Europe Economics

RAMBØLL

Scoping Study for an Impact Evaluation of the RMR

September 2013

Europe Economics
Chancery House
53-64 Chancery Lane
London WC2A 1QU

Tel: (+44) (0) 20 7831 4717
Fax: (+44) (0) 20 7831 4515

Ramboll
60 Newman Street
London W1T 3DA
United Kingdom

Tel: (+44) (0) 20 7631 5291
Fax: (+44) (0) 20 732 4645

www.europe-economics.com

Contents

1 Executive Summary

1.1 Introduction

Ofgem's Retail Market Review (RMR) proposals represent the most significant intervention in the energy retail market since the sector was opened up to competition more than a decade ago. Ofgem is concerned that the retail market is not currently working in the interests of consumers as effectively as it could and has proposed policy measures that it considers will make the energy retail market simpler, clearer and fairer for domestic consumers. Europe Economics and Ramboll have prepared this scoping study to consider options for an impact evaluation of the RMR.

1.2 Choosing an Evaluation Design

The evaluation design will have a bearing of the level of robustness, fidelity, and the scope for generalising, the results generated. We summarise the main variables affecting the choice of evaluation design.

We consider that the RMR program is likely to be relatively stable and that it is possible to develop clear, evidence-based, intervention logics for the different policy areas. In addition it would be possible to document intervention fidelity (i.e. the extent to which the policies were implemented as Ofgem intended) for most of the policies, though some policies are more prescriptive than others.

The key challenge lies in the extreme difficulty of establishing a counterfactual and / or a control group, whether based on random or non-random assignments. In the absence of these, we conclude that attribution methods such as Randomised Control Trials or Matching would not be possible for the RMR impact evaluation. Instead, we suggest evaluation strategies which address the contribution of the RMR policies and identify causal relationships through a mixed-mode evaluation approach. Given the relatively good baseline data already in place it would be possible to design forms of time series analysis and apply different analytical strategies, such as multivariate analysis and qualitative assessments, to explore causal relationships between the interventions, outputs and outcomes.

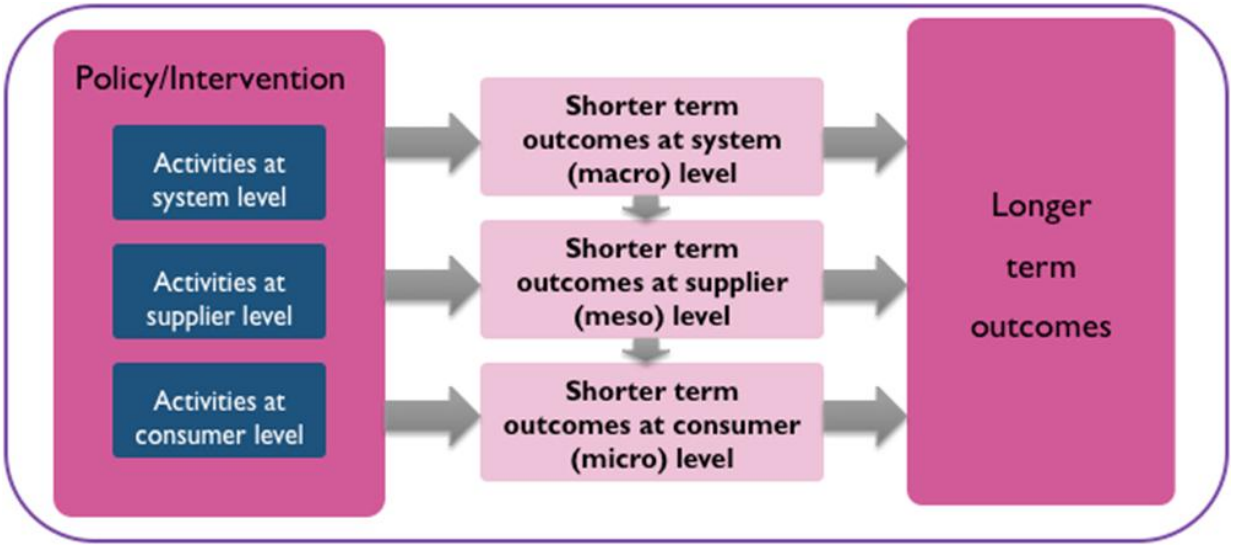
If only the existing data sources are relied on, however, there would be challenges related to bias, coverage and comparability across the indicators that we have put forward. We therefore suggest a tailored tracker survey of consumers, a separate supplier survey and additional qualitative consumer research in order to build a robust evidence base which would allow for more comprehensive time series analysis.

The impact evaluation should link closely with a process evaluation of the RMR. The process evaluation would provide important evidence for the impact evaluation, such as which factors critically contributed to a successful outcome, or whether a failure can be attributed to a poor initial hypothesis (theory failure) or to poor delivery (implementation failure). Whilst this study focuses on the design for the impact evaluation, section 7.3 sets out certain considerations for a process evaluation.

1.3 Overview of the Evaluation Framework

The following figure illustrates how interventions such as the RMR may influence units at different levels from individual consumers (micro-level), suppliers (meso-level) and the overall market (macro-level). The proposed evaluation approach would consider impacts on each of these levels.

Figure 1.1: Levels at which potential outcomes may be generated



In light of these different levels of impact, the evaluation framework that we set out in this report is focused on testing the contribution of policies towards the broad objectives of the RMR rather than evaluating every detail of the proposals.

Ofgem has structured its RMR consultations and Impact Assessments around the energy market features that individual policy measures are designed to influence. These are:

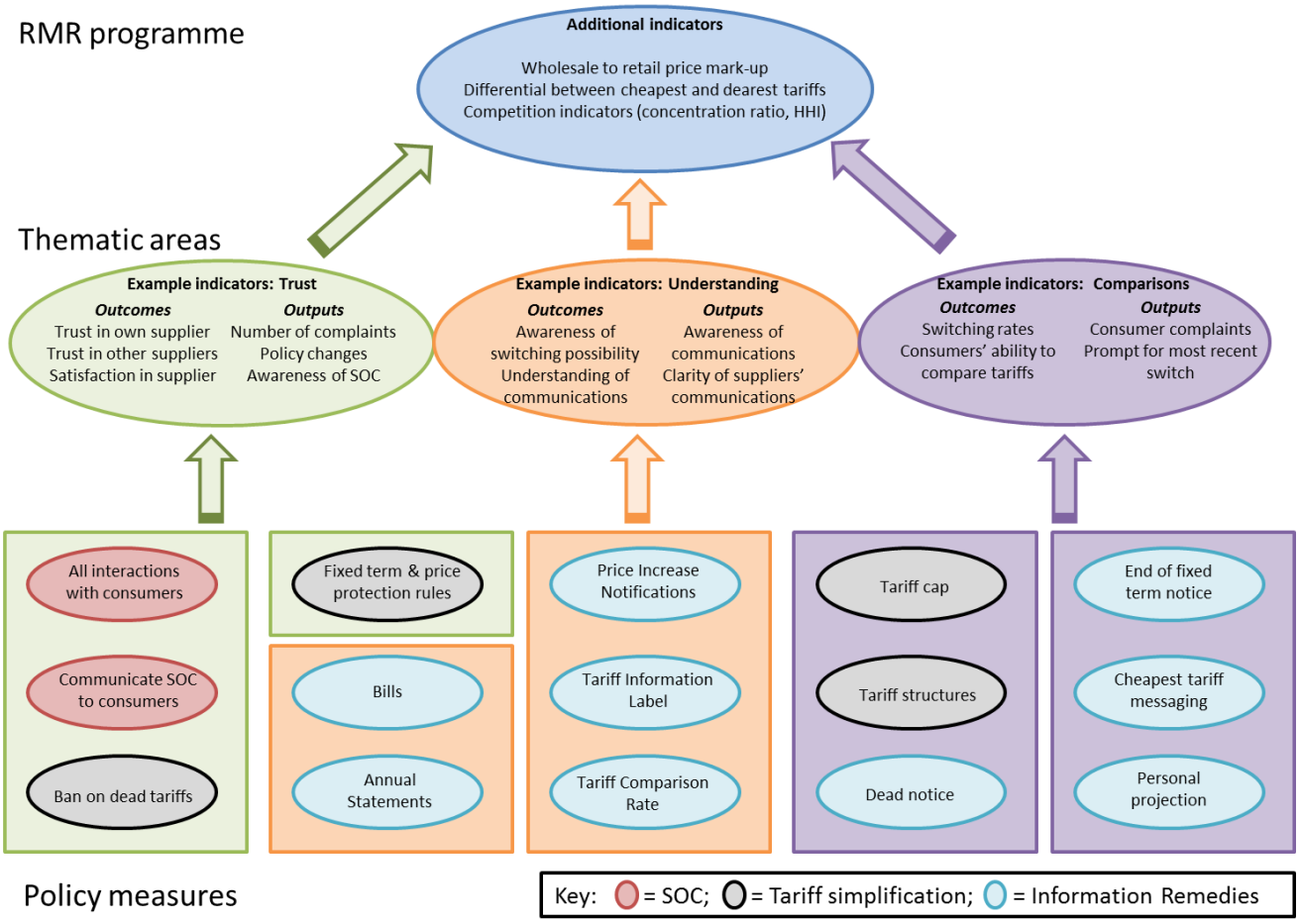
- treatment of customers, influenced by Standards of Conduct (SOC);
- quality of suppliers' communications, influenced by Information Remedies; and
- tariff complexity, influenced by the Tariff Simplification proposals.

We considered the details of each of the individual policy measures presented by Ofgem and concluded that for the purposes of an impact evaluation it is appropriate to group policy measures by objective rather than by affected market feature. Accordingly we have split the policy measures into the following three categories:

- building trust;
- improving understanding of the energy market; and
- simplifying tariff choices.

Our approach was inspired both by a critical review of the individual policy measures proposed by Ofgem and by high-level consumer journey diagrams produced by the Ofgem policy teams during the course of this project. The figure below summarises how we have mapped the RMR policies across to this framework, having prioritised the objectives of each policy measure.

Figure 1.2: Evaluation Framework



The above diagram should not be read as implying any absence of inter-dependency between these themes. For example, the building trust proposals can be considered to influence both the consumer's decision to invest time achieving a clearer understanding of both their energy tariff and of the market, and also whether (or not) to take any action in response to a trigger point. Nonetheless, creating a distinction between the key objectives of the different policy measures is an essential first step in establishing the indicators against which the RMR will be evaluated.

1.4 Key Recommendations

It is important to note that we propose a 'top-down' approach for the evaluation: because of the interdependent nature of the policy measures it would be fraught with difficulty to attempt to focus the evaluation on identifying the impact of each of the policy measures in isolation. Rather, the proposed evaluation design will enable the assessment of how policies, often in combination, contribute to the three broad objectives: trust, understanding and simplifying tariff choice.

The indicators suggested operationalise the broad objectives and focus the monitoring activities on the factors which matter the most. If coupled with strong evaluation norms or performance criteria, it will be possible to assess and compare the relative success or failure of the policies' contribution to each of the three broad objectives in a cross-section analysis.

We summarise our proposed evaluation options in Table 1.1. These options have been designed such that Ofgem will have a range of approaches to choose from, taking into account the required rigour of the analysis, the resources available for completing the evaluation and the data available to Ofgem. We do not

propose a single preferred option but simply outline three appropriate evaluation options given the trade-offs described. No option includes economic evaluation in the form of a formal cost-benefit analysis (by monetising the costs and benefits of the RMR package as whole) as we conclude that it is not feasible to undertake.

Table 1.1: Summary of evaluation options

Type of option	Analytical method	Main data sources	Description
High-cost option	Comprehensive time series analysis	Repeated tailor-made consumer survey with a comprehensive set of independent variables	The tailor-made survey would enable to collect data on outcomes and output indicators as well as consumers underlying attitudes. The latter are extremely important for a multivariate and sub-group analysis. Repeating the survey over time would enable to address the issue of time-lags in implementation of the policy and consumers' response.
		Suppliers' data: comprehensive summary statistics of activities	Working with suppliers to regularly obtain a comprehensive set of summary statistics related to the indicators, where relevant.
		Supplier survey	Experiences of introducing and implementing new policies
		Repeated independent assessments and qualitative consumer research	Complementing statistical data with more qualitative measures would enable the exploration of contribution and causal relationships
Medium-cost option	Limited time series or before-and-after analysis	Before-and-after tailored tracker consumer survey with a comprehensive set of independent variables; or Repeated tailor-made consumer survey with limited set of independent variables	These options are variations of the high-costs option. Because the designing of the questionnaire (with a comprehensive set of variables) will be the most costly part of the methodology, it is expected that a one-off tailor-made consumer survey with a comprehensive set of independent variables will be more costly than a repeated tailor-made consumer survey with limited set of independent variables. The latter would focus on collecting outcome and output indicators as identified in this report. Based on this it would be possible to undertake some multivariate analysis.
		Suppliers' data: limited summary statistics of activities	Working with suppliers to regularly obtain a limited set of summary statistics related to the indicators, where relevant.
		One-off targeted independent assessments and qualitative consumer research	Complementing statistical data with more qualitative measures would enable the exploration of contribution and causal relationships
Low-cost option	Descriptive analysis	Existing survey and consumer data	This approach would focus on using existing information in a systematic and consistent way overtime. Limited multivariate (or correlation) analysis might also be possible.

2 Introduction

Ofgem's Retail Market Review (RMR) proposals represent the most significant intervention in the energy retail market since the sector was opened up to competition more than a decade ago. Ofgem is concerned that the retail market is not currently working in the interests of consumers as effectively as it could. While it considers that the relatively light-touch remedies that were introduced following the Energy Supply Probe led to some improvements for consumers, it still has concerns about the extent to which consumers are engaged with the energy market.

Ofgem intends the RMR to make the energy retail market simpler, clearer and fairer for domestic consumers. The proposals aim to simplify the market by limiting the number of energy tariffs that can be offered by each supplier and requiring each tariff to have a 'standing charge plus unit rate' structure. The RMR aims to make the market clearer by specifying the manner in which discounts can be applied to tariffs and introducing a Tariff Comparison Rate (TCR) to make it easier for consumers to compare the cost of energy tariffs. Ofgem also seeks to make the information provided on suppliers' communications easier for consumers to understand. Finally, binding Standards of Conduct (SOC) will be introduced to ensure that energy suppliers treat their customers fairly.

2.1 The Contribution of this Report

Ofgem asked Europe Economics and Ramboll to undertake a scoping study to consider options for an impact evaluation of the domestic RMR.[1]

Evaluating the success or failure of the RMR will be challenging. The Review was announced in autumn 2010 and a number of consultations have been issued since that date. A number of suppliers appear to have responded positively to some of Ofgem's concerns already. For example, a number of suppliers have started to re-design their energy bills to make the information more accessible to consumers and the industry appears to be actively working towards re-building trust with consumers. Moreover, some suppliers have reduced the number of tariffs that they offer and / or have amended the structure of some tariffs.

The fact that the industry has responded to some of the themes of the RMR before the proposals have been introduced complicates the methodology that would be required to evaluate the RMR package, if it is introduced following statutory consultation. The reason for this is that policy impacts may be misrepresented if it is assumed that suppliers respond only once the RMR proposals are implemented.

One important contribution of this report, therefore, is the proposal of an evaluation approach that can account for the long and open consultation process that has led to Ofgem's final proposals. This report also identifies key indicators which have the potential to demonstrate the extent to which the RMR succeeded in its various objectives.

It is generally not possible to isolate entirely the impact of individual policy measures. In part, this is because numerous individual measures share the same objective but further difficulties concern separating the impacts of an RMR policy from those actions that suppliers would have taken even in the absence of the RMR and the impacts of media coverage and other external factors on consumers' perceptions of the

[1] The project was undertaken in close collaboration with Ofgem and made significant use of its knowledge and expertise.

energy market. Instead, through careful categorisation of policy measures into sets (see section 2.2) and considered design of the proposed indicators, we have sought to ensure that each identified indicator will primarily capture the impacts of one of these sets of policy measures.

We propose evaluation options that we consider to be feasible, given the numerous challenges of evaluating the RMR. In forming each of these options, we have considered the baseline data that are available and also the additional information that Ofgem would need to gather in the future if the impacts of the RMR are to be identified.

2.2 Categorisation of Policy Measures

Ofgem has structured its RMR consultations and Impact Assessments around the energy market features that individual policy measures are designed to influence. These are:

- treatment of customers, influenced by Standards of Conduct (SOC);
- quality of suppliers' communications, influenced by Information Remedies; and
- tariff complexity, influenced by the Tariff Simplification proposals.

We have reviewed the details of each of the policy measures presented in Ofgem's consultation documents and Impact Assessments and have concluded that, for the purpose of evaluation, it would be appropriate to group policy measures by objective rather than affected market feature.

As part of the analysis to support this study, Ofgem identified 16 individual policy measures that could be subjected to future evaluation. We have combined the three Standards of Conduct policy measures into two and divided the revised total of 15 policy measures into three categories, defined by policy objective:

- building trust;
- improving understanding of the energy market; and
- simplifying tariff choices.

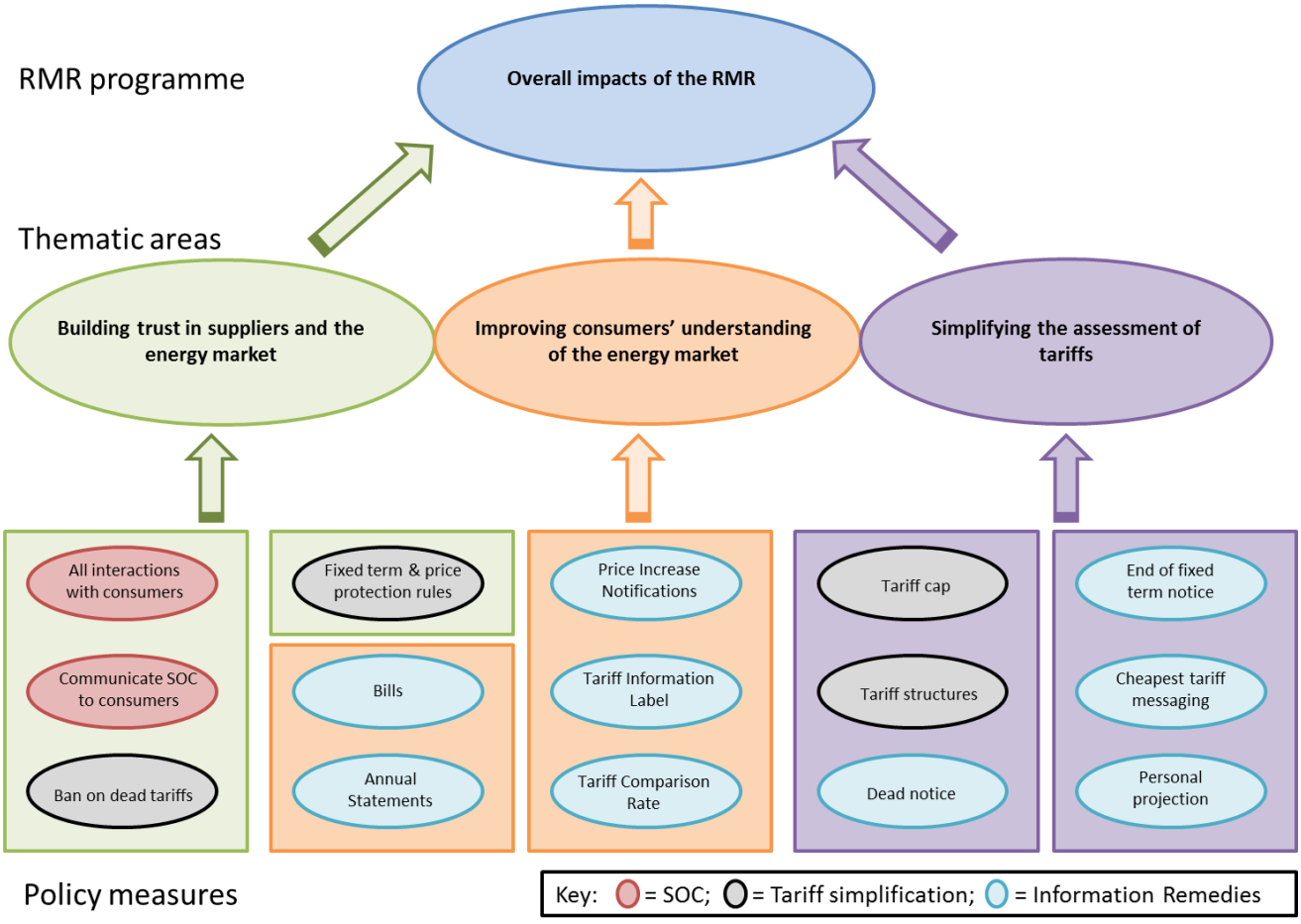
Our approach was inspired both by a critical review of the individual policy measures proposed by Ofgem and by high level consumer journey diagrams produced by the Ofgem policy team during the course of this project.

2.3 Framework for Scoping Study

As indicated above, we consider that grouping Ofgem's policy proposals by objective rather than by affected market feature creates a clear distinction between the types of impacts that should be expected of each group of proposals. To enable policy measures to be grouped in this way, it was necessary first to identify the key objective of each policy measure (notwithstanding the fact that some measures have multiple objectives).

Before turning to a detailed description of the policy measures, Figure 2.1 summarises how we have mapped Ofgem's policies across to our framework. This reflects our prioritisation of the objectives of each policy measure.

Figure 2.1: Framework for scoping study



As indicated in the figure above, the proposals to build trust, improve understanding of the energy market and simplify tariff choices work together and contribute to the overall RMR policy objective of creating a simpler, clearer and fairer energy market. In particular:

- the building trust proposals can be considered to influence both the consumer's decision to invest time achieving a clearer understanding of their energy tariff and of the market and also whether (or not) to take any action in response to a trigger point;
- the proposals to improve consumers' understanding of the energy market should act to increase the proportion of consumers that understand their current energy tariff and consumption and so could use this information when reviewing the market; and
- the proposals to simplify tariff choices would act to increase the proportion of consumers that are able to effectively assess their tariff options during a review of the market and so should lead to an increase in the proportion of consumers that switch following a market review, especially if the proposals significantly increase trust in the market.

This analysis identifies some of the interdependencies between the policies which we have allocated to the different objectives. We describe the interaction between policies in more detail in section 2.4.

2.4 Policy Measures

2.4.1 Building trust

We consider that a number of policy measures contribute towards building trust in the energy market. Trust can be seen as a pre-requisite for consumers to participate in the energy market. If consumers believe that energy suppliers are "all as bad as each other" or do not trust the information provided by suppliers (or their representatives) it is unlikely that they will consider participating in the market.[2] On this basis, the proposals that aim to build trust would act, in part, to reduce the number of consumers that choose to 'do nothing' in response to any given trigger point.

The policy measures that we consider primarily contribute towards building trust in the energy market are described below.

Establishing binding and enforceable Standards of Conduct and communicating these to consumers

The introduction of binding Standards of Conduct (SOC) aims to ensure that suppliers and their representatives treat consumers in a manner that is fair, honest, transparent, appropriate and professional at all points of interactions. There is also a process element, e.g. suppliers will be required to make it easy for consumers to contact them and ensure that arrangements are fit for purpose.

The SOC should ensure that suppliers can be held to account by consumers. These proposals aim, therefore, to increase the level of consumer trust in suppliers, both through improved supplier conduct and consumer awareness that suppliers will be subject to binding and enforceable SOC.

It is noted that while we have defined the key objective of the SOC to be building trust with the energy market, it will also contribute towards improving consumers' understanding of the energy market. Indeed, suppliers will be required to provide information that is: complete, accurate, and not misleading; communicated in plain and intelligible language; relates to products or services that are appropriate for the customer; and fair in terms of its content and in terms of how it is presented.

The SOC will therefore influence the quality of communications that suppliers will send to consumers but it will not be possible to identify the influence of the SOC on communications separately from Ofgem's specific policy measures for suppliers' communications (indeed, these proposals have been designed with the SOC in mind). Therefore, we consider that it is appropriate to evaluate the SOC in terms of its core objective of building trust, notwithstanding the fact that they will influence numerous other RMR policy measures.

Ban on dead tariffs

The ban on dead tariffs aims to protect consumers that are currently on a dead tariff (i.e. those that are no longer open to new customers), improve the transparency of dead tariffs and improve the comparability of dead tariffs with those that are open to new consumers. The policy measure aims to build trust among those consumers that are on a dead tariff through such enhanced consumer protection.

This measure is likely to affect primarily consumers who are comparatively less engaged in the energy market (as identified by previous switching behaviour), as those on dead tariffs are less likely to have switched supplier and/or tariff recently. Therefore, building trust is likely to be a particularly important first step towards increasing the level of engagement for these consumers.

[2] Ipsos MORI (2012), "Consumer engagement with the energy market, information needs and perceptions of Ofgem", page 29

In light of these observations, we consider that the primary objective of the ban on expensive dead tariffs is increased consumer protection, which should lead to increased trust in the energy market. This measure may have a side-effect of simplifying the assessment of tariffs for those that currently have a dead tariff but we understand that this is not the core reason for Ofgem's intervention.

Fixed term and price protection rules

This policy measure aims to increase consumer protection for those customers that are on a fixed term tariff or switch to a new fixed term tariff. Through enhanced consumer protection, the policy measure aims to build trust among those consumers that have chosen a fixed term tariff, for example by building recognition among consumers that they are not at risk of being 'ripped-off' by their supplier when the fixed term tariff comes to an end.

Consumers that are on a fixed term tariff will, at some point, have made an active decision to switch to this tariff. Therefore, the fixed term and price protection rules are likely to affect primarily consumers who are (or at least have been) comparatively more engaged in the energy market.

2.4.2 Improving understanding of the energy market

Even if consumers have trust in the market (which can be seen as a pre-requisite for full engagement), there may be other barriers that prevent them from participating in it. One such barrier would be a lack of understanding of the market, which can lead to a lack of self-confidence in a consumer's ability to make correct decisions. A number of the policies proposed by Ofgem seek to improve consumers' understanding of the energy market, in part by making information currently provided by suppliers more accessible and in part by providing additional information to consumers in a user-friendly format. These proposals should increase the number of consumers that are able to understand both their own energy tariff and consumption, and also to review the market more broadly.

The policy measures that we consider primarily contribute towards improving understanding of the energy market are described below.

Improving the quality of regular communications

The proposals to improve the quality of suppliers' regular communications with consumers aim to ensure that consumers engage with key information and better understand their current tariff, consumption and other relevant circumstances (such as existing payment method and meter type). The affected communications include Price Increase Notices (PINs), Annual Statements and Bills.[3]

The PIN is a letter sent to each consumer when their contract is changing, whether due to an increase in price or another change which is to their disadvantage. It is designed to inform consumers of any disadvantageous change(s) to their tariff, how it will affect them and the steps they can take to manage the change. In light of evidence from a review of supplier practice, language experts and consumer research that found that PINs are not fulfilling these purposes, Ofgem has proposed new rules for what the PIN must contain and how the information is to be presented. This policy measure is designed to build understanding of adverse changes to tariffs amongst all affected consumers.

Annual Statements aim to facilitate quality engagement in the market by providing a consumer with key information about their current tariff and energy consumption. However, Ofgem considers that this communication has not worked as effectively as it could have done and so is proposing additional requirements which aim to ensure that the Annual Statement achieves its purpose. The measure aims to

[3] The End of Fixed Term Notice is discussed in Section 2.4.3.

improve the quality and accessibility of information provided on the document and to introduce some consistency between suppliers. This policy is designed to build understanding amongst all consumers.

Bills are the regular communication that consumers are most likely to notice and read (with the exception of consumers that are on fixed term contracts, who are most likely to notice and read the End of Fixed Term notice). This policy measure aims to provide all consumers with key information about their current tariff so as to build understanding. The information could be used if the consumer wishes to consider switching tariff.

It is important to note that while we have defined the key objective of the regular communications policy measures to be 'building understanding', each communication contains certain information that primarily aims to make it easier for consumers to compare tariffs and engage with the energy market. For example, the Cheapest Tariff Message (CTM), Tariff Comparison Rate (TCR) and Personal Projection (PP) will appear on regular communications and are designed to simplify tariff choices (and so are discussed in Section 2.4.3).

However, for the purposes of evaluation it is important to focus on the 'unique' elements of the communications so as to avoid confusing an analysis of the success of regular communications with the impacts of other policy measures. The unique aspect of the PINs, Annual Statements and Bills proposals is the improved clarity of the communications (which is achieved both by simplifying language and, in some cases, improving the layout of the communications). Therefore, the evaluation should focus on the extent to which the communications have become more accessible to consumers, notwithstanding the fact that they will contain other policy measures (and, indeed, will be a key means of communicating those other policy measures to consumers).

Tariff Information Label

The Tariff Information Label (TIL) is a new tool which aims to provide consumers with key tariff information in a format that is consistent across suppliers. The consistent format will make it easier for consumers to compare numerous features of different tariffs and so should improve the quality of consumers' engagement with the energy market.

The TIL will be provided whenever customers enter into a new contract or there is a change in their contracts. A version will also appear on Annual Statement and will be available on suppliers' websites. Evidence suggests that initially the TIL will have higher levels of uptake (and therefore associated impact) among consumers that are relatively engaged with the energy market.[4] The same evidence suggests that the TIL may be seen as too detailed for some currently disengaged consumers who are reluctant to spend time reviewing information relating to their energy arrangements.[5] This is likely to include vulnerable consumers who are disproportionately represented among consumers who are currently disengaged.[6]

Tariff Comparison Rates

Tariff Comparison Rates (TCRs) have been designed to allow 'at a glance' comparisons between tariffs. The TCR will be based on the consumption of a typical consumer and hence can only be used as a guide, in much the same way as APRs are used in the financial services sector. The TCR would be communicated to consumers through the regular communications of suppliers and may appear in best buy tables and advertising campaigns.

---

[4] SPA Future Thinking (2012), "Price Increase Notification Letters, Summary Box on Bills, Tariff Information Labels and Annual Statements", page 13

[5] SPA Future Thinking (2012), "Price Increase Notification Letters, Summary Box on Bills, Tariff Information Labels and Annual Statements"

[6] Ipsos MORI (2013), "Customer Engagement with the Energy Market - Tracking Survey 2013"

This policy measure aims to build awareness amongst consumers of the possibility of switching tariffs and the fact that savings are available through switching. Therefore, it is hoped that it will act as a prompt for customers to seek a Personal Projection. The policy measure is likely to affect primarily relatively disengaged consumers;[7] engaged consumers are generally already aware of the possibility of switching to save money on fuel bills and have strategies to facilitate such comparisons (e.g. using price comparison services).[8]

2.4.3 Simplifying tariff choices

It is hoped that consumers with an improved understanding of the market and their own energy circumstances will be more likely to have the confidence and inclination to review their tariff options. At this point, complex information and difficulties of comparison may discourage consumers from taking further action or could lead to consumers making sub-optimal choices relative to that which they would make if it were easier to compare options. Ofgem aims to make it easier for consumers to assess alternative tariffs, in part by simplifying the structure of tariffs and requiring suppliers to use a common methodology when providing quotations. By requiring suppliers to provide personalised tariff comparisons in their regular communications with consumers, Ofgem seeks to encourage greater engagement with the energy market, particularly for consumers that are particularly disengaged at present. This set of proposals should increase the number of consumers that are able to accurately assess their options, both with their current and with alternative suppliers. The policy measures that we consider primarily contribute towards simplifying tariff choices are described below.

Tariff cap

The tariff cap aims to limit the number of tariffs a consumer is faced with to allow them to assess their options more effectively. Suppliers will be required to offer no more than four core electricity and four core gas tariffs per meter type.

This policy measure is likely to have a greater impact on consumers that are currently relatively disengaged with the market because those that are relatively engaged have demonstrated that the number of tariffs in the market does not represent an absolute barrier to comparing their tariff options and / or switching tariff or supplier. This policy measure aims to address an important perception-based barrier to engagement that is present among currently less engaged consumers, i.e. that there are too many tariffs in the market for simple comparison to be possible.[9] Given that vulnerable consumers are disproportionately represented among less engaged consumers, this policy measure may address an important perception-based barrier among this group. However, it is not possible to predict how this will interact with additional resource-based barriers more vulnerable consumers may also encounter (e.g. no internet access, barriers to switching due to supplier debt / pre-payment meter status, etc.).

The tariff cap and tariff structure proposals (see below) are designed to work together, and reinforce each other, with the aim of making it easier for consumers to compare tariffs by changing the nature of the products offered by suppliers.

Tariff structures

This policy will require suppliers to simplify and standardise tariff structures in order to make it easier for consumers to compare tariffs. Suppliers will be required to recover all charges for a supply activity through a standing charge and unit rate. In addition, suppliers will only be able to offer certain non-contingent cash discounts. Product bundles will be required to comply with rules to limit the number and type on offer.

---

7 Among whom more vulnerable customers are disproportionately represented.
8 Ipsos MORI (2012), "Consumer views on Tariff Comparison Rates", Page 7
9 Ipsos MORI (2013), "Customer Engagement with the Energy Market - Tracking Survey 2013", page 40

Previous consumer research has identified that those who find it most difficult to compare tariffs with multiple unit rates (and are most likely to 'drop out' from attempting to engage in the market as a result) are not proactive consumers. Therefore, this policy measure aims to have a greater impact on consumers that are currently relatively disengaged with the market. Evidence suggests that vulnerable customers are more likely to struggle with complex pricing structures and tariff information than non-vulnerable customers and so this policy measure aims to improve outcomes for this consumer segment as well.[10]

Cheapest Tariff Message

The Cheapest Tariff Message (CTM) will require suppliers to provide personalised savings messages to consumers by comparing a consumer's current tariff to other tariffs offered by that supplier. It will feature on suppliers' regular communications, including Bills and Annual Statements, and aims to prompt engagement amongst consumers.

The policy is targeted at consumers that are relatively less engaged with the energy market (including vulnerable customers, who are more likely to have low levels of engagement in the market). Those that are relatively engaged are likely to prefer to consider tariffs offered by other suppliers in addition to those offered by their current supplier. For relatively engaged consumers, the CTM may instead act as reassurance that they are already getting a good deal.

Personal Projection

The Personal Projection (PP) will require suppliers to use a consistent methodology when providing estimates of annual costs to consumers. The PP will allow consumers to make an accurate comparison of the cost of different tariffs and so should improve the quality of switching.

Suppliers will be required to use the PP when speaking to or writing to a customer (or potential customer) about their tariff options. The PP will also appear on regular communications and the PP methodology may be used by switching services to estimate the annual cost of a tariff.

The PP is more likely to be used explicitly by customers who wish to compare tariffs offered by different suppliers. Therefore, initial uptake at least is likely to be among more engaged consumers who are willing to invest time in cross-market comparisons. It is to be expected that less engaged customers who are less willing to invest time in this way are more likely to use the CTM when considering switching tariffs. Consequently, while the methodology used to calculate the CTM is the same as that for the PP, the policy measures should be considered separately.

End of fixed term notice

The end of fixed term notice aims to ensure that consumers have the information they need to make an appropriate tariff choice at the end of their fixed term tariff. It also aims to prompt consumers to take action at the end of their fixed term tariff, or make an informed choice to default onto a different tariff offered by their current supplier.

Only those consumers who have at least switched once (and to a fixed term contract) will receive an end of fixed term notice and hence the policy measure will affect relatively engaged consumers. The 2013 Omnibus Survey shows that those consumers who receive an end of fixed term notice are more likely to engage with this communication than they are with any other communication from their supplier.[11]

[10] Creative Research (2011), "Tariff Comparability Models Volume 1 - Consumer qualitative research findings", page 10 and Ipsos MORI (2011), "Consumer reactions to varying tariff comparability", page 7

[11] Ipsos MORI (2013), "Customer Engagement with the Energy Market - Tracking Survey 2013", page 8

Dead notice

The dead notice aims to ensure that consumers on dead tariffs are informed of all RMR-related changes to a tariff that is no longer open to new customers. The dead notice will inform those consumers of the changes that will occur to their tariff and will provide a range of switching information (including the CTM and TCR) to encourage affected consumers to take action in response to the RMR-induced changes.

All consumers who are affected will receive one Dead Notice at the end of the process. It is to be expected that the majority of consumers that are on dead tariffs that will be affected by the RMR will be relatively disengaged and consequently are likely to include vulnerable customers.

3 Evaluating the RMR

3.1 Introduction

In considering options for an impact evaluation of the RMR, the following design principles have been adopted:

- The evaluation should enable Ofgem to infer judgment using explicit criteria. The criteria could be loosely defined around tracked progress and improvements or could be quantified based on expected levels of change against indicators.
- The evaluation should involve systematic data collection to build an evidence base for identifying the impact of the RMR.
- The evaluation should produce useful information for Ofgem and its stakeholders. As such, it is important to clarify and prioritise overall evaluation objectives (i.e. accountability, learning or other).
- Finally, the evaluation should encompass both formative and summative elements.[12]

This last point is discussed in more detail below.

3.1.1 Formative and summative evaluation

Formative elements are covered in process evaluations. Process evaluations help to answer descriptive or normative evaluation questions by looking at *why* and *how* the intervention worked or did not work and in what way. As such, a process evaluation will generate important learning to continuously drive improvements in the implementation of the RMR. It will also provide important evidence for the impact evaluation, since it will help evidence what critical factors contributed to a successful outcome or whether a failure can be attributed to a poor initial hypothesis (theory failure) or poor delivery (implementation failure). This is particularly critical in those areas where the implementation of the RMR policies is less prescriptive. In fact, the less prescriptive the intervention the greater the complexity associated with structuring the evaluation. The nature of such non-prescriptive interventions are often more complex to define and the outcomes expected are often further removed from the actual measure itself with little overall control of the environment exercised through the implementing organisation or individuals.

In the context of the RMR evaluation, policies such as SOC are less prescriptive. The impact evaluation would seek to address the core question of how SOC (the intervention) affect levels of trust (the expected outcome). To track the impact of the SOC, it would be particularly important to understand process-related issues of when and how suppliers have implemented SOC. In addition, because of the fact that

[12] The following provides a definition of formative and summative evaluation types: "*Formative evaluations* strengthen or improve the object being evaluated — they help form it by examining the delivery of the program or technology, the quality of its implementation, and the assessment of the organizational context, personnel, procedures, inputs, and so on. *Summative evaluations*, in contrast, examine the effects or outcomes of some object — they summarize it by describing what happens subsequent to delivery of the program or technology; assessing whether the object can be said to have caused the outcome; determining the overall impact of the causal factor beyond only the immediate target outcomes; and, estimating the relative costs associated with the object". Source: William M.K. Trochim, Cornell University, 2006: Research Methods Knowledge base

many other aspects influence trust, the impact assessment would need to cover — either in quantitative or qualitative ways — how other factors have influenced consumer levels of trust in suppliers over time. In this regard, the formative element of the RMR evaluation would be instrumental in developing knowledge about the implementation and qualitative assessments of how activities have contributed to outputs and outcomes over time.

Summative elements are usually covered in impact evaluations. An impact evaluation can be described as an "assessment of how the intervention being evaluated affects outcomes, whether these effects are intended or unintended.[13] It may also include a quantification of the economic impact of the intervention and the value-for-money delivered by the introduced changes. The most rigorous impact evaluations require quantitative attribution and a counterfactual of what those outcomes would have been in the absence of the intervention. However, in practice, it is not always possible to reach this level of rigour; the nature of the intervention, its context and the feasibility of data collection are some of the key considerations which need to be taken into account when choosing an evaluation design and determining the degree to which it is possible to attribute a cause or effect directly to that intervention.

The focus of this report is to consider the appropriate design for the impact evaluation of the RMR, given the specific methodological challenges related to the RMR policies, the context in which they are to be introduced and the feasibility of data collection. In terms of conducting an impact evaluation of the RMR, it will be necessary to conduct an end evaluation no later than 2017. However, it is important to design the evaluation now in order to establish the baseline and shape the on-going monitoring activities. We also recommend assessing the impact of the RMR in periodic intervals from the inception of the policies later this year to the delivery of the impact evaluation no later than 2017.

As mentioned above, a robust evaluation of the RMR would need to include both a formative (process) element and a summative (impact) element. Although the focus in this study is on the impact evaluation we provide some overall recommendations regarding the process evaluation in section 7.3 and highlight throughout this document where we see the need for the on-going monitoring of supplier implementation.

The rest of this chapter outlines the methodological considerations that affect the design of the RMR impact evaluation.

3.2 Choosing the Evaluation Design

The choice of evaluation design will have a bearing of the level of robustness, fidelity and the scope for generalising the results generated. We have made use of a decision tree in order to:

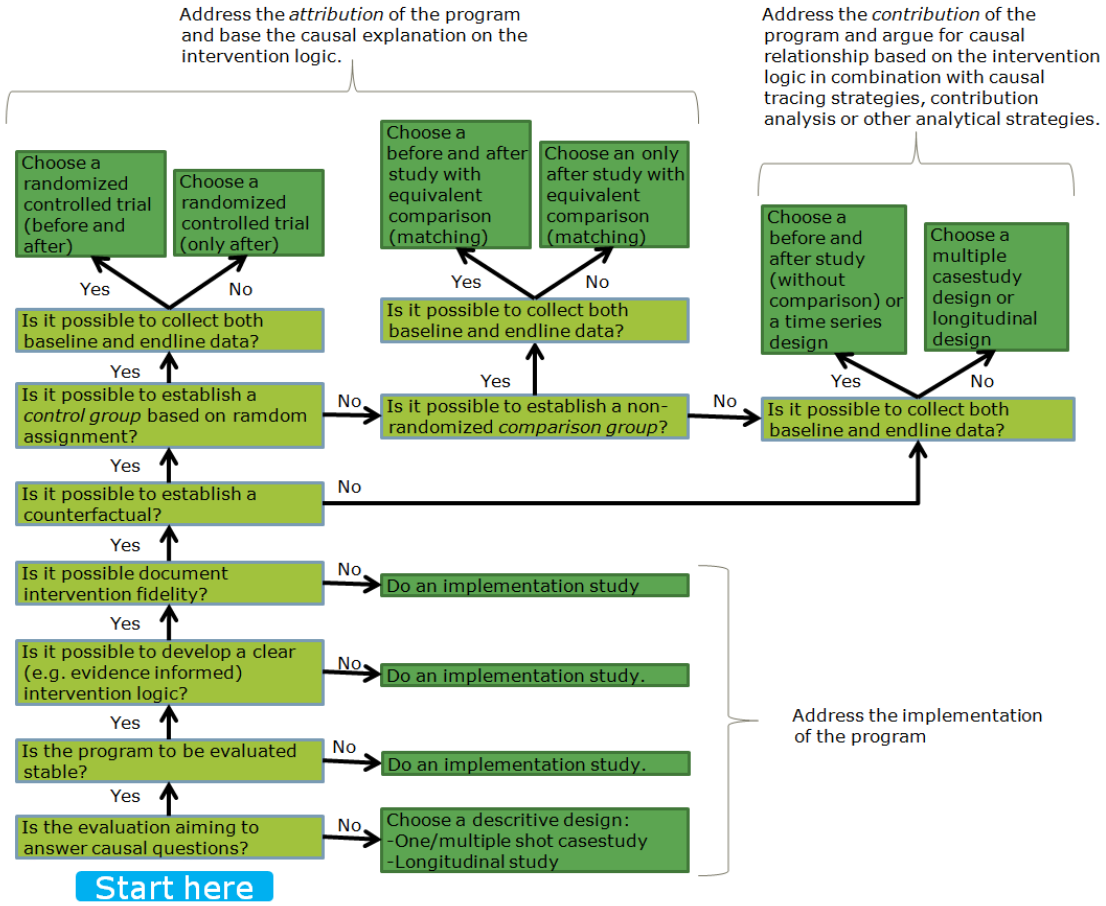
- guide the assessment of the methodological challenges and possibilities related to the evaluation of RMR; and
- consider the appropriate design for the RMR evaluation and the extent to which it is possible to address causal relationships in terms of attribution or contribution.

The figure below sets out the decision tree for impact evaluation design.[14]

[13] OECD: Outline of principles of impact evaluation (no publishing date provided)

[14] Framework adapted by Ramboll from impact assessment ("Effektmåling") Nielsen, C., Dinesen, P. T., Benjaminsen, L and Bonke, J, The Danish National Centre for Social Research).

Figure 3.1: Decision tree for impact evaluation design



3.2.1 Step one: Is the evaluation aiming to answer causal questions?

The first step in considering evaluation design options is to establish if the evaluation is aiming to answer causal questions. The answer here is clear: the purpose of the RMR impact evaluation is to answer, as robustly as possible, causal questions (see table below).

Evaluation questions set out what is expected from an evaluation and as such provide the focus for framing an appropriate design for the evaluation. The table below sets out three basic types of evaluation questions and provides some specific examples related to the RMR.

Table 3.1: Evaluation questions

	Descriptive questions	Normative questions	Causal questions
The focus of the questions	Understanding and documenting intervention implementation: this may describe the types and quantities of services delivered, the beneficiaries of those services, the resources used to deliver the services, the practical problems encountered, and the ways such problems were resolved.	To what extent have programmes been implemented and delivered in the way that they were meant to?	Identifying and exploring cause-and-effect relationships, including: • The strength of the effect: what are the intended/unintended outcomes of the intervention • The type of relation: did the intervention cause or contribute to the observed effect? • The nature of the cause-and-effect relationship: in what way, for whom, and under what circumstances did the intervention work?
Example questions for RMR policies	Did suppliers' change licence conditions, processes and practices with regard to customer interactions since the implementation of the SOC?	To what extent have changes been in line with SOC provisions?	To what extent did the SOC contribute to enhancing consumers' trust in their own energy suppliers?

Evaluation questions are usually operationalized into sub-questions, from which indicators can be derived. Indicators provide information on the type of data that needs to be collected and how it should be analysed. This helps to determine the specific evaluation design, including decisions on how the evaluation will cover the RMR policies at an individual level and as a set of interlinked policies.

Though the core focus of the impact evaluation is on the causal questions, a comprehensive evaluation design of the RMR should also encompass descriptive and normative questions in order not only to understand how and the extent to which policies have been implemented, but also to put findings related to the causal questions into context.

As shown in the decision tree above, a descriptive design should be chosen if the evaluators only seek to answer descriptive and normative questions. While it will be critical to understand how the RMR policies were interpreted and implemented by suppliers and which groups benefited from this (i.e. descriptive questions) and the extent to which the RMR polices have been implemented as envisaged (normative questions), it is important to move beyond this to also explore cause-and-effect relationships (causal questions).

3.2.2 Step two: Is the program to be evaluated stable?

The next step is to consider the extent to which the RMR program is stable. While we assume that the RMR program will remain stable, it is important for the Ofgem evaluation team to reassess the indicators suggested in this report (see 4.2, 5.2 and 6.2) when the proposals are fully finalised. Once launched, it will also be important to follow implementation closely and re-assess evaluation strategies if there are any changes made during the course of 2014-2017. In case of substantial changes in one or more of the individual policies, it might be necessary to isolate the policy in question from the overall impact evaluation and undertake a separate implementation study focused on that particular policy measure.

3.2.3 Step three: Is it possible to develop a clear intervention logic?

The RMR policies are underpinned by a large body of research and reflect current knowledge of how consumers interact with the energy market and the drivers for improving consumer trust, understanding

and simplifying tariff choices. Working closely with Ofgem clear intervention logics were developed during this project for each of the individual policies under RMR. These were set out as theories of change and can be found, in summarised form, in Table 4.1, Table 5.1 and Table 6.1.

While it has been possible to develop clear intervention logics for each policy measure, doing so has been far from trivial. There are a number of reasons for this, which all pertain to the design of an impact and process evaluation:

- Complexity of the policy instruments**:** Taken together, the RMR policies seek to tackle heterogeneous issues by targeting a diverse set of consumers through a variety of interventions. These aim to achieve a wide range of objectives some of which may support, reinforce or even hinder one another. Examples include:
  - Support: SOCs support the majority of other policy measures by ensuring that consumers will have a better experience when interacting with their energy suppliers.
  - Reinforce: the tariff cap and tariff structures proposals reinforce each other because both contribute towards a reduction in the complexity of comparing tariffs.
  - Hinder: the proposals to build trust (e.g. SOC) could lead to a greater increase in consumers' trust of their own supplier than in trust in other suppliers. Therefore, it could reduce the likelihood that a consumer would consider switching supplier and hence hinder policies that aim to increase engagement with the market and prompt switching.
- Complexity of market dynamics**:** How and why consumers interact (or do not interact) with their energy consumption and tariff is a particularly complex question. This has been shown extensively in the research and evidence reviews undertaken by Ofgem and other agencies.[15] The factors that shape energy literacy, decision-making around energy consumption behaviour and market engagement are influenced by many systemic factors (e.g. demographic, institutional, economic, social and cultural) as well as idiosyncratic ones (e.g. knowledge, values, attitudes and emotions). Whilst the RMR policies are not explicitly intended to address wider energy literacy and pro-active environmental behaviours, they build on and interlink with other policies that seek to address these broader issues. As such, the impact of the RMR policies will be affected by external and internal factors such as the introduction of other policies, and general developments in both consumer behaviour and in the market. It follows that the RMR evaluation will need not only to encompass the indicators related to the theories of change in the narrow sense but also identify alternative explanatory variables (such as energy prices or media coverage).

Given this complexity, the causal effects set out in the intervention logic might not always materialise due to inhibitors, drivers or contextual factors that are not under the control of the policy maker. For example, given the current levels of disenchantment of consumers with the energy market, even positive experiences with the current or new supplier may not change perceptions of the overall market. There is also the possibility that increased trust or understanding may lead to more complaints as consumers demand better/fairer treatment.

Similarly, there are many instances where the policies are only one of many triggers to facilitate the achievement of the intended outcomes, and observed changes (both at output and outcome levels) are likely to have additional explanatory factors other than the changes associated with the policies in this area. In terms of improved understanding, for example, there are means through which consumers'

[15] See, for example, Ofgem, (2011), "What can behavioural economics say about GB energy consumers?"

understanding of the energy market may be improved other than through written communications (e.g. word of mouth,[16] customer service staff taking the time to explain concepts verbally, etc.).

Therefore, it will be important to contextualise quantitative data, such as the number of complaints, with qualitative information about the nature of the intervention and the way in which the policy was implemented and developed. This means on-going qualitative feedback is needed from consumers and suppliers alike to be better able to map and explain the full range of factors underpinning the outcome and output variables. Some form of contribution analysis would then be possible.

To summarise, although the policy instruments are complex and nested in an open system where other factors are likely to influence the achievement of the goals, it has been possible to map out clear intervention logics for the policies and to identify alternative logics at play in the customer journey. This enables evaluation to move beyond an implementation study only.

3.2.4 Step four: Is it possible to document intervention fidelity?

Intervention fidelity may be defined as the extent to which the delivery of an intervention adheres to the protocol or program model originally developed.[17] This necessitates, to a degree, prescribed implementation models in the policy interventions. If the RMR allows for too much variation in how suppliers interpret and implement the changes — and how consumers experience these — it will challenge the assessment of policy impact and comparisons across different groups.

While it is possible to document intervention fidelity for most of the RMR policies, there are some exceptions. In relation to the SOC, for example, the open-ended implementation approach means that standards may be implemented to varying degrees. This could mean that very little change is experienced with certain suppliers, and could hinder like-for-like comparisons. In these areas it is therefore particularly critical for evaluators to: track progress closely; identify the range of ways in which suppliers interpret and choose to implement such non-prescriptive policies; and evaluate impact in a way which allows for a comparison of effectiveness across different supplier approaches to implementation.

3.2.5 Step five: Is it possible to develop a counterfactual?

To identify the impact of a policy robustly, a counterfactual should be specified. The counterfactual measures how the variable subject to policy intervention would have evolved in the absence of such intervention. In a laboratory setting, the trial conditions are controlled such that any difference between the control and treatment groups is a change against the counterfactual. In non-laboratory settings, the counterfactual needs to be defined before policy impacts can be measured in an unobjectionable manner.

In general, an appropriate counterfactual is a pre-requisite for monetising the costs and benefits of policy interventions. If a counterfactual cannot be defined, alternative evaluation approaches must be employed.

In constructing the counterfactual, it should not be assumed that the 'do nothing' / 'do minimum' option implies a continuation of the status quo, although this may turn out to be the appropriate definition in some cases. This is because various developments may take place even with no changes in this area of policy. In the context of the energy retail market such developments may include the increasing importance of collective switching schemes, the introduction of smart meters and associated time-of-use tariffs and so on.

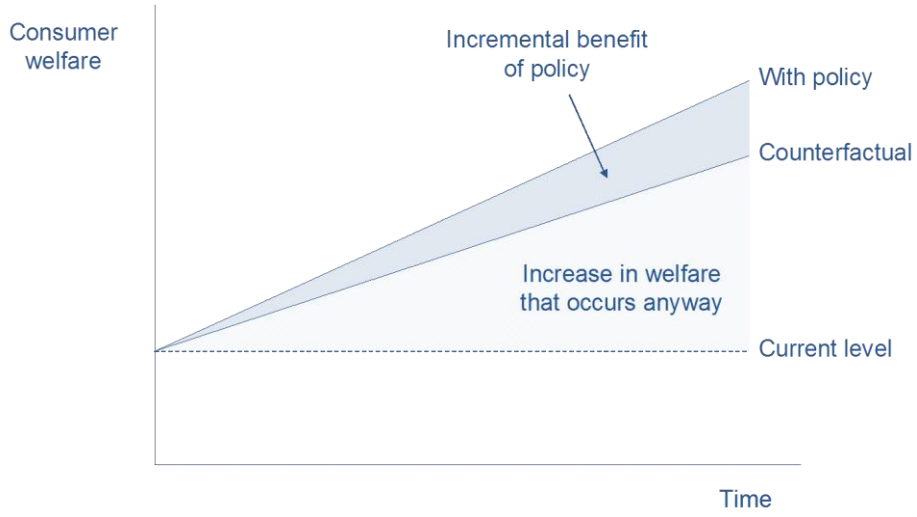
The importance of defining the counterfactual is illustrated by Figure 3.2, which takes the example of a policy aimed at increasing the welfare of consumers in a particular market. The diagram shows that

[16] As indicated in the Ipsos MORI report "Consumer engagement with the energy market, information needs and perceptions of Ofgem, Findings from the Ofgem Consumer First Panel Year 4: second workshops (held in March 2012)", 30.08.2012.

[17] American journal of evaluation, 2003: Fidelity Criteria: Development, Measurement, and Validation

consumer welfare is expected to increase through time under the counterfactual scenario. If the impact evaluation were to ignore this and assume that any increase in consumer welfare above its current level was attributable to the policy, then the benefits of the policy would be significantly over-estimated.

Figure 3.2: The Importance of Defining a Counterfactual



Note: This diagram assumes that it is possible to identify a single start date

The RMR presents numerous difficulties for defining the counterfactual. Some of the reasons for this relate to the inherent methodological challenges discussed above. The difficulty of defining the counterfactual is increasing in the number of factors that influence the variable(s) of interest and the generality of a policy measure. It is also difficult to define a single counterfactual if a single policy intervention date cannot be defined.

A further complication relates to the fact that many of the RMR policy measures have similar objectives and there is some overlap between policy options. While this is one of the strengths of the RMR, it does mean that it is not feasible to define a robust counterfactual for individual policy measures: doing so would require the extrapolation on pre-RMR trends in addition to estimates of the relative impact of related RMR policies, discounting the interaction effect amongst policy measures. With regards to the latter, we consider that far too many assumptions would be required to estimate the impacts of other RMR policies for a counterfactual for specific policy measures to be credible.

Extrapolating previous trends may be feasible in theory but, in practice, defining this type of counterfactual may prove difficult because of the significant amount of change that is currently occurring in the energy market, some of which has been influenced by the RMR. The counterfactual typically takes the decisions of other parties to be fixed and not in the control of the policymakers. Since the RMR was announced, however, DECC and other stakeholders have taken actions that have in some ways built on the work undertaken by Ofgem during the RMR (e.g. see DECC's consultation document entitled "Ensuring a better deal for energy consumers").[18] These proposals would arguably not have arisen in the absence of the RMR, although something similar may have been proposed. Unfortunately, a formal counterfactual could not be defined without considering how other stakeholders would have acted in the absence of the RMR.[19]

Given these challenges, we believe that it would not be possible, in a robust manner, to attempt to predict the extent to which the behaviours of suppliers and consumers would have differed in the absence of the

[18] DECC (2012), "Ensuring a better deal for energy consumers"

[19] In principle, Ofgem could define theoretical counterfactuals based on, for example, an assumption that DECC would have taken no policy action in the absence of the RMR. However, there would be little evidence to support such an assertion and the validity of defining a theoretical counterfactual for evaluation purposes is questionable.

RMR and, importantly, how this would affect outcomes of trust, understanding, and simplifying tariff choices.

On this basis, we consider that it would not be possible for a future evaluation to robustly specify the state of the world in the absence of the RMR. Therefore, we consider that evaluation options that rely on a robust definition of the counterfactual are inappropriate for evaluating the RMR.

According to the decision tree above, this points to the overall conclusion that the impact evaluation of RMR would not be able to address rigorously the attribution of the program and base the causal explanation on the intervention logic explicitly. Instead, the evaluation would seek different means of addressing the contribution of the program and would argue for causal relationships between interventions, outputs and outcomes of the RMR based on the intervention logics in combination with causal tracing strategies through forms of multivariate analysis and time series design. The sections below investigate this in more detail. Before doing so, however, it is useful to consider the final two steps in the decision tree: the possibilities of developing a control group (step six) and the feasibilities of collecting both baseline and "end-line" data (step seven).

3.2.6 Step six: Is it possible to develop a control group

The RMR policies are, as a whole, universal in nature, i.e. the policies will be implemented for the benefit of all domestic consumers. While there are specific policies which are primarily targeted at sub-groups (e.g. fixed-term tariff customers, or currently 'disengaged' consumers), the majority of these policies — for the purposes of the evaluation — are taken to cover all consumers.

In order to identify a randomised or non-randomised control group, the suppliers would need to treat a random subset or several subsets of customers differently. Given the statutory requirements of the policies, this would not allow suppliers to allocate consumers randomly into compliant and non-compliant groups and/or to stagger compliance. Even if this were possible, it is unlikely to be supported through existing administrative and staff procedures or IT-systems (e.g. for accounting, monitoring, Client Relationship Management). It would also be very difficult to blind the treated and non-treated populations, and such piloting would require the cooperation of suppliers who may have conflicting interests.

As it is not possible to establish a randomised or non-randomised control group for the policy area, it is not possible to meet the requirements of higher attribution methodologies such as Randomised Control Trials or Matching. Sections 3.3.1 and 3.3.2 below discuss this in more detail.

3.2.7 Step seven: Is it possible to develop both baseline and end-line data?

The final step in the decision tree requires an exploration of the possibility of developing both baseline and end-line data. As mentioned above, Ofgem has developed the RMR proposals though a highly consultative process. The RMR was first announced in autumn 2010 and has been subject to numerous rounds of consultation. (The statutory consultation on most of Ofgem's domestic RMR proposals has recently closed. The statutory consultation on Ofgem's final SOC proposals was conducted earlier than for the other policy measures.) Given the openness of Ofgem's process — and the fact that some suppliers may already have implemented changes to policies and procedures in anticipation of the introduction of RMR — it will be important to set the baseline earlier than the actual implementation and track impacts of the policy over time against several milestones.

We therefore suggest that the monitoring should include several baselines covering retrospectively the time from when we assume that the RMR would have begun to have effects:

- from the date on which the Review was announced (autumn 2010);
- from the date on which Ofgem's initial detailed proposals were announced (December 2011);

- from the date on which Ofgem's updated proposals were announced (October 2012); and
- from the date — or dates — of the actual implementation of specific policy measures.

With regards to the latter point, there are particular policies which have a longer lead-time before compliance with the post-RMR Licence conditions is required. While some of the policy measures — such as the SOC — will be implemented from late summer 2013, the majority will come into force in late 2013 (or March 2014 in the case of policies such as Bills, Annual Statement, Price Increase Notification, End of Fixed Term Notice, CTM and Personal Projection). The difference in lead times between policy measures means that different policies in the RMR package will not necessarily be comparable in terms of maturity of implementation at any given time.

Considering the need to build a baseline from 2010 onwards, the existing monitoring data — in particular the Ipsos MORI CAPI omnibus survey and the GfK Energy Satisfaction Monitor — provide a relatively good baseline for many of the indicators that we propose (see presentation of these in 4.2, 5.2 and 6.2). There are, however, some indicators for which a baseline dating back to autumn 2010 is not available. For example, there is no baseline in place for consumers' confidence in the accuracy of information provided by suppliers (indicator 4 below).

Where existing baseline information is not available it will need to be gathered as soon as possible, either through the existing monitoring mechanisms or through a tailored tracker survey (see section 3.3.3 below for more information about such a survey). However, in developing this baseline prior to implementation it is important to bear in mind that the RMR may have had impacts on the energy market between 2010-2013 and any such impacts might already be partly reflected in that baseline.

Importantly, many of the indicators of the success or failure of RMR policies rely on consumers' (and to some extent suppliers') self-assessment (e.g. consumers' confidence that they are getting a fair deal or consumers' awareness of different tariff types and suppliers). There are significant frame of reference and measurement biases related to self-assessment of this nature as it challenges the ability to measure consistently across different groups and over time. To counter-balance this, it is important to measure for alternative explanatory factors and to complement quantitative monitoring activities with data from suppliers and qualitative research.

Overall, however, we consider it to be possible to collect both baseline and end-line data. This means that it would be possible to identify the contribution of the program through various forms of time series analysis. As such, the most rigorous methodology for the RMR impact evaluation should address the contribution, rather than attribution, of the program and explore cause-and-effects through multivariate analyses coupled with qualitative assessment of the interrelationships. Section 3.3 assesses the options for impact evaluation design, based on the methodological considerations above.

3.3 Assessing Options for the Impact Evaluation Design

3.3.1 Randomised Control Trials

The highest level of attribution analysis is achieved through Randomised Control Trials (RCTs). As described by a recent report by the Cabinet Office[20] "*[RCTs] are the best way of determining whether a policy is working…What makes RCTs different from other types of evaluation is the introduction of a randomly assigned control group, which enables you to compare the effectiveness of a new intervention against what would have happened if you had changed nothing*".

[20] Cabinet Office, 2012: Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials

The introduction of a control group eliminates many of the problems related to establishing counterfactuals and reducing biases which generally challenge the evaluation process and analytical process. RCTs would, if feasible, be able to answer with the very highest level of attribution the core question of the extent to which the RMR policies have delivered the outcomes expected. As a method it would also have helped to establish the counterfactual(s) with a high level of certainty and provide answers to questions such as: what would happen to levels of consumer understanding of the energy market if the RMR had not been introduced? Comparing the outcome under the RMR (i.e. the treatment group) with the change that would have occurred in the absence of the RMR (i.e. the control group) would show the impact of the RMR on consumers' understanding of the energy market.

However, given the challenges discussed above — in particular the fact that it is not possible to establish a control group based on random assignment — it is not possible for the evaluation of RMR package, or any individual policies within the RMR, to reach this "gold-standard" of RCT impact evaluation.

3.3.2 Matching

Matching is a method of assessing the contribution of a program. This is done by exploring behaviours, attitudes and outcomes related to two non-randomised groups (a treatment and a control group) and comparing the effects in these while ensuring that differences in the characteristics of the two groups are accounted for. This is achieved by ensuring observable characteristics in the two groups are the same prior to the treatment. The two groups are therefore comparable and the evaluation is able to isolate the effect from the intervention and to achieve a more efficient estimate.

As set out above it is not possible to establish a control group based on a random assignment. For similar reasons, it is not possible to establish a non-randomised comparison group either: the RMR policies are universal in nature and cover the whole population, without any pilot phase in the implementation. Because of this we can rule out matching as a possible means of evaluating the RMR.

3.3.3 Time series or before-and-after studies with multivariate analysis

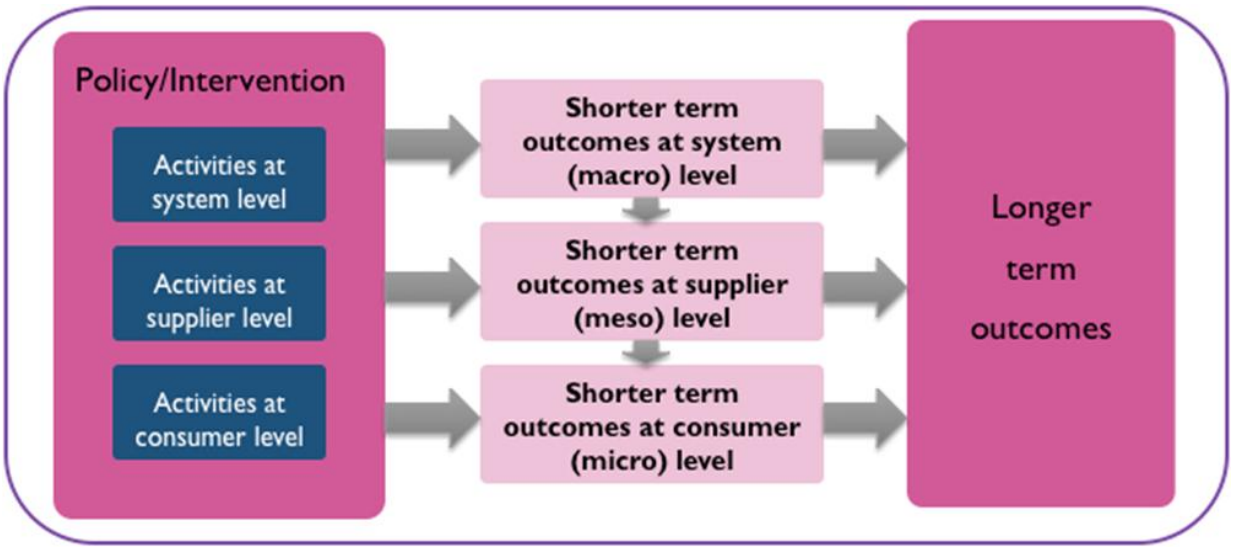
While it is not possible to adopt attribution methodologies of RCTs and matching, it is possible to collect both baseline and end-line data which enables before and after studies or, better still, time series design.

A time series study allows for multivariate analysis to be applied in order to explore correlations between output and outcome variables over time, provided that units can be compared and other factors controlled for. Units of comparison could be:

- area-based (e.g. regions);
- group-based (e.g. sub-group of the population);
- institution-based (e.g. type of supplier); or
- individual-based (e.g. segments of consumers).

The following figure illustrates how different RMR policies are targeted to influence units at different levels from individual consumers (micro-level), suppliers (meso-level) and the overall market (macro-level). In order to enable multivariate analysis on the time series data, it is important that monitoring covers all three of these levels.

Figure 3.3: Levels at which potential outcomes may be generated



Monitoring activities at all these levels will require a comprehensive set of monitoring data. As noted, there is a relatively good baseline against many of the indicators that we propose for an evaluation of the RMR policies. However, we recommend supplementing the existing monitoring data with a tailored tracker survey to enable more sophisticated time series analysis. There are three key reasons for this:

- The GfK Energy Satisfaction Monitor surveys the same panel over time. While this provides interesting cohort tracking possibilities, the survey itself introduces a bias by the fact that repeatedly asking questions about consumers' behaviour in the retail energy market may have an impact on respondents' behaviour.
- The baseline data do not cover all the indicators. While it would, of course, be worth exploring the possibility of adding specific questions to the existing surveys, it may be difficult to include all the dimensions needed for a time series analysis which enables sophisticated sub-group analysis and multivariate analysis.
- The baseline data are drawn from a number of different sources and are collected with different levels of frequency. This could pose significant challenges in the time lag before data can be accessed. While it might be possible to overcome this by altering the frequency of monitoring, it might be more appropriate to develop a tailored survey instrument instead.

The tailored tracker survey of consumers would, in its most rigorous form, contain a comprehensive set of questions relating to background characteristics, behavioural traits and attitudes. The questions would relate both to the indicators suggested across the whole of the RMR and alternative explanatory factors such as awareness of negative media coverage or wholesale energy costs. We suggest that the tracker survey should cover the same questions over time, but with a different sample in order to avoid the possibility of respondents changing their behaviour as a result of participation in the survey.

Repeating the survey over time would address the issue of time-lags in implementation of the policy and in consumers' response to it. Importantly, the survey would also enable further segmentation analysis, where results could be used to develop clusters of consumers based not just on background characteristics, but on their past switching behaviour and attitudes. It is particularly important to gain a sophisticated understanding of different consumer groups above and beyond vulnerable, engaged and disengaged. Existing consumer research by Ofgem and others suggests that there are, within each of these categories, different characteristics and drivers for levels of engagement, disengagement and vulnerabilities: these different drivers will have a bearing on the impact of the RMR policies on these different consumer groups.

For example, 'disengaged' consumers might include both those who have particular challenges in understanding the energy market and those who are not interested in engaging with household bills and are less inclined to manage household finances tightly. The first group are naturally more likely to be influenced by policies regarding bills, whilst the latter group will likely be less receptive to the RMR policies and may well report lower levels of impact overall.

Cluster analysis based on survey questions on both the core indicators and other background characteristics, behavioural and attitudinal indicators would help Ofgem to examine in greater depth the developments in cluster sizes over time *and* different levels of impact for different consumer groups. Additional indicators such as 'general time spent reviewing household bills every month' or 'tendency to 'shop-around' for the best deal when purchasing goods and services' might be useful proxy variables to include in a survey in order to segment consumer groups by means of cluster analysis and examine other explanatory factors by means of multivariate analysis.

It is important to note that a survey of this nature would require a large sample size and thus would require significant resources. Therefore it might be necessary to consider cheaper alternatives such as a one-off tailor-made consumer survey with a comprehensive set of independent variables or a repeated tailor-made consumer survey with only a limited set of independent variables. The sample size would depend on a number of factors, though the level of sub-group analysis required is the main issue to consider. We suggest that the survey should include a minimum of 6000 respondents, though it might be necessary to consider increasing this number to allow for further sub-group analysis which stays within acceptable confidence intervals for margin for error.

3.3.4 Descriptive analysis

The form of correlation analysis set out in the previous section inherently has a high level of bias, especially if it is deemed infeasible to complement the existing baseline data with a comprehensive tailored tracker survey. If the tailor-made tracker survey does not go ahead, we consider that only an analytical strategy of descriptive analysis only can be applied to the RMR. This approach would not in itself address the issue of the contribution of RMR policies to the achievement of outcomes, but instead would monitor and describe developments in the output and outcome indicators over time.

In sections 3.4, 4.4 and 5.4 below, we suggest that descriptive analysis could be adopted as the least resource intensive (but also least rigorous) option for Ofgem to consider. This option would primarily involve the use of existing data sets for the descriptive analysis, with only a few modifications. However, if this option is chosen, it would be advisable to complement the existing monitoring data with qualitative evidence which, along with the theories of change, can be used to argue for causal relationships. For instance, quantitative surveys might help identify general levels of trust, but qualitative methods such as interviews or focus groups would help test the intervention logics and explore the reasons behind reported levels of trust.

3.3.5 Working towards an evaluation design

In summary, the decision tree for impact evaluation design is a useful tool for guiding the assessment of the feasibility of different evaluation approaches for the RMR. While it is not possible to attribute changes to the RMR through means of RCTs and Matching, it is possible to adopt a timeline series design without assessment of the counterfactual but with application of multivariate analysis. This approach would, however, need to be complemented with qualitative consumer research, descriptive supplier data and independent assessments in a mixed mode evaluation approach, which would enable triangulation and qualitative analysis of the causal relationships. There are of course variations across each of the policy areas and no one method or approach can be applied uniformly.

In the following three chapters, we consider each group of policies in turn and present the overarching theories of change developed in workshops with Ofgem staff. In those workshops, and through subsequent work by Ofgem staff, a large number of possible output and outcome indicators were considered and put forward. Following a critical review, we have selected a smaller set of indicators which, taken together, pass the so-called CREAM test of establishing relevant indicators:

- **C**lear: is the indicator clear?
- **R**elevant: is the indicator relevant to measure?
- **E**conomic: does it warrant the effort to measure the indicator?
- **A**dequate: does the indicator cover the different areas?
- **M**onitorable: is the indicator possible to monitor?

It is important to note that the indicators suggested do not define the evaluation norms[21] or performance criteria against which it will be possible to assess success or failure of the RMR. Such criteria are not just important for the impact evaluation but should also guide the process evaluation.

[21] Evaluation norms can be cut-off points or intervals which delineate success from failure for the programme. There is, for many of the indicators suggested below, a baseline which can help shape the evaluation norms. However, given the nature of the indicators suggested below, it may be advisable for Ofgem to define evaluation norms against which success or failure could be judged. Given the complex and market sensitive nature of the proposals Ofgem may wish to engage with stakeholder (consumer advocacy groups and suppliers) in validating the indicators and developing the norms.

4 Building Trust in Suppliers and the Energy Market

4.1 Policy Intent

According to recent research, several factors contribute to consumers' sense of trust (or distrust) in the energy market.[22] Prominent influences on trust include: perceptions of excess profit-making by energy suppliers, perceptions of tariff complexity, perceptions of ease of comparing tariffs and past experience of poor service from an energy supplier. In other words trust is not only the result of actual experience at the individual level but is also affected by overall perceptions of the market that may or may not be entirely separate from own experience.

The RMR package contains a number of proposals that are designed to build trust in suppliers and the energy market, largely by requiring improvements in suppliers' interactions with their customers. Trust will be influenced by several other RMR policies but the impact of those policies on trust would be indirect; the policies listed here, however, are directly focused on building trust.

4.1.1 Theories of change

As discussed in Section 2.4.1, we have grouped policy measures that share the primary objective of building trust in the energy market. These are:

- establishing binding and enforceable SOC and communicating these to consumers;
- banning dead tariffs; and
- expanding fixed term and price protection rules.

In Table 4.1, we summarise the theories of change that Ofgem has developed for each of these policy measures.

[22] Insight Exchange (2012), "Consumer research and collaborative engagement on the proposed Standards of Conduct – Domestic Customers"

Table 4.1: Summary of overarching theories of change for policy measures aimed at building trust

Policy measure	Policy detail	Activities		Intended outcomes		Target population
		Supplier	*Consumers*	*Direct*	*Indirect*	
Enforceable SOC and communication of SOC to consumers	Enforceable SOC applicable to all interactions between suppliers (incl. representatives) and consumers; requirement to communicate SOC to consumers	Amends policies and processes to align with SOC; communicates SOC to consumers	Demands fairer treatment by suppliers at various stages of the consumer journey	Improved supplier conduct and customer focus	Increased consumer trust	All consumers
Ban on dead tariffs	Transfers consumers off expensive dead tariffs; provides greater transparency and comparability for consumers that remain on dead tariffs	Do not add new dead tariffs; make all dead tariffs RMR compliant; annual check for customers still on dead tariffs	Affected consumers transfer off expensive dead tariffs	Better protection for consumers on dead tariffs; financial benefit to consumers moved off expensive dead tariffs	Increased consumer trust	Consumers on dead tariffs (likely to be disengaged and potentially vulnerable)
Fixed term and price protection rules	Reinforce consumer protections to make it easier and fairer for consumers who switch to a fixed term tariff	Ensure that consumers have the information they need to make an appropriate choice at the end of their fixed term tariff	May consider the merits of a fixed term tariff due to increased protection for consumers on these tariffs	Fairer treatment of consumers by suppliers	Increased consumer trust	Consumers on fixed term tariffs (likely to be relatively engaged)

4.1.2 Outside influences and unintended consequences

Table 4.1 focuses on how the policy measures are intended to work. There are, however, factors that could influence the impact of the RMR proposals and lead to unintended consequences and/or the breakdown in the theories of change described above: either effect would cause the positive impact of the policies to be lower than expected, or even wholly negated.

With respect to the RMR proposals that seek to build trust, adverse media coverage of the sector could affect consumers' beliefs concerning the trustworthiness of suppliers. Such an outside influence could affect the observed increase in trust following RMR implementation.[23] Similarly, it is conceivable that the proposals that seek to build trust in suppliers and the energy market could lead to a reduction in the

[23] In some cases, 'outside' influences may have been driven by the RMR and in such cases the influence should be seen as an unintended consequence (either positive or negative). For example, increased media coverage of Ofgem's enforcement action against suppliers due to their being in breach of the SOC would be an unintended positive consequence of the RMR rather than a pure outside influence.

proportion of consumers that would consider switching supplier, and so to a reduction in competition between suppliers. This unintended consequence would arise if, for example, the impact of the SOC on consumers' trust in their own supplier exceeds the impact on their trust in other suppliers. Such an effect could increase the required financial saving that would be the tipping point for consumers to switch suppliers and, all else being equal, would reduce the number that do so.

It would be important for a future evaluation to consider the extent to which outside influences have affected the operation of the theories of change following the implementation of the RMR. In Section 4.4 we provide some examples of alternative explanatory variables which could be added to consumer surveys as well as qualitative monitoring activities that can explore alternative logics at play.

4.2 Indicators

4.2.1 Outcome indicators

The overall policy objective to increase trust is reflected in the following outcome indicators:

1. Consumers' overall satisfaction with the service received from their energy supplier.
2. Consumers' trust in their own energy supplier: the extent to which consumers trust their own energy supplier.
3. Consumers' trust in energy suppliers overall: the extent to which consumers trust energy suppliers overall.
4. Consumers' confidence that the information provided by their own and others energy suppliers is accurate.

The responses to the outcome indicators are likely to be dependent on type of interaction. The output indicators suggested below will provide data which will enable Ofgem to investigate patterns and relationships across output and outcome variables. Ofgem would also be able to assess correlations between the different outcome indicators themselves.

For example, while satisfaction is not equivalent to trust, per se, we consider that it is reasonable to assume that trust is linked with satisfaction which, in turn, is affected by the behaviour of suppliers. Measuring satisfaction across all consumers would be useful over time as it would provide an indication of whether trust in suppliers might be increasing. However, in order to explore underlying reasons for any changes in trust it will be important for a process evaluation to undertake qualitative research to capture actual consumer experiences and explore factors that influence levels of trust (or its absence).

4.2.2 Output indicators

A key element of trust relates to customers' experiences when interacting with suppliers and so there is a need to reflect this in the output indicators. However, consumers may not have had any direct recent interaction with their supplier absent a specific issue or problem. It is therefore relevant to focus on instances of actual interaction and the perception of that experience including:

5. Number of consumers on dead and fixed term tariffs: Total numbers of consumers on dead tariffs; total number of consumers on fixed term tariffs.

6. Consumers' interaction with customer service: Total number of consumer interaction with customer service per year/month, broken down by suppliers, reason for interaction and type of interaction; Response times to phone/email communications.[24]
7. Consumers' complaints: total number of complaints per year/month, broken down by suppliers and type of complaint.
8. Consumers' level of satisfaction with interactions with their suppliers (excluding complaints): level of satisfaction with regard to information received, how the request was handled and the timeliness of response.
9. Consumers' level of satisfaction with treatment of complaints by their supplier: level of satisfaction with regard to the information received; level of satisfaction with regard to the timeliness of treatment; level of satisfaction with outcome/resolution.
10. Suppliers' introduction and implementation of new policies: Suppliers' self-reported changes in licence conditions and changes in processes and practices according to SOC.[25]
11. Consumers' awareness of SOC: Proportion of consumers reporting to be aware of the SOC.

In addition to the indicators specified above, it will be important to understand more about the experiences of consumers on dead and fixed term tariffs. We suggest that this could be monitored through sub-group analysis of some of the indicators suggested in this chapter and in subsequent chapters (i.e. broken down by types of tariffs the consumers are on non-fixed term, dead and fixed-term tariffs). This approach would, for example, enable monitoring of:

- fixed-term tariff consumers' understanding of the key features of the tariff they are on (link with indicator 14);
- the extent to which fixed term tariff consumers at the end of their tariff understand the options available to them (link with indicator 15);
- the extent to which fixed term tariff consumers at the end of their tariff make an active switching decision following tariff end (link with indicator 25); and
- the extent to which those who remain on dead tariff feel able to compare their tariff with live ones (link with indicator 24).

4.3 Defining the Evidence Base

4.3.1 Existing monitoring activities

Ofgem has previously gathered some data that can serve as a baseline against which the impact of the RMR could potentially be evaluated, given the output and outcome indicators that are identified above. These are presented in the following table.

[24] With regards to process indicators such as telephone/email response times, it is necessary to investigate the extent to which suppliers hold this data in a form which evaluators are able to collate and analyse, in practice, it may not be feasible to include.

[25] As understanding develops over time, it will likely generate additional output indicators that may be relevant to monitor.

Table 4.2: Baseline data

Indicator	Baseline question / data item	Source of baseline data	Collection frequency	Sample size	Key sample segmentations
Consumers' satisfaction with their energy supplier (1)	Numerous questions including: overall satisfaction with service, ease of contact, value for money, effort to retain customer.	GfK Energy Satisfaction Monitor	Quarterly	<12,000	Current supplier, fuel, whether customer has switched supplier, payment method, whether payment method has changed and several socio-economic variables[26]
Consumers' trust in own and others' energy supplier (2,3)	Please tell me the extent to which you trust or distrust energy suppliers to be open and transparent in their dealings with consumers?	Omnibus survey (CAPI)	Annually	1,500	Fuel, payment method, switching/ engagement behaviours, several socio-economic variables
	How much, if at all, do you trust your energy supplier to do the following [categories specified]?	DECC Public Attitudes Tracker	Quarterly	Approx. 2,000	Income, type of property, main heating fuel, on/off gas grid, presence of pensioner(s), whether disabled/infirm person in household
Consumers' confidence in accuracy of information (4)	No baseline	N/A	N/A	N/A	N/A
Number of consumers on dead and fixed term tariffs (5)	Data provided to Ofgem following a request for information in 2012	Suppliers	Ad hoc	N/A	Supplier, fuel
Number and nature of consumer interactions with customer service (6)	No baseline	N/A	N/A	N/A	N/A
Consumer complaints (7)	Resolution statistics	Ofgem	Annually	N/A	N/A
	Contacts regarding the level of charges	Citizens Advice	Annually	N/A	N/A
	Complaints made with reference to the SOC	Consumer Futures	Annually	N/A	N/A
Consumers' level of satisfaction with interactions with their suppliers (8)	Have you suffered from shock billing (a bill that was larger than you were expecting)?	GfK Energy Satisfaction Monitor	Quarterly	<12,000	See above
	How satisfied are you with the				

[26] These are: Age; Gender; Social Class; Working Status; Region; Size of household; Presence of children & how many; Number of adults; Household make up; Life stage; Internet availability; Internet usage; Type of dwelling; Number of rooms; Household income; Rent or own home.

Indicator	Baseline question / data item	Source of baseline data	Collection frequency	Sample size	Key sample segmentations
	ease of contacting your supplier? If you have had your meter read in the last month, how would you rate your overall satisfaction with the visit?				
Consumers level of satisfaction with treatment of complaint by their supplier (9)	How satisfied were you, overall, with the way in which your complaint was handled by [supplier]? And how satisfied were you with the resolution to your complaint?	Complaints handling telephone survey	Every two years	3,000 complain-ants	Current supplier, several demographic variables, level of literacy/numeracy
Suppliers' introduction and implementation of new polices (10)	No baseline	N/A	N/A	N/A	N/A
Consumers' awareness of SOC (11)	No baseline	N/A	N/A	N/A	N/A

4.3.2 Recommendations for additional monitoring activities

Considering the current evidence base, there is a relatively strong set of baseline data which can be linked to the indicators in this area. The table below sets out a snapshot review of:

- existing monitoring data against the recommended indicators;
- suggested amendments to the existing data collection; and
- additional monitoring activities to be considered.

Table 4.3: Recommendations for monitoring activities

Indicator	Consideration of existing monitoring data	Suggested amendment to existing monitoring	Additional monitoring activities to be considered
Consumers' satisfaction with their energy supplier (1)	GfK Energy Satisfaction Monitor provides strong data	None	Tailored consumer tracker survey Further focus groups/workshops on trust building on qualitative research undertaken in 2012
Consumers' trust in own and others' energy supplier (2,3)	Good baseline, though respondents will have different underlying reasons for their self-assessment of trust Unclear if Omnibus survey cover questions of trust in *own* supplier	Ensure that omnibus survey separately covers trust in *own* supplier and suppliers in general	Tailored consumer tracker survey Consumer trust related focus groups with different customer segments
Consumers' confidence in accuracy of information (4)	No data available	Add questions (own and others) to existing Omnibus	Tailored consumer tracker survey Consumer trust related focus groups with different customer segments
Number of consumers on dead and fixed term tariffs (5)	Data provided to Ofgem following a request for information in 2012	N/A	Aggregated supplier statistics
Number and nature of consumer interactions with customer service (6)	No data available	Obtain data from suppliers	Tailored consumer tracker survey Aggregated supplier statistics
Consumer complaints (7)	Strong data from Energy Ombudsman, Ofgem, Citizens Advice, and Consumer Futures	None	Tailored consumer tracker survey
Consumers' level of satisfaction with interactions with their suppliers (8)	Some non-specific evidence available from GfK Energy Satisfaction Monitor	Add question to GfK monitor that focuses on interactions with suppliers	Tailor-made consumer tracker survey Aggregated supplier statistics Consumer focus groups
Consumers' level of satisfaction with treatment of complaint by their supplier (9)	GfK Energy Satisfaction Monitor and complaints handling telephone survey provides strong data, though the latter is only undertaken every two years	Adapt timing and focus of complaints handling survey in line with RMR implementation	Tailored consumer tracker survey
Suppliers' introduction and implementation of new policies (10)	N/A	N/A	Supplier tracker survey
Consumers' awareness of SOC (11)	N/A	N/A	Tailored consumer tracker survey

4.4 Impact Evaluation Options and Limitations

4.4.1 Methodological considerations

As set out in Chapter 3, there are a number of methodological challenges to take into account when considering the options for evaluating the impact of the RMR on trust. Having ruled out the possibility of RCT and Matching methodologies as evaluation options (see Section 3.3) the discussion below focuses specifically on the possibility of undertaking combinations of time series analysis and qualitative studies.

- The evaluation activities in this area will seek to facilitate an assessment of the developments in consumer trust over time and the degree of supplier compliance with the Licence conditions introduced (or amended) by the RMR. The indicators will guide a data collection and analytical strategy to assess if trust, overall, has improved. There are a number of specific methodological challenges in relation to this: Given the interconnected nature of the policies, it might be challenging to identify the relative contribution of individual measures, though larger-scale investment in quantitative and qualitative monitoring activities would enable contribution analysis and tracing outcomes back to individual policies. For example, it would be possible to explore through consumer surveys, the extent to which consumers are aware of the SOC. This could then (with other output variables, say, consumer complaints and interactions with customer services) be subjected to multivariate analysis to explore the degree to which awareness of the SOC influences overall levels of trust and/or satisfaction.
- The outcome indicators in this area all rely on consumers' self-assessment, which brings with it an inherent frame of reference bias; different consumers are likely to have different definitions of what constitutes 'trust' and 'satisfaction'. Qualitative consumer focus groups or interviews could help to deepen the understanding of the different aspects of consumer trust (for example, trust that they will be treated fairly or trust that an interaction with a supplier will meet the consumer's expectations) and why and how consumer trust and satisfaction is affected by the RMR. Taken together, the quantitative and qualitative data can help determine the extent to which the theory of change can be validated.
- In relation to the SOC, the open-ended implementation approach (i.e. standards may be implemented to varying degrees) could mean that very little changes in certain suppliers' conduct. It will therefore be important to understand if, how and when suppliers have introduced changes in order to contextualise the findings from the survey. Engagement with suppliers, though relying on suppliers' own account of changes made, could help to develop a comprehensive picture of implementation.
- It is also important to bear in mind the multifaceted nature of trust and what factors influence it. These factors may differ significantly between different types of consumers. Because of this, the causal effects set out in the theory of change might not always materialise: given the current levels of disenchantment of consumers with the energy market, even positive experiences with the current or new supplier may not change perceptions of the overall market. There is also the possibility that increased trust may lead to more complaints as consumers demand better/fairer treatment. A tailored tracker survey of consumers could allow alternative explanatory variables (for example, energy prices and negative media coverage involving suppliers) to be examined. The survey would also permit a sophisticated level of sub-group analysis based on clustering to be conducted with the categorisation of consumers based on a range of attitudinal, behavioural, and socio-demographic variables. It would also be important to contextualise quantitative data, such as the number of complaints for example, with qualitative information about the nature and context.

While these factors pose significant challenges for an impact evaluation based on attribution, the evaluation activities in the area of building consumer trust will benefit from a strong set of baseline information. There is also a good body of knowledge around how trust is defined in the context of the energy market

which has provided a strong evidence-based theory of change. We set out below three options for evaluating those RMR polices which are primarily aimed at building consumer trust.

4.4.2 Comprehensive time series analysis

A comprehensive time series analysis would be the most rigorous evaluation option for this area, though also the most resource intensive. As indicated in the discussion above, we suggest that this option should combine the following sources of data:

- The existing relevant baseline data from the GfK Energy Satisfaction Monitor; the CAPI Omnibus survey; the DECC Public Attitudes tracker; aggregated supplier statistics; and the Complaints Handling telephone survey.
- A tailor-made tracker survey of consumers, repeated every year from 2013-2016/17: The survey should include questions on all the consumer-related output and outcome indicators described above, supplemented with indicators related to a broader set of consumer socio-demographic, attitudinal and behavioural indicators. This would allow for sophisticated customer segmentation based on cluster analysis and the exploration of alternative explanatory factors based on multivariate analyses.
- Consumer trust related workshops (in addition to the 2012 workshops undertaken), ideally undertaken yearly but — at a minimum — once before/at the start of the implementation and once at the point of evaluation in 2016/17.
- Annual engagement with suppliers to monitor if, how and when suppliers have introduced changes related to the SOC and to track developments in the number of consumers on dead and fixed term tariffs.
- Market monitoring to track developments in the overall energy market (the macro level as set out in Figure 2.1 above). Thus would include indicators which can provide evidence of more global objectives of the RMR (e.g. increasing competition in the market). It would also enable an analysis of the extent to which developments in macro-related indicators have affected developments in trust. As discussed in Chapter 7, macro indicators could include wholesale to retail price mark-up rates; differentials between the cheapest and dearest tariffs and standard indicators of retail market competition.

These five different strands of research would enable a triangulation of different quantitative and qualitative consumer data, supplier data and market monitoring data. A tailor-made tracker survey allows for the development of a baseline in those areas where this is currently not available[27] and provides, with the additional monitoring activities suggested, the basis for multivariate analysis based on time series in order to identify relationships between treatment variables, outputs and outcomes.

To illustrate the layers of analysis possible:

- Basic time series analysis enables a mapping of how levels of trust in own supplier (indicator 2) develops over time over time. This will not, however, provide any information about what factors have contributed to this development over time.
- Multivariate analysis would enable the exploration of the relative explanatory powers of different factors, such as the level of satisfaction with the interactions with their suppliers (indicator 8) and meso-related indicators such as the number of consumers on dead and fixed term tariffs' (indicator 5)

[27] There is no baseline currently for consumer-related indicators 4: consumers' confidence in accuracy of information; indicator 5: number of consumers on dead and fixed term tariffs; and indicator 6: number and nature of consumer interactions with customer service.

and consumers' awareness of SOC (indicator 11) As such, it would enable the analysis of the extent to which different factors contribute to possible increases in the levels of trust.

- Coupled with qualitative insights into how these relationships are experienced by consumers, a comprehensive time series analysis of this nature would enable a rich contribution analysis of the extent to which the introduction of the SOC, restriction on dead tariffs and fixed term and price protection rules have helped to build trust in suppliers and the energy market.

4.4.3 Limited time series analysis or before-and-after study

While the comprehensive time series analysis is the most rigorous evaluation option, a limited time series analysis or a before-and-after study would be a less resource intensive option. This option builds on the same five data sources as above, though fewer additional monitoring activities over and above the existing baseline data would be needed:

- The tailored tracker survey of consumers could be conducted as before-and-after research, (i.e. end 2013 and end 2016 rather than the four times suggested above). The survey could include all the questions suggested above and/or it could be more limited in scope and include fewer socio-demographic, attitudinal and behavioural indicators.
- The consumer trust related workshops could be conducted once at the end of the evaluation 2016/17, and interim tracking of impact during implementation would rely on the existing consumer trust related qualitative work (undertaken by Ofgem 2012) and the strength of the theories of change in order to argue the case of causal explanations.
- We suggest that the engagement with suppliers should still be repeated every year in this option as, for both the impact and the process evaluation, it would be difficult to track outcomes against activities if there is insufficient evidence about how suppliers have introduced changes.
- Similarly, we suggest that market monitoring should remain in place so that developments in the overall energy market are tracked continuously. However, it might be feasible to undertake this monitoring as a before-and-after study only in order to limit the research to two points in time only.
- Naturally, this option still assumes that the existing relevant baseline data is used in order to monitor developments over the course of the implementation 2013-2017.

This more limited time series study would still enable data triangulation and some level of multivariate analysis, though it would be more limited in scope if the survey is shortened. This option would, however, rely on Ofgem's existing monitoring activities (in particular the Ipsos MORI CAPI survey and the GfK survey) for on-going tracking of impact during implementation.

4.4.4 Descriptive analysis

If it was deemed infeasible to undertake tailored consumer and supplier surveys, a more limited descriptive analysis could be adopted making use of the existing monitoring data. This type of analysis, however, would not constitute an impact evaluation as it would not include an analysis of causality. Instead, it would rely on the theories of change framework, and ideally some or all of the qualitative activities above, to argue for causal relationships.

If this option is chosen we suggest a few tweaks to the existing monitoring data:

- Existing surveys could have additional questions that would focus on indicators such as the 'extent to which consumers are confident that information provided by suppliers is accurate' (indicator 4) or 'consumers' level of satisfaction with the interactions with their suppliers' (indicator 8).

- It could also mean changes in frequency of monitoring activity, such as repeating the Complaints handling survey every year rather than every two years as is currently the case.

4.5 Potential for Economic Evaluation

Economic evaluation seeks to estimate the costs and benefits of individual policy measures and/or the combined impact of a group of policy measures, where possible monetising them. This type of analysis is only robust if it is possible to specify a counterfactual and if it is possible to attribute observed changes to the RMR separately from other influencing factors.

As discussed in Chapter 3, it is difficult to define a counterfactual for the RMR and we consider that it is unlikely to be feasible to do so with any degree of confidence. Therefore, we consider that the potential for a formal economic evaluation by monetising the costs and benefits of the RMR package as whole is limited.

A similarly negative conclusion applies to potential for monetising the impact of the RMR policy measures that seek to build trust in the energy market. To our knowledge, there is no single agreed definition of trust and no agreed scale on which trust is measured, let alone an agreed set of assumptions about the monetary value of an additional unit of trust. This suggests that even if it were possible to robustly identify an increase in, say, the number of consumers that state that they trust their energy supplier and to attribute this increase to the RMR, it is unlikely to be feasible to measure the benefit of this improvement in monetary terms.

Overall, we consider that it would be extremely difficult — if not impossible — to monetise the impacts of the RMR on consumers' trust in the energy market. Any attempts to monetise such impacts would require a significant number of critical assumptions to be made and so the estimates are unlikely to be particularly credible. On this basis, we consider that a future evaluation of the RMR should not seek to monetise the impacts of building consumers' trust in the energy market.

5 Improving Understanding of the Energy Market

5.1 Policy Intent

A lack of understanding of the market can lead to a lack of confidence among consumers in their own ability to make correct decisions about energy tariffs. Ofgem has proposed a range of policy measures that aim to improve consumers' understanding of the energy market, defined as better cognitive knowledge of basic energy concepts, better understanding of one's own tariff features and better understanding of the options available for energy consumers.

5.1.1 Theories of Change

As discussed in Section 2.4.2, we have grouped policy measures that share the primary objective of improved understanding of the energy market. These are:

- improving the quality of regular communications (specifically bills, Annual Statements and Price Increase Notices);
- the tariff information label; and
- tariff comparison rates.

In the table below, we summarise the theories of change that Ofgem has developed for each of these policy measures.

Table 5.1: Overarching theories of change for policy measures aimed at improving understanding

Policy measure	Policy detail	Activities		Intended outcomes		Target population
		Suppliers	*Consumers*	*Direct*	*Indirect*	
Price Increase Notice (PIN)	New rules to ensure consumers receive clear information when their prices increase or there is any other adverse unilateral variation	Send PIN to consumers as a standalone mailing within specified time period. Provide personalised information on the PIN and present price information in a clear, standardised format	Receive PIN, read and understand it	Improve consumers' understanding of the change(s) that will take place to their tariff, the impact and steps to manage the change	Enables consumers to review options and engage in market	All consumers
Annual Statement (AS)	Similar AS across suppliers, improving quality and accessibility of information provided	Improve AS, send out standalone (with no marketing material) and clearly distinct from bills	Receive an AS at a specified time in a specified format, read it and understand it	Improve consumers' understanding of features of current tariff, energy consumption etc.	Enables consumers to review options and engage in market	All consumers
Bills	Provide consumers on a regular basis with the information they need to switch and a prompt to do so	Provide clear personalised information on bills incl. information about cheapest tariffs from current supplier, key characteristics of the tariff and the TCR of the tariff	Receive, read and understand regular bills with clearer information and signposting	Higher levels of consumer understanding and engagement with the bill as a result of more informative content of the bill	Prompt consumers to review energy options and switch tariff and/or provider	All consumers
Tariff information label (TIL)	Presents key tariff information in a consistent format	Present tariff information to a common standard, provide on request and publish on website	Request and read TIL, understand it, consider using information to compare tariffs	Improve consumers' understanding of energy tariff features	Prompt consumers to review energy options and switch tariff and/or provider	All consumers
Tariff comparison rate (TCR)	Require suppliers to calculate p/kWh for typical consumer using specified methodology and assumptions	Include TCR in marketing as well as in their routine communications	Hear of TCR in marketing or read about it in regular communications; use TCR to make 'at a glance' comparisons	Prompt consumers to review energy options and switch tariff and/or provider	Improved consumer understanding of price of available tariffs and how they compare in terms of price	All consumers, particularly the less engaged and more vulnerable

5.1.2 Outside influences and unintended consequences

Table 5.1 focuses on how the policy measures are intended to work. As we have noted in the previous chapter there are also factors that could influence the impact of the RMR proposals and lead to unintended consequences and / or the breakdown in the theories of change described above: either effect would cause the positive impact of the policies to be lower than expected, or even wholly negated.

With respect to the RMR proposals that seek to improve consumers' understanding of the energy market one such unintended consequence could arise if some consumers find it difficult to understand the new terminology that is introduced by the RMR (e.g. the TCR). In this case, there is a risk that a policy that is designed to improve consumers' understanding of the market has increased the level of confusion experienced by some consumers. There is also a risk that some consumers may fail to understand that the TCR is based on comparative costs for a *typical* consumer, and as such should only be used as an *indication* of tariff price and a prompt to look into a Personal Projection. This could mean that some consumers would switch to a more expensive tariff.

Similarly, the smart meter rollout will require consumers to become familiar with additional energy terminology and this could affect the extent to which consumers feel that they understand the energy market and feel able to participate in it. This outside influence could, therefore, limit the extent to which the consumers feel able to use the information on regular communications even if they understand the information provided on those specific communications.

Finally, we explained above that trust is a pre-requisite for consumer engagement with the energy market. If the RMR proposals to build trust in the market do not succeed, for whatever reason, the proposals to build understanding with the market are less likely to succeed as fewer consumers will be willing to incur the costs associated with building understanding (i.e. the opportunity cost of time to reading communications from energy suppliers). This would especially be the case if consumers do not regard the information that they receive from their suppliers to be transparent and accurate.

5.2 Indicators

5.2.1 Outcome indicators

Outcome measures for this set of policies are both direct and indirect in nature in that improved understanding and awareness of consumers' own energy consumption is expected to generate changes in attitudes/values and/or changes in behaviours/intentions. Thus, these policies are likely to contribute to some of the wider outcomes expected also from other RMR policies. The outcome indicators we propose are:

12. Consumers' understanding of the retail energy market: % of consumers who have a better understanding of the energy market overall; the extent to which consumers understand the features of other tariffs supplied by their own supplier; the extent to which consumers understand the features of other tariffs supplied by other suppliers.
13. Consumers' understanding of their current consumption: % of consumers who have a better understanding of their own energy consumption compared to the same time last year; % of consumers that are aware of how much energy they consume per year, % of consumers reporting they know the price they pay for their energy consumption.
14. Consumers' understanding of their current tariff: % of consumers who have a better understanding of the features of their own tariff, compared to the same time last year; the extent to which consumers

understand the features of their own tariff; the extent to which consumers know the payment terms related to their tariff.

15. Consumers' awareness of possibilities of switching: the extent to which consumers are aware of the possibility of switching tariff or supplier: % of consumers who do not believe it is possible to change tariffs; % of consumers reporting that they don't believe it is possible to change supplier.
16. Consumers' understanding of regular communications.

These outcome indicators will, according to the overall intervention logic of the RMR, affect macro-level outcomes for the sector as a whole, ultimately including levels of competition (if improved information prompts or supports market engagement). These kinds of indicators are discussed in Section 7.

5.2.2 Output indicators

The output indicators relate more specifically to measuring the activities associated with the specific policies in this area, including the extent of implementation of the policies by suppliers and degree of awareness of new concepts amongst consumers:

17. Consumers' awareness of suppliers' communications: % of consumers that claim they have received a TIL, PIN, AS at least once; % of consumers that are aware that their bill contains new information (about Personal Projection, cheapest tariffs from current supplier, TCR, etc.).
18. Consumers' level of engagement with suppliers' communications, and reported impact on behaviours: % who have read the communications; the % who report that suppliers' communications have made them consider their energy options; % who report that they have taken action as a result of reading the information.
19. Clarity of suppliers' communications: assessment of the clarity of suppliers' regular communications (including bills, AS, PINs and TIL) by language / linguistics experts.
20. Suppliers' degree of compliance to improve regular communications: the number of suppliers who comply with their regular communication obligations, including bills, AS and PINs.
21. Consumers' complaints regarding communications: total number of complaints regarding communication per year/months, broken down by suppliers: % of complaints regarding communication against the total number of complaints

5.3 Defining the Evidence Base

5.3.1 Existing monitoring activities

Ofgem has previously gathered some data that can serve as a baseline against which the impact of the RMR could potentially be evaluated, given the output and outcome indicators that are identified above. These are shown in Table 5.2.

Table 5.2: Baseline data

Indicator	Baseline question / data item	Source of baseline data	Collection frequency	Sample size	Key sample segmentations
Consumers' understanding of energy market overall (12)	Findings from general discussion on experiences with energy market and activity in it	Qualitative focus group / deliberative research	Ad hoc	N/A	N/A
	To what extent would you say you understand the range of different energy tariffs available to you?	Omnibus survey (CAPI)	Annual	1,500	Fuel, payment method, several socio-economic variables, level of engagement
Consumers' understanding of own energy consumption (13)	Findings from general discussion on experiences with energy market and activity in it	Qualitative focus group / deliberative research	Ad hoc	N/A	N/A
Consumers' understanding of current tariff (14)	Findings from general discussion on experiences with energy market and activity in it	Qualitative focus group research	Ad hoc	N/A	N/A
Consumers' awareness of possibility of switching tariff or supplier (15)	Did you know it was possible to change to a different tariff or method of payment with your current supplier? (asked of those that have not switched tariff/payment method in previous year) Did you know it was possible to switch to a different gas or electricity supplier? (asked of those that have never switched supplier)	Omnibus survey (CAPI)	Annual	<1,500 (depends on number that have not switched tariff/ payment method in previous year / never switched supplier)	See above
Consumers' understanding of regular communications (16)	Extent to which those that read bills, annual statements, PIN and end of fixed-term notice say they understood the information	Omnibus survey (CAPI)	Annual	<1,500	See above
	How satisfied are you that your bills are easy to understand	GfK Energy Satisfaction Monitor	Quarterly	<12,000	See above
Consumers' awareness of suppliers' communications (17)	Please tell me if you recall receiving any of the following in the last year? You may have received these via post or email: AS; bill/statement of account; PIN; end of fixed term letter	Omnibus survey (CAPI)	Annual	1,500	See above
Consumers' level of engagement with suppliers' communications, and reported impact on behaviours (18)	Extent to which those that read bills, annual statements, PIN and end of fixed-term notice say: (a) the information made them consider their energy options (b) they took action as a result	Omnibus survey (CAPI)	Annual	<1,500	See above

Indicator	Baseline question / data item	Source of baseline data	Collection frequency	Sample size	Key sample segmentations
	of reading this information (e.g. looked into, or actually switched supplier, tariff or payment method)				
Clarity of regular communications including bills, AS and PINs (19)	Academic research found that energy-related terminology is not part of consumers' everyday vocabulary. It also found weaknesses in suppliers' current communications with consumers with respect to language, terminology and presentation.	Lawes Consulting[28]	Single point in time (2011)	N/A	N/A
Suppliers' degree of compliance in improving regular communications (20)	No baseline[29]	N/A	N/A	N/A	N/A
Consumers' complaints regarding communications (21)	What was your recent complaint about?	Complaints handling telephone survey Ofgem / Consumers Futures / Ombudsman / Citizens Advice statistics	Every two years Monthly	3,000 complain-ants N/A	Current supplier, several demographic variables, level of literacy/numeracy

5.3.1 Recommendations for additional monitoring activities

Considering the current evidence base, there is a relatively strong set of baseline data which can be linked to the indicators in this area. The table below sets out a snapshot review of:

- existing monitoring data against the recommended indicators;
- suggested amendments to the existing data collection; and
- additional monitoring activities to be considered.

[28] Lawes Consulting &Lawes Gadsby Semiotics (November 2011),"Retail Market Review – energy bills, annual statements and price rise notification advice on layout and the use of language: A research report for Ofgem"

[29] This refers to new rules; it might be, however, that there is relevant evidence in the data related to the annual monitoring of supplier compliance, done by Ofgem on an annual basis since 2009

Table 5.3: Recommendations for monitoring activities

Indicator	Consideration of existing monitoring data	Suggested amendment to existing monitoring	Additional monitoring activities to be considered
Consumers' understanding of energy market overall (12)	Good evidence behind Theories of Change from qualitative focus groups Omnibus survey question on range of energy tariffs is useful	Add questions to Omnibus for all sub-indicators suggested	Tailor-made consumer tracker survey Consumer awareness related focus groups with different customer segments
Consumers' understanding of own energy consumption (13)	No data available	Add questions to Omnibus survey for all sub-indicators suggested	Tailor-made consumer tracker survey Consumer awareness related focus groups with different customer segments
Consumers' understanding of current tariff (14)	Good qualitative evidence base but no quantitative evidence	Add questions to Omnibus survey for sub-indicators suggested	Tailor-made consumer tracker survey Consumer awareness related focus groups with different customer segments
Consumers' awareness of possibility of switching tariff or supplier (15)	Good baseline	None	Tailor-made consumer tracker survey Consumer awareness related focus groups with different customer segments
Consumers' understanding of regular communications (16)	Good baseline, though respondents will have different definitions of understanding and may not in practice differentiate between different types of communications	None	Tailor-made consumer tracker survey Independent assessment of consumer understanding through interviews or focus groups
Consumers' awareness of suppliers' communications (17)	Good baseline from Omnibus survey	None	Tailor-made consumer tracker survey
Consumers' level of engagement with suppliers' communications (18)	Good evidence from Omnibus survey	None	Tailor-made consumer tracker survey Independent assessment of consumer understanding through interviews or focus groups
Clarity of regular communications including bills, AS and PINs (19)	Good evidence from Lawes Consulting report	None	Independent assessment
Suppliers' degree of compliance in improving regular communications (20)	No baseline	N/A	Independent assessment Supplier Tracker Survey
Consumers' complaints regarding communications (21)	Complaints handling telephone survey provides strong data, though it is only undertaken every two years	Adapt timing and focus of complaints handling survey in line with RMR implementation	Tailor-made consumer tracker survey

5.4 Impact Evaluation Options and Limitations

5.4.1 Methodological considerations

As set out in Chapter 3, RCT and Matching methodologies are not appropriate evaluation options, largely due to the fact that it is not possible to establish a counterfactual or a control group. The discussion below therefore focuses specifically on options for undertaking forms of time series analysis coupled with longitudinal descriptive analyses in order to evaluate the impact of RMR policies targeted to improve the understanding of the energy market. The indicators above will guide a data collection and analytical strategy. There are, however, a number of specific methodological challenges in relation to this:

- Many of the individual policies work in concert, e.g. the TCR would be included in bills and annual statements. Similarly some of the policies in the other thematic areas (trust and tariff choice) contribute to an improved understanding, e.g. cheapest tariff messaging (CTM). It would be possible to explore these issues both qualitatively and through multivariate analysis correlating consumers' level of engagement with different types of regular communications (indicator 18) with outcome indicators.
- As is the case for the trust-related outcome indicators, all of the outcome indicators and some of the output indicators in this area all rely on consumers' self-assessment, which brings an inherent frame of reference bias: consumers are likely to have different definitions of what constitutes good or poor levels of understanding. Qualitative consumer focus groups or interviews could enable a more objective assessment of levels of understanding and provide important information about what aspects of the communications are particularly valuable to consumers and why. As noted previously, taken together, the quantitative and qualitative data can help determine the extent to which the intervention logic can be validated.
- Some of the policies in this area are non-prescriptive in that they allow suppliers to maintain a level of flexibility to tailor communications to maintain own corporate standards. It is therefore important to break down consumer survey results by supplier and include an expert independent assessment of the quality of communications.
- There is not necessarily a direct linear process between better awareness and / or understanding and attitudinal and/or behaviour change. A time series repeated tracker survey will provide a picture of developments in the levels of understanding over time and qualitative consumer groups will help explore the actual consumer journeys as they are experienced following the implementation of the RMR.
- The theories of change set out the intervention logic related to RMR policies in this area, but there are additional means through which consumers' understanding is raised, other than through written communications and physical tools (e.g. word of mouth,[30] customer services staff taking the time to explain concepts verbally in person or over the phone, etc.). Hence, observed changes (both at output and outcome levels) are likely to have additional explanatory factors other than the changes associated with the policies in this area. A comprehensive consumer tracker survey could include other indicators or broader response categories to enable an assessment of questions of what prompted better understanding (i.e. suppliers' written communications, suppliers' verbal communication or word of mouth) which would enable the assessment of other explanatory factors. In addition, qualitative feedback from consumers and suppliers alike could map and explain the full range of factors underpinning the outcome and output variables in more detail.

[30] As indicated in the Ipsos MORI report "Consumer engagement with the energy market, information needs and perceptions of Ofgem, Findings from the Ofgem Consumer First Panel Year 4: second workshops (held in March 2012)", 30.08.2012.

While these factors pose significant challenges for an impact evaluation based on attribution, the evaluative activities in the area of building understanding benefit from some strong baseline information in many areas. Below we set out three options for evaluating those RMR polices which are primarily aimed at improving understanding of the energy market.

5.4.2 Comprehensive time series analysis

A comprehensive time series analysis would be the most rigorous evaluation option for this area, though also the most resource intensive. We suggest that this option combines the following sources of data:

- The existing relevant baseline data from the GfK Energy Satisfaction Monitor; the CAPI Omnibus survey; the complaints handling survey; Ofgem, Consumer Futures, Ombudsman and Citizens Advice Statistics; the Lawes report and qualitative focus group research.
- A tailor-made tracker survey of consumers, repeated every year from 2013-2016/17: The survey should include questions related to all the consumer-related output and outcome indicators above, supplemented with indicators related to a broader set of consumer socio-demographic, attitudinal and behavioural indicators. This would enable a sophisticated customer segmentation to be completed based on cluster analysis and would allow for the exploration of alternative explanatory factors through multivariate analyses.
- Consumer awareness and understanding related workshops (in addition to the existing workshops undertaken by Ofgem). These would ideally be undertaken yearly but — at a minimum — once before/at the start of the implementation and once at the end of the evaluation 2016/17.
- Supplier tracker survey repeated every year (at a minimum) to monitor if, how and when suppliers have introduced changes in their communication.
- Market monitoring to track developments in the overall energy market (the macro level as set out in Figure 3.3 above). As the case for the other areas, this would include indicators which can provide evidence of more global objectives of the RMR. It would also enable an analysis of the extent to which developments in macro-related indicators have affected developments in consumers' experiences of the communication from suppliers. As discussed in Chapter 7 below, macro indicators could include wholesale to retail price mark-up rates; differentials between the cheapest and dearest tariffs and standard indicators of retail market competition.

The five different strands of research enable a triangulation of different quantitative and qualitative consumer data; supplier data and market research. A tailor-made tracker survey allows for the development of a baseline in those areas where this is currently not available[31] and provides, with the additional monitoring activities suggested, the basis for multivariate analysis based on time series in order to identify relationships between treatment variables, output and outcomes. To illustrate the layers of analysis possible:

- Basic time series analysis enables a mapping of, say, how consumers' understanding of their current tariff changes over time (indicator 14). This, however, will not provide any information about what factors have contributed to this change in understanding.
- Regression analysis would enable the exploration of the relative explanatory powers of different factors, such as consumers' level of engagement with suppliers' communication (indicator 18), or

[31] There is no baseline currently for consumer-related indicators: Consumers' understanding of own energy consumption (13) and Suppliers' degree of compliance in improving regular communications (20). In addition, there is no quantitative baseline for Consumers' understanding of current tariff (14)

consumers' level of awareness of suppliers' communication (indicator 17). Coupled with qualitative assessments of the clarity of suppliers' communications (indictor 19) or suppliers' degree of compliance (indicator 20) it would be possible to undertake a rich contribution analysis of the extent to which the changes in communications as a result of RMR has helped to improve consumer understanding.

- The clarity of suppliers' communications (indicator 19) can be assessed in line with best practice (such as robust discourse analysis) by independent experts. It is possible to assess the evolution of changes from before the first RMR consultation through to implementation and post-implementation as all suppliers are expected to be able to retrieve old communications.
- To enable the assessment of consumers' perceptions of their ability to understand these communications it is important that the suggested tailored tracker survey of consumers has a large enough sample to assess the data by supplier and region since some key concepts are somewhat dependent on regional characteristics (e.g. the TCR).

5.4.3 Limited time series analysis or before-and-after study

While the comprehensive time series analysis is the most rigorous evaluation option, a limited time series analysis or a before-and-after study would be a less resource intensive option. This option uses similar data sources to the previous option, though with more limited suggested new monitoring activities:

- The suggested tailor-made tracker survey of consumers could be conducted as a before-and-after (i.e. end 2013 and end 2016 rather than the four times suggested above). The survey could include all the questions suggested above and/or it could be more limited in scope and include fewer socio-demographic, attitudinal and behavioural indicators.
- The consumer awareness related workshops could be conducted once at the point of evaluation in 2016/17. In this case, interim tracking of impact during implementation would rely on previous qualitative work (undertaken by Ofgem in 2012) and the strength of the theories of change in order to argue the case of causal explanations.
- The independent assessments could be conducted as a retrospective analysis at the end of the evaluation, though — for formative evaluation purposes — it is recommended that these assessments are undertaken more frequently throughout implementation.
- We suggest that the supplier tracker survey should still be repeated every year in this option as, for both the impact and the process evaluation it would be difficult to track outcomes against activities if there is insufficient evidence about how suppliers have introduced changes.
- Similarly, we suggest that market monitoring should be completed on a regular basis so that developments in the overall energy market are tracked continuously. However, it might be feasible to undertake this monitoring as a before-and-after study only in order to limit the research to two points in time.
- Naturally, this option still assumes that the existing relevant baseline data is used in order to monitor developments over the course of the implementation 2013-2017.

This type of more limited time series study would still enable data triangulation and some level of multivariate analysis, though it would be more limited in scope if the survey is shortened. This option would however, rely on the existing monitoring activities (in particular the Ipsos MORI CAPI survey and the GfK survey) for on-going tracking of impact during implementation.

5.4.4 Descriptive analysis

Ofgem may decide that the types of monitoring activities outlined above are not feasible given potential budget constraints. In this case, the existing monitoring data could be applied in a descriptive design with a few suggested tweaks:

- Existing surveys could have additional questions on those indicators that are not currently covered in existing surveys (e.g. 'consumers' understanding of own energy consumption (13) and consumers' understanding of current tariffs (14)).
- It could also mean changes in frequency of monitoring activity, such as repeating the Complaints Handling survey every year rather than every two years as is currently the case.
- Inclusion of some or all of the qualitative activities set out above.

As indicated in Chapter 4, this would not constitute an impact evaluation as it would not be possible to attribute changes in trends to the interventions. As such, it would rely on the theories of change framework and additional qualitative consumer research to argue for causal relationships.

5.5 Potential for Economic Evaluation

As discussed in Chapter 3, it is difficult to define a counterfactual for the RMR and we consider that it is unlikely to be feasible to specify a counterfactual with any degree of confidence. Therefore, we consider that the potential for a formal economic evaluation monetising the costs and benefits of the RMR package as whole is limited.

With respect to the RMR proposals that seek to improve consumers' understanding of the energy market, we similarly consider that it will be difficult to monetise the impacts of the policy measures. One reason for this is the fact that it will be extremely difficult to specify a robust counterfactual for this set of policy measures and hence we consider that it would not be possible to attribute observed changes to the RMR. However, there are additional difficulties that are specific to this policy objective.

In particular, improving consumers' understanding of the energy market should lead them to feel more in control of their affairs and better able to participate in the energy market, should they choose to do so. It is extremely difficult to capture these benefits on a continuous quantitative scale, however, which is an important first step towards completing an economic evaluation. Furthermore, we are not aware of any studies that have attributed a monetary value to increasing an individual's understanding of the energy market.

As with the previous objective (i.e. trust described in Chapter 4) we consider that a future evaluation of the RMR should not seek to monetise the impacts of building consumers' understanding of the energy market.

6 Simplifying Tariff Choices

6.1 Policy Intent

Complex information and difficulties in comparing tariffs can discourage consumers from engaging in the energy market and can lead to poor switching decisions. Ofgem aims to make it easier for consumers to compare energy tariffs and so increase the number of consumers that can make effective comparisons. The policy measures aim to make it easier for consumers to check both whether they are on the best energy tariff for their circumstances and to identify a superior tariff if they consider that their current deal does not fully meet their needs. The policy measures may lead to an increase in switching rates (both within and between suppliers), and should certainly lead to an improvement in the average quality of a switch (i.e. the benefit that consumers derive from the switch).

6.1.1 Theories of change

As discussed in Section 2.4.3, we have grouped policy measures that share the primary objective of simplifying tariff choice. These are:

- the tariff cap;
- tariff structures;
- the cheapest tariff message;
- the personal projection;
- the end of fixed term notice; and
- the dead notice.

In Table 6.1 we summarise the theories of change that Ofgem has developed for each of these policy measures.

Table 6.1: Overarching theories of change for policy measures that aim to simplify tariff choices

Policy measure	Policy detail	Activities		Intended outcomes		Target population
		Suppliers	*Consumers*	*Direct*	*Indirect*	
Tariff cap	Limits the number of tariffs a consumer is faced with to allow them to assess their options more effectively	Offer no more than four core electricity and four core gas tariffs per meter type.	Are presented with fewer tariffs in regular and specific communications, at suppliers' or price comparison websites, etc. Face fewer choices when making tariff comparisons	Reduce the complexity of market to enable consumers to compare and assess their options more easily and effectively	Increase consumer engagement (by reducing concerns about, and perception of 'hassle' involved in picking most appropriate tariff for their needs and removing the barrier of being overwhelmed by choice)	All consumers, in particular currently disengaged consumers
Tariff structures	Simplifies and standardises tariff structures; sets rules for use of discounts and bundles.	Adhere to new tariff structures, limiting the number and form of discounts applied to energy tariffs that may 'lock' in customers or unduly influence switching decisions.	Are presented with simpler tariffs in regular and specific communications, at suppliers' or price comparison websites, etc. Face a standardised tariff structure when making tariff comparison	Increase in number of consumers that feel able to compare tariffs	Increase consumer engagement (through fairer and more transparent tariffs); improved quality (and possibly quantity) of switching	All consumers, in particular currently disengaged consumers
Cheapest tariff message	New set of rules to provide personalised savings messages to consumers	Use CTM in routine communications	Read about cheaper tariffs in suppliers specific communications	Increase awareness of current suppliers' alternative tariffs and savings available	Increase consumer engagement / switching	All consumers, in particular currently disengaged consumers
Personal projection	New tool: creating a consistent method for providing a projection of annualised costs to consumers	Calculate, for each consumer, an estimated annual cost (to be referred to as a Personal Projection)	Provided with an accurate, personalised figure for estimated cost of current and alternative tariffs	Improved indicator for consumers to compare / assess their options and budget for future energy costs	Increase consumer engagement (through greater transparency of projected costs on current or alternative tariffs)	All consumers
End of fixed term notice	Notice to consumers 42-49 days ahead of the end of their fixed term	Write to consumers, including relevant switching info on CTM, PP, TCR and guarantee of	Receive notice, read it, understand it and potentially use it to compare tariff options	Consumers understand their options and what will happen if they take	Consumers engage with the market at the end of their fixed term contract	All consumers on fixed term contracts

Policy measure	Policy detail	Activities		Intended outcomes		Target population
	contract	offer of relevant cheapest evergreen if no consumer action		no action at the end of a fixed term contract		
Dead notice	Notice to consumers on tariffs no longer open to new customers about RMR-related changes to their contract	Explain changes to dead tariffs to customers and present relevant switching information in a clear language and easy-to-compare fashion	Receive notice, read it and potentially use it to compare tariff options	Consumers understand what changes are made to their dead tariff, and why. Also know how they can switch tariffs or suppliers	Affected consumers engage in the market	Consumers on dead tariffs (more likely to be disengaged and potentially vulnerable)

6.1.2 Outside influences and unintended consequences

Table 6.1 focuses on how the policy measures are intended to work. As we have noted in the previous chapters there are also factors that could influence the impact of the RMR proposals and lead to unintended consequences and / or the breakdown in the theories of change described above: either effect would cause the positive impact of the policies to be lower than expected, or even wholly negated. It would be important for a future evaluation to consider the extent to which outside factors such developments in energy prices have influenced the operation of the theories of change following the implementation of the RMR.

With respect to the RMR proposals that seek to simplify the assessment of tariffs, one such unintended consequence could arise if the cheapest tariff message leads to a reduction in the number of consumers that consider tariffs offered by other suppliers. This effect may suggest that suppliers would have a lower risk of losing customers post-RMR and so may indicate that the RMR had led to a reduction in competition between suppliers. Similarly, there is a risk that the tariff cap and tariff structures proposals may frustrate some currently engaged consumers — particularly those that can no longer receive a particular discount or additional feature which they value — to such an extent that the proposals would not increase overall engagement with the market.

6.2 Indicators

6.2.1 Outcome indicators

The overall policy objective to simplify tariff choices is reflected in the following outcome indicators.

22. Consumers' level of tariff comparison activity and reasons for this: % of consumers that have reviewed options and compared tariff over the past year/month; % of consumers reporting that comparison activity was prompted by bills, AS, PINs, end of fixed term notice, dead notice etc.
23. Use of comparison services: web traffic on comparison/switching websites; % of consumers that use price comparison/switching sites; % of consumers that use telesales/other price comparison

services.

24. Consumers' confidence that they can compare tariffs easily: % of consumers reporting they can compare tariff easily.
25. Consumer switching by supplier, meter type, payment and energy type: intra-supplier and inter-supplier: % of consumers that have switched tariff at least once (survey data); number of consumers that have switched tariff over the past year/month by supplier, meter type, payment and energy type: intra-supplier and inter-supplier. Consumers' perceptions of whether they made monetary savings by switching tariff or supplier: % consumers reporting that they saved money by switching tariff or supplier.
26. Consumers' confidence that they are getting the best energy deal possible for their needs: % of consumers that consider they are on the best deal for their circumstances.

6.2.2 Output indicators

Identifying the impact of these policies will require consideration of the steps which suppliers have taken to enable this to happen and of what has prompted some customers' to switch. Consumer complaints in this area will also provide indicators of impact.

27. Consumers' prompt for most recent switch: % consumers reporting that the following made them consider switching tariff and/or supplier: bills, AS, PINs, CTM, End of fixed term notice, dead notice; % consumers reporting they used CTM, PP tools to compare tariffs during tariff comparison/switch.
28. Suppliers' degree of compliance to simplify the assessment of tariffs: the extent to which suppliers comply with their tariffs obligations and can document it.
29. Consumers' complaints regarding loss of discounts per year/month: total number of complaints regarding loss of discounts per year/month, broken down by suppliers; % of complaints regarding loss of discounts in the total number of complaints.
30. Consumers' complaints regarding tariffs (or comparability of tariff): total number of complaints regarding tariff per year/month, broken down by suppliers; % of complaints regarding tariff in the total number of complaints.
31. Suppliers' data on impact of CTM: % customers that have switched in response to the CTM; % of consumers that have switched to a similar tariff (narrow CTM); % of consumers that have switched to a different tariff ('wide' CTM).

6.3 Defining the Evidence Base

6.3.1 Existing monitoring activities

Ofgem has previously gathered some data that can serve as a baseline against which the impact of the RMR could potentially be evaluated, given the output and outcome indicators that are identified above. These are shown in the table below.

Table 6.2: Baseline data

Indicator	Baseline question / data item	Source of baseline data	Collection frequency	Sample size	Key sample segmentations
Consumers' level of tariff comparison activity and reasons for this (22)	Extent to which readers of bills, AS, PINs, end of fixed term notices and dead notices think the information made them consider their energy options and / or took action	Omnibus survey (CAPI)	Annual	<1,500	Fuel, payment method, several socio-economic variables, level of engagement
	In the last 3 months did you seriously consider changing your energy supplier?	GfK Energy Satisfaction Monitor	Quarterly	<12,000	Current supplier, fuel, whether customer has switched supplier, payment method, whether payment method has changed and several socio-economic variables[32]
Use of comparison services (23)	Thinking about the last time you switched gas/electricity supplier, how did you find out about the deals offered by the supplier you switched to? (asked of those that have ever switched)	Omnibus survey (CAPI)	Annual	<1,500	See above
	How did you switch/change your energy supplier? (asked of those that have ever switched)	Omnibus survey / GfK Energy Satisfaction Monitor	See above	See above	See above
Consumers' confidence that they can compare tariffs easily (24)	How easy or difficult do you believe it is to compare different tariffs for electricity or gas? Thinking about the number of different tariffs available to you, in your view, are there the right number of energy tariffs available, or are there too many tariffs or too few tariffs?	Omnibus survey (CAPI)	Annual	1,500	See above
Consumer switching by supplier, meter type, payment and energy type (25)	Did you switch your gas/electricity supplier? Did you change the type of tariff you are on or the method by which you pay for your energy?	Omnibus survey / GfK Energy Satisfaction Monitor	See above	See above	See above

[32] These are: Age; Gender; Social Class; Working Status; Region; Size of household; Presence of children & how many; Number of adults; Household make up; Life stage; Internet availability; Internet usage; Type of dwelling; Number of rooms; Household income; Rent or own home.

Indicator	Baseline question / data item	Source of baseline data	Collection frequency	Sample size	Key sample segmentations
	Gains and losses by supplier	Suppliers	Monthly	N/A	Supplier, payment method, region, online/offline
Consumers' confidence that they are getting the best energy deal possible for their needs (26)	No baseline	N/A	N/A	N/A	N/A
Consumers' prompt for most recent switch (27)	Thinking about the last time you switched your gas/electricity supplier, what was the main trigger causing you to switch? (unprompted) Thinking about the last time you changed your tariff or payment method (without switching supplier) what was the main trigger causing you to change it? (unprompted)	Omnibus survey (CAPI)	Annual	<1,500	See above
	What made you decide to switch tariffs?'	GfK Energy Satisfaction Monitor	Quarterly	<12,000	See above
Suppliers' degree of compliance to simplify the assessment of tariffs (28)	No baseline	N/A	N/A	N/A	N/A
Complaints regarding loss of discounts (29)	What was your recent complaint about?	Complaints handling telephone survey	Every two years	3,000 complain-ants	Current supplier, several demographic variables, level of literacy/numeracy
Complaints regarding tariffs / tariff comparability (30)	What was your recent complaint about?	Complaints handling telephone survey	Every two years	3,000 complain-ants	See above
Suppliers' data on impact of CTM (31)	No baseline	N/A	N/A	N/A	N/A

6.3.2 Recommendations for additional monitoring activities

Considering the current evidence base, there is a relatively strong set of baseline data which can be linked to the indicators in this area. The table below sets out a snapshot review of:

- existing monitoring data against the recommended indicators;
- suggested amendments to the existing data collection; and
- additional monitoring activities to be considered.

Table 6.3: Recommendations for monitoring activities

Indicator	Consideration of existing monitoring data	Suggested amendment to existing monitoring	Additional monitoring activities to be considered
Consumers' level of tariff comparison activity and reasons for this (22)	Good baseline	None	Tailor-made consumer tracker survey Consumer confidence related focus groups/workshops with different customer segments Aggregated supplier data on number of calls/email contacts related to reviewing tariff options
Use of comparison services (23)	Good baseline based on consumer research; limited data from comparison services	Gather more data from comparison services	Monitoring of website traffic and contacts to telephone switching services Web-based feedback survey placed on comparison websites
Consumers' confidence that they can compare tariffs easily (24)	Strong baseline and possibility for on-going monitoring via CAPI Omnibus survey	None	Tailored tracker survey Consumer confidence related focus groups with different customer segments (could be included in consumer trust focus groups as suggested above)
Consumer switching by supplier, meter type, payment and energy type (25)	Good baseline, including reasons behind switching	None	Tailored consumer tracker survey Qualitative interviews with engaged consumers exploring reasons for their switch
Consumers' confidence that they are getting the best energy deal possible for their needs (26)	No data available	Add question to Omnibus survey	Tailor-made consumer tracker survey Consumer confidence related focus groups/workshops with different customer segments
Consumers' prompt for most recent switch (27)	Good baseline	Could consider specifying potential prompts in surveys	Tailored tracker survey Qualitative interviews with engaged consumers exploring reasons for their switch
Suppliers' degree of compliance to simplify the assessment of tariffs (28)	N/A	N/A	Independent assessment Ofgem monitoring
Complaints regarding loss of discounts (29)	The complaints handling survey provides good baseline, though it is only undertaken every two years	Adapt timing focus of complaints handling survey in line with RMR implementation	None
Complaints regarding tariffs / tariff comparability (30)	The complaints handling survey provides good baseline, though it is only undertaken every two years	Adapt timing focus of complaints handling survey in line with RMR implementation	None
Suppliers' data on impact of CTM (31)	N/A	N/A	Aggregated supplier statistics

6.4 Evaluation Options

6.4.1 Methodological considerations and limitations

With the possibility of RCT and Matching methodologies as evaluation options ruled out, the discussion below focuses specifically on the possibility of undertaking combinations of time series analysis and qualitative studies. The evaluation activities in this area will seek to facilitate an assessment of whether the policy measures listed above have simplified tariff choices for consumers. The indicators will guide a data collection and analytical strategy to assess if tariff choices have, overall, been simplified. In doing so, there are a number of specific methodological challenges to consider:

- Most of the outcome indicators in this area rely on consumers' self-assessment[33] which brings with it an inherent measurement bias. However, it is worth noting that some of the indicators are more objective in nature such as the 'level of activity' (indicator 22) and switching[34] (indicator 25). At the other end of the scale, indicators such as 'confidence in getting the best deal possible' (indicator 26) are more subjective. Qualitative consumer focus groups or interviews can help to deepen the understanding of how consumer confidence and behaviours are affected by the policy measures and so provide important causal tracing strategies.
- The theories of change set out the intervention logic related to RMR policies in this area, but observed changes in relation to the indicators will not just be a function of the success or failure of the RMR policy measures. Indeed, there are other factors which encourage consumer engagement such as price changes, media coverage of energy suppliers, marketing by comparison websites, etc. Through a comprehensive consumer tracker survey it would be possible to include indicators which would enable the assessment of other explanatory factors. For example, including proxy measures which relate more broadly to indicators of consumer engagement in other sectors (for example, use of insurance or other price comparison sites over the past year) could help identify the extent to which consumers' increased use of energy comparison sites are a function of a general trend of increased consumer engagement instead of a specific impact of RMR policies. In addition, qualitative feedback from consumers and suppliers alike could map and explain the full range of factors underpinning the outcome and output variables in more detail.

That said, there is a good baseline in this area and the indicators — particularly those relating to behaviours —– are clearly defined and based on sound theories of change. Below we set out three options for evaluating those RMR polices which primarily aim to simplify tariff choices.

6.4.2 Comprehensive time series analysis

A comprehensive time series analysis would again be the most rigorous evaluation option, though also the most resource intensive. As discussed above, this option should combine the following data sources:

- The existing relevant baseline data from the GfK Energy Satisfaction Monitor; the CAPI Omnibus survey; aggregated supplier statistics; and the Complaints Handling telephone survey.

[33] With exception of "use of comparison sites" (indicator 22) which could be monitored through website traffic and statistics from telephone

[34] Though switching in relative terms is less subject to self-reporting bias than more subjective indicators, it is not free from it. Supplier data on switching currently shows a very different picture of actual switches than that reported by consumers who are not likely to report all their switching behaviour, especially if switching did not happen within the near future. Consequently, it is important to consider this measure as consumer perception rather than an objective fact.

- A tailor-made tracker survey of consumers, repeated every year from 2013-2016/17: The survey should include questions on all the consumer-related output and outcome indicators described above, supplemented with indicators related to a broader set of consumer socio-demographic, attitudinal and behavioural indicators. This would allow for sophisticated customer segmentation based on cluster analysis and the exploration of alternative explanatory factors based on multivariate analyses.
- Consumer confidence related workshops, ideally undertaken yearly but — at a minimum — once before/at the start of the implementation and once at the point of evaluation in 2016/17. These workshops could also cover issues related to trust and understanding.
- Market monitoring to track developments in the overall energy market (the macro level as set out in Figure 3.3 above). This would include indicators which can provide evidence of more global objectives of the RMR (e.g. increasing competition in the market). As discussed in Chapter 7, macro indicators could include: wholesale to retail price mark-up rates; differentials between the cheapest and dearest tariffs and standard indicators of retail market competition.

These four different strands of research would enable a triangulation of different quantitative and qualitative consumer data; supplier data and market research. A tailor-made tracker survey would allow for the development of a baseline in those areas where this is currently not available[35] and provide, with the additional monitoring activities suggested, the basis for multivariate analysis based on time series in order to identify relationships between treatment variables, outputs and outcomes.

The layers of analysis possible would include:

- Basic time series analysis enables us to map over time the extent to which consumers switch tariffs and/or suppliers (indicator 25). This, however, will not provide any information about what factors have prompted switching behaviours.
- Multivariate analysis would enable the exploration of the relative explanatory powers of different factors, such as consumers' confidence that they can compare tariffs easily (indicator 24) and consumers' prompt for most recent switch (indicator 27). As mentioned above, it could also be worthwhile analysing switching behaviours against other micro-level proxy indicators like switching behaviour in other markets or, in the macro level, developments in pricing. This would enable an analysis of the extent to which levels in switching are down to the introduction of the RMR policies or can better be explained by developments in consumer behaviours in general and/or broader market developments.
- Coupled with qualitative insights into how these relationships are experienced by consumers, a comprehensive time series analysis of this nature would enable a rich contribution analysis of the causal relationships between the RMR policies and switching.

6.4.3 Limited time series analysis or before-and-after studies

While the comprehensive time series analysis is the most rigorous evaluation option, a limited time series analysis or a before-and-after study would be a less resource intensive option. This option builds on the same four data sources as above, though fewer new monitoring activities would be conducted:

- The tailor-made tracker survey of consumers could be conducted as before-and-after research, (i.e. end 2013 and end 2016 rather than the four times suggested above). The survey could include all the

[35] There is no baseline currently for consumer-related indicators 4: consumers' confidence in accuracy of information; indicator 5: number of consumers on dead and fixed term tariffs; and indicator 6: number and nature of consumer interactions with customer service.

questions suggested above and/or it could be more limited in scope and include fewer socio-demographic, attitudinal and behavioural indicators.

- The consumer confidence related workshops could be conducted once at the end of the evaluation 2016/17, and interim tracking of impact during implementation would rely on the strength of the theories of change in order to argue the case of causal explanations.
- Similarly, we suggest that the market monitoring should remain in place so that developments in the overall energy market are tracked continuously. However, it might be feasible to undertake this monitoring as a before-and-after study only in order to limit the research to two points in time.
- Naturally, this option still assumes that the existing relevant baseline data is used in order to monitor developments over the course of the implementation 2013-2017.

This more limited time series study would still enable data triangulation and some level of multivariate analysis, though it would be more limited in scope if the survey is shortened. This option would, however, rely on Ofgem's existing monitoring activities (in particular the Ipsos MORI Omnibus survey and the GfK Satisfaction Monitor) for on-going tracking of impact during implementation.

6.4.4 Descriptive analysis

In the options above, descriptive analysis should supplement the activities. It may also serve as a stand-alone monitoring strategy, if it is not deemed feasible to undertake tailored tracker surveys. Descriptive analysis would describe trends across the indicators but without any implied analysis of causality. As a stand-alone option, qualitative contribution tracing analyses should complement these observations to link outputs and outcomes (or impacts). This analysis could build on the existing monitoring data with a few suggested tweaks:

- existing surveys could have additional questions that would focus on the indicator where there is currently no baseline, such as 'Consumers' confidence that they are getting the best energy deal possible for their needs' (26);
- independent assessment would also need to be put in place to monitor areas, not currently covered in existing monitoring strategies such as 'Suppliers' degree of compliance to simplify the assessment of tariffs' (28); and
- the frequency of monitoring activity could be changed in some cases, such as repeating the Complaints handling survey every year rather than every two years as is currently the case.

As mentioned above, this type of descriptive analysis would not constitute an impact evaluation as it would not be possible to attribute changes in trends to the interventions. As such, it would rely on the theories of change framework and any qualitative studies to argue for causal relationships.

6.5 Potential for Economic Evaluation

As discussed in Chapter 3, it is difficult to define a counterfactual for the RMR and we consider that it is unlikely to be feasible to specify a counterfactual with any degree of confidence. Therefore, we consider that the potential for a formal economic evaluation monetising the costs and benefits of the RMR package as whole is limited.

With respect to the RMR measures that seek to simplify the assessment of tariffs, we again consider that it is unlikely to be possible to measure the costs and benefits of the policy measures against a formal counterfactual. As in previous chapters, this conclusion is based on the difficulty of specifying a counterfactual and hence the limited potential for attributing observed changes to the RMR.

However, we consider that it would be possible for a future evaluation to include some evidence of how monetary figures have developed following the introduction of the RMR.

For example, it should be possible to obtain data from suppliers on the average amount saved through an intra-supplier switch.[36] Comparing the change in this between the pre- and post-RMR periods would provide a measure of how the benefit of switching has changed over time. We understand that it would not be possible to obtain similar data for inter-supplier switches, but online comparison websites should hold those data for switches that were completed through the website. These statistics would not, however, be sufficient to attribute any change to the RMR.

Overall, we consider that it should be feasible for a future evaluation to provide an assessment of how certain monetary variables have changed over time. We consider that such an analysis would necessarily be illustrative in nature and note that it would not be possible to attribute observed changes to the RMR.

[36] We note that Ofgem can identify differences in tariff prices using its own database. However, it cannot identify the switches made by individual consumers nor the actual saving made by consumers. Suppliers will have those data for intra-supplier switches.

7 Overall Evaluation Framework

7.1 Macro-level indicators

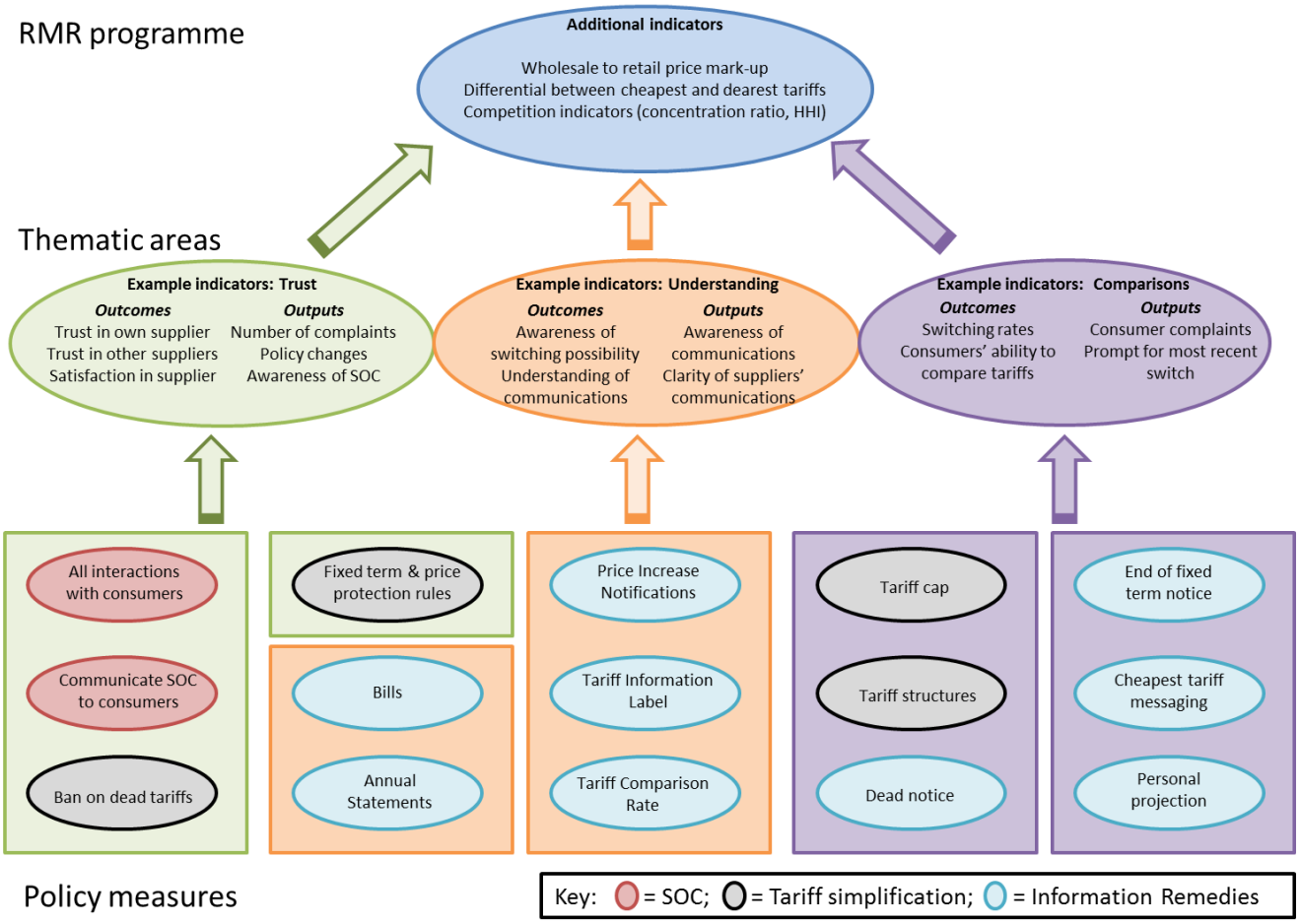
Outcomes associated with RMR policies can occur at different levels of aggregation: consumer-level, supplier-level, or overall market (societal) level. In previous chapters we have identified indicators that could be used to assess the extent to which the RMR has achieved its three key objectives, where our focus has been mainly on consumer- and supplier-level impacts. However, it would also be necessary to look at overall macro indicators that can provide some evidence that the global objectives of the RMR (e.g. increasing competition amongst energy suppliers) are being met. These macro-indicators may also provide context and alternative explanatory factors in a contribution analysis. We therefore consider that there are a number of additional impact indicators that it will be important to look at as part of the RMR evaluation. These include:

32. Wholesale to retail price mark-ups — it will be important to assess the relationship between wholesale energy prices and the extent to which price fluctuations in these are reflected differently in the mark-up on consumer tariffs pre- and post- RMR. Given data on the number of consumers on each energy tariff and the price of each tariff, it would be possible to identify how the prices charged by energy suppliers have changed over time. An observation of falling average prices relative to estimated costs (e.g. as estimated in Ofgem's weekly Supply Market Indicators) could indicate that consumers, as a whole, have captured some of the suppliers' profit margins over time. Again, however, this could not be attributed to the RMR but it would go some way towards understanding the dynamics of the overall market and the extent to which the RMR has facilitated an increase in competition and led to downward pressure on prices.
33. Differential between the cheapest and dearest (or average) tariff — if this narrows as a result of the RMR, the incentive to switch would fall and, for some consumers, may fall below their perceived inter-supplier search and switching costs. This could lead some consumers to switch within supplier (since such switches presumably impose fewer costs on the consumer) and could lead some to choose to stick with their current tariff. The observation of lower tariff differentials – particularly if combined with increased intra-supplier switches and reduced inter-supplier switches – may indicate that the RMR had led to a reduction in competition between suppliers.
34. Standard indicators of retail market competition such as concentration ratios and Herfindahl-Hirschman Indices to assess whether the market share of the largest suppliers is lower following the introduction of the RMR than it was previously.

7.2 Overall Framework for Impact Evaluation

Combining these macro-indicators with the consumer- and supplier-level indicators outlined in previous Chapters, it is possible to define an overall framework for the RMR evaluation, as illustrated in Figure 7.1.

Figure 7.1: Evaluation Framework



In a future evaluation of the RMR we would expect that all available evidence according to each policy objective would be synthesised and graded according to robustness and conclusions drawn with regards to the evaluation questions (to be defined by Ofgem). As already outlined in Chapter 3 it will not be possible to answer questions of causality through robust attribution analysis (such as RCTs or Matching). However, bringing the overall body of evidence together, particularly with comprehensive time series information, it will be possible to use multivariate analysis, coupled with qualitative assessment, to infer some judgement about the overall effectiveness of the RMR.

As discussed in Chapter 2, many policy measures have more than one objective and so there is quite some overlap between different measures, both within and between thematic areas. While the suggested evaluation design is anchored in the three thematic areas, it is important for the evaluation team to be cognisant that overlaps between policy measures exist when interpreting the results of the evaluation.

The indicators suggested operationalize the broad objectives and focus the monitoring activities on the factors which matter the most. If coupled with strong evaluation norms or performance criteria, it will be possible to assess and compare the relative success or failure of the policies' contribution to each of the three broad objectives in a cross-section analysis.

We conclude below that it is not feasible to undertake economic evaluation in the form of a formal cost-benefit analysis (by monetising the costs and benefits of the RMR package as whole).

7.3 Considerations for the Process Evaluation

In addition to the impact evaluation it would be important for Ofgem to complete a formative evaluation of RMR implementation. Process evaluation is an essential complement to the evaluation of impact as implementation will affect observed consumer-level effects and will help explain how the policies have worked and the reasons for either success or failure. In the context of the building trust proposals, the process evaluation is particularly important given the fact that some of the policies are non-prescriptive and allow for variance in implementation. The process evaluation would also be instrumental in informing an assessment of both the underpinning theories of change (theory failure) and the actual implementation (implementation failure).

The process evaluation should seek to identify:

- Descriptive elements of the actual changes made by suppliers as a result of introducing the RMR policies. This should include assessment of the extent to which implementation takes place in the form of a "tick-box" mentality among suppliers rather than genuine culture change.
- Consumer experiences of changes made: the factors driving consumers' experiences of and engagement with their suppliers and the energy market as a whole.
- Supplier behaviours over time, possibly guided by hypotheses about expected behaviours by typology of supplier. This should also explore the unintended consequences of introducing the policies.
- Identify best practices and areas of improvement. Reflections on progress made against the indicators with a view to identify best practices and weaknesses (particularly for the non-prescriptive parts of the policies), and areas which needs further attention and improvement.
- Relevance: Stakeholders' perceptions of the extent to which the interventions are relevant to the needs, problems and issues to be addressed.

In the overall synthesis and triangulation of results and impacts, the process evaluation will allow a judgement to be made on whether the impacts documented in the impact evaluation can reasonably be explained by the quality of the initial intervention logic or the quality of suppliers' implementation. Understanding the fidelity of implementation is also important if Ofgem is to be able to explain observed differences in experience across consumers that appear to have identical or very similar profiles.

8 Coherent Strategy for Data Collection

8.1 Anticipated Resource Intensity of Evaluation Options

In previous chapters we have discussed the evaluation options for each of the three core objectives (trust, understanding and simplifying tariff choices). Each option would enable contribution analysis but would require different levels of resource intensity. Because of the nature of the output and outcome indicators defined in this report, all options build on a consumer survey, whether tailored and/or existing surveys. The data gathered through the surveys on consumer perceptions could be analysed alongside aggregated supplier data, independent qualitative assessment and qualitative consumer research.

We summarise our proposed evaluation options in Table 8.1.

Table 8.1: Summary of evaluation options

Type of option	Analytical method	Main data sources	Description
High-cost option	Comprehensive time series analysis	Repeated tailor-made consumer survey with a comprehensive set of independent variables	The tailor-made survey would enable to collect data on outcomes and output indicators as well as consumers underlying attitudes. The latter are extremely important for a multivariate and sub-group analysis. Repeating the survey over time would enable to address the issue of time-lags in implementation of the policy and consumers' response.
		Suppliers' data: comprehensive summary statistics of activities	Working with suppliers to regularly obtain a comprehensive set of summary statistics related to the indicators, where relevant.
		Supplier survey	Experiences of introducing and implementing new policies
		Repeated independent assessments and qualitative consumer research	Complementing statistical data with more qualitative measures would enable the exploration of contribution and causal relationships
Medium-cost option	Limited time series or before-and-after analysis	Before-and-after tailored tracker consumer survey with a comprehensive set of independent variables; or Repeated tailor-made consumer survey with limited set of independent variables	These options are variations of the high-costs option. Because the designing of the questionnaire (with a comprehensive set of variables) will be the most costly part of the methodology, it is expected that a one-off tailor-made consumer survey with a comprehensive set of independent variables will be more costly than a repeated tailor-made consumer survey with limited set of independent variables. The latter would focus on collecting outcome and output indicators as identified in this report. Based on this it would be possible to undertake some multivariate analysis.
		Suppliers' data: limited summary statistics of activities	Working with suppliers to regularly obtain a limited set of summary statistics related to the indicators, where relevant.
		One-off targeted independent assessments and qualitative consumer research	Complementing statistical data with more qualitative measures would enable the exploration of contribution and causal relationships
Low-cost option	Descriptive analysis	Existing survey and consumer data	This approach would focus on using existing information in a systematic and consistent way overtime. Limited multivariate (or correlation) analysis might also be possible.

The tailored tracker survey would require a careful sampling strategy and we recommend a representative, random sample of at least 6,000 respondents, although a greater number is needed to enable very detailed sub-group analysis. The survey would also include a range of broader indicators, over and above the output and outcome indicators suggested above. This would allow the analysis to control for the most important independent variables that explain consumers' attitudes. Other things being equal, the more variables that are to be taken into account, the greater the cost of preparing the analysis. However, given the need to maximise response rates and accuracy, the survey should be designed so that respondents would be able to complete it in no more than 15 minutes.

For time series analysis, consistency of the survey instrument and sample profile over time is critical, but we suggest that each round of the tailored tracker survey would have different cohorts so that respondents' answers and behaviours are not influenced by having completed the survey previously. In short, we believe that there is a bias introduced by the fact that repeatedly asking questions about consumers' behaviour on the retail energy market may have an impact on respondents' behaviour. Another way to follow cohorts would simply be to use suppliers' information and analyse, for example, switching behaviours over time, but there may be legal restrictions to accessing such data at an individual level.

The proposed evaluation options do not allow for strong inference of causality or attribution. Because of this, qualitative consumer research and independent assessments would strengthen the contribution analysis and trace links of causality. At the very least, suppliers' implementation of the policy over time should be carefully monitored. This type of descriptive analysis will be a critical cornerstone of all three options.

8.2 Recommendations for Ofgem's Future Data Collection Strategy

In earlier chapters we identified 31 key indicators that have the potential to measure the extent to which the RMR achieved its objectives, if included in a future evaluation of the policy. We also considered the extent to which these indicators are already collected by Ofgem (or other parties).

In this section, we provide a number of recommendations on Ofgem's future data collection strategy. Our recommendations are designed to ensure that there will be no duplication of effort between different policy teams as Ofgem decides on the type of data that it should collect for the purposes of enhanced monitoring and impact evaluation. Where possible, we suggest relying on observed data rather than, say, the self-reported actions of consumers due to the greater reliability of such data. However, it is essential to secure consumers' perspectives for the majority of indicators.

As discussed in Chapter 3, we note that certain RMR proposals will be implemented prior to other proposals. We consider that this does not present any significant difficulties for the data gathering strategy, subject to the condition that there is a clear understanding of the status of each policy measure at the point of data collection. Given this understanding, some indicators gathered during 2013 and 2014 would be treated as baseline observations while others would be treated as early impact observations. Our recommendations are summarised in Table 8.2.

Table 8.2: Recommendations for Ofgem's future data collection strategy

Source	Indicator (Thematic area)	Baseline data to be gathered	Required frequency	Required segmentation
Omnibus survey	Consumers' trust in own and other energy suppliers (trust) Consumers' understanding of energy market overall (understanding) Consumers' awareness of possibility of switching tariff and/or supplier (understanding) Consumers' understanding of regular communications (understanding) Consumers' awareness of suppliers' communications (understanding) Consumers' level of engagement with suppliers' communications (understanding) Consumers level of activity (simplification) Use of comparison services (simplification) Consumers' confidence that they can compare tariffs easily (simplification) Consumers' prompt for most recent switch (simplification) Consumers' perceptions of making monetary savings from switching tariffs (simplification)	No additional data to be gathered from this source	Annually	Fuel, payment method, vulnerability, several socio-economic variables
Qualitative research	Can be used to support assessment of causality for numerous indicators Can also be used to provide key evidence for the following indicators: Consumers' understanding of energy market overall (understanding) Consumers' understanding of own energy consumption (understanding) Consumers' understanding of current tariff (understanding)	No additional data to be gathered from this source	Ad hoc	N/A
Tailored consumer tracker survey	Should ideally include all consumer-related indicators plus independent variables to allow for detailed sample segmentation. Priority additional indicators to gather though	N/A	Annually	Supplier, fuel, payment method, several socio-economic variables, vulnerability,

Source	Indicator (Thematic area)	Baseline data to be gathered	Required frequency	Required segmentation
	this survey (which are not currently collected elsewhere) are: Consumers' understanding of own energy consumption (understanding) Consumers' understanding of current tariff (understanding) Consumers' confidence in accuracy of information received from suppliers (trust) Consumers' confidence that they are getting the best energy deal possible for their needs (simplification)			level of engagement
Complaints handling survey	Proportion of complaints by category (trust, understanding, simplification) Satisfaction with treatment of complaint (trust)	No additional data to be gathered from this source	Annually (rather than every two years)	Supplier, fuel, payment method, reason for complaint
DECC Public Attitudes Tracker	Trust in energy supplier (trust)	No additional data to be gathered from this source	Quarterly	Supplier, fuel, payment method, several socio-economic variables
Suppliers	Switching statistics: intra- and inter-supplier (simplification) Expected saving per intra-supplier switch, based on Personal Projection (PP) (simplification) Number switches in response to Cheapest Tariff Message (CTM), by original meter type (simplification) Number of complaints (trust, understanding, simplification) Number of consumers on dead and fixed term tariffs(trust) Number and nature of interactions with customer service (trust) Suppliers' introduction and implementation of new policies (trust)	Intra-supplier switching statistics Typical saving per intra-supplier switch N/A Complaints data No additional data to be gathered No additional data to be gathered No additional data to be gathered	Monthly	Fuel, payment method, previous supplier (if known), CTM/not
Price comparison services	Number of unique searches / contacts (simplification) Number of searches / contacts that led to a switch (simplification)	As much historical as possible	Monthly	Fuel, payment method, old and new supplier

Source	Indicator (Thematic area)	Baseline data to be gathered	Required frequency	Required segmentation
	Expected saving per switch, based on PP (simplification)			
GfK Energy Satisfaction Monitor	Consumers' satisfaction with their energy supplier (trust) Consumers' level of satisfaction with interactions with their suppliers (trust) Shock billing (trust)	No additional data to be gathered from this source	Quarterly	Supplier, fuel, payment method, whether customer has switched supplier
Ofgem	Number of complaints (trust, understanding, simplification)	No additional data to be gathered from this source	Annually	Supplier, fuel, payment method, reason for complaint
Energy Ombudsman	Number of complaints (trust, understanding, simplification)	No additional data to be gathered from this source	Annually	Supplier, fuel, payment method, reason for complaint
Citizen's Advice	Number of contacts (trust, understanding, simplification)	No additional data to be gathered from this source	Annually	Supplier, fuel, payment method, reason for complaint
Consumer Futures	Number of complaints (trust, understanding, simplification)	No additional data to be gathered from this source	Annually	Supplier, fuel, payment method, reason for complaint
Academic studies	Clarity of regular communications including bills, Annual Statement and Price Increase Notices (understanding)	No additional data to be gathered from this source	Single point in time	N/A

Appendix: Sample Size for the Proposed Tailored Tracker Survey of Consumers

A number of factors need to be taken into account when considering the sample size needed for a survey, including the overall levels of confidence required, the analytical strategy (i.e. the preferred number of sub-groups and intended type of analysis) and the budget.

The tailored tracker survey suggested for the evaluation of the RMR would enable sub-group analysis based on the following layers:

- Individual-based (e.g. segments of consumers). We envisage no more than six segments, based on cluster analysis.
- Type of tariffs. The number of tariffs varies by suppliers; following the introduction of the RMR proposals, some suppliers will offer up to four core tariffs for each of the three payment methods, but other suppliers will offer fewer tariffs.
- Type of suppliers. Ofgem's data shows that there are currently 18 electricity and 16 gas suppliers active in the domestic market. We note that the majority of customers are supplied by one of the Big 6 and hence the combined market share of other suppliers is currently small in comparison. For that reason, we assume that the emphasis of any future analysis would be on customers of the larger suppliers.
- Area-based. There are 14 regions in the UK. These relate to the old monopoly electricity supply regions and are the standard regional definition used in the energy supply industry.

The suggested survey sample size would generate 6,000 respondents to the survey. This recommendation is based on an assumption that the main type of sub-group analysis will be based on segments of consumers. With a maximum of six segments a survey sample of 6,000 would, if evenly spread, provide 1,000 respondents per sub-group. This would generate results of a very acceptable confidence interval of +/- 3.1%

In reality, however, the segments of consumers would not be evenly spread across each cluster. Assuming that a cluster has a minimum of 500 respondents, this would generate results within the confidence interval of an acceptable range of +/- 4.4 per cent.

Ofgem may require even more detailed breakdowns, such as segments of consumers by type of tariff. Assuming there are 10 main tariff / payment method combinations and 6 segments of consumers (evenly spread, for arguments' sake), each sub-group would contain 100 respondents. This would generate results within the confidence interval of +/- 9.8 per cent.

Another example of breakdown would be segments of consumers by type of supplier. Where similar results would be generated if they analysis is based on 10 clusters of suppliers (such as the six big and four other clusters of smaller suppliers) and 6 consumer segments.

Going one level further, such as segments of consumers by the full set of suppliers and region, would begin to challenge the robustness of results. A much larger survey sample would be required to go to that depth of analysis and, overall, we consider that this would not generate significant additional insights which would warrant the level of investment required.

On the basis of the above, we consider that a random sample size of 6,000 would be the most cost-effective way in which to meet Ofgem's analytical needs.