Ex post impact evaluation framework

December 2018

Contents

How to navigate this document onscreen

returns you to the contents list

1 Foreword

1.1 Our Mission[1] makes it clear that we want to be more transparent and accountable to the public, and that evaluation is a critical part of getting our interventions right.

1.2 Finding out whether our past interventions have been effective and what impact they have had is an important part of developing a strong evidence base to guide our decisions – both our decisions over which issues to prioritise, and our decisions over how best to intervene to tackle harm. It is also widely recognised – including by the National Audit Office – as best practice for regulators.

1.3 We consulted on a proposed framework for ex post impact evaluation earlier in the year. This document reflects the feedback we received. It outlines how we approach measuring the impact of our interventions. It explains why we do ex post (after the event) impact evaluation, how we choose specific interventions to study, and how we ensure that our evaluations are robust, impartial, and therefore credible.

1.4 Studying the impact of our interventions is not new for us – and this document contains examples of impact evaluation work by the FCA and FSA before it. However, this is the first time we have set out a framework of this kind, and committed to a programme of ex post impact analysis.

1.5 We intend to keep our approach to measuring impact after our intervention up to date and evolving over time. We continue to be open to views from our stakeholders on how we approach these evaluations.

Robin Finer
Acting Chief Economist

1 Our Mission: www.fca.org.uk/news/press-releases/fca-publishes-mission-business-plan-2017-18

2 Executive summary

2.1 Our Mission, published in April 2017, provided a framework for our strategic choices. It explained how we serve the public interest by improving how financial markets function and how firms conduct their business. It also explained how we regulate to deliver good outcomes for a wide range of users of financial services, including individuals and businesses.

2.2 At the heart of our Mission is a decision-making framework that starts with identification of harm, moves on to use of diagnostic tools and remedy tools, and finishes with evaluation of our decisions.

Figure 1. Our decision-making framework

2.3 This document focuses on the evaluation part. It sets out how we intend to use ex impact evaluation to assess the impact our interventions have had on consumers firms and markets. Our evaluation work will feed back into our decision-making and how best to use our diagnostic and remedy tools.

2.4 Ex post impact evaluations are one way in which we assess the impact of our interventions. They focus on quantifying the impact. Other types of evaluation approaches that we use, such as post-implementation reviews, may also include some elements of quantification of impact.

2.5 Our ex post impact evaluations are focusing initially on measuring the impact of discrete interventions: (i) rules and guidance; (ii) remedies from market studies; and (iii) market wide recommendations stemming from thematic reviews. We are still considering the best way to measure the impact of our broader regulatory activities.

2.6 In the Feedback on Our Mission there was support for more post-event evaluations and measuring outcomes as an important way to promote our accountability. We have started an evaluation programme. This included a pilot of ex post impact evaluations of 3 interventions published in 2018:

- bringing additional benchmarks into the regulatory and supervisory regime
- guaranteed asset protection insurance remedies
- reducing barriers to entry in banking

2.7 In the Mission Feedback Statement, we committed to undertake at least 1 ex post evaluation per year. We will keep this number under review. We will select the interventions we evaluate by considering several factors:

- whether enough time has passed since the intervention for the effects to be observable
- the scale or strategic significance of the intervention
- whether the intervention was novel, and the learning potential offered by an evaluation
- the likely tractability and reliability of an ex post impact evaluation – whether we expect to get a meaningful result

2.8 There are 3 key challenges in undertaking ex post evaluation of our impact.

2.9 The first is in ensuring enough time has passed since the intervention to allow any remedies to have been implemented and identifiable changes in behaviour to take effect: yet, not so long to have passed that things have moved on too much.

2.10 The second is whether we can quantify and attribute any observable impact to our interventions rather than to other changes in the market. We aim to address this challenge by embedding evaluation into our policy-making – being clear from the start about what we need to be able to measure in order to understand the impact of the intervention. However, sometimes events will intervene and it will be impossible to distinguish the effect of our interventions from other market developments.

2.11 The third is to ensure the robustness and credibility of our evaluations. We will do this by either commissioning them externally or, where carried out by FCA staff, ensuring they are peer reviewed by external experts. We expect to publish most, if not all, of our evaluations, so they will also be subject to external scrutiny.

2.12 By addressing these challenges and being transparent about the effectiveness of our work and learning from our past, ex post impact evaluations will contribute to a body of 'what works' evidence across regulation and public policy. They will be a critical element of our commitment to being an effective regulator.

3 Why and what we evaluate

Introduction

The role of evaluation in our Mission

3.1 Our Mission commits us to using our regulatory tools to prevent and reduce harm in a cost effective way (creating public value). We have adopted a decision-making framework (see Figure 1) to focus decisions on public value.

3.2 We identify and diagnose harm, and decide whether it is cost effective to use remedial tools available to us to address that harm.

3.3 Evaluation is central to the decision-making framework because we must understand the impact of our work in order to improve our performance. It allows us to learn from this process by understanding whether remedies have been effective. Over time, evidence from evaluation research will inform our prioritisation decisions, and allow us to focus on remedies that are more likely to be effective.

3.4 The Mission highlighted a three-tier approach to measure how we are doing.

1. **Operational efficiency:** Measuring our operational efficiency by assessing all our internal operations against a Value for Money (VfM) framework based on the National Audit Office's (NAO's) criteria of economy, efficiency and effectiveness.

2. **Impact of our interventions:** Measuring the impact of the interventions we make using our available tools, to identify what works and what does not. We will also keep the indirect effects of our interventions on the way markets work under review to build up a more detailed picture over time.

3. **Outcomes in markets:** Looking at markets and sectors as a whole to identify common root causes, which we can tackle at both a market and firm-by-firm level. This will help us to define the ways we want a sector to improve in both the short and long term and how we will measure the results.

3.5 This document focuses on the second tier of the three-tier approach. We set out how we intend to approach an aspect of our formal evaluation work: ex post impact evaluations.

Structure of this document

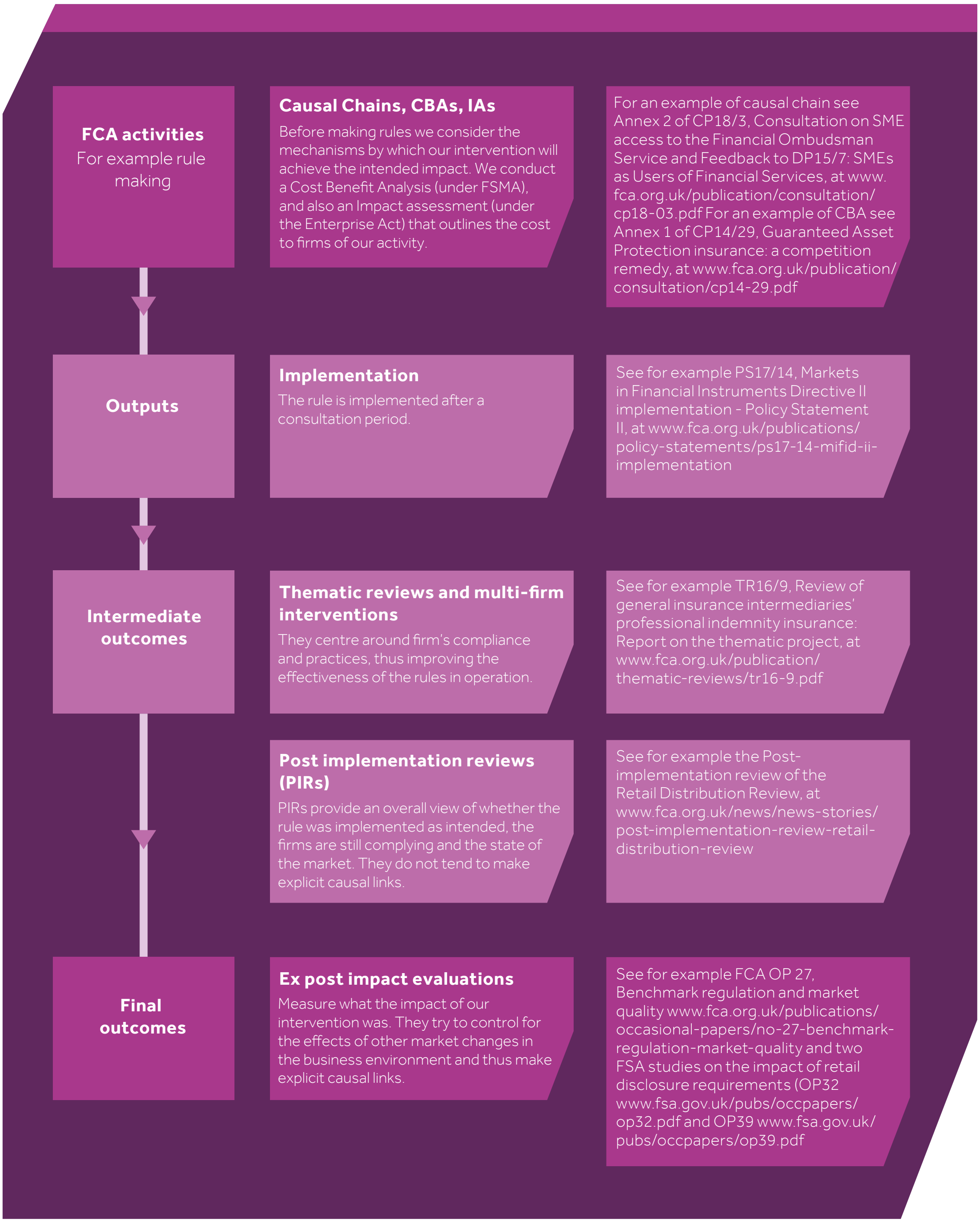
3.6 The paper builds on previous evaluation work, including other regulators' experience.[2] It describes our intended approach to ex post impact evaluation by outlining:

- what we mean by 'ex post impact evaluation'
- why ex post impact evaluation is important for us
- the scope and scale of ex post impact evaluations
- how we will select which interventions to evaluate
- how we will conduct ex post impact evaluations
- the key challenges

2 See for example: Office of Fair Trading (2006) Evaluation strategy for market studies, OFT 862; and various evaluations by the CMA including on BAA market investigation remedies: assets.publishing.service.gov.uk/media/57399d43ed915d152d00000b/evaluation_of_baa__market_investigation_remedies.pdf. See also recommendations by the NAO on measuring effectiveness including: NAO (2014) Regulating financial services www.nao.org.uk/wp-content/uploads/2015/03/Regulating-financial-services.pdf; NAO (2016) Financial services mis-selling: regulation and redress www.nao.org.uk/wp-content/uploads/2016/02/Financial-services-mis-selling-regulation-and-redress-Summary.a.pdf; NAO (2016) Performance measurement by regulators www.nao.org.uk/wp-content/uploads/2016/11/Performance-measurement-by-regulators.pdf

What we mean by 'ex post impact evaluation'

3.7 There is a range of ways in which we can measure the impact of an intervention. The following diagram shows how they fit together.



3.8 Ex post impact evaluations represent the focus of this framework (see also Annex 1 for more detail).

3.9 Ex post impact evaluations look to measure the impact of specific FCA interventions after the event. Impact evaluations focus on assessing what effects interventions have had on consumers and other buyers of financial services[3], firms and market outcomes.

3 Throughout the document when we refer to 'consumers' we refer to both retail consumers and other buyers of financial services, for example companies who buy financial services, or pension fund and charity trustees.

They do this as rigorously as possible by attempting to establish a counterfactual[4] and so control for the effects of other material changes in the business environment. They assess whether the intended outcomes have been achieved, what has worked, and if something has not worked, why not.

3.10 Several of our activities are devoted to ensuring that our interventions are having an impact and understanding potential reasons why that might not be the case (for example, assessing compliance with a new rule). Our impact evaluations build on this ongoing work.

3.11 The key difference between ex post impact evaluations and other types of review work is the focus on quantifying the impact (and with it, where possible, the costs and the benefits) of the intervention. That said, we will generally use a mix of quantitative and qualitative tools (such as feedback from firms or consumers). It is important that we do not over-focus on those impacts that are most straightforward to quantify. In some cases it may be appropriate for other reviews to include an element of ex post impact evaluation.

3.12 These evaluations do not focus on internal process nor on whether we came to the right judgements. Those are covered by our Internal Audit and Risk functions.

Why ex post impact evaluation is important for us

3.13 Ex post impact evaluation makes us a better and more effective regulator in a number of ways:

- It helps us to make better decisions in the future. It helps us **learn** what works and why. It informs the choice of intervention methods and tools. It also helps inform our prioritisation. If we find that the issues identified in a market are still occurring and our interventions have not had the intended effect, or had a negative effect, we will consider our next steps and whether to take further action or to change or remove the intervention if the effects have been negative.
- It provides a strong **evidence base** for ensuring market changing regulations are proportionate and also for repealing ineffective regulation.
- It increases **transparency** regarding our effectiveness in delivering our objectives. We want to be clear to our stakeholders and organisations such as the National Audit Office about how our actions have added public value.
- It helps us **demonstrate, as well as understand, our net impact**. Under the Enterprise Act we have a duty to report on the direct costs we impose on firms. Those costs need to be set against the benefits we bring to consumers and society through our operations and policies.[5]

4 The counterfactual attempts to define what would occur in the absence of an intervention.
5 See NAO (2016) The Business Impact Target: cutting the cost of regulation www.nao.org.uk/wp-content/uploads/2016/06/The-Business-Impact-Target-cutting-the-cost-of-regulation.pdf

The scope and scale of ex post impact evaluations

3.14 Our ex post evaluation programme will focus on identifying **market wide** (rather than firm specific) effects on outcomes (for example the extent of consumer benefits achieved from introducing competition remedies), which have occurred because of our interventions. Specific firm interventions are not currently in scope, nor are ongoing regulatory activities such as authorisations or enforcement. We are still considering ways to assess the impact of such activities.

3.15 The initiatives initially covered by the ex post impact evaluation effort will fall into 3 broad categories:

- rules and guidance
- interventions arising from market studies
- market wide recommendations stemming from thematic reviews

3.16 We are committed to undertake at least 1 ex post impact evaluation per year, but we will keep this number under review. We have started an evaluation programme of work that included a pilot of ex post impact evaluations of 3 interventions: bringing additional benchmarks into the regulatory and supervisory regime (see box 1 and FCA OP 27[6]), guaranteed asset protection insurance remedies (see box 3), and the 2013 reducing barriers to entry in banking.[7] We have now completed and published those evaluations.

6 FCA (2017) OP 27: Benchmark regulation and market quality www.fca.org.uk/publications/occasional-papers/no-27-benchmark-regulation-market-quality

7 See Bank of England and Financial Services Authority 'Review of requirements for firms entering into or expanding in the banking sector' (2013), at www.fsa.gov.uk/static/pubs/other/barriers-to-entry.pdf

4 How we evaluate impact

How we will select which interventions to evaluate

4.1 Undertaking ex post impact evaluations requires us to use our resources and often to make ad hoc data requests from firms. We are mindful of the need to consider value for money in the work we do. We will undertake them only when it is proportionate to do so, for a subset of our interventions. As explained in Our Mission, where it is less cost-effective to conduct detailed analysis, we will monitor key market indicators that help show the impact of our interventions.

4.2 We will select and prioritise evaluations according to the following criteria[8]:

- **Time since the intervention.** Enough time needs to have elapsed to observe change since the intervention (see paragraphs 5.2-5.6).
- **Scale** of the intervention. We are more likely to evaluate an intervention which addresses a significant harm, generates market upheaval or has large ongoing cost to firms.
- **Novelty of the intervention,** or uncertainty over outcomes at the time. It is likely to be more valuable to evaluate the impact of uncertain or novel interventions. Where a remedy has been trialled, for example, and we have strong ex ante (before the event) evidence as to impact we may be less likely to evaluate ex post or, if we do, to focus on particular aspects.
- **Learning potential.** We are more likely to evaluate an intervention where there is potential to act on lessons learnt from the evaluation and those lessons may have relevance to future work (such as extensions to policies) or other markets. This means we are likely to consider both those interventions that are perceived to have worked well and those that are perceived to have worked less well.
- **Likely reliability** of the evaluation (or strength of counterfactual). This affects the type of evaluation we could adopt, as the key constraint is often our ability to identify relevant counterfactuals against which impacts can be measured. If a market has experienced a range of external shocks or there have been multiple interventions, or data have not been collected at implementation, it will be very difficult or impossible to establish the impact of a particular intervention. See paragraphs 5.7-5.12 for a discussion of these challenges.
- **Data availability.** We will also consider whether the data needed to conduct the evaluation are likely to be available or easily attainable. We are mindful that data requests can create a burden on firms. We will consider the best ways to collect the information to undertake the evaluation and balance the rigour gained against the costs this would entail.

8 The selection criteria are broadly based on those proposed by NERA for the FSA in 2004, as well as criteria for selecting ex post evaluations used by a few government departments. See NERA Economic Consulting (2004) The FSA's Methodology for Cost-Benefit Analysis.

How we will conduct ex post impact evaluations

4.3 Our work will include both quantitative and qualitative aspects, as is best practice in evaluation across UK government and internationally.[9]

4.4 We will look to measure the outcome of our interventions in the market against what would have happened without the intervention (or counterfactual). This will allow us to understand our impact robustly and learn any lessons to inform future interventions. A few methodological approaches can help us measure with some confidence the extent to which we have caused any effects in the treated market. For example, 'difference in difference', a statistical technique that studies the differential effect of a treatment on a 'treated' group versus an 'untreated' or comparison group.[10] Attributing the changes observed is a central consideration for the evaluation design. See Annex 2 for a brief description of research designs and methods we can use.

4.5 We will seek to value the benefits of the intervention and relate these to the cost of the intervention. Where feasible, we will make a comparison with ex ante costs and benefits. This will allow us to identify areas where we systematically and materially under or overestimate outcomes of an intervention, and inform key assumptions typically used in CBAs (cost benefit analyses).[11]

4.6 The aim of our ex post impact evaluations will be to assess the impact of our interventions in the market. In particular, whether we have made a positive difference. We will do this by focusing on the fundamental changes we expected. We do not propose that ex post evaluations consider all detailed aspects of the interventions, nor, for example, re-run a CBA.[12]

4.7 For our analysis, we will generally start by using any regulatory data[13] we already routinely collect from firms and existing consumer surveys, such as the Financial Lives Survey. Where these data are not sufficient for us to be able to estimate the impact of an intervention, we will consider what further data need to be gathered and how frequently.

4.8 In addition to the quantification described above, we will also seek to understand the mechanisms by which the intervention has influenced the outcomes observed. We will often include a 'process evaluation' element[14], for example a check that remedies have been implemented as intended based on qualitative discussions with firms or consumers.

4.9 Quantitative or qualitative evidence may reveal previously unknown, unintended consequences. We will examine significant unintended consequences that might transpire from the evidence gathered during an impact evaluation.

9 See for example the Independent Evaluation Group, World Bank at http://ieg.worldbankgroup.org/methodology and also the evaluation work of the European Bank for Reconstruction and Development at www.ebrd.com/what-we-do/evaluation-methodology.html. In the UK, see evaluations produced by the Competition and Markets Authority.

10 See Box 1 for an example and FCA Occasional paper 27, 'Benchmark Regulation and Market Quality', 2017, for more detail.

11 See also Department for Transport (2016) Strengthening the links between Appraisal and Evaluation, by Frontier Economics www.gov.uk/government/uploads/system/uploads/attachment_data/file/540733/strengthening-the-links-between-appraisal-and-evaluation.pdf

12 This is in line with NERA 2004: 'We recommend that ex post analysis investigate specific policy impacts in depth, rather than necessarily attempting an evaluation of the total net impact. For example, compliance costs should only be addressed in ex post analysis where these are thought to be significant to the overall cost-benefit balance, and where ex post analysis is likely to yield significantly improved estimates over ex ante analysis.' See NERA Economic Consulting (2004) The FSA's Methodology for Cost-Benefit Analysis.

13 On the importance of using administrative data see also the NAO report on Evaluation in Government, 2013, at www.nao.org.uk/report/evaluation-government/

14 See HM Treasury (2011) The Magenta Book, Guidance for Evaluation www.gov.uk/government/uploads/system/uploads/attachment_data/file/220544/magenta_book_part_b.pdf

4.10 Box 1 provides an example of how we will use both quantitative and qualitative analysis.

Box 1: Bringing additional benchmarks into the regulatory and supervisory regime

While we have regulated the London InterBank Offered Rate (LIBOR) since 2013, in April 2015 7 additional benchmarks (the ICE Swap Rate, the ICE Brent Index, SONIA, RONIA, WM/Reuters 4pm London Closing Spot Rate, LBMA Gold and Silver Prices) were brought into the regulatory and supervisory perimeter for the first time.

By using a mixture of quantitative and qualitative techniques we construct a detailed picture of the effects of bringing the additional benchmarks into the regulatory perimeter.

For the ICE Swap Rate we carried out an in depth statistical analysis (published in Occasional Paper 27[15]). We used a 'difference in difference' approach to compare outcomes for swap tenors for which a regulated benchmark is calculated to tenors for which it is not calculated, controlling for a number of other factors. The outcomes for tenors for which a benchmark is not calculated represent our 'counterfactual': for this subset of tenors there should not be an effect of the regulation. Therefore, the difference in behaviour between these 2 classes of tenors can be attributed to the regulation itself. We focused our analysis on a number of market quality measures such as spreads and depth in the underlying market on the representativeness of the benchmark itself. In the paper, we estimated direct benefits stemming from the regulation in the region of $4m-7m. These only accounted for a single tenor and a single trading venue. Overall benefits are likely to be significantly larger.

For forex benchmarks (WM/Reuters 4pm London Closing Spot Rate) we used a similar technique. Our rules apply only to the regulated benchmark calculated at 4pm but we calculated synthetic benchmarks and compared liquidity at other times of the day as a counterfactual. We also extended the work of the ICE Swap Rate assessment by producing new measures of benchmark effectiveness – representativeness, attainability, and robustness. We use a mix of simulation analysis and empirical techniques to measure changes in each of these measures.

For benchmarks for which constructing a counterfactual is not possible (eg SONIA and RONIA which are based on all trades that take place in a day) we used a different approach. We relied on desk based research and interviews with a number of market participants to assess what happened to the market as a result of our intervention. By discussing the changes with benchmark administrators, market participants and users of the benchmarks we can construct a reasonably detailed picture of the perceived effects of the regulatory change. Although the results are not as robust statistically they still shed light on the changes experienced in the market.

Embedding a culture of measurement of impact

4.11 We will strengthen the links between ex ante CBA and ex post evaluation. This will make our ex post impact evaluations as robust and effective as possible and help ensure we have a continuous and viable programme of ex post impact evaluations.

15 Financial Conduct Authority (2017) Benchmark regulation and market quality, Occasional Paper 27 www.fca.org.uk/publications/occasional-papers/no-27-benchmark-regulation-market-quality

4.12 We have a legal obligation to undertake CBA ahead of imposing new regulatory requirements. CBAs, as well as the analysis undertaken in market studies, help provide a reliable baseline for ex post impact evaluations.

4.13 We will share any lessons learnt through the evaluations and maintain that knowledge, so that decision-making is informed by the evidence produced by ex post impact evaluations.

Box 2: Strengthening the links between ex ante CBA and ex post evaluation

Planning evaluation at early stages (eg prioritisation and CBA) is beneficial in 3 key ways16: First, it clarifies thinking around the anticipated outcomes that should be included in the ex ante CBA. Constructing a causal pathway17 which sets out the causal chain of events from implementation to outcomes is an effective way to visualise the theory of change underpinning the CBA. This enables us to critically evaluate each link in the chain in a systematic manner. It ensures that the benefits discussed in the CBA are well-founded and evidence-based. Often more than one possible outcome can be expected (for example lower prices or expanded product range or higher quality or a combination of these are all potential positive outcomes of an intervention aimed to improve competition) and the causal chain will reflect this.

Second, it can help ensure consistency in ex ante CBA and ex post impact evaluation. For example, it could help to align the 'without intervention' case in the CBA (often referred to as the 'do minimum') and the 'without intervention' (or counterfactual) case in the evaluation.

Third, it can prompt us to store and collect the data needed for the evaluation, where a decision is made that such an evaluation is likely to meet the selection criteria in Section 7. It is important to remember the key assumptions underlying the CBA modelling. This will allow future evaluators to explore the causes of any discrepancies between outcomes in the ex post impact evaluation and the outcomes expected in the CBA.

Further, by identifying the key indicators needed for the evaluation against what data are available, evaluators can identify gaps in the data that require bespoke data collection, including baseline data, and monitoring. It is crucial that this is done at the ex ante CBA stage because any additional data must be collected before implementation, and in some cases before the Consultation Paper is published. This is due to potential 'anticipation' effects - for instance, prices or sales strategies responding immediately after the announcement. Our recent consultations on overdrafts (CP18/42) and Rent to Own (CP18/35) have signaled our intention to continue to collect data to allow us to be in a position to evaluate such interventions in future.

4.14 Box 3 provides an example of an intervention where the ex post impact evaluation mirrors closely the ex ante CBA, while attempting to provide additional information on the relative outcomes of 2 remedies.

16 See Department for Transport (2016) Strengthening the links between Appraisal and Evaluation, by Frontier Economics www.gov.uk/government/uploads/system/uploads/attachment_data/file/540733/strengthening-the-links-between-appraisal-and-evaluation.pdf

17 A causal chain, or pathway, or logic model, describes the causal mechanisms by which an intervention addresses the identified market failure and reduces the harm, leading to costs and benefits. It does this by linking the intended intermediate and final outcomes with the intervention inputs, activities, processes and theoretical assumptions (see Logic Model Development Guide, WK Kellogg Foundation, 2004, www.bttop.org/sites/default/files/public/W.K.%20Kellogg%20LogicModel.pdf)

Box 3 - Guaranteed Asset Protection

In July 2014[18], we published the general insurance add-on products market study final report. The market study identified significant consumer harm in the sale of add-on Guaranteed Asset Protection (GAP) insurance[19] products. It estimated total consumer overpayment for add-on GAP products to be around £76-121 million (out of an estimated market size of £152 million).

This overpayment was attributed to the point of sale advantage held by add-on sellers, as well as a lack of consumer understanding of GAP insurance products. We considered that this provided add-on sellers with a significant competitive advantage over standalone sellers. To address this, we intervened in 2015 by: making it mandatory for vehicle sellers to provide sufficient information to consumers; and requiring a pause in the sale ('deferred opt-in'). This means that vehicle sellers can start the sales process but cannot conclude the GAP insurance sale for 2 clear days. We believed that having both time and information would enable consumers to decide whether they need GAP insurance, and to shop around if they do.

We expected:

- Improved competition between add-on and standalone sellers
- Better outcomes during the purchasing process, including: an overall decrease in add-on GAP insurance sales, given our concern about consumers buying, potentially, unsuitable add-on products; and more consumers shopping around and buying GAP insurance from standalone providers

To assess the extent to which these impacts occurred, and if these impacts arose due to our intervention, our ex post evaluation looked at the change in the following indicators, many of which were considered in the ex ante CBA:

- the price of add-on GAP insurance
- the quantity of add-on GAP insurance sold
- the marginal cost of supplying add-on GAP
- the share of add-on GAP insurance sales to total GAP sales and to car sales
- the share of GAP insurance sales between add-on and standalone providers
- consumers' engagement with the process of buying GAP insurance (eg shopping around)

Our evaluation included the following key elements:

- work to understand what had happened in the market since our intervention through insight gathered from discussions with industry trade bodies and an analysis of market-level industry data
- an econometric analysis of firms' data to diagnose and isolate the impact of our interventions on the GAP insurance market
- a commissioned survey of those consumers who have recently purchased cars, and so had the opportunity to buy GAP insurance following our intervention

18 www.fca.org.uk/publications/market-studies/general-insurance-add-ons-market-study

19 GAP insurance is predominantly sold as an add-on when someone buys a vehicle. It provides cover for a financial shortfall that can happen when: i) a customer's vehicle is written off or stolen; and ii)the motor insurance pay-out does not pay back its original value at purchase or the remaining finance value (if the vehicle was bought on finance). It can be purchased from the car seller (ie as an add-on) or from a standalone provider.

5 Key challenges

5.1 Ex post impact evaluation can be complex and challenging. Experience to date and from other regulators shows that sometimes findings from evaluations are mixed or unclear. There may be uncertainty around the quantified net benefits, or there may be positive elements of interventions even when they have not worked as anticipated overall, or the impact of some interventions simply cannot be evaluated with any degree of robustness. In this section, we explain some of the key challenges and how we will address them.

Time since the intervention

5.2 While we have undertaken many activities since we were established in 2013, there is often a lag between our activities and resulting market changes. For example, it may take time for market study remedies to be implemented as rules, for compliance systems to be put in place and for identifiable changes in behaviour to take effect (although some behaviours might change in anticipation of rule changes). Even after implementation, some effects might happen immediately (eg pay day price cap), while others will take longer (eg consumers' learning and plans for firms to innovate or enter the market) so we may only expect the full benefits of an intervention to come about some time later.

5.3 Undertaking an evaluation too soon may lead to no impact being identified and an inability to quantify any benefits to consumers, even though these materialise later on. On the other hand, waiting too long before evaluating an intervention:

- makes it too difficult to attribute impact to the intervention, as other factors also affect the market (eg it is difficult to evaluate the impact of Retail Distribution Review, in light of the activities brought about by the Financial Advice Market Review)
- may lead to memory recall problems when asking firms and consumers about the intervention
- delays remedial action to address shortcomings in the intervention

5.4 If we are interested in the exact impact of a measure, we may look at impacts right around the intervention date so that we do not conflate with broader trends and dynamic responses. However, if the remedy was introduced with dynamic interactions between firms and consumers in mind, we may be more interested in looking at the outcome after enough time has passed for these effects to have taken place. Both are valid approaches in evaluation.[20]

5.5 Planning for evaluation at an early stage helps to address this timing challenge. For example, by collecting and monitoring any necessary data, and then undertaking the final impact evaluation assessment at the appropriate later stage.

20 See also Financial Stability Board (2017) Proposed Framework for Post-Implementation Evaluation of the Effects of the G20 Financial Regulatory Reforms www.fsb.org/wp-content/uploads/Framework-for-the-post-implementation-evaluation-of-the-G20-financial-regulatory-reforms.pdf

5.6 If monitoring of key measures in the causal chain reflects the theory and/or our initial expectations, then monitoring may often be sufficient evidence that our interventions achieve what was intended. So, an impact evaluation may not be needed. A good example of ways in which we are trying to improve monitoring is in the future monitoring of the retail banking sector following the Strategic Review of Retail Banking Business Models.

Quantifying and attributing impact to the FCA - ie whether the intervention has caused the change in outcomes

5.7 The counterfactual may be difficult to identify (for example our evaluation of reducing barriers to entry in banking). This makes impact difficult to quantify, as markets may change over time due to factors external to the FCA (such as innovation).[21] Establishing the right counterfactual is crucial to estimating impact correctly. For example, a lack of positive change in outcome measures after an intervention may not be a bad result. There might have been a negative change if the intervention had not taken place. Other regulatory changes happening around the same time also need to be taken into account.

5.8 There may be instances where we do not have a strong counterfactual, or where this cannot be identified in advance. This is inevitable because we want to focus on evaluating areas that matter, not just those where the counterfactual is most robust. In some cases we may judge it is important to go ahead with an evaluation because we can learn from it even if the counterfactual is not as robust as we would like.

5.9 We may not always be able to carry out an intended impact evaluation of interventions. Intervening events (for example actions by other agencies) can change market conditions significantly. This can mean it is no longer possible to isolate the impact of a specific intervention.

5.10 Sometimes we need to accept that it may not be possible to fully quantify the impact downstream of an intervention upstream. For example, the total net impact on consumers resulting from reducing barriers to entry in banking through a change in the authorisation process. In addition, some outcome measures might be difficult to quantify (eg quality, innovation) and we need to rely on proxies.

5.11 Very often, our interventions include a package of inter-related remedies, rather than a single remedy, and we need to decide whether we want to understand individual or combined effects. For example, with Guaranteed Asset Protection (GAP) (see box 3) we asked firms to provide information at point of sale and decoupled the sale of the car from the sale of GAP insurance. In the evaluation, we tried to learn about the relative efficacy of those 2 remedies. In other instances, we may only be interested in the combined effect. An evaluation will want to demonstrate the value and effectiveness of the policy but also learn about whether it needs refining (as also explained in Section 3).

5.12 In some cases, there will be various actors involved (eg HMT, the BoE, the European Commission or the CMA). It may be hard to separate the work of the FCA from that of others (for example, our changes to remuneration in banking occurred alongside

21 See Office of Fair Trading (2006) Evaluation strategy for market studies, OFT 862

the EC's bonus cap; see also our evaluation of the impact of bringing additional benchmarks into the regulatory and supervisory regime). In those instances, we propose to review particular outcomes of an intervention ex post rather than attempt to measure every aspect or the whole intervention. For example, we may decide to focus on whether there are signs of improved conduct and whether the market is 'cleaner', rather than attempting:

1. to disentangle the effect of each policy requirement on firms and consumers[22]

2. to separate contributions of the FCA from other organisations, if it is reasonable to suppose our intervention had a part to play

Ensuring robustness and credibility

5.13 To be credible, evaluation work needs to be robust and done with sufficient independence. We have discussed how we intend to deliver robust ex post impact evaluations by embedding evaluation at paragraphs 4.11-4.14. We anticipate the robustness of our evaluations will increase over time.

5.14 We are committed to independent ex post impact evaluations. We will either externally commission them or, when we undertake them internally, building on knowledge and expertise from across the FCA, we will seek input from objective external experts. They will peer review the method and the quality of evaluations.[23] The impact evaluations we have published so far have all benefited from the advice and peer review of academics who are experts in evaluation techniques and/or the sectors we were evaluating. Any ex post impact evaluation will be led by a different team from the one that introduced the intervention being evaluated.

5.15 We will continue to consider further ways to ensure objectivity in our impact evaluations. This may include, for example, double peer reviews, an academic panel and/or commissioning work externally.

5.16 We will consider whether to publish our evaluations on a case by case basis, to take into account potential commercial sensitivities. We expect to publish our evaluations in most cases. This will ensure transparency and credibility of the work and will allow it to contribute to the body of public policy evidence on effective regulatory interventions.

22 Jill Rutter in her paper 'Evidence and Evaluation in Policy Making' acknowledges this difficulty by stating that "the more natural reaction was to accept it was not possible [to test a policy] rather than working out what parts of the policy could be tested". See www.good-government.info/files/evidence_and_evaluation_in_template_final_0.pdf

23 Other regulators, including the CMA, use the same approach to ensure independence of their evaluations.

Annex 1
Activities relevant to the measurement of impact of interventions ex post

1. The table below shows that different types of evaluations might be useful in different instances. For example, the cumulative effects analysis might be useful for advocacy purposes and Post Implementation Reviews could give us a good idea of the state of the market and whether future action is needed.

Measuring the impact of the FCA	What it involves	How to use for measuring impact
Monitoring relevant metrics *	Develop indicators to help monitor the development of financial markets on a regular basis, with a focus on what matters to consumers, such as product quality or suitability of financial advice – eg granular data on product prices, volumes and quality, and personal characteristics of the buyer. It may also include looking at trends over time (eg from the Financial Lives Survey). It helps identify issues eg remaining market imperfections or consumer harms that could subsequently be examined in detail during evaluations.[26]	To provide data that enable ex post impact evaluations. Indicators need to be examined and framed carefully, to avoid mis-interpreting negative or positive movements (eg an increase in volume of complaints may be due to an increase in awareness, rather than the worsening of a problem).
Market research*	Undertake qualitative or quantitative research to gather consumers' own views on topics of interest, such as the impact of new regulation. It brings in the consumer view and adds useful insight in terms of consumer behaviour and attitudes in relation to financial services.	An important research tool often used in ex post impact evaluations to provide consumer context.
Thematic reviews*	Review to assess firms' compliance (eg regarding suitability of advice) and practices, for instance in how firms treat consumers or undertake sales. They highlight areas of non-compliance that may lead to enforcement action or guidance to firms that need help in complying, so improving the effectiveness of the rules in operation. (When thematic reviews provide recommendations, then the effectiveness of those recommendations could also become the object of an ex post evaluation.)	To provide information on firms' compliance and practices in response to our interventions. This can feed into ex post impact evaluations.

24 See Financial Stability Board (2017) Proposed Framework for Post-Implementation Evaluation of the Effects of the G20 Financial Regulatory Reforms www.fsb.org/wp-content/uploads/Framework-for-the-post-implementation-evaluation-of-the-G20-financial-regulatory-reforms.pdf

Post-implementation reviews (PIRs)	Captures all kinds of ex post reviews but often means assessing the impact through discussion with firms or consumers, after an intervention has been implemented. It provides useful information on compliance, implementation issues and the state of the market after an intervention, but says less on whether the state of the market is a direct result of our intervention.	Where counterfactual currently difficult to establish, eg RDR.
Ex post impact evaluations	A subset of PIRs that attempts to establish the counterfactual (what would occur in the absence of an intervention). They measure the impact of interventions on outcomes in a way that controls for the effects of material changes in the business environment. It can include qualitative discussions with stakeholders to help understand why results are as they are. It can also include studies which show that both intermediate and final outcomes have changed in the direction we expected along an agreed causal pathway (from the implementation of the policy to the realisation of benefits). Data analysis, coupled with views from stakeholders, helps establish a counterfactual.	For tightly focused interventions (eg Guaranteed Asset Protection rules at Box 3).
Measuring the impact of types of activity	Focuses on a particular impact and group of similar activities eg our deterrent effect. An example of this are the OFT studies on the deterrent effect of competition enforcement (OFT 962, OFT 1391)[27]. OFT 1391 showed that sanctions and enforcement in the UK appear to have a substantial deterrent effect. It reported that for every cartel investigation, 28 cartel cases are deterred. In the case of other commercial agreements and abuse of dominance for every OFT investigation, 40 and 12 cases were deterred respectively. This type of study shows the value of a group of activities, by surveying key players such as competition lawyers or heads of risk or compliance.	For example to measure the value of our competition enforcement.
Analysing cumulative impacts**	Measuring market cleanliness (benefit). Assessing the cumulative burden of FCA regulation (costs).	For example to measure trust in markets.

* These are not types of evaluations but they provide important information to enable measurement of impact ex post.

** This type of studies is of particular interest where there have been multiple changes within the market (eg changes to regulation, multiple regulatory interventions from different authorities etc).

25 Office of Fair Trading (2007) The deterrent effect of competition enforcement by the OFT, by Deloitte, OFT 962; Office of Fair Trading (2011) The impact of competition interventions on compliance and deterrence, OFT 1391

Annex 2
Research designs we may use

1. There are several research designs and methods that can be used in ex post evaluation. The following tables[26] outline those we are more likely to adopt.

Designs and methods	Description	Example of use
Unconstrained designs		
Quasi experimental (QE) designs	Use a comparison group that is 'as good as random' or obtained in a way that allows selection bias to be modelled. Various options for obtaining a comparison group to ensure the treated and untreated groups are as similar as possible.	When a policy can be introduced in a staggered fashion.
Regression Discontinuity Design	Examines the boundary between the 'only just eligible' and the 'not quite eligible'. Results only apply directly to those at the boundary.	We have used this in ex ante CBA (price cap on high-cost short-term credit).
Difference in difference method	A method for analysing QE data. Compares how trends in outcomes change between treated and untreated groups over a time period relevant to the intervention. Unobserved factors might affect the outcome, but if they do not affect trends in the outcome then the trends for both groups in the absence of a policy will be the same.	Used in our Occasional Paper 27 to study the effect of bringing the ICE swap rate under the regulatory umbrella.
Interrupted time series design	No comparison group is available. The counterfactual is estimated from a projection of the outcome measure before the intervention.	When alternative causes for changes in outcomes can be eliminated and the impact is large compared with the error inherent in forecasting.

Designs and methods	Description	Example of use
Constrained designs		
Natural experiments and instrumental variables	Comparisons with a naturally occurring comparison group can be made even though none was present by design.Or an external factor can be identified, which influences the likelihood of being exposed to a policy and does not in itself affect outcomes.	When the policy has already been implemented and the opportunity to put an evaluation design in place at implementation was missed.
Before and after studies	An outcome is measured before and after an intervention but there is no comparison group.	When the intervention is the only thing that could reasonably be expected to influence the result.
Use of process evaluation information	Draws upon the findings of studies of the implementation and delivery of an intervention, often using qualitative methods, including case studies. Eg front line staff often have a good feel for whether an intervention is effective or not.	When quantitative measures of impact are weak or not available. May capture a direction of change.

26 From HMT (2011) 'The Magenta Book – Guidance for evaluation' – please see for a more extensive and detailed description of evaluation research designs and methods, www.gov.uk/government/publications/the-magenta-book.

2. For our ex post impact evaluations, we will use the most robust design that we can, given the data and any other constraints related to the intervention we wish to evaluate.

3. This means that our ex post impact evaluations may include studies where a robust counterfactual is available. For example, a comparison group similar in all respects to the group affected by the policy, but for being subject to the intervention. There are some examples in financial services.[27]

4. They may include studies where it can be difficult to identify a counterfactual, but data may nevertheless be revealing. For example, where a market shrinks or the price jumps post-regulation as theory would predict. To come to a view about causality, those before and after studies will typically consider interim as well as final outcomes, and triangulate from a variety of sources, to consider if other factors have influenced the results or data. Other regulators and agencies, such as the CMA, have used similar techniques in the past.

Only studies whose counterfactual is above a certain 'robustness threshold' are usually considered in systematic reviews[28] that summarise the evidence on 'what works' in a particular sector. We believe that in some instances studies just below that threshold (such as those included in the table, under 'constrained designs') can provide valuable lessons on our impact, as long as the limitations and caveats on any findings are made clear.

All our publications are available to download from 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 or write to: Editorial and Digital team, Financial Conduct Authority, 12 Endeavour Square London E20 1JN

27 See FCA (2017) OP 27: Benchmark regulation and market quality www.fca.org.uk/publications/occasional-papers/no-27-benchmark-regulation-market-quality. Two published Financial Services Authority (FSA) studies found evidence of retail disclosure requirements having measurable impacts (FSA (2009) Did life and pensions 'disclosure' work as expected? Occasional Paper 32 www.fsa.gov.uk/pubs/occpapers/op32.pdf; FSA(2009) Regulatory reform and the cost of retail investing through life offices: 1988-2006, by 1776 Consulting, Occasional Paper 39 www.fsa.gov.uk/pubs/occpapers/op39.pdf). FSA OPs 27 and 35 found no beneficial effects on mortgage purchases from the introduction of mortgage regulation in 2004 (see FSA (2007), Market impacts of MCOB, Occasional Paper 27 www.fsa.gov.uk/pubs/occpapers/op27.pdf; FSA (2009) Has MCOB regulation affected the suitability of mortgage sales to borrowers with impaired credit histories? Occasional Paper 35 www.fsa.gov.uk/pubs/occpapers/op35.pdf).

28 Such as those undertaken by the network of What Works Centres, which was created to provide robust and comprehensive evidence that will guide decision-making on public spending (see www.esrc.ac.uk/collaboration/collaboration-oportunities/what-works-centres/). A systematic review is defined as: 'A review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyse data from the studies that are included in the review' (Cochrane Collaboration, 2014).

Pub ref: 005846

FINANCIAL
CONDUCT
AUTHORITY

12 Endeavour Square London E20 1JN
Telephone: +44 (0)20 7066 1000
Website: www.fca.org.uk