure the true purpose of the experiment in order to observe real behaviour or unconscious processes. For example, explaining that an experiment will be testing participants' understanding and memory of documents may encourage participants to put in more effort to remember and interpret those documents than they would in real life. Telling participants that that the experiment is about something else and presenting the documents as a peripheral activity may elicit more accurate results.

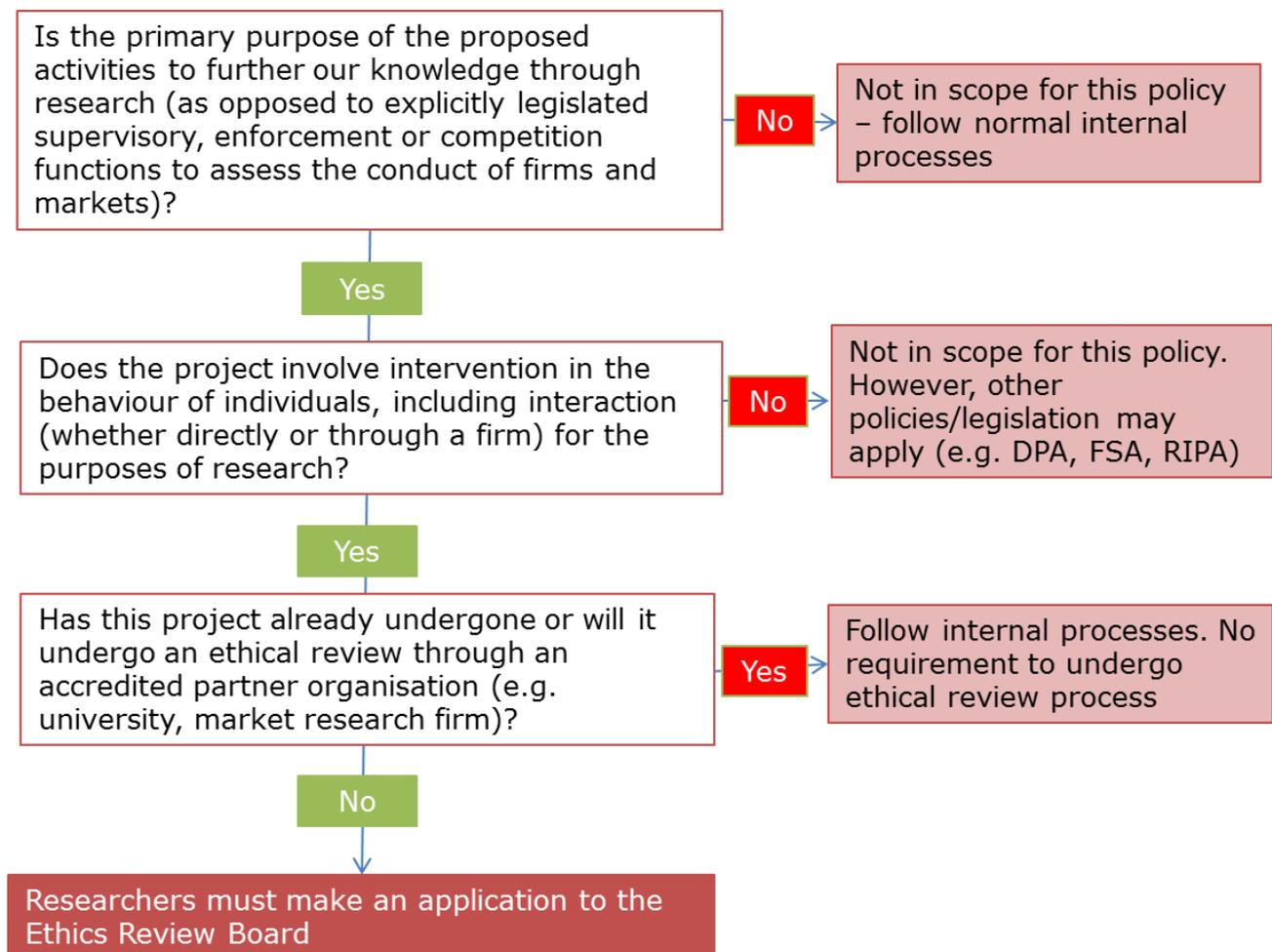
In experiments involving psychological manipulation including deception, researchers should acquire informed consent to participate in the experiment and should give reasons for the use of psychological manipulation in the trial protocol (though not necessarily to the participants) as well as considering any associated risks.

### 3. Ethics procedure

The proposed process is based on our review of the literature on research ethics and the procedures of other institutions. It takes into account the need to ensure that ethics is properly considered and risks are appropriately mitigated, while balancing this against the need for a process which is workable and flexible. The procedure was chosen from a number of alternative options as the one which best balanced the need for flexibility to allow important and insightful research with active consideration and mitigation of risks to participants.

#### 3.2.1 Scope

The first stage of the ethics procedure involves the researcher determining whether their proposal is in scope of the formal review procedure, using the flowchart below. A large proportion of research will be outside the scope of the formal procedure and will require no further work.



#### 3.2.2 Application

The application to be submitted will consist of the trial protocol (which includes questions on ethics; see Annex 1) and any supporting documents thought to be relevant by the researchers. This will usually include examples of the stimuli/treatments and details of any collaborators. Reviewers may ask researchers to provide more information as necessary and appropriate.

#### 3.2.3 Review

### Standard cases

Review will be carried out by an independent FCA manager or senior manager who is not managing the project. In some cases, the reviewer may decide to involve another manager to review the application.

### High risk or contentious cases

In cases that are judged by the reviewer/s to be high risk or in need of further scrutiny, the application will be escalated to two members of an agreed pool of internal and external reviewers (see below). At least one reviewer should be external to the FCA. If the two members fail to agree or come to a conclusion, a further member will be brought in.

If, following review and appropriate modifications, the two pool members are not able to reach agreement that the proposal meets ethical standards according to the policy, the proposal will be escalated to the director of MIDA. Taking into account the comments and submissions of the initial reviewers and the pool members and any other relevant material as appropriate, the director will then make the final decision.

The pool consists of:

- Internal: GCD representative, MIDA representative, Communications representative, CED representative, Policy representative
- External: Academic, Consumer Panel member

Project managers are responsible for finding two or more reviewers from the Board and should approach reviewers in alphabetical order, starting off from where the last project finished on the list.

Pool members agree to review projects within 5 working days of receipt of the application, unless the project is particularly unusual or complex or requires more than two reviewers, in which case reviewers should inform researchers of any expected delay.

If external to the FCA, pool members should sign an NDA with the FCA if necessary, which allows information to be shared.

Pool membership should be reconsidered annually, to ensure that the composition remains appropriate.

### **3.2.3 Follow up**

All decisions should be recorded in an ethics log.

All applications and their decisions should be taken to a quarterly review meeting with the relevant director for follow-up discussion. This is to provide a check on consistency and to make sure the procedure is working well and is fit for purpose, as well as to ensure that the relevant director has an overview of all cases, including standard ones.

## **References**

### *Academic papers*

Edwards, SJL; Lilford, RJ; Braunholtz, DA; Jackson, JC; Hewison, J; & Thornton, J. (1998) Ethical issues in the design and conduct of randomised controlled trials. *Health Technology Assessment Vol 2, No 15.*

Haynes, L., Service, O., Goldacre, B. & Torgerson, D. Test, learn, adapt: developing public policy with randomised controlled trials. *Cabinet Office*

Hutchinson, D. & Styles, B. A guide to running randomised controlled trials for educational researchers. *National Foundation for Educational Research*

Meyer, Michelle N. (2015) Two cheers for corporate experimentation: the A/B illusion and the virtues of data-driven innovation. *13 Colo. Tech L.J.* 273

Miller, Franklin G. & Kaptchuk, Ted J. (2004) Sham procedures and the ethics of clinical trials. *Journal of the Royal Society of Medicine* 97 (12)

Oakley, A., Strange, V., Toroyan, T., Wiggins, M., Roberts, I. & Stephenson, J. (2003) Using random allocation to evaluate social interventions: three recent UK examples. *American Academy of Political and Social Science.*

Ziliak, S.T. & Teather-Posadas, E.R. (Forthcoming.) The unprincipled randomisation principle in economics and medicine. *Oxford University Press Handbook on Professional Economic Ethics* (New York: Oxford University Press, 2014), Eds. George DeMartino and Deirdre N. McCloskey

### *University procedures*

[Harvard University](#)

[London School of Economics](#)

[Massachusetts Institute of Technology](#)

[Sheffield University](#)

[Teesside University](#)

[University of Edinburgh](#)

[University of Cambridge](#)

[University College London](#)

[Yale University](#)

### *Other organisations*

[Abdul Latif Jameel Poverty Action Lab](#)

[British Psychological Society](#)

[British Sociological Association](#)

[Social Research Association ethics guidelines](#)

*Other references*

Anon. [The ethics of experimentation](#)

Harford, T. [The random risks of trials.](#)

Metcalf, Jacob. [Getting the formula right: Social trust, A/B testing and research ethics](#)

Neal, David. [The ethical review process for clinical trials in the European Union.](#) *NHS National Research Ethics Service.*

Zoido-Oses, Paula. [The problem with nudge policies is that they threaten our freedom to choose to act well](#)

All our publications are available to download from [www.fca.org.uk](http://www.fca.org.uk). If you would like to receive this paper in an alternative format, please call 020 7066 9644 or email: [publications\\_graphics@fca.org.uk](mailto:publications_graphics@fca.org.uk) or write to: Editorial and Digital team, Financial Conduct Authority, 25 The North Colonnade, Canary Wharf, London E14 5HS