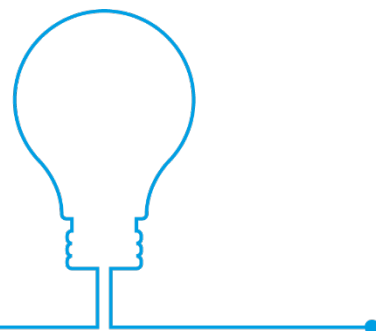




Department for
Business, Energy
& Industrial Strategy

EVALUATION OF THE TRANSITIONAL ARRANGEMENTS FOR DEMAND SIDE RESPONSE

Phase 2 – Appendices



December 2017

Contents

| | |
|--|----|
| Contents | 1 |
| Appendix 1. Introduction to the Capacity Market and Transitional Arrangements for DSR and small-scale generation | 3 |
| Introduction to the TA and CM | 3 |
| Testing requirements | 7 |
| Baseline methodology for DSR CMUs | 11 |
| Fulfilment of obligations | 12 |
| Appendix 2. Initial theoretical framework | 14 |
| Appendix 3. Revised theoretical framework | 20 |
| Appendix 4. Generative causation approaches | 29 |
| Applying contribution analysis within realist evaluation | 29 |
| Contribution tracing with Bayesian updating | 32 |
| Appendix 5. Contribution tracing paper | 34 |
| 1. Introduction | 34 |
| 2. Methodology for Phase 2 contribution tracing tests | 37 |
| 3. Revisions since Phase 1 contribution tracing | 38 |
| 4. How we have applied these tests | 54 |
| 5. Summary of contribution tracing test findings | 60 |
| 6. Sensitivity of results to probability assumptions and coding | 68 |
| 7. How the contribution tracing findings have been used | 73 |
| Appendix 6. Methodology for email survey of TA participants in autumn 2016 | 74 |
| Purpose | 74 |
| Email survey | 74 |
| Email survey findings | 75 |
| Appendix 7. Methodology for in-depth interviews with TA participants and aggregator clients in spring 2017 | 76 |
| Introduction | 76 |
| Sampling and recruitment | 76 |
| Interview approach | 78 |

| | |
|---|-----|
| Analysis approach _____ | 80 |
| Limitations _____ | 81 |
| Appendix 8. Research instruments for the email survey and in-depth interviews _____ | 83 |
| Email survey _____ | 84 |
| Invitation email for in-depth interview _____ | 90 |
| Pre-interview preparation _____ | 91 |
| Introductory script _____ | 91 |
| Phase 2 master topic guide – TA 2016-17 delivery year participants _____ | 93 |
| Appendix 9. Methodology for DSR and SPD case studies _____ | 107 |
| Purpose _____ | 107 |
| Approach _____ | 107 |
| Data sources _____ | 108 |
| Case study findings _____ | 108 |
| Conclusion _____ | 109 |

Appendix 1. Introduction to the Capacity Market and Transitional Arrangements for DSR and small-scale generation

Introduction to the TA and CM

The TA is a pilot and forms part of the Capacity Market (CM) for security of electricity supply, within the government's Electricity Market Reform (EMR) programme. The TA aims to support BEIS's overall objectives of promoting growth and energy security, while ensuring affordability of the energy supply.

The TA aims to encourage development of Demand-Side Response (DSR) and small-scale distribution-connected generation that is increasingly needed to balance supply and demand in a decarbonised electricity grid¹. In this report we used the CM definition of DSR: the activity of reducing the metered volume of imported electricity of one or more customers below an established baseline, by means other than a permanent reduction in electricity use. By this definition, DSR may be achieved through any combination of onsite generation, temporary demand reduction or load-shifting. We refer to the last two activities as 'turn-down' DSR.

The TA scheme involves two auctions for specific types of capacity within the CM, the first for delivery of capacity in the 2016/17 delivery year² and the second for delivery of capacity in 2017/18. These TA auctions are additional to the main CM auctions: the main four-year ahead auctions (T-4) and the smaller one-year ahead auctions (T-1) which will deliver capacity from 2018/19 onwards and the Early Auction which BEIS introduced to deliver capacity in 2017/18³.

The TA has three main objectives, which we have used as the basis of three project hypotheses to be tested by the evaluation:

1. To contribute to security of electricity supply to help with short-term forecasted system tightness (winter 2016/17 and winter 2017/18).

¹National Infrastructure Commission (2016) *Smart Power: A National Infrastructure Commission Report*. Available at: <https://www.gov.uk/government/publications/smart-power-a-national-infrastructure-commission-report>. Accessed 27/7/2016

²The delivery year runs from 1st October of one year through to 30th September of the following year.

³<https://www.gov.uk/government/news/government-announces-capacity-market-auction-parameters>

Appendix 1. Introduction to the Capacity Market and Transitional Arrangements for DSR and small-scale generation

2. To develop a stock of flexible capacity⁴ that can be available for the one year ahead (T-1) auction in 2017 for delivery in 2018/19, thereby contributing to liquidity in this and subsequent year-ahead auctions.
3. To encourage enterprise and develop experience, confidence and understanding so that DSR and embedded generation will be able to realise their potential and ultimately compete with larger generation assets in the CM.

The TA is designed to be a stepping stone to the main CM for flexible capacity⁵ that might have difficulty in competing in the main CM. The TA is open to both direct participants, putting forward their own capacity, and aggregators, putting forward capacity on behalf of clients. To partake in the first TA auction, Capacity Market Units (CMUs) have to be between 2 MW and 50 MW in size, but may comprise multiple components across different sites and organisations. The minimum CMU size was reduced from 2MW to 500KW for the second TA auction. As in the main CM, TA participants are required to prequalify and, if successful, will bid in the TA auction to indicate the price at which they would be willing to make this capacity available when required by the System Operator. This is a 'pay as clear' auction, in which all participants who are successful receive the auction clearing price⁶. Unproven DSR and New Build CMUs must, as a condition of their prequalification, submit credit cover for their CMUs.

Participants successful in the auction are awarded a capacity agreement for their CMU(s) which sets out their obligations to deliver capacity if there is a CM system stress event⁷, either by providing generation capacity or by reducing demand below their baseline. They have to meet CM requirements for metering and DSR testing, and risk their agreement(s) being terminated if they fail. If one or more CM stress events happen during the delivery year, TA participants have the option of delivering their capacity obligation or paying penalties. Conditions in the TA are softer than the main CM, to encourage new entrants: the level of credit-cover collateral is set at £500/MW, which is 90% lower than the main CM, and there is a time-banded option, which only requires response to stress events occurring between 9-11am and 4-8pm on winter week days. The penalty regime is the same as the main CM: penalties cannot exceed the original CM payment set by the auction clearing price.

While the TA does not automatically lead on to future CM participation, it aims to build capacity and confidence so that providers of DSR and small-scale generation are better placed to compete in future CM auctions. A schematic representation of the different steps in the TA is shown below.

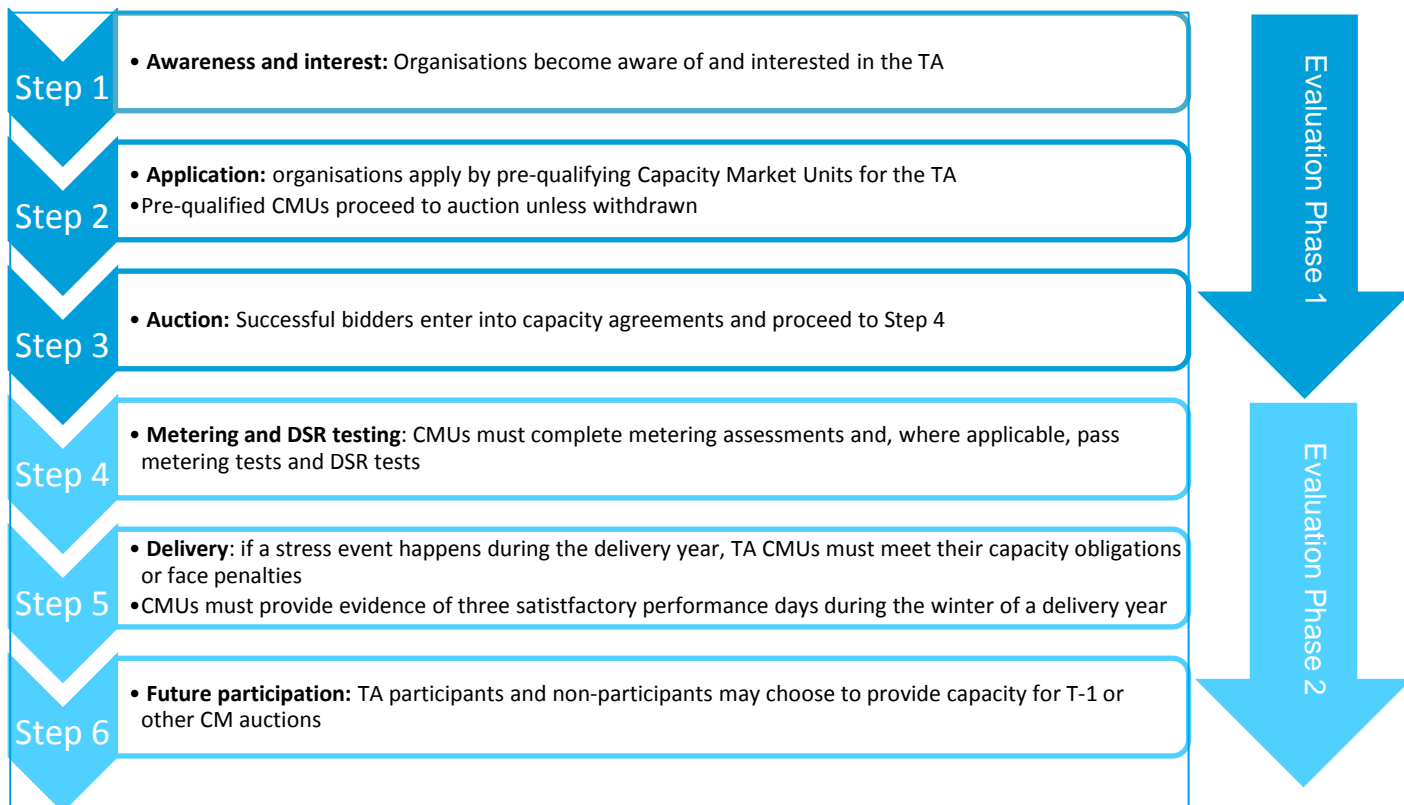
⁴ Flexible capacity means electricity generating capacity and demand that is able to increase or decrease in response to signals, to help balance supply and demand of electricity across the GB grid.

⁵ Ofgem defines flexibility as 'modifying generation and/or consumption patterns in reaction to an external signal (such as a change in price) to provide a service within the energy system'.

⁶ Bidding strategies are likely to differ between 'pay as clear' auctions (where participants tend to bid their own supply costs, knowing that they will receive the clearing price if successful) and 'pay as bid' auctions (where participant bids are influenced by their estimate of the bid price for the last unit likely to clear the auction).

⁷ A stress event is a period in which the electricity supply/demand balance is too tight (as determined by National Grid algorithms).

Figure A1.1 Steps in TA process



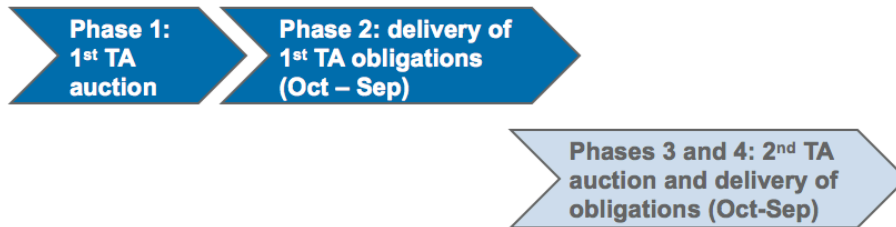
The first TA auction, which is evaluated in this report, was open to providers of DSR and small-scale distribution-connected generation services. In May 2016, BEIS announced changes to the CM, including narrowing the second TA auction to ‘turn-down’ DSR only⁸. The second TA will be researched in Phases 3 and 4 of the evaluation.

The timeline for the first and second TA and other capacity market auctions is shown in Figure A1.2 below.

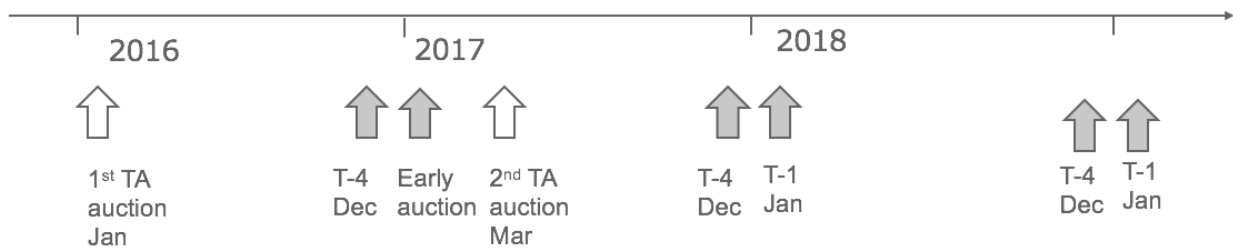
⁸ Department of Energy and Climate Change (2016) *Government Response to the March 2016 Consultation on Further Reforms to the Capacity Market*. Report 16D027. Available at: <https://www.gov.uk/government/consultations/consultation-on-reforms-to-the-capacity-market-march-2016>. Accessed 27/7/2016

Figure A1.2 Timeline for TA auctions and evaluation

Phases of this evaluation



Capacity market auctions



Testing requirements

Following the award of capacity agreements for the first TA auction, participants had to pass standard CM tests to confirm their capacity. The tests are summarised in Figure A1.3 and explained further below.

Figure A1.3 Summary of Capacity Market testing requirements

Metering assessments are required for all Capacity Market Units (CMUs) to determine which metering option applies to each of their sites. Three metering options qualify, as follows:

- (a) Supplier settlement metering;
- (b) Bespoke metering; and
- (c) Balancing services metering

Metering tests are required for sites using metering options (b) and (c), but not option (a).

DSR tests are required for unproven DSR CMUs to demonstrate that they can deliver the required demand reduction against a measured baseline of demand. The 'proven' capacity of the CMU reflects the outcome of DSR testing.

Three 'satisfactory performance days' (SPDs) are required for all CMUs to demonstrate that their capacity remains available through the winter delivery period.⁹

About metering assessments

All sites within a CMU required a metering assessment, as part of the operational readiness checks prior to Capacity Market participation. The purpose of the metering assessment is to ensure that each metering set-up accurately reflects the energy use on site so that the performance of the capacity obligation can be observed; and that metered data is appropriately assured and regularly submitted to the settlement body in a suitable format.

The CM rules stipulate accurate metering. For generation, this is metering of output; for demand reduction, this is metering of a demand reduction against a measured baseline. Three metering options qualify, as follows:

- (a) Balancing and Settlement Code (BSC) Supplier or Balancing Mechanism Unit (BMU) metering, generally referred to as 'supplier settlement metering';
- (b) Bespoke metering; and
- (c) Balancing services metering.

⁹ The winter delivery period is defined as 1st October 2016 to 30th April 2017. Demonstration of satisfactory performance days would not be required if there had been three stress events within this period.

Option (a) is the default metering for electricity market settlement. It is well understood and governed by industry codes of practice. Whilst all participants will have some form of settlement metering, they might not want to use these meters for the CM (e.g. because the settlement meter measures capacity at the boundary of a site, net of onsite loads). In these cases they may wish to install bespoke metering (option (b)), or use metering specifically in place for participation in National Grid-run balancing services (option (c)).

Accurate metering is also required for any renewable generation assets behind the meter that receive government subsidy (e.g. the Feed-In-Tariff). Separate metering is required for these assets so that subsidised renewable generation can be netted off the capacity offered to the CM, avoiding double-subsidy. The metering requirements for the CM are more demanding than those required for the Feed-In-Tariff.

Options (b) and (c) require a metering test for each meter (i.e. multiple tests per CMU if multiple CMU components were using these options). This is because accuracy could not be taken for granted as industry codes do not govern these metering options.

For CMUs that know their metering arrangements in advance (i.e. existing generation and proven DSR¹⁰), metering assessments are submitted as part of the pre-qualification process. New-build generation must submit them when operational, while unproven DSR can defer their assessments by up to one month prior to the delivery year. In the pre-qualification process for the first TA auction, all unproven DSR CMUs deferred their metering assessments.

Metering statements

CMU components that require a metering test must complete a metering statement. The requirements of a metering statement are detailed in Schedule 6 of the CM regulations.

The whole metering system encompasses the meter device, current, voltage and power transformers, data collection systems and communication system. All these elements contribute to overall metering accuracy.

There is evidence (presented in the main report) that many participants found it challenging to meet these accuracy requirements, particularly for metering options (b) and (c).

Our review of Schedule 6 requirements suggests that participants may have struggled with these requirements, not because of the accuracy requirements per se, but because of the type of metering systems to which the requirements were being applied. So, for example, settlement metering systems (option a) would tend to be installed for settlement purposes, and be specified for the required accuracy. They would be designed as a package to log readings and communicate with settlements in the appropriate data format.

In contrast, retrospective application of the same requirements to sub-metering systems (options (b) or (c)) presents more challenges, as sub-metering systems were not intended for such a role. These systems may have disparate components for measuring, logging and communicating data in various formats. Accuracy

¹⁰ No proven DSR CMUs were submitted to the first TA auction. All the DSR CMUs were classed as unproven.

may be difficult to establish without visual inspection and / or on-site checks. Moreover, it may simply be extremely time-consuming to provide evidence of meter accuracy for multiple sub-metering component parts.

Metering tests

For participants successful in the first TA that required a metering test, a test had to be completed and a certificate had to be issued by 31 August 2016 – one month prior to the delivery year for the first TA. Once the metering statement has been completed, the test itself was simply a desk-based review of the metering statement submissions, which included a CSV file from the day of meter commissioning.

Electricity Market Reform Settlements Limited (EMRS) could nominate sites for a site visit for further validation of, for instance, the location of meters and associated equipment. Interview evidence indicated that site access could be complex and time-consuming to arrange owing to health and safety requirements, particularly if the site housed critical infrastructure and/or processes needed to be shut down to allow inspection. Interview data indicated that replacing metering systems or equipment was sometimes more cost-effective than providing documentation and arranging site visits for old systems.

About the DSR test

The DSR test is used to verify that all CMUs can achieve their capacity output. The test establishes a baseline and then calculates the capacity reduction of a DSR component by comparing a test data point against the baseline. The following rules apply for the timing of the tests:

- The DSR test can be conducted prior to the prequalification window, in which case the applicant, if successful, can apply as a proven DSR CMU; or
- After the award of the capacity agreement but no later than one month prior to the delivery year, in which case the applicant can apply as an unproven DSR CMU.
- A DSR test may not take place during the prequalification assessment window.

The DSR test process is as follows:

1. The CMU must submit to the System Operator (i.e. National Grid):
 - Metering Point Administration Number(s) (MPANs) of the meters for the site for all components.
 - A metering test certificate or confirmation that the CMU has a Capacity Market (CM) approved meter configuration.
2. Historic test – 3 x 30mins¹¹ over the last two years can be evidenced from balancing services; or
3. Live test – two working days' notice of the CMU's intention to test the DSR CMU, together with the Settlement Period in which the activation will be carried out.

¹¹ Settlement Period or DSR Alternative Delivery Period (i.e. 30mins that is not on the hour or half hour)

4. National Grid has 5 days from receipt of meter data from the Settlement Body to calculate:
 - Baseline Demand (over the 6 week baseline period);
 - the DSR evidenced (which can be zero); and
 - the Proven DSR Capacity
5. Following a successful test National Grid must provide a DSR test certificate in 5 days. CMUs have the option for a further retest.

A DSR Test Certificate remains valid for so long as the components in a DSR CMU remain the same. Where they do not, the certificate will be invalidated and the CMU will be deemed to be an Unproven DSR CMU until such time as a new DSR Test Certificate has been issued.

About the Satisfactory Performance Days (SPDs)

Satisfactory Performance Days are intended to check during the delivery period that the CMUs are still available to achieve their capacity output. Each CMU must nominate three half-hour settlement periods, on different days within the winter period (between 1st October and 30th April), when they were delivering their full capacity. For DSR CMUs, delivery for a system stress event over the winter period can also count as an SPD even if the load following capacity obligation is lower than the full capacity obligation; similarly delivery of capacity in response to a request for a balancing service can also count as an SPD for DSR CMUs.

As noted above, the baseline methodology for SPDs is the same as that of the DSR test (see below). However, the participants can retrospectively nominate any half hour periods of their choice within the winter period. The intention is to minimise disruption to the participants, in that they can choose a time when the DSR asset is in the required operational state for other reasons. For example, generation assets might be being operated for Triad or turn down assets might be switched off during a holiday period.

If a CMU fails to demonstrate 3 SPDs over winter, the CMU's capacity payments is suspended until 3 SPDs have been met.

If a CMU fails to deliver output of 1kWh during system stress events in 2 or more months, the CMU is required to demonstrate 6 SPDs over winter, instead of 3.

Credit cover

Participants with unproven or new-build CMUs in the first TA were required to provide £500/MW credit cover. If a participant with a prequalified unproven DSR CMU nominates a lower bidding capacity or fails to deliver 90% of the bidding capacity DSR tests and SPDs, then the credit cover is lost (100% for joint tests). This mechanism is designed to dissuade speculative bids into the auction.

Credit cover in the main CM is significantly higher: for new-build CMUs it is now £10,000/MW and for unproven DSR CMUs it is currently £5,000/MW.

Joint DSR Test and SPDs

A rule change was introduced in 2016 (Rule 13.2B) that allows several CMUs that have the same Capacity Provider to be tested together for the purposes of DSR tests and SPDs.

If a capacity provider fails to meet its combined obligation, it will not receive a DSR test certificate. The CMU is able to have one retest. Each CMU under joint test has its own DSR test certificate. Any changes to the composition of the CMUs involved in a joint DSR test will result in the certificate becoming invalid and a new test will be needed.

In contrast to tests for single CMUs, where participants must demonstrate 90% of their auction capacity in order to avoid losing credit cover (see above), 100% of the capacity must be demonstrated under a joint test.

Baseline methodology for DSR CMUs

Baseline methodology for DSR tests and SPDs

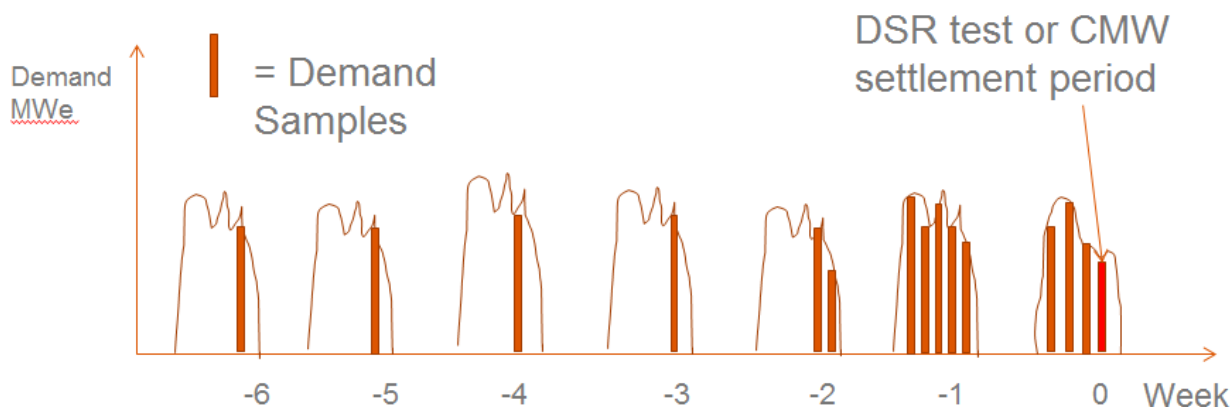
A reduction in energy demand for a DSR CMU cannot be measured directly, only estimated by comparing actual demand against what demand would have been under the same conditions (i.e. establishing a counterfactual). The baseline methodology seeks to provide a fair representation of how a DSR asset would have performed in the absence of the DSR test, SPD or stress event. The methodology must balance issues of accuracy, integrity (avoiding gaming), simplicity and alignment to the goals of the programme.

The baseline is calculated as the average of half hourly Demand Samples relative to the nominated test Settlement Period, with the Demand Samples selected as follows:

- the same Settlement Period on the same day of the week for each of the last six weeks (if a sample falls on Non-Working day i.e., a Bank Holiday, then that sample is disregarded); and
- where the Settlement Period for which the baseline is being calculated is on a Working Day, on the last ten Working Days; and
- where the Settlement Period for which the baseline is being calculated on a Non-Working Day, on the last ten days that are a Non-Working Day,

Depending on the date, up to 6 of the 16 data samples can overlap. The greatest overlap occurs if a test or stress event is on a Saturday in a period without bank holidays. If the event or test is on a Working Day, as is mostly commonly the case, then there are two 2 overlaps and 14 unique measurements.

Figure A1.4 Example of baseline half-hour samples on a Working Day



Baseline methodology for stress events

The baseline methodology for stress events depends on whether the component of the CMU has responded to a balancing services call or not. If it has, then an 'Adjusted Demand Sample' is used, whereby the baseline is adjusted using a Pre-Capacity Market Warning (CMW) Adjustment. The Pre-CMW Adjustment is calculated as the average of the difference between the provisional baseline and the actual demand during the 6 Settlement Periods before the capacity market notice Settlement Period. The Pre-CMW adjustment is expressed as a positive number if actual Demand is greater than the Provisional Baseline and as negative number if it is less. This can result in a positive or negative adjustment. The baseline is then calculated as:

$$\text{Adjusted Baseline} = \text{Provisional Baseline Demand} + \text{the Pre-CMW Adjustment}$$

If the component has not responded to a balancing services call, then the baseline methodology used for the DSR test and SPDs applies (the 'Provisional Baseline Demand').

Fulfilment of obligations

Participants with CMUs in the TA must deliver against their Capacity Obligation at any time of system stress during the Delivery Year, or face a financial penalty. A 'System Stress Event' means a Settlement Period in which a System Operator Instigated Demand Control Event occurs where such event lasts at least 15 continuous minutes.

TA participants are required to deliver the 'Adjusted Load Following Capacity Obligation' (ALFCO) for all of their units during a 'System Stress Event'. The ALFCO is a period of (involuntary) load reduction, by voltage reduction or demand disconnection, which is necessary to maintain the security of the system in the event of a shortage of generation.

Capacity Market Notices (CMNs) are issued by National Grid when a shortage of generation is anticipated. The CMN is a signal to all providers that system stress is anticipated. Capacity providers are not 'called upon' to deliver capacity or receive an individual despatch instruction. The Capacity Market Notice is a signal to all providers that system stress is anticipated (although may not materialise). Four hours after the issue of

Appendix 1. Introduction to the Capacity Market and Transitional Arrangements for DSR and small-scale generation

the Capacity Market Warning, if a System Stress Event occurs, any participant who fails to deliver their ALFCO will be subject to penalties.

At the time of this research, two CMNs had been issued:

- One issued on 31st October at 12.06pm, which was live from 4.30pm to 7.00pm. The cancellation notice was issued at 6.53pm.
- One issued on 7th November at 12.06pm. This was due to go live at 4.30pm but was cancelled at 3.07pm.

Neither CMN developed into a System Stress Event.

After a live CMN, the delivery body determines whether a System Stress Event has occurred. This is determined retrospectively, by examining the balance between supply and demand in the GB electricity system.

If a System Stress Event has occurred, the settlement body compiles meter data for all CMUs which had a capacity obligation at the time, and uses this to assess whether each CMU met its Adjusted Load Following Capacity obligation (ALFCO). Subsequent penalties/over-delivery payments are determined once all data has been submitted, which is no later than 9 working days after the end of the month the System Stress event takes place in.

Penalties for failure to deliver for a specific System Stress Event are related to a provider's Capacity Market Payment. Penalties for repeated failure to deliver are capped at 100% of a Capacity Provider's annual Capacity Market Payment with respect to a CMU, and at 200% of a CMU's monthly Capacity Market Payment.

Appendix 2. Initial theoretical framework

The theoretical framework from Phase 1, as reported in the Phase 1 evaluation report, was rationalised and updated at the start of Phase 2 by the project team, based on findings from Phase 1 evidence. The initial theoretical framework for Phase 2 of the TA evaluation comprises the following parts:

Figure A2.1: Delivery theory: Testing and fulfilment

Figure A2.2: Participation theory

Figure A2.3: Non-participation theory

Figure A2.4: Additionality theory for hypotheses H1-H3

Figure A2.5: Overall map of high-level TF, showing how these elements fit together.

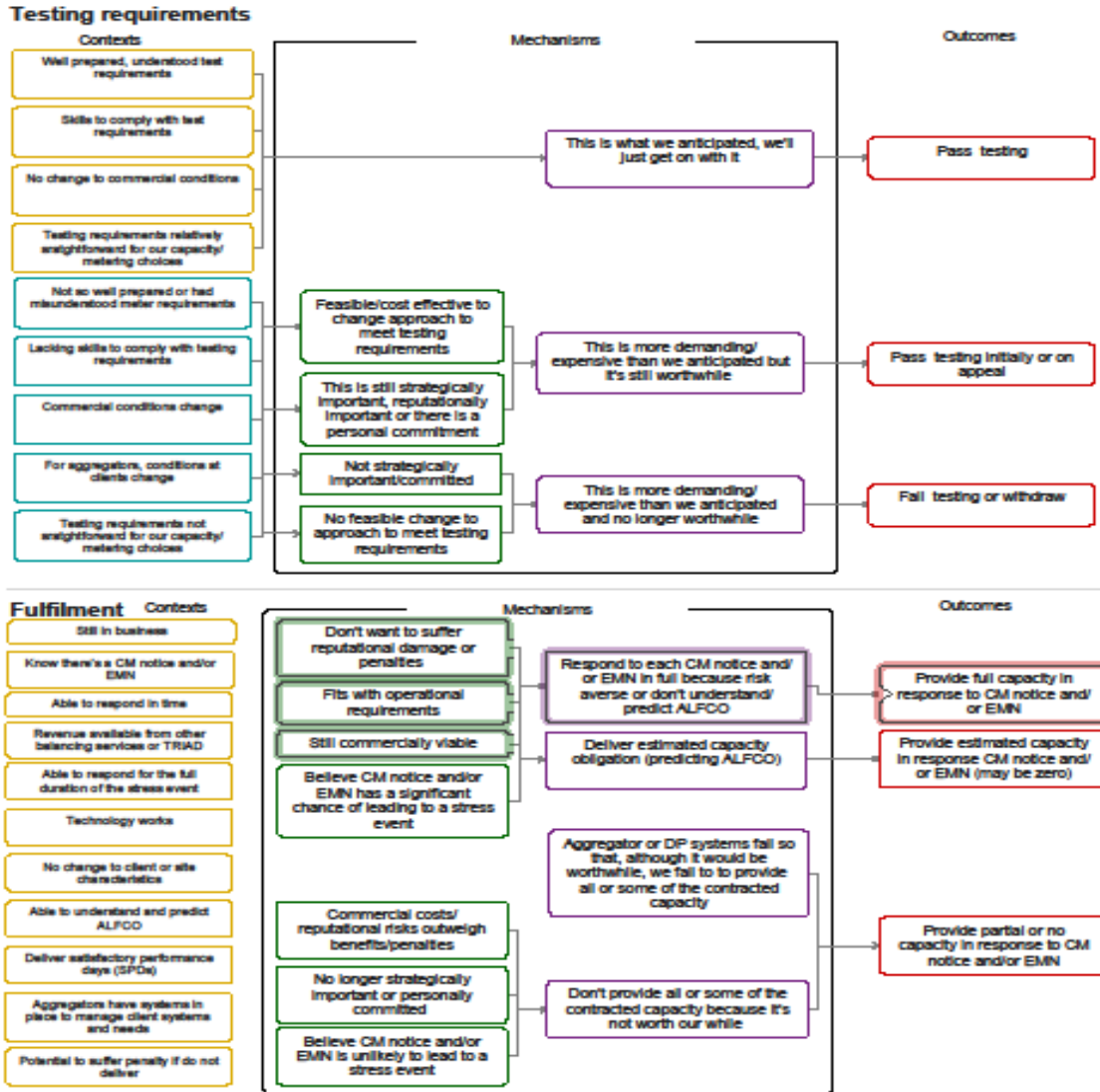
These elements of theory relate to the Phase 1 theoretical framework as follows:

| Phase 1 theory | Phase 2 theory |
|-------------------------------------|--|
| Step 1 – awareness and interest | Participation and non-participation theory |
| Step 2 – application | |
| Step 3 – auction | |
| Step 4 – metering and DSR testing | Testing theory |
| Step 5 – delivery | Fulfilment theory |
| Step 6 – future participation in CM | Additionality theory for H2 and H3 |

Each theory is set out below in realist terms, showing the different ‘contexts’ that are expected to trigger various ‘mechanisms’ on the part of TA participants, leading to a range of ‘outcomes’ for the TA.

Figures A2.1 to A2.3 should be read from left to right. The participation and non-participation outcomes are implicit in Figures A2.2 and A2.3. Figure A2.4 shows additionality contexts and mechanisms on the left (read from left to right) and non-additionality contexts and mechanisms on the right (read from right to left), with the relevant outcome in the centre of the diagram. Figure A2.5 is an overarching theory diagram, showing how the participation, non-participation and additionality theories fit together.

Figure A2.1 Delivery theory: Testing and fulfilment



Delivery theory [5]

Figure A2.2 Participation theory

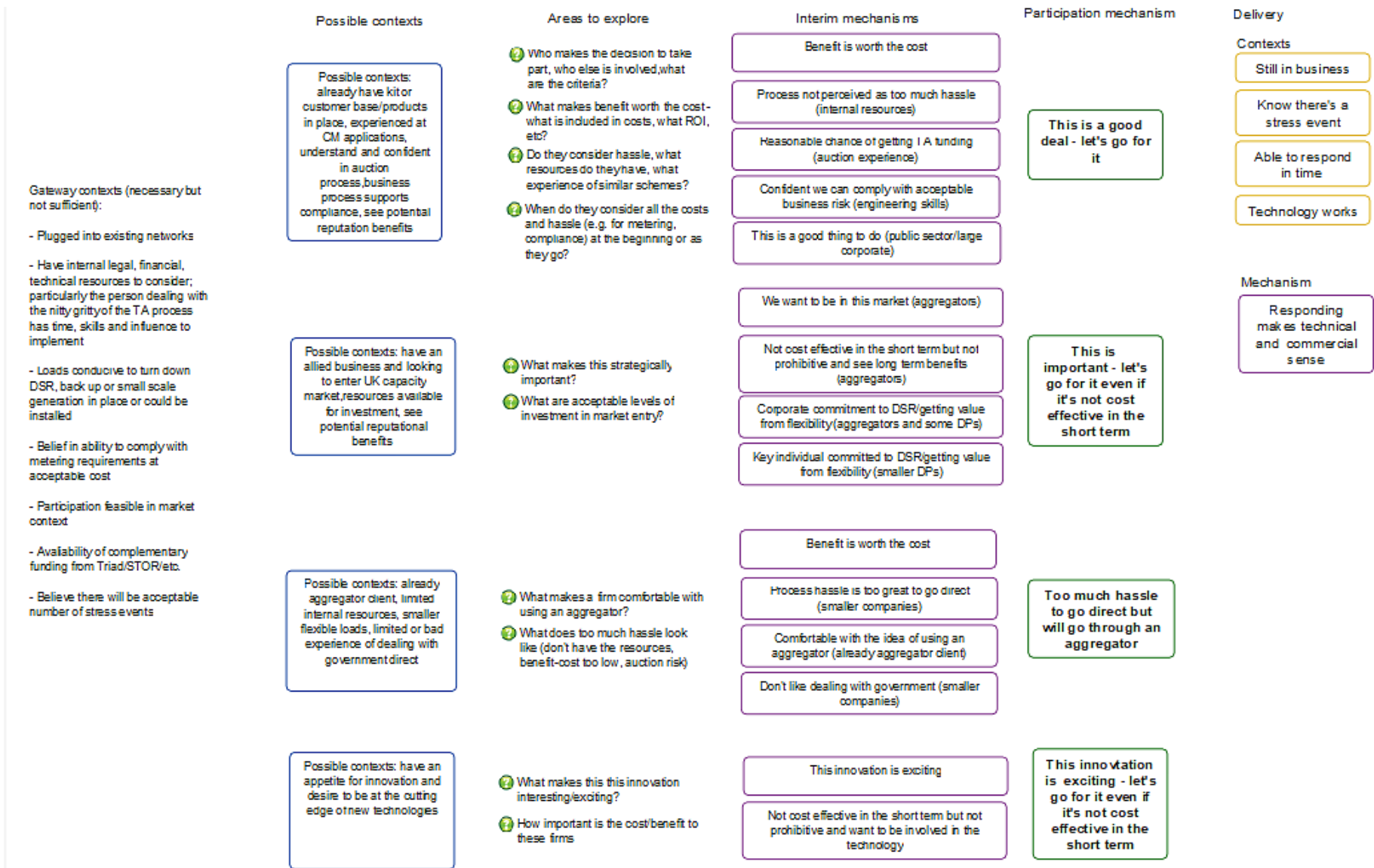
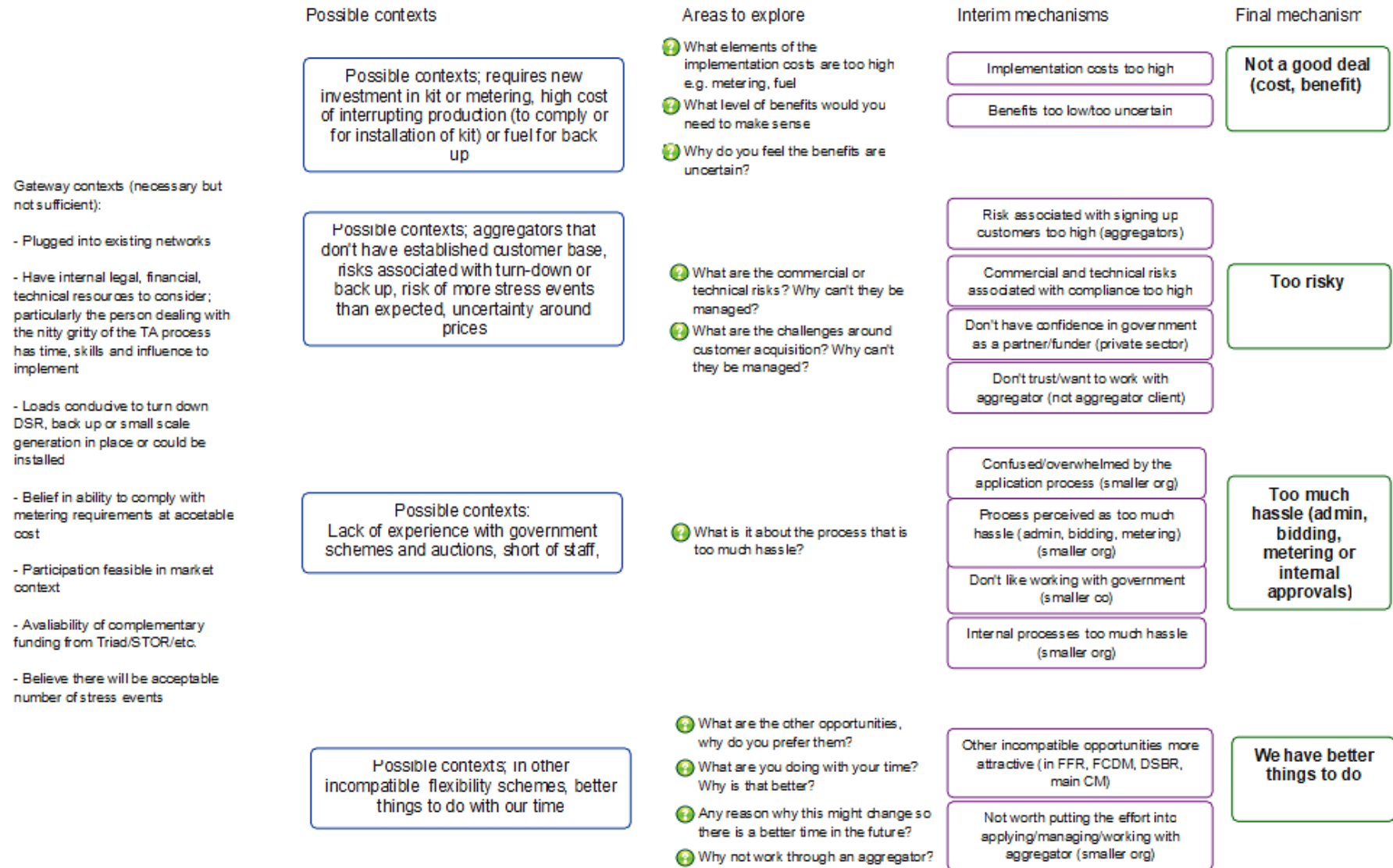


Figure A2.3 Non-participation theory



Appendix 2. Initial theoretical framework

Figure A2.4 Additionality theory for hypotheses H1-H3

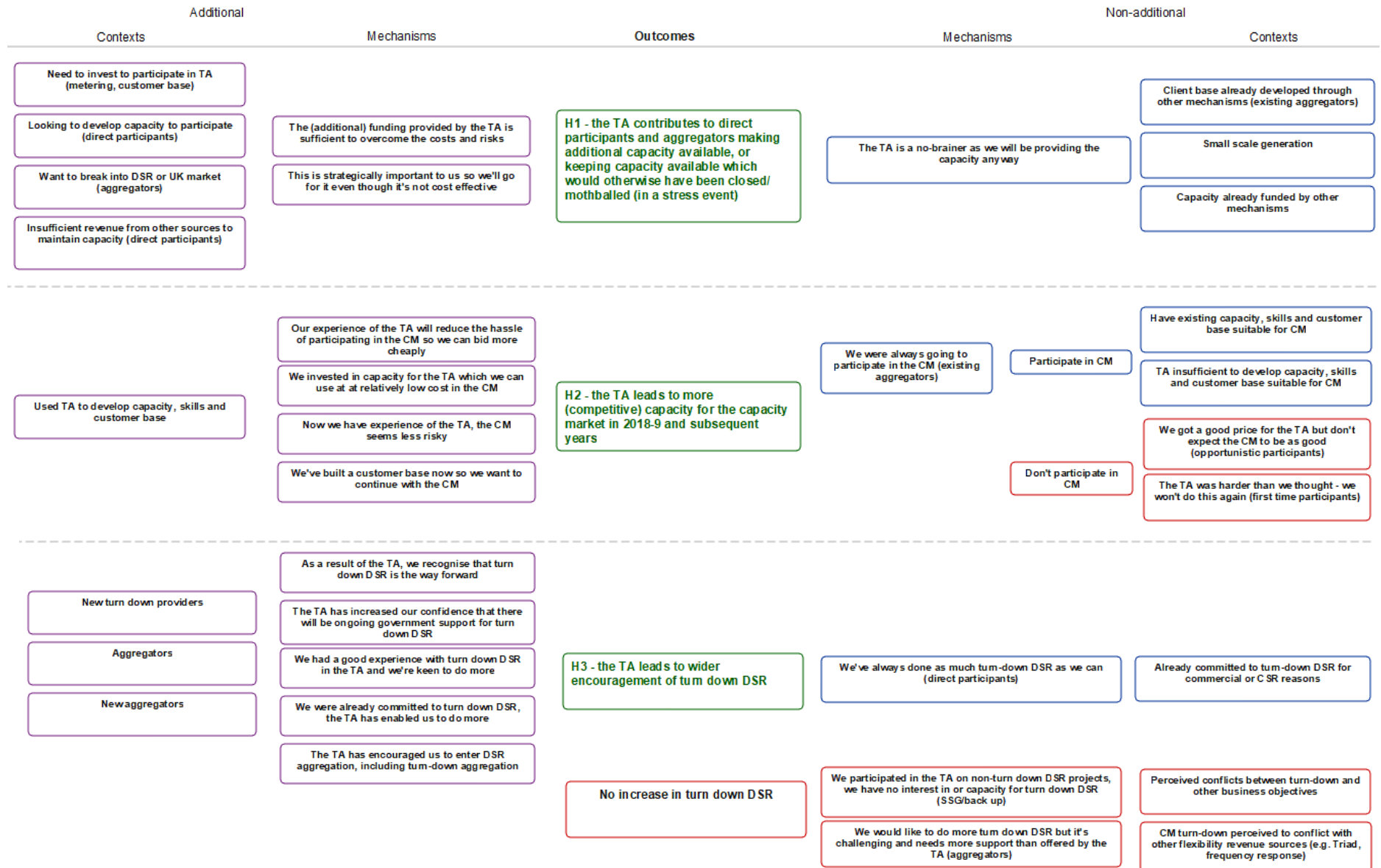
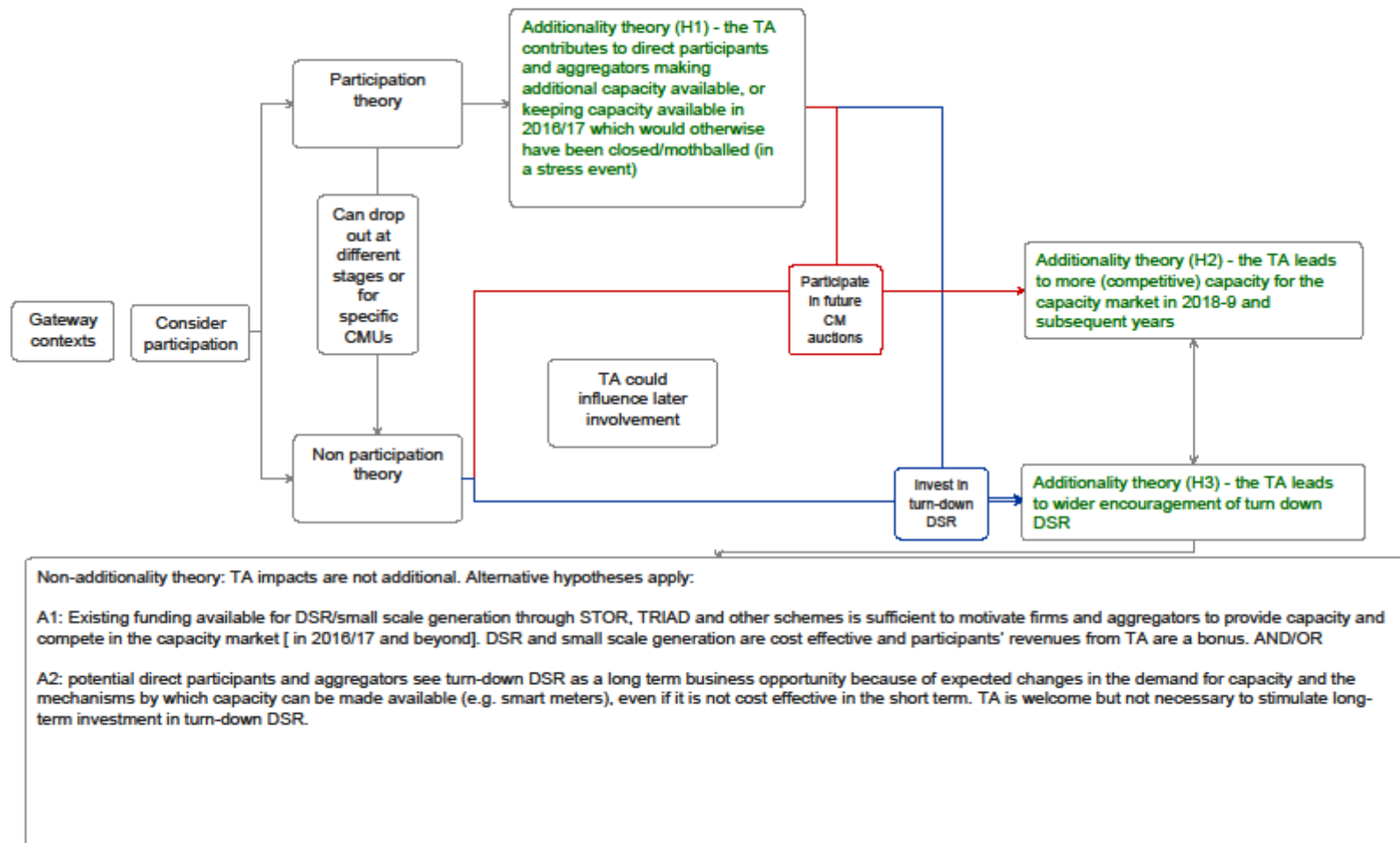


Figure A2.5 Overall theory map



Appendix 3. Revised theoretical framework

The initial theoretical framework from Phase 2, as reported in Appendix 2 above, was updated at the end of Phase 2, based on findings from Phase 2 evidence. Both Phase 1 and Phase 2 of the evaluation have found that the behaviour of aggregator clients within the TA is an important influence on TA outcomes. The combined evidence from interviews with aggregator clients during Phases 1 and 2 allowed us to develop two new elements of the theoretical framework: participation theory for aggregator clients, and delivery theory for aggregator clients. At the moment we have limited evidence from aggregator clients who have loads too small to participate in the TA directly, but we have hypothesised CMOs for this group as we anticipate that they may be observed further in later phases of the evaluation (see Figure A3.1).

Overall theory on TA participation and non-participation (Figure A2.2 and Figure A2.3 in Appendix 2) are unchanged from Phase 1 and so are not included here. The following figures presented below are those newly developed or revised in light of the findings from Phase 2:

Figure A3.1 Aggregator client participation theory

Figure A3.2 Testing stage theory

Figure A3.3 Response to CMNs

Figure A3.4 Aggregator client response to CMNs

Figure A3.5 Additionality theory for hypothesis H1

Figure A3.6 Additionality theory for hypothesis H2

Figure A3.7 Additionality theory for hypothesis H3

Figure A3.8 High level theory map, showing how additionality hypotheses fit together.

Figure A3.1 Aggregator client participation theory

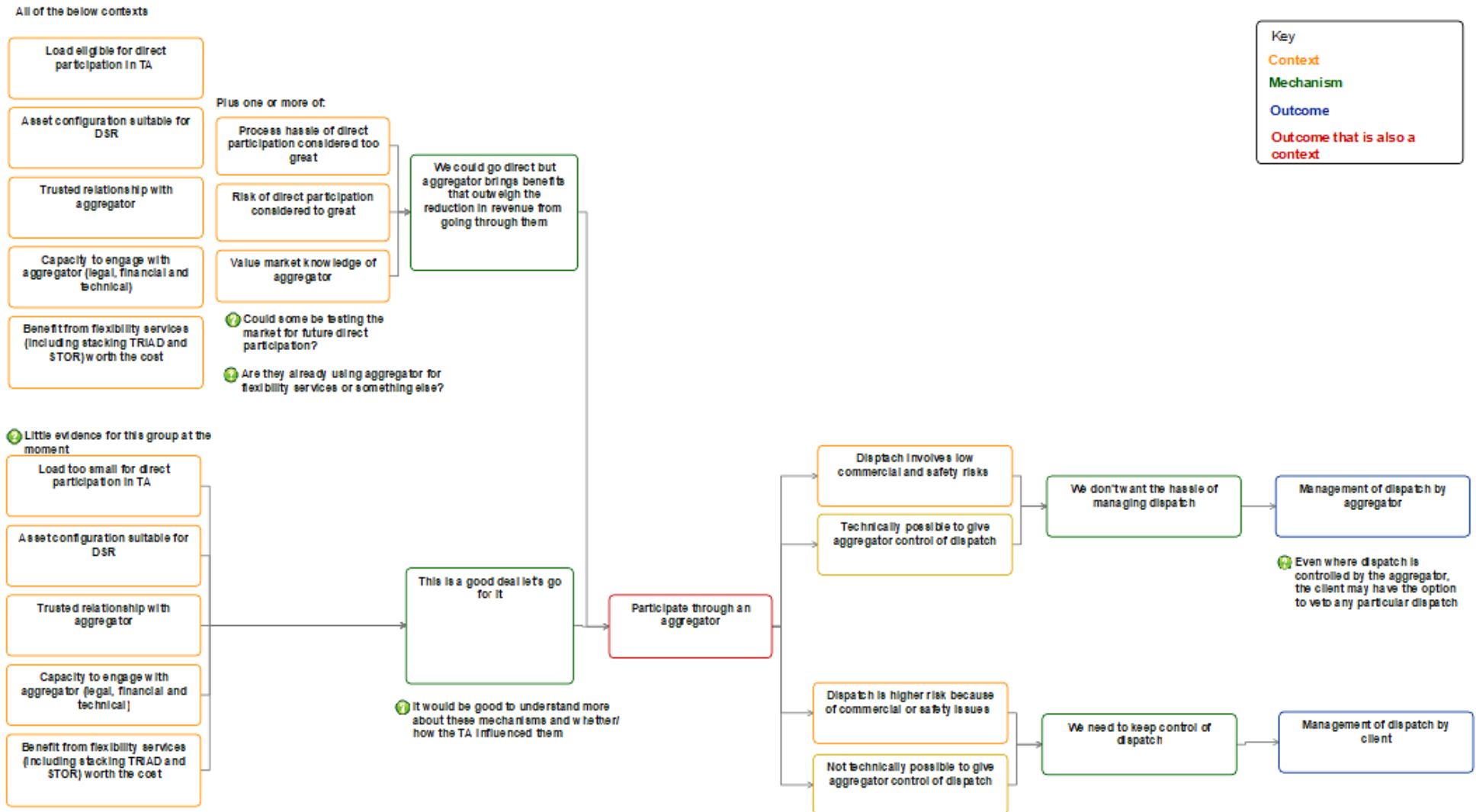


Figure A3.2 Testing stage theory

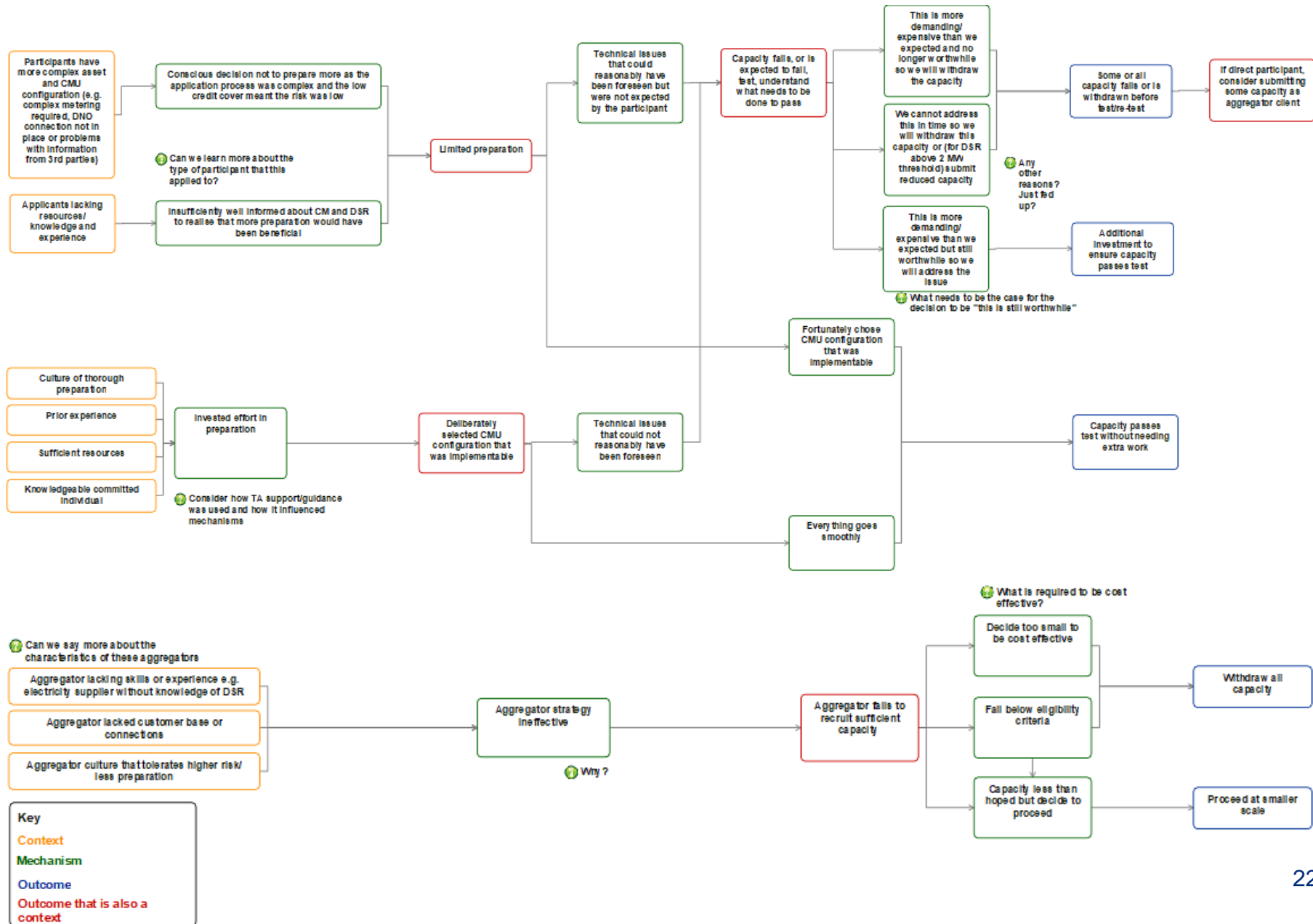
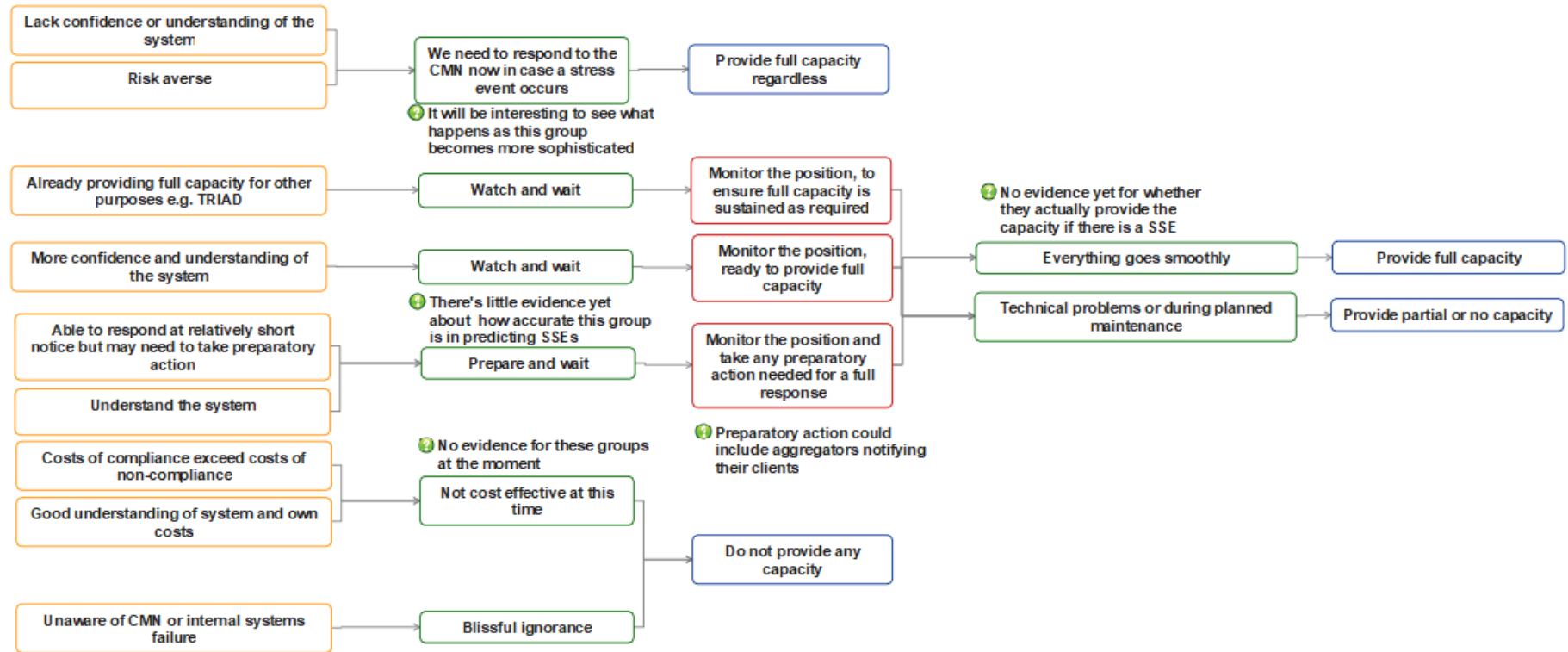


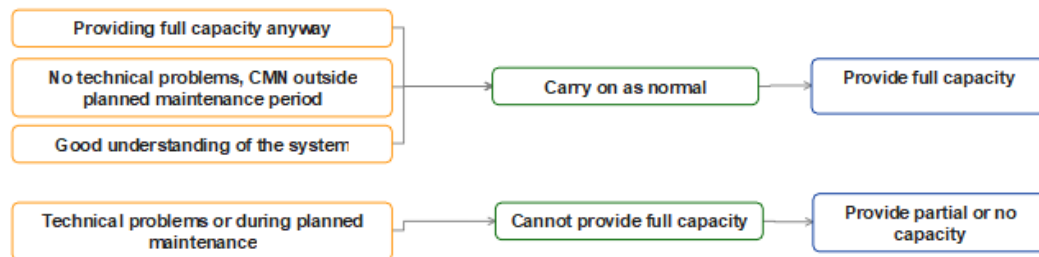
Figure A3.3 Response to CMNs

Turn down or back up DSR

Do the below groups consider cost of non-compliance vs compliance at this stage?



Small scale generation



Key

- Context
- Mechanism
- Outcome
- Outcome that is also a context

Figure A3.4 Aggregator client response to CMNs

Client response to aggregator call for delivery of turn down or back up DSR (e.g. in response to CMN)

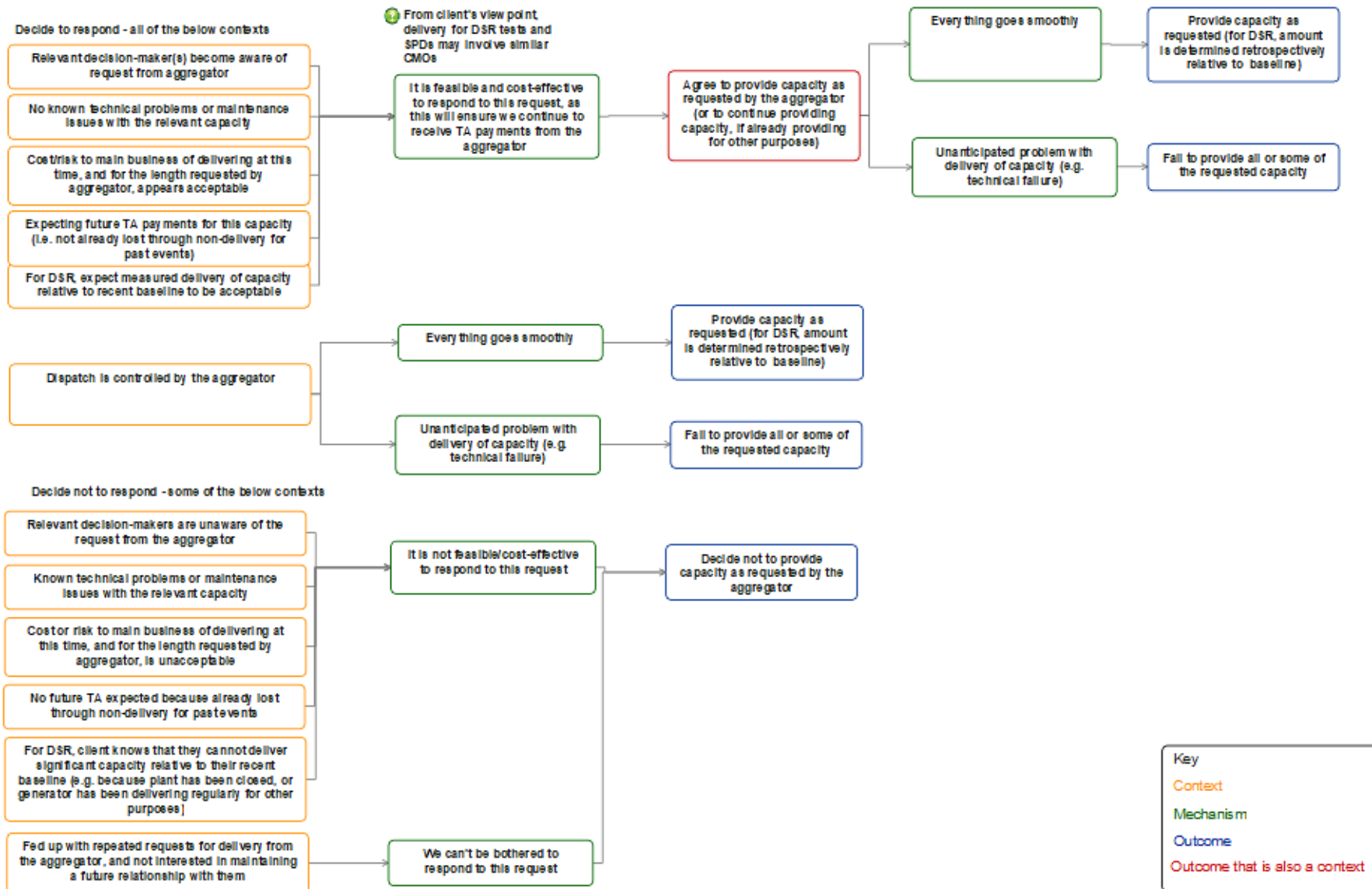


Figure A3.5 Additionality theory for hypothesis H1

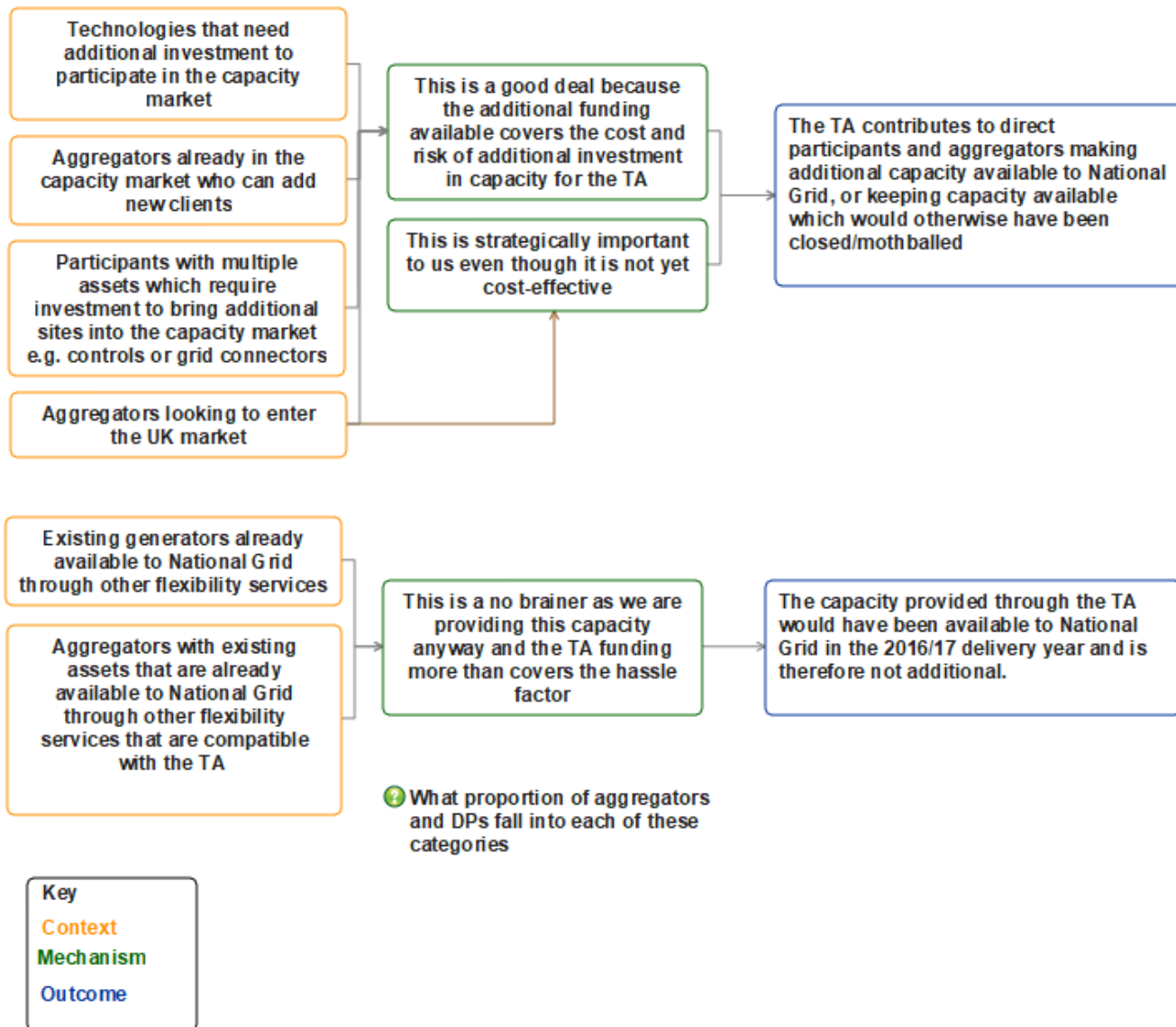


Figure A3.6 Additionality theory for hypothesis H2

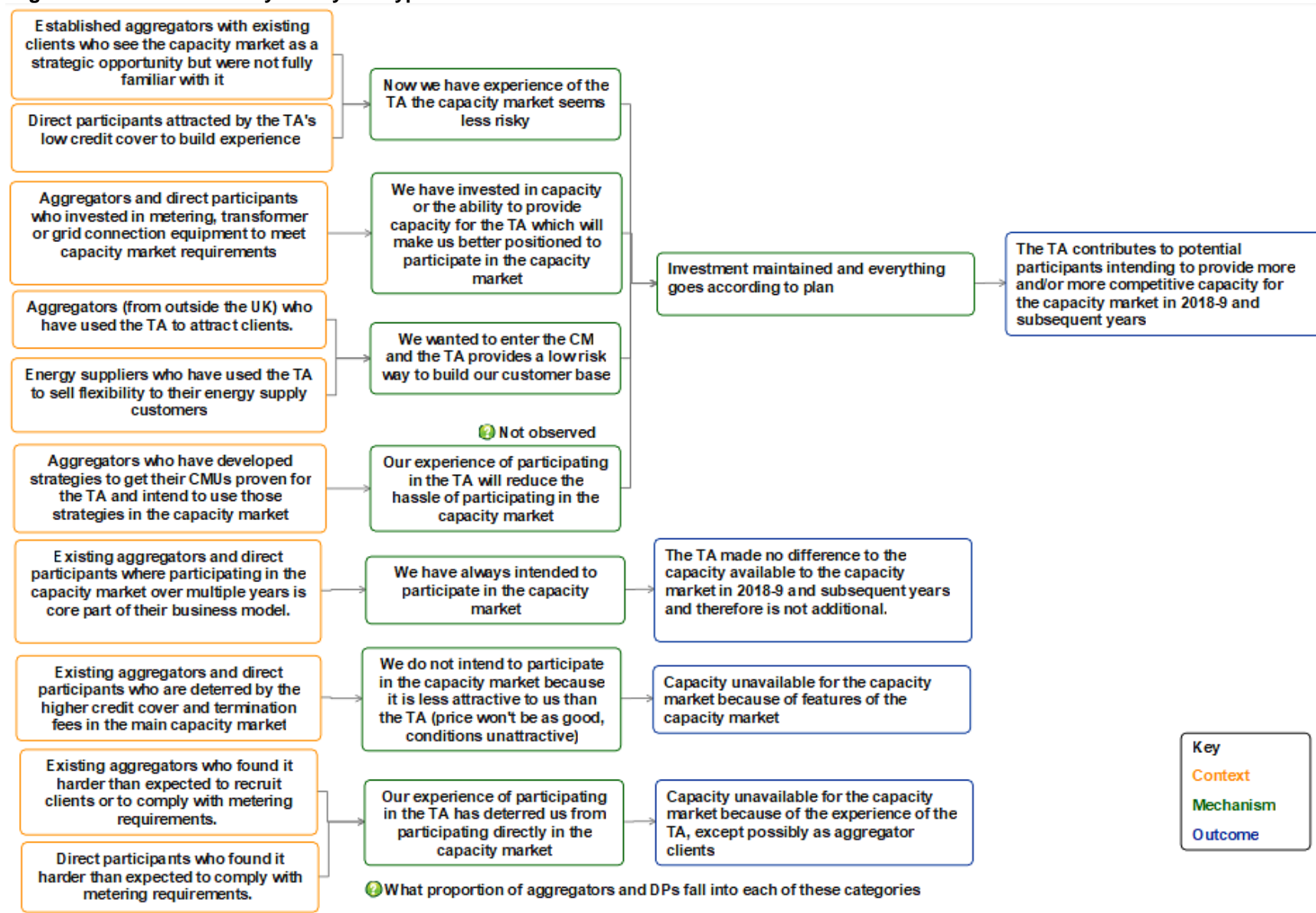


Figure A3.7 Additionality theory for hypothesis H3

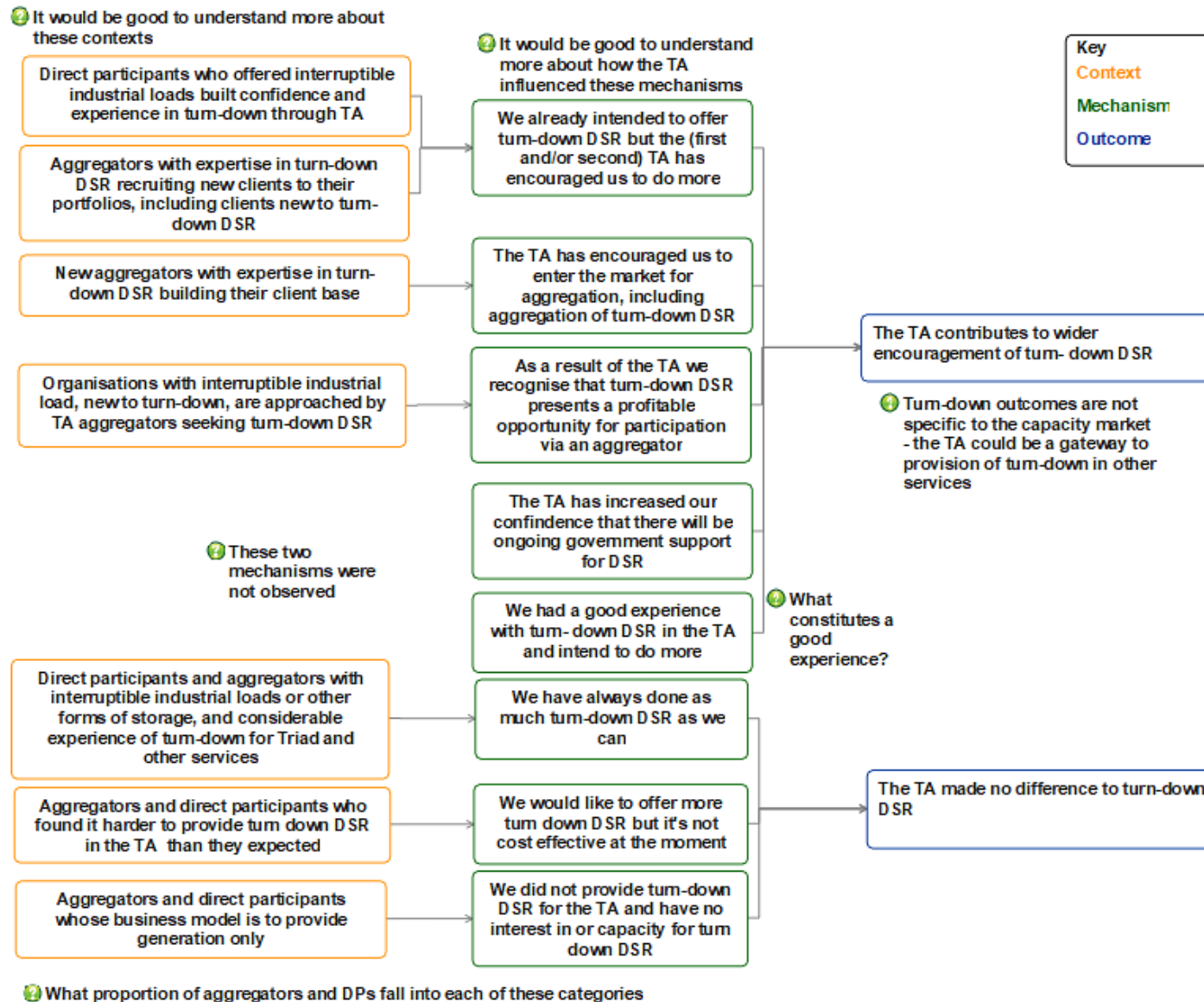
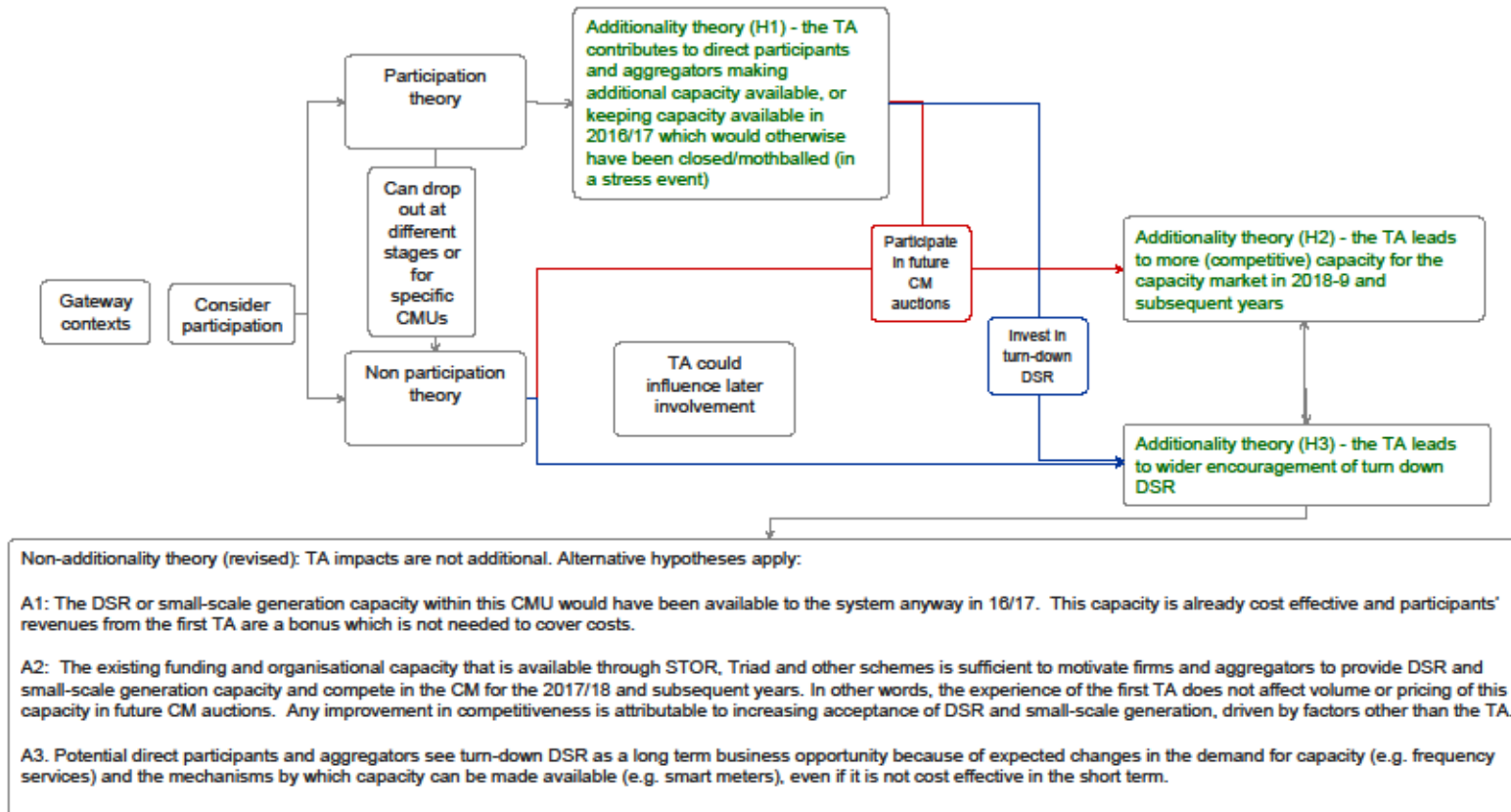


Figure A3.8 High level theory map, showing how additionality hypotheses fit together.



Appendix 4. Generative causation approaches

Applying contribution analysis within realist evaluation

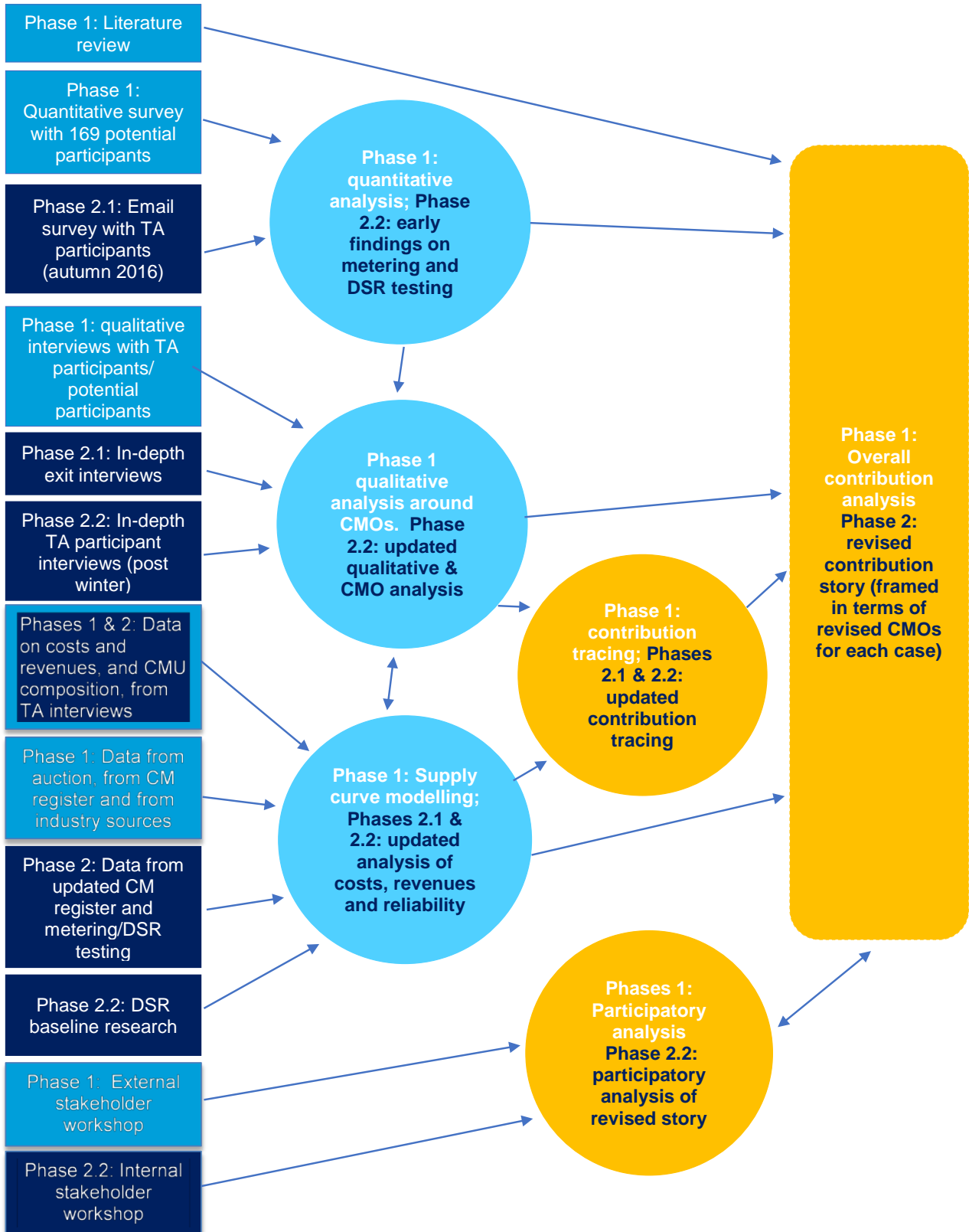
Our main approach to this evaluation has been to use contribution analysis on a case by case basis, to assess the contribution of the TA to causality in different contexts and circumstances. Contribution analysis involves the specification of a theory of change, assessment of the evidence base, gathering of new evidence, theory testing and then refinement of the theory of change. While contribution analysis is often used to assess the 'average' contribution of an intervention, across a scheme as a whole, we have applied this method using a realist approach and have assessed the TA's contribution on a case by case basis.

Our approach has involved the analysis, testing and refinement of context-mechanism-outcome (CMO) configurations on a case by case basis, using evidence from a range of sources. The sources of data on which we have drawn during this 'realist' contribution analysis process included:

- Qualitative research findings from TA participants in Phase 2 (in-depth interviews, supplemented by email correspondence and email survey responses where appropriate).
- Qualitative research findings from TA participants and non-participants in Phase 1.
- Observed behaviour in the TA and other CM auctions, for TA participants and other players.
- The latest information on TA CMUs published in the Capacity Market Register.
- Information on TA testing outcomes provided by the delivery body and BEIS.
- Analysis of exit prices for the first TA auction compared to modelled supply costs, based on supply curve modelling undertaken in Phase 1.
- Case study analysis of DSR and SPD meter data for a sample of CMUs with turn-down DSR components, during Phase 2.
- Published articles and statements by TA participants and industry commentators.

Figure A4.1 below shows how the different strands of evidence and analysis contributed to the overall contribution analysis. Pale blue boxes show Phase 1 evidence; dark blue boxes show Phase 2 evidence; turquoise circles show conventional analysis and orange shapes show generative causation techniques.

Figure A4.1 Relationship of evidence sources, evaluation phases, conventional analysis and generative causation methods



The process that we used for 'realist' contribution analysis was as follows:

1. We refined the theoretical framework at the start of Phase 2, on the basis of evidence gathered during Phase 1, and identified the topics and methods we would use to gather new evidence during Phase 2 to test this framework.
2. We used the theoretical framework to inform the design of the topic guides for qualitative fieldwork, with a view to using these interviews to test and refine the framework.
3. Following our data collection activities, we organised the new and existing data into readily accessible spreadsheets.
4. As part of this process, we created a CMO coding spreadsheet for each element of the theory that we sought to test from the initial theoretical framework for Phase 2 (e.g. testing theory). The coding spreadsheets had rows for every CMU, group of CMUs or organisation that required different CMO coding, with columns for 'contexts', 'mechanisms' and 'outcomes' across the spreadsheet.
5. We then coded contexts, mechanisms and outcomes for each case, based on evidence from all sources and all phases of the evaluation. Where the most recent qualitative evidence superseded evidence from earlier phases, we gave priority to evidence gathered in Phase 2. But we reviewed evidence from all the phases as part of the analysis process: where there were gaps in Phase 2 evidence, or inconsistencies between the phases, we endeavoured to understand the reason for any apparent inconsistencies and based our coding on the most relevant evidence from Phase 1, Phase 2 and the email survey. Where possible we cross-checked qualitative evidence with observed behaviour (e.g. whether an organisation had obtained capacity agreements for generation or DSR CMUs in recent CM auctions). Where motivations were not clear, we checked participants' websites or public statements for other insights into their rationale.
6. If the initial theory did not exactly fit the observed evidence for a given case, we suggested revisions to the CMOs. We cross-checked that our assessment of outcomes was consistent with the findings of contribution tracing (see below).
7. The analysis and refinement of CMOs for each case was an iterative process, as we were able to see patterns emerging as we analysed successive cases. The coding was undertaken by two analysts within the team, and discussed (in non-disclosive form) with the wider team and BEIS at a policy review workshop.
8. The detailed coding spreadsheets formed the basis of our findings in the synthesis report, and were also used to inform revision of the theory at the end of Phase 2.

Contribution tracing with Bayesian updating

We used contribution tracing with Bayesian updating to test the strength of evidence for the additionality hypotheses, and to cross-check our coding of CMOs for these hypotheses. The method we used is the same as 'process tracing', but we use the term 'contribution tracing' here because we have applied the method to assess the TA's impacts or contribution to the additionality hypothesis, rather than to trace successive steps in a process. The method is described in the CECAN working note 2.1 by Barbara Befani on 'Testing Contribution Claims with Bayesian Updating' (December 2016).¹²

Contribution tracing involves the formulation and testing of competing hypotheses which could explain observed outcomes. The method involves explicit assumptions about the weight attached to different types of evidence, and aims to increase the transparency and replicability of qualitative analysis. We applied contribution tracing on a case by case basis, which was consistent with our realist approach to analysis and synthesis.

To undertake contribution tracing, we first assessed the 'prior' probability of each of the additionality hypotheses being true (i.e. made an initial assessment of the likelihood of each of them being true), based on findings from Phase 1 and on a peer reviewer's expert assessment. Similarly, we assessed the 'prior' probability of each of the competing hypotheses being true, that could provide an alternative causal explanation for observed outcomes.

We then specified a set of evidence tests for each of the three additionality hypotheses and three competing hypotheses. These evidence tests defined the 'clues' that we would look for, to help distinguish between support the additionality or competing hypotheses. These clues drew on all the evidence sources we expected to collect, ranging from interview statements to auction behaviour and public statements. The evidence clues were assigned different weights in the analysis depending on the estimated probability that a given clue would be observed when a particular hypothesis held. The set of evidence tests and their assumed probabilities were agreed with BEIS and with a peer reviewer who had in-depth knowledge of the TA and DSR market.

A process called Bayesian updating was then used to update the probability that each hypothesis was true for each case, taking into account whether we had observed the specified clues for that case. The evidence tests drew on everything we knew about a participant, not just what they said in research interviews. For example, they used information from their interview statements in Phases 1 and 2 of the evaluation, their responses to the email survey in Phase 2, their initial TA application, their auction behaviour and their company websites as well as estimates of costs and revenues from the supply curve work. Analysis of additionality for H2 and H3 was undertaken at participant level (i.e. for an organisation), but analysis of additionality for H1 was analysed at CMU level because there were significant differences between evidence for different CMUs from some organisations.

¹² <http://www.cecan.ac.uk/resources>

This process provided an assessment of whether or not, for a particular case, the additionality hypothesis or competing hypothesis was likely to be true. The competing hypotheses were not direct converse of the project hypotheses, so in some cases the evidence supported both hypotheses or neither of them. We took account of the implications of this in the formulation of CMOs within our realist synthesis (see above). Further details of contribution tracing, including the evidence tests, the probability assumptions and the process for Bayesian updating of probabilities are set out in Appendix 5.

Appendix 5. Contribution tracing paper

1. Introduction

This paper builds on the approach developed in Phase 1: contribution tracing with Bayesian updating. By ‘contribution tracing’ we mean the application of a ‘process tracing’¹³ approach to hypotheses about the TA’s contribution to its objectives. ‘Bayesian’ updating refers to the specification of ‘prior’ probabilities for each hypothesis, and to the updating of these to ‘posterior’ probabilities, based on certain evidence tests.

The formula used to update the ‘prior’ probability for a given hypothesis (T), in relation to a particular evidence test (E) is as follows:

$$P(T|E) = P(T) * P(E|T) / \{[P(T) * P(E|T)] + [P(\sim T) * P(E|\sim T)]\}$$

Where

P(T) is the ‘prior’ probability of the hypothesis being true, before the evidence is observed

P(\sim T) is the prior probability of the hypothesis not being true (= 1-P(T))

P(E|T) is the ‘sensitivity’ of the evidence test (i.e. the probability of the evidence being observed if the hypothesis is true)

P(E| \sim T) is the ‘Type 1 error’ of the evidence test (i.e. the probability of the evidence being observed if the hypothesis is not true)

P(T|E) is the resulting ‘posterior’ probability of the hypothesis being true, after the evidence is observed

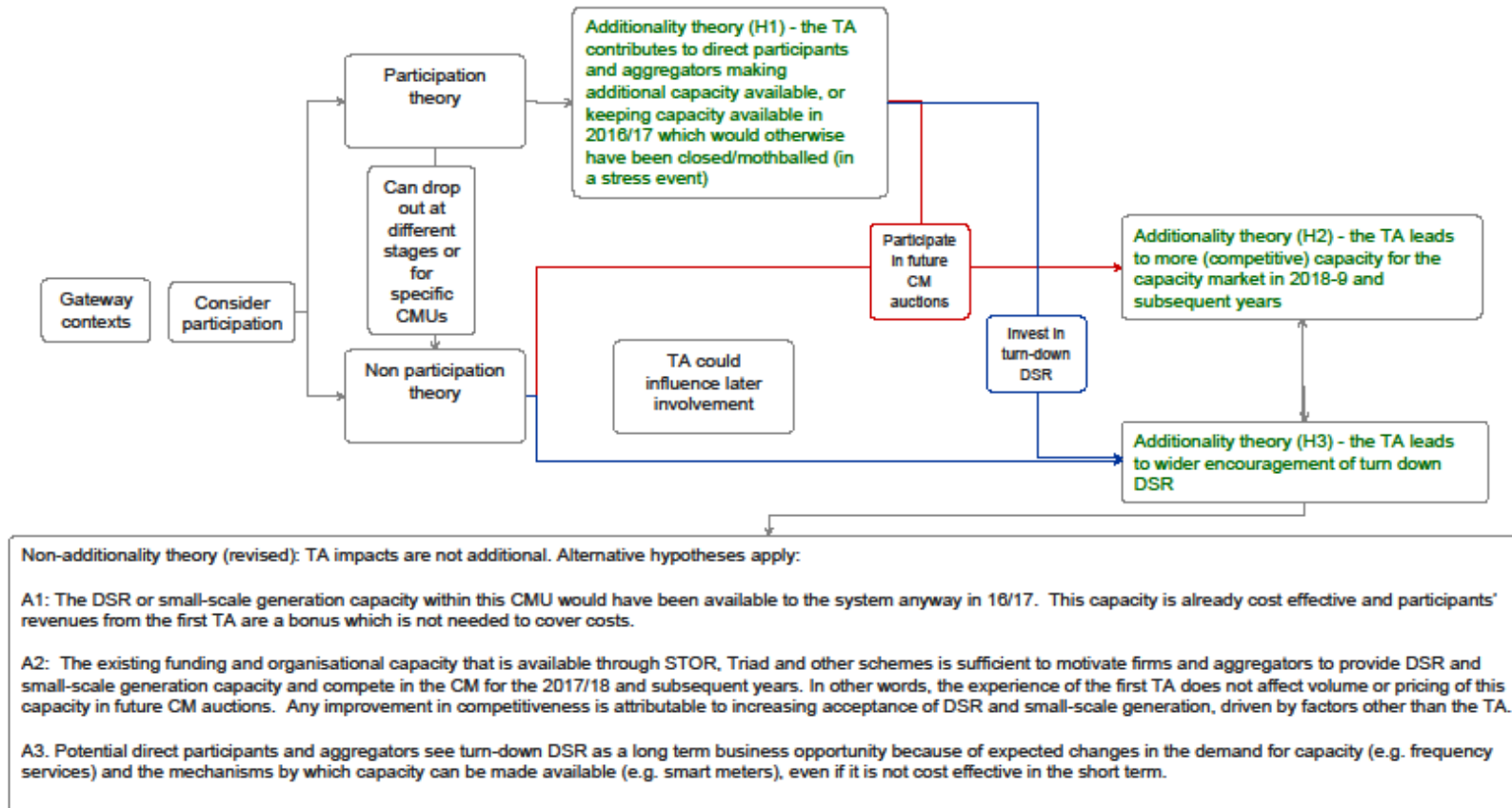
Source: [Dr Barbara Befani's Bayes Formula Confidence Updater Spreadsheet](http://www.cecan.ac.uk/resources) (www.cecan.ac.uk/resources)

¹³ The approach is based on Befani, B and G. Stedman-Bryce (2017) “Process Tracing and Bayesian Updating for Impact Evaluation”, Evaluation, Vol 23, pp42-60; Befani, B., D’Errico, S., Booker, F. and A. Giuliani (2016) “Clearing the Fog: New Tools for Improving the Credibility of Impact Claims”, IIED Briefing, April; and CECAN (2016) “Testing contribution claims with Bayesian Updating”, Evaluation and policy practice note 2.1, winter 2016.

Process/contribution tracing is intended to bring a rigorous approach to setting and analysing questions to test causality. We are using it as just one input to the evaluation's overall realist approach alongside contribution analysis, primarily to test the strength of evidence for the additionality of TA outcomes. The contribution tracing enables the combined assessment of qualitative and quantitative evidence supporting each of the additionality hypotheses and competing hypotheses in the Phase 2 theoretical framework, developed at the start of Phase 2. The evidence tests and probability assumptions used in contribution tracing make the analysis traceable and explicit, at least in theory. The methodology helps to provide an understanding of the support offered by different 'cases' (i.e. different TA participants, or different Capacity Market Units (CMUs)) for each of the hypotheses in the theoretical framework. The contribution tracing findings have informed our case by case analysis of the additionality of outcomes, feeding in to our development of case by case 'context-mechanism-outcome' configurations within the overall realist contribution analysis process.

A schematic diagram of the theoretical framework as a whole is shown below. This highlights the three additionality hypotheses, which relate to the three objectives of the first TA auction. The diagram also shows three alternative hypotheses which compete with each of the additionality hypotheses. The additionality hypotheses and competing alternative hypotheses are described in more detail in the next chapter, together with the specification of the evidence tests that we have used to assess each hypothesis.

Figure A5.1 High-level map of theoretical framework for Phase 2 of the Transitional Arrangements



One limitation of the contribution tracing approach – as we have applied it – is that it is complex to understand the probability calculations and the interaction between different evidence tests in generating the contributing tracing results. Particularly when applied across a large number of cases, and across a large number of evidence tests, the calculations themselves are complex and opaque – even though the tests and probability assumptions are set out explicitly.

A further limitation on the methodology as we have applied it is that it has focused primarily on evidence for the TA's role in influencing outcomes, rather than on the contexts or mechanisms for change, and therefore needs to be interpreted within the wider context of realist analysis within the overall synthesis process. However, even more evidence tests would have been needed if we had attempted to test for 'context' and 'mechanism'¹⁴ in a similar way, and the whole process would have become unmanageable.

Also, great care is needed in interpreting contribution tracing results as they rely to a large extent on coding complex findings into binary 'yes/no' evidence tests. The Phase 1 contribution findings were sensitive to small changes in probability assumptions and to the interpretation of particular pieces of evidence. However, we have run some sensitivity tests on the findings here (presented in section 6 below), using hypotheses H3 and A3 as examples of hypotheses to test. In contrast to Phase 1, the sensitivity tests show that findings for hypotheses H3 and A3 are relatively robust to small changes in assumptions. This demonstrates the robustness of the evidence tests.

And, finally, we have drawn our evidence from respondents in those organisations who obtained a capacity agreement following the 1st TA auction, referred to here as 'TA participants'. We have not used evidence from respondents in organisations participating in the TA as clients of aggregators, to avoid double-counting of the impacts between clients and their aggregators. But we have included evidence from respondents in organisations that obtained capacity agreements in the first TA but subsequently left the scheme, because their future actions may have been influenced by TA participation.

2. Methodology for Phase 2 contribution tracing tests

The methodology set out below is similar to that applied in 'process tracing' during Phase 1 of the evaluation, which was informed by Barbara Befani's papers on this topic.¹⁵ We are now using the term 'contribution tracing' rather than 'process tracing', as this is a more accurate description of the

¹⁴ Realist theory suggests that mechanisms are not directly observable, other than through qualitative statements, so evidence testing for mechanisms per se might be problematic.

¹⁵ The approach is based on Befani, B and G. Stedman-Bryce (2017) "Process Tracing and Bayesian Updating for Impact Evaluation", Evaluation, Vol 23, pp42-60; Befani, B., D'Errico, S., Booker, F. and A. Giuliani (2016) "Clearing the Fog: New Tools for Improving the Credibility of Impact Claims", IIED Briefing, April; and CECAN (2016) "Testing contribution claims with Bayesian Updating", Evaluation and policy practice note 2.1, winter 2016.

testing process, which focuses primarily on the TA’s contribution to its objectives. The proposed revisions to the tests have been reviewed by BEIS, and probabilities have been further reviewed by BEIS and by Dr. Jacopo Torriti of the University of Reading. The results from applying these tests to evidence from Phases 1 and 2 are presented in section 5 below.

3. Revisions since Phase 1 contribution tracing

In the Phase 1 contribution tracing, we found that the alternative hypothesis tests were picking up some cases where the desired outcome was not observed, as well as cases where the desired outcome was observed but appeared to be attributable to factors other than the TA. We therefore revised the testing procedure for Phase 2 so that we screened for desired outcomes before applying the attribution tests for a given outcome. This paper uses a screening test for each desired outcome and uses this screening test to identify those cases in which that outcome occurred. We then applied contribution tracing tests to the cases that passed the screening test for each outcome. Analysis of the reasons for the desired outcome not being observed in certain cases was undertaken separately as part of the overall qualitative analysis and contribution analysis.

Table A5.2 Outcomes and hypotheses

| Desired outcome | Attribution hypothesis | Non-attribution hypothesis |
|--|---|--|
| 1. DSR and small-scale generation capacity is made available by TA participants in the 16/17 delivery window. | H1 - The first TA leads to direct participants and aggregators making additional capacity available, or keeping capacity available that would otherwise have been closed/mothballed. This capacity contributes to security of supply and/or meeting the reliability standard in 16/17 delivery window. | Revised A1 - The DSR or small-scale generation capacity within this CMU would have been available to the system anyway in 16/17. This capacity is already cost effective and participants’ revenues from the first TA are a bonus which is not needed to cover costs. |
| 2. Increased volumes of (and more competitive) DSR and small-scale generation capacity enter the CM in 2017/18 and subsequent years. | Revised H2 - The first TA leads to more (competitive) capacity for the CM in 2017/18 and subsequent years. | New A2 (amendment of A1) – The existing funding and organisational capacity that is available through STOR, TRIAD and other schemes is sufficient to motivate firms and aggregators to provide DSR and small-scale generation capacity and compete in the CM for the 2017/18 and subsequent years. In other words, the experience of the first TA does not affect volume or pricing of this capacity in future CM auctions. Any improvement in competitiveness is attributable to increasing acceptance of DSR and small-scale generation, driven by factors other than the TA. |
| 3. Increased commitment to providing turn-down DSR capacity in the CM or other | H3 - The first TA leads to wider encouragement of turn-down DSR | Revised A3 (old A2) - Potential direct participants and aggregators see turn-down DSR as a long term business opportunity because of expected changes in |

| Desired outcome | Attribution hypothesis | Non-attribution hypothesis |
|---|------------------------|---|
| services in the longer term, from existing ¹⁶ and/or potential providers of turn-down. | | the demand for capacity (e.g. frequency services) and the mechanisms by which capacity can be made available (e.g. smart meters), even if it is not cost effective in the short term. . |

A second change to the testing process is that we tested hypotheses relating to the first outcome at CMU rather than TA participant level. This allowed us to make better use of detailed evidence relating to the content of particular CMUs, and the bids made for these CMUs at auction. It also helped us to use contribution tracing results to develop estimates of the volume of capacity brought forward by the TA in the 2016/17 delivery year. As the previous analysis was undertaken at TA participant level (i.e. for aggregators and for direct participants), the CMU-level analysis in Phase 2 was at a level equivalent to, or lower than, the previous analysis. We did not currently have enough evidence about the behaviour of aggregator clients to undertake sub-CMU level analysis for aggregated CMUs.

We added some new tests below to bring in emerging evidence from CM auctions which were observed during Phase 2. We also adjusted the tests so that there was an initial screening test and then a set of tests for the attribution hypothesis and the related alternative non-attribution hypothesis. We assigned probabilities to the new tests and have in places reviewed probabilities for existing tests. We based our assumptions about prior probabilities for each hypothesis on the assumptions made in the Phase 1 contribution tracing. These probabilities have been reviewed by a technical peer reviewer and by the policy and evaluation teams at BEIS. The tests are set out in the rest of this section, for each outcome in turn.

¹⁶ While Phase 2 of the evaluation primarily provided evidence about existing providers of turn-down DSR (since it does not include interviews with non-participants), we will also look for evidence about providers of small-scale or back-up generation becoming interested in providing turn-down DSR as well.

Note on definition of 'sensitivity' and 'Type 1' probabilities

Bayesian analysis involves assigning probabilities to the likelihood of each piece of evidence being seen if the hypothesis is true (known as the 'sensitivity') and if it is false (known as the 'Type 1 error') prior to conducting any research or gathering any evidence.

In Bayesian analysis, the strength of each evidence test is often categorised using the following terms:

- **Hoop tests** (high sensitivity)– weaken the hypothesis if not found but not sufficient to confirm the hypothesis; these are pieces of evidence that we would '**expect to see**' if the given hypothesis is true.
- **Smoking gun** (low Type 1 error)– strengthens the hypothesis if observed but does not weaken the hypothesis if not observed; these are pieces of evidence that we would '**like to see**' if a given hypothesis is true.
- **Double-decisive** (high sensitivity and low Type 1 error) – strengthens the hypothesis if observed and if not observed the hypothesis is weakened (these are pieces of evidence that are expected but are also confirmatory of the hypothesis).
- **Straw-in-the-Wind** (moderate sensitivity and moderate Type 1 error) – evidence that is 'nice to have' but not sufficient to confirm the hypothesis if observed or to reject the hypothesis if not observed.

We have categorised the evidence tests using these definitions in the tables below.

Tests for outcome 1: DSR and small-scale generation capacity is made available by TA participants in the 16/17 delivery window.

Screening test: This DSR or small-scale generation CMU has a Capacity Agreement from the first TA auction that is ongoing, and not terminated or withdrawn.

Level of testing: tests for outcome 1 will be applied at CMU level. Where the tests require evidence about an organisation's broader intentions, this evidence will be applied across all the CMUs put forward by that organisation.

Attribution hypothesis 1: The first TA leads to direct participants and aggregators making additional capacity available, or keeping capacity available within this CMU that would otherwise have been closed/mothballed. This capacity contributes to security of supply and/or meeting the reliability standard in 16/17 delivery window.

Non-attribution hypothesis 1: The DSR or small-scale generation capacity within this CMU would have been available to the system anyway in 16/17. This capacity is already cost effective and participants' revenues from the first TA are a bonus which is not needed to cover costs.

Prior probabilities: We suggest that prior probabilities for the attribution hypothesis should be kept unchanged from the Phase 1 analysis (since we will effectively be rerunning and refining this analysis, with limited new evidence). So the prior probability of the attribution hypothesis being true would be **30%**. This is relatively low because it is thought likely that many participants are not offering additional capacity. The prior probability of the non-attribution hypothesis being true would therefore be **70%**.

The proposed tests for these two hypotheses are set out below. Some of the tests are presented as alternatives to each other, to ensure that the evidence tests are independent of each other.

Table A5.3 Tests for attribution hypothesis 1 (H1). The first TA leads to direct participants and aggregators making additional capacity available, or keeping capacity available within this CMU that would otherwise have been closed/mothballed. This capacity contributes to security of supply and/or meeting the reliability standard in 16/17 delivery window.

| Test | Evidence | Type 1 (i.e. probability that this evidence is observed if outcome is NOT attributable to TA) | Sensitivity (i.e. probability that this evidence is observed if outcome is ATTRIBUTABLE to TA) | Type of test | Evidence source | Rationale for probabilities |
|------|---|---|--|-----------------|-------------------------------|--|
| H1a | Decisions made since September 2014 (the date of the announcement of the TA) to invest in new capacity or keep capacity available for this CMU, that would otherwise have been closed/mothballed. (do not include if H1b found) | 0.8 | 0.95 | Hoop | Interview or public statement | If the attribution hypothesis is true, it is almost essential that participants have invested in or maintained capacity. However, participants are also very likely to have done so if the hypothesis is not true. |
| H1b | The participant saying in the interview that the TA contributed to their decision to invest in capacity or keep capacity available for this CMU (do not include if H1c is found). | 0.1 | 0.9 | Double-decisive | Interview | If the attribution hypothesis is true, it is highly likely that participants will confirm this in interview (although not all may be willing to 'credit' the TA). If the hypothesis is not true, a few may give a 'false positive' – but the low number of participants claiming a TA contribution in Phase 1 evidence suggests that lying is not widespread. |
| H1c | Evidence from public statements that the TA made a contribution to the decision to invest in capacity or keep it available for this CMU. | 0.05 | 0.3 | Smoking gun | Public statement | Public statements and business cases may be available for a small proportion of participants. It is possible that a wider public statement might be made by a trade body, but we do not propose to include this evidence as it would be difficult to attribute to individual CMUs. Where the attribution hypothesis is untrue, it is very unlikely that this evidence will be found. |
| H1d | TA participants saying in interview that they had considered taking part in the CM but did not do so because the TA required lower credit cover. | 0.1 | 0.5 | Smoking gun | Interview | We would expect participants to have considered the CM but only a proportion will have been put off by the credit cover (thus the sensitivity |

Appendix 5. Contribution tracing paper

| Test | Evidence | Type 1 (i.e. probability that this evidence is observed if outcome is NOT attributable to TA) | Sensitivity (i.e. probability that this evidence is observed if outcome is ATTRIBUTABLE to TA) | Type of test | Evidence source | Rationale for probabilities |
|-------------|--|---|--|-------------------|------------------------|--|
| | | | | | | of 50%). But we're unlikely to see this evidence if the attribution hypothesis is untrue. |
| H1d.2 (NEW) | TA participants saying in interview that they had considered taking part in the CM but did not do so because the TA offered one-year ahead contracts. | 0.1 | 0.5 | Smoking gun | Interview | Some but not all participants were put off earlier T-4 auctions (before the first TA) by the four-year lead time. TA participants are unlikely to claim that they required a one-year ahead contract if this is untrue. |
| H1e | The participant being a price maker and the exit price in the auction for this CMU being higher than the hassle costs identified in the supply curve analysis. | 0.5 | 0.8 | Hoop | TA auction behaviour | If the attribution hypothesis is true, participants will generally need to 'profit' from the TA and cover the costs of making the capacity available (although a few may have chosen to bid low for strategic reasons). But if it is false, they may still choose to submit a high exit price to maximize their revenue, so we're still fairly likely to see this piece of evidence. |
| H1f (NEW) | Participant says in interview that it was difficult to meet the DSR/metering tests but they pushed through because they really wanted to obtain TA funding. | 0.4 | 0.6 | Straw in the wind | Interview | If the attribution hypothesis is true, participants may have stronger motivation to remain in the TA, in spite of difficulties – although some may not have seen the tests as problematic. If the attribution hypothesis is false, some may still be motivated to retain TA revenues anyway. To apply this test, we need to ask first whether the DSR/metering tests were difficult to meet. |
| H1g | The participant isn't penalized for non-delivery during the delivery year. | 0.5 | 0.8 | Hoop | Stress event behaviour | If the attribution hypothesis is true, participants are likely to try to retain TA revenues to cover the costs of making capacity available, by avoiding penalties – although the |

Appendix 5. Contribution tracing paper

| Test | Evidence | Type 1 (i.e. probability that this evidence is observed if outcome is NOT attributable to TA) | Sensitivity (i.e. probability that this evidence is observed if outcome is ATTRIBUTABLE to TA) | Type of test | Evidence source | Rationale for probabilities |
|-----------|---|---|--|--------------|---|---|
| | | | | | | timing of stress events may mean that some do not manage to avoid all penalties. But if the attribution hypothesis is false, a participant may still try to avoid most of the penalties and retain TA income. |
| H1h (NEW) | The participant meets satisfactory performance day requirements for this CMU. (Do not use if H1g found) | 0.5 | 0.8 | Hoop | Satisfactory performance day requirements met | Same rationale as for H1g. |
| H1i (NEW) | The participant states in interview that they have operational plans in place to ensure adequate capacity will be available during a stress event. (Do not use if H1g or H1h found) | 0.5 | 0.8 | Hoop | Interview evidence | Same rationale as for H1i. |

Table A5.4 Tests for non-attribution hypothesis 1 (A1):The DSR or small-scale generation capacity within this CMU would have been available to the system anyway in 16/17. This capacity is already cost effective and participants’ revenues from the first TA are a bonus which is not needed to cover costs.

| Test | Evidence | Type 1 (i.e. probability that this evidence is observed if outcome is ATTRIBUTABLE to TA | Sensitivity (i.e. probability that this evidence is observed if outcome NOT attributable to TA) | Type of test | Evidence source | Rationale for probabilities |
|------|--|--|---|-------------------|---|---|
| A1a | TA participants state in the interview that they would have invested in, or maintained, capacity for this CMU in 2016/17 regardless of the TA | 0.1 | 0.9 | Double-decisive | Interview | This is the converse of H1b, so the probabilities are determined by H1b probabilities (viz Type 1 probability for A1a = 1-sensitivity for H1b; and Sensitivity for A1a = 1- type 1 for H1b) |
| A1b | Participants being a price taker or the exit price for this CMU being at or below the hassle costs in the TA auction | 0.2 | 0.5 | Straw in the wind | TA auction behaviour | This is the converse of H1e, so probabilities are determined by those for H1e (as described for test A1a). |
| A1c | Participants in TA claim in interview that without TA there is sufficient funding (from STOR etc.) to justify investing in or keeping this CMU's capacity available. | 0.1 | 0.6 | Smoking gun | Interview | This statement would confirm the non-attribution hypothesis but some participants may be reluctant to agree with it, even if true, because of lobbying bias. However, if it is false they will be happy to disagree so the chances of type 1 error are small. |
| A1d | The participant is penalized for non-delivery at some point during the delivery year. | 0.2 | 0.5 | Straw in the wind | Stress event behaviour | This is the converse of H1g, so probabilities are determined by those for H1g (as described for test A1a). |
| A1e | The participant fails to meet satisfactory performance day requirements for this CMU. (Do not use if A1d found) | 0.2 | 0.5 | Straw in the wind | Satisfactory performance day requirements met | Same rationale as for A1d. |
| A1f | The participant states in interview that they are concerned they may not have adequate capacity available to cover stress event. (Do not use if A1e or A1f found) | 0.2 | 0.5 | Straw in the wind | Interview evidence | Same rationale as for A1d. |

Tests for Outcome 2: Increased volumes of (and more competitive) DSR and small-scale generation capacity enter the CM in 2017/18 and subsequent years.

Screening test (similar to old H2a/H2a.1): Organisations that obtain capacity agreements in the first TA auction state their intention to make DSR and small-scale generation capacity available to the CM in 2017/18 and subsequent years.

Level of testing: tests for outcome 2 will continue to be applied at organizational level, as they were in the Phase 1 contribution tracing, as none of the tests are CMU-specific. They refer to an organisation's reasoning, behaviour and intentions.

Attribution hypothesis 2: The first TA leads to more (competitive) capacity for the CM in 2017/18 and subsequent years.

Non-attribution hypothesis 2: The existing funding and organisational capacity that is available through STOR, TRIAD and other schemes is sufficient to motivate firms and aggregators to provide DSR and small-scale generation capacity and compete in the CM for the 2017/18 and subsequent years. In other words, the experience of the first TA does not affect the volume or pricing of this capacity in future CM auctions.

Prior probabilities: We have slightly increased the probability for the attribution hypothesis, based on strong support for this hypothesis in Phase 1 analysis. The prior probability of the attribution hypothesis being true has been increased from 50% to **60%**, with the probability of the non-attribution hypothesis now being **40%**. This represents an expectation that, on balance, the TA will contribute to increases in volumes (or decrease in prices) for DSR and small-scale capacity offered by TA participants in future CM auctions.

Table A5.5 Tests for attribution hypothesis 2 (H2): The first TA leads to more (competitive) capacity for the CM in 2017/18 and subsequent years.

| Test | Evidence | Type 1 (i.e. probability that this evidence is observed if outcome is NOT attributable to TA) | Sensitivity (i.e. probability that this evidence is observed if outcome is ATTRIBUTABLE to TA) | Type of test | Evidence source | Rationale for probabilities |
|-----------|---|---|--|--------------|-----------------|---|
| H2a (NEW) | The participant saying they now have more confidence in being able to meet CM rules and regulations/be competitive in the other CM auctions as a result of the TA | 0.3 | 0.9 | Hoop | Interview | We are very likely to see this evidence if the attribution hypothesis is true. But a few participants may agree with this statement even if it's not true. |
| H2b | The participant saying in interview that they have developed or invested in assets (e.g. controls), markets (e.g. building a client base, entering the UK market) or skills (e.g. delivering DSR for clients) for the TA that will be used in one or more main CM auctions. | 0.4 | 0.8 | Hoop | Interview | We are likely to see this evidence if the attribution hypothesis is true, however, some can already participate in the wider CM without investing in assets, markets or skills. If the hypothesis is not true the participant could invest in assets intended for other CM auctions (not the TA) so this evidence could still be seen. |
| H2c | The participant did not obtain capacity agreement for TA-type capacity in the T-4 auction, prior to the first TA auction. | 0.3 | 0.8 | Hoop | T-4 behaviour | If the attribution hypothesis is true, a participant is not likely to have obtained a capacity agreement for TA-types capacity in the first two T-4 auctions (so we are likely to see this evidence). If it is false, it is still possible that a participant would not have obtained a capacity agreement for TA-type capacity in T-4. |
| H2d | The participant saying in interview that they intend to bid at a lower price in one of the main CM auctions than they would otherwise have done. (Do not include if H2d.1 is found) | 0.1 | 0.3 | Smoking gun | Interview | If the attribution hypothesis is true, we would expect to see this evidence in some cases. However, many might not have formed a plan on how to bid yet, or might not be willing to discuss future bidding |

Appendix 5. Contribution tracing paper

| | | | | | | |
|-------------|---|-----|-----|------|--|---|
| | | | | | | strategies. If the hypothesis is untrue we could still see this evidence as participants may expect the main CM to be more competitive. |
| H2d.1 (NEW) | Participant actually bids at a lower price in one of the main CM auctions, on DSR or small-scale generation capacity, than they did in the first TA | 0.6 | 0.8 | Hoop | Behaviour in Early Auction, 2 nd TA or (2016) T-4 | Participants may bid lower in the main CM than the TA for other reasons, even if the TA did not make their bids more competitive, so the Type 1 error is 0.6. But if the TA did contribute to competitiveness, we would generally expect to see them bid lower into main CM. However, there might be exceptions to this (e.g. if the TA had generated experience that showed their costs were actually higher than previously thought; or if they had bid particularly low into the TA for other reasons – e.g. learning). So the sensitivity is set at 0.8 rather than 0.9 or above. |
| H2e | Participant agrees in an interview that one or more future CM auctions is likely to be more competitive as a result of the TA. | 0.6 | 0.9 | Hoop | Interview | Participants are likely to agree with this statement if the attribution hypothesis is true, but are also fairly likely to agree with this general statement if it's false. |

Table A5.6 Tests for non-attribution hypothesis 2 (A2): The existing funding and organisational capacity that is available through STOR, Triad and other schemes is sufficient to motivate firms and aggregators to provide DSR and small-scale generation capacity and compete in the CM for the 2017/18 and subsequent years. In other words, the experience of the first TA does not affect the volume of capacity or bid levels in future CM auctions.

| Test | Evidence | Type 1 (i.e. probability that this evidence is observed if outcome is ATTRIBUTABLE to TA) | Sensitivity (i.e. probability that this evidence is observed if outcome NOT attributable to TA) | Type of test | Evidence source | Rationale for probabilities |
|-----------|---|---|---|-----------------|--|--|
| A2a (NEW) | Respondents stating in the interview that they would have invested in, or maintained, capacity for future CM auctions regardless of the TA | 0.2 | 0.8 | Hoop | Interview | If the non-attribution hypothesis is true, participants are likely to agree with this. They are unlikely to agree with it if the non-attribution hypothesis is false (i.e. the attribution hypothesis is true). |
| A2b (NEW) | Respondents claim in interview that without TA there is sufficient funding (from STOR etc.) to justify investing in or keeping DSR and small-scale generating capacity available for future CM auctions | 0.05 | 0.5 | Smoking gun | Interview | Even if the non-attribution hypothesis is true, respondents may not agree with this statement because of lobbying bias. If it is not true, we are very unlikely to see this evidence. |
| A2c (NEW) | Market-wide: organisations obtain capacity agreements for DSR and small-scale generation capacity in future CM auctions without having participated in the TA. | 0.1 | 0.8 | Double-decisive | Results of EA and 2016 T-4: non-TA organisations obtaining capacity agreements for small-scale generation and DSR CMUs | This is a single piece of evidence, which would be applied across all the cases. We are likely to see this evidence if the non-attribution hypothesis is true, and very unlikely to see it if the non-attribution hypothesis is false. |

Test for Outcome 3: Increased commitment to providing turn-down DSR capacity in the CM or other services in the longer term, from existing¹⁷ and/or potential providers of turn-down.

Screening test (new): Organisations that obtained capacity agreements in the first TA state in interview that they have some interest in/ commitment to turn-down DSR capacity.

Level of testing: tests for outcome 3 will be applied at organisational level, as none of the tests are CMU-specific. They refer to an organisation's reasoning, behaviour and intentions.

Attribution hypothesis 3: The first TA leads to wider encouragement of turn-down DSR.

Non-attribution hypothesis 3: Potential direct participants and aggregators see turn-down DSR as a long-term business opportunity because of expected changes in the demand for capacity (e.g. frequency services) and the mechanisms by which capacity can be made available (e.g. smart meters), even if it is not cost effective in the short term.

Prior probabilities: We suggest that prior probabilities for the attribution hypothesis should be kept unchanged from the Phase 1 analysis (since we will effectively be rerunning and refining this analysis, with limited new evidence). So the prior probability of the attribution hypothesis being true would be **55%**¹⁸, with the probability of the non-attribution hypothesis being **45%**. This indicates that, where organisations indicate an increased commitment to turn-down DSR, this is slightly more likely than not to be attributable to the TA rather than non-TA factors.

Note: In principle, the issue of whether CM affects long-term plans to deliver turn-down could trigger different responses between aggregators and direct participants. This is because the latter are often much more constrained in terms of what type of DSR they can offer. We should review the results of contribution tracing to see if this effect is observed.

¹⁷ While Phase 2 of the evaluation will primarily provide evidence about existing providers of turn-down DSR (since it does not include interviews with non-participants), we will also look for evidence about providers of small-scale or back-up generation becoming interested in providing turn-down DSR as well.

¹⁸ Our peer reviewer, Jacopo Torriti, has commented that the prior probability for this hypothesis being true could increase to 60%, if there was higher certainty around the future of the CM and TA.

Table A5.7 Tests for attribution hypothesis 3 (H3): The first TA leads to wider encouragement of turn-down DSR.

| Test | Evidence | Type 1 (i.e. probability that this evidence is observed if outcome is NOT attributable to TA) | Sensitivity (i.e. probability that this evidence is observed if outcome is ATTRIBUTABLE to TA) | Type of test | Evidence source | Rationale for probabilities |
|------|---|---|--|-------------------|-------------------|---|
| H3a | Participants confirm in the interview that their long term strategic commitment to turn down DSR strengthened as a result of the TA funding. OR | 0.4 | 0.9 | Hoop | Interview | If the attribution hypothesis is true, participants are likely to agree with this in an interview. If the hypothesis is untrue, they may already be committed to turn-down DSR and hence may disagree with this statement, although some may agree anyway if they perceive this to be the 'right' answer. |
| H3b | Participants cite TA funding in public statements as a reason for their commitment to turn down DSR (do not include if H3a is found). | 0.1 | 0.4 | Smoking gun | Public statements | Participants are less likely to make public claims but if they do they are more likely to be true. |
| H3c | Participants in TA say in interview that they are implementing turn down DSR for the first time | 0.01 | 0.2 | Smoking gun | Interview | Given the lead-times involved in getting buy-in for turn-down DSR, participants are unlikely to be implementing turn-down DSR for the first time in the TA, even if the attribution hypothesis is true in the longer term. But they are very unlikely to invest in new turn down DSR for the first TA if the hypothesis is false. |
| H3e | Participants in TA say in interview that they are considering bidding turn down DSR projects into the CM or other services in future | 0.6 | 0.7 | Straw in the wind | Interview | If the TA is successful in encouraging turn down DSR then we would expect to see participants planning to bid projects in the CM and other services. However, some may still be planning to regardless of the TA (e.g. because of incentives offered by frequency related services). |

Table A5.8 Non-attribution hypothesis 3: Potential direct participants and aggregators see turn-down DSR as a long-term business opportunity because of expected changes in the demand for capacity (e.g. frequency services) and the mechanisms by which capacity can be made available (e.g. smart meters), even if it is not cost effective in the short term.

| Test | Evidence | Type 1 (i.e. probability that this evidence is observed if outcome is ATTRIBUTABLE to TA) | Sensitivity (i.e. probability that this evidence is observed if outcome NOT attributable to TA) | Type of test | Evidence source | Rationale for probabilities |
|------|---|---|---|--------------|-----------------|---|
| A3a | Respondents state in the interview that the first TA made no difference to their commitment to turn down DSR | 0.1 | 0.6 | Smoking gun | Interview | This is the converse of H1a, so probabilities are determined by H1a (see A1a for details). |
| A3b | Respondents say in interview that turn down DSR projects are considered able to compete effectively in the CM. | 0.1 | 0.3 | Smoking gun | Interview | If the non-attribution hypothesis is true then some participants are likely to consider the CM a viable source of revenue for turn-down DSR; if untrue they are unlikely to do so. But Phase 1 evidence suggests that few respondents will say this. |
| A3c | Respondents say in interview that turn down DSR projects are considered cost effective because of the existing non-CM revenues available to them | 0.1 | 0.6 | Smoking gun | Interview | If the non-attribution hypothesis is true, then some participants are likely to be planning to use the non-CM sources of revenue. Phase 1 evidence suggests that this is more likely to be observed than A3b because of the shorter time windows required for frequency services. If the non-attribution hypothesis is untrue, respondents are unlikely to agree with this statement. |
| A3d | Respondents say in interview that turn down DSR projects that exploit smart grids and other new developments are considered likely to be commercially viable in the long run. | 0.05 | 0.3 | Smoking gun | Interview | If the non-attribution hypothesis is true, this long term potential opportunity may motivate some to see turn-down DSR as a long term business opportunity. This evidence is very unlikely to be seen if the non-attribution hypothesis is false |

Appendix 5. Contribution tracing paper

| Test | Evidence | Type 1 (i.e. probability that this evidence is observed if outcome is ATTRIBUTABLE to TA) | Sensitivity (i.e. probability that this evidence is observed if outcome NOT attributable to TA) | Type of test | Evidence source | Rationale for probabilities |
|------|---|---|---|--------------|----------------------------------|--|
| A3e | Marketing material and websites for TA aggregators indicate that they are actively marketing other services for turn-down DSR in preference to the TA ¹⁹ (where those other services are apparently not compatible with the TA – e.g. frequency-related services). | 0.1 | 0.3 | Smoking gun | Websites and publicity material. | If the non-attribution hypothesis is true, we might possibly see this evidence, but it is not very likely. If the non-attribution hypothesis is false (i.e. if the TA is contributing to encouragement of turn-down), we are unlikely to see this evidence. Not ' <u>very</u> unlikely' since it's possible that aggregators may attract customers using another service but then bring them into the TA. (NB This test would not be applied to direct participants) |

¹⁹ This test would effectively apply to both TA auctions, since current marketing materials are likely to relate to the second rather than first TA.

4. How we have applied these tests

We compiled a spreadsheet containing databases of evidence codes for each case, for each of the attribution and alternative hypotheses.

For H1 and A1, relating to outcome 1, there were 46 cases. Each case was a single CMU that has gone through to delivery in the first TA.

For H2/A2 and H3/A3, relating to outcomes 2 and 3, each case is an organisation that was awarded a capacity agreement in the first TA, including those who have exited the TA since the auction (since their choices and behaviour may still be influenced by their experience of the TA):

- There were 23 cases for H2/A2, out of the 24 organisations that obtained capacity agreements in the first TA: one organisation did not pass the screening test because the outcome of participating in the future CM was not observed (and hence the additionality of this outcome could not be analysed).
- There were only 15 cases for H3/A3 because 9 out of the 24 TA participants had business models based around generation and therefore no interest in delivering turn-down DSR (and, again, the additionality of this outcome could not therefore be analysed, since the outcome was not observed).

The evidence codes are based on review of the following sources of evidence:

- Phase 1 interview findings and supply curve modelling (e.g. evidence on hassle cost levels and exit prices), which provide a census of TA participants.
- Information compiled from the latest TA Capacity Market register and test results provided by BEIS and National Grid (e.g. for volumes of capacity going forward, and for metering, DSR and SPD test results);
- The results of other CM auctions, including earlier T-4 auctions and the recent T-4 auction and Early Auction (e.g. evidence of who participated in these auctions).
- Findings from the email survey of TA participants which was undertaken in autumn 2016 (17 responses).
- Findings from Phase 2 interviews and email exchanges, providing different levels of information on 23 out of 24 TA participants. We obtained detailed interviews with all but one of the aggregators, but had to rely on email exchanges (each covering 3-4 key topics) from four participants who were not sufficiently motivated to respond to the interview request. The key topics were tailored to these participants, and were chosen to fill specific gaps in our understanding of these participants' experience of and behaviour in TA delivery.

- Marketing information and public statements made by TA participants, where we are aware of these. We proactively checked website statements where we were unclear about aggregator's motivations.

The evidence database for a particular hypothesis includes contextual information on each case, together with codes for each evidence test, together with notes explaining the rationale for the coding and the source(s) of evidence on which it is based.

Each evidence test for each hypothesis is coded as 'Yes' (i.e. this evidence has been observed), 'No' (i.e. this evidence has not been observed) or 'Omit' (we do not currently have access to the information required to assess the evidence for this test). Where a test is coded as 'Omit' for a particular case, the prior probability of the relevant hypothesis is not influenced by the result of the test for that case.

In most cases, the coding is quite straightforward. For example, 'was a particular statement made in interview?' (yes/no), or was the exit price from a CMU above or below our modelled costs? (yes/no); or 'did this organisation bid DSR into the recent T-4 or Early Auctions?' (yes/no).

However, some coding decisions are not straightforward. Many of the evidence tests refer to statements being made in interview, but sometimes we have coded 'Yes' where such a statement is implied (e.g. by other related answers) rather than being made explicitly. There can also be subtleties about how to code in cases where an interviewee has explicitly supported a given statement (e.g. 'our capacity is not dependent on TA revenue'), but has not explicitly denied its converse (e.g. 'we are confident in having enough revenue from other sources to maintain our capacity'). Where in doubt, we have coded converse statements as opposites of each other (unless we have specific evidence to the contrary). We have tried to avoid overusing the 'Omit' code, because it is effectively a 'non-code', but we have used this in some cases where we have very little evidence about a particular point (e.g. because an interview was cut short or an interviewee was unwilling to answer a particular question; or because we do not have access to bid prices from the recent T-4 and Early Auction).

Given these uncertainties, we have highlighted marginal coding decisions in the coding databases and have reviewed and adjusted them as part of our quality assurance process and as part of the sensitivity analysis below.

Summary of observed evidence

This section summarises the types of evidence observed from these various sources. To avoid disclosure, this evidence is not identified by case. The tables below provide a brief commentary on the evidence observed. These summaries are presented to give an indication of the weight of evidence observed in relation to each test, rather than to provide meaningful quantitative evidence.

The rows highlighted in pale blue summarise evidence which supports the attribution hypothesis in each case. 'Yes' for these evidence tests supports the attribution hypothesis. Conversely, 'Yes' for evidence tests in the pink shaded rows (relating to the alternative hypothesis) supports the competing 'non-attribution' hypothesis, while a 'No' for the alternative hypothesis indirectly supports the attribution hypothesis.

Table A5.9 Summary of evidence observed for H1 and A1 evidence tests

(46 CMUs passed the screening test, because this is the number of CMUs that obtained capacity agreements in the 1st TA auction.)

| Test | Evidence | Yes | No | Omit | Commentary |
|-------------|---|-----|----|------|---|
| H1a | Decisions made since September 2014 (the date of the announcement of the TA) to invest in new capacity or keep capacity available for this CMU, that would otherwise have been closed/mothballed. (do not include if H1b found) | 0 | 46 | 0 | Not generally observed. In practice, this evidence test has been of limited use as public announcements of intentions to invest or close would only be made for very large plant. |
| H1b | The participant saying in the interview that the TA contributed to their decision to invest in capacity or keep capacity available for this CMU (do not include if H1c is found). | 23 | 23 | 0 | Mixed evidence |
| H1c | Evidence from public statements that the TA made a contribution to the decision to invest in capacity or keep it available for this CMU. | 0 | 46 | 0 | Not generally observed – see note for H1a. |
| H1d | TA participants saying in interview that they had considered taking part in the CM but did not do so because the TA required lower credit cover. | 3 | 40 | 3 | Not generally observed |
| H1d.2 (NEW) | TA participants saying in interview that they had considered taking part in the CM but did not do so because the TA offered one-year ahead contracts. | 21 | 22 | 3 | Mixed evidence |
| H1e | The participant being a price maker and the exit price in the auction for this CMU being higher than the hassle costs identified in the supply curve analysis. | 19 | 27 | 0 | Mixed evidence |
| H1f (NEW) | Participant says in interview that it was difficult to meet the DSR/metering tests but they pushed through because they really wanted to obtain TA funding. | 35 | 8 | 3 | Generally observed |
| H1g | The participant isn't penalized for non-delivery during the delivery year. | 0 | 0 | 46 | Insufficient evidence because no stress event |
| H1h (NEW) | The participant meets satisfactory performance day requirements for this CMU. (Do not use if H1g found) | 43 | 3 | 0 | Generally observed |
| H1i (NEW) | The participant states in interview that they have operational plans in place to ensure adequate capacity will be available during a stress event. (Do not use if H1g or H1h found) | 42 | 0 | 4 | Generally observed |

Appendix 5. Contribution tracing paper

| | | | | | |
|-----|--|----|----|----|--|
| A1a | TA participants state in the interview that they would have invested in, or maintained, capacity for this CMU in 2016/17 regardless of the TA | 23 | 23 | 0 | Mixed evidence |
| A1b | Participants being a price taker ²⁰ or the exit price for this CMU being at or below the hassle costs in the TA auction | 27 | 19 | 0 | Mixed evidence |
| A1c | Participants in TA claim in interview that without TA there is sufficient funding (from STOR etc.) to justify investing in or keeping this CMU's capacity available. | 18 | 27 | 1 | Mixed evidence |
| A1d | The participant is penalized for non-delivery at some point during the delivery year. | 0 | 0 | 46 | Insufficient evidence because no stress event |
| A1e | The participant fails to meet satisfactory performance day requirements for this CMU. (Do not use if A1d found) | 3 | 43 | 0 | Not generally observed – supportive of attribution hypothesis. 3 CMUs currently suspended while SPD issues resolved. |
| A1f | The participant states in interview that they are concerned they may not have adequate capacity available to cover stress event. (Do not use if A1d or A1e found) | 0 | 42 | 4 | Not generally observed – supportive of attribution hypothesis |

²⁰ Existing generation CMUs were classed as 'price-takers' in the auction and were not allowed to exit the auction above a specified price, unless they had specific permission from National Grid to be treated as 'price-makers'. DSR and new-build CMUs were treated as 'price-makers' by default, and allowed to submit exit bids at any price below the auction starting price. The effect of this rule was that some 'price-takers' did not enter an exit bid before the close of the auction, and just accepted the clearing price.

Table A5.10 Summary of evidence observed for H2 and A2 evidence tests

(23 cases out of the 24 TA participants passed the screening test - each case represents an organisation that participated in the first TA and is contemplating future CM participation)

| Test | Evidence | Yes | No | Omit | Commentary |
|-------------|---|-----|----|------|--|
| H2a (NEW) | The participant saying they now have more confidence in being able to meet CM rules and regulations/be competitive in the other CM auctions as a result of the TA | 17 | 1 | 5 | Generally observed - a few gaps in coverage where full interviews were not possible during Phase 2 research |
| H2b | The participant saying in interview that they have developed or invested in assets (e.g. controls), markets (e.g. building a client base, entering the UK market) or skills (e.g. delivering DSR for clients) for the TA that will be used in one or more main CM auctions. | 21 | 2 | 0 | Generally observed (includes development of skills and knowledge base for CM) |
| H2c | The participant did not obtain capacity agreement for TA-type capacity in the T-4 auction, prior to the first TA auction. | 12 | 11 | 0 | Mixed evidence – some had already participated in T-4 prior to the first TA |
| H2d | The participant saying in interview that they intend to bid at a lower price in one of the main CM auctions than they would otherwise have done. (Do not include if H2d.1 is found) | 1 | 19 | 3 | Generally not observed – little information disclosed on bidding intentions. |
| H2d.1 (NEW) | Participant actually bids at a lower price in one of the main CM auctions, on DSR or small-scale generation capacity, than they did in the first TA | 0 | 0 | 23 | No information on bid prices |
| H2e | Participant agrees in an interview that one or more future CM auctions is likely to be more competitive as a result of the TA. | 17 | 2 | 4 | Generally observed |
| A2a (NEW) | Respondents stating in the interview that they would have invested in, or maintained, capacity for future CM auctions regardless of the TA | 8 | 14 | 1 | Mixed – more evidence against than for the alternative hypothesis (e.g. because the TA has prompted aggregators to expand their client base or test new markets; or because a few direct participants specifically needed one-year ahead income) |
| A2b (NEW) | Respondents claim in interview that without TA there is sufficient funding (from STOR etc.) to justify investing in or keeping DSR and small-scale generating capacity available for future CM auctions | 7 | 14 | 2 | Mixed – more evidence against than for the alternative hypothesis (e.g. because of concerns about unreliability of future DSR revenues from various sources) |
| A2c (NEW) | Market-wide: organisations obtain capacity agreements for DSR and small-scale generation capacity in future CM auctions without having participated in the TA. | 23 | 0 | 0 | A few organisations obtained agreements for DSR in T-4 and Early Auctions without having participated in the TA. Not included as cases but provide indirect evidence for A2 across all cases. |

Table A5.11 Summary of evidence observed for H3 and A3 evidence tests

(15 cases because 15 out of 24 TA participants passed the screening test – these were TA participants expressing some interest in turn-down DSR)

| Test | Evidence | Yes | No | Omit | Commentary |
|------|---|-----|----|------|--|
| H3a | Participants confirm in the interview that their long term strategic commitment to turn down DSR strengthened as a result of the TA funding. OR | 3 | 11 | 1 | Not generally observed – many participants were already committed to turn-down. |
| H3b | Participants cite TA funding in public statements as a reason for their commitment to turn down DSR (do not include if H3a is found). | 1 | 14 | 0 | Not generally observed – few public statements available. |
| H3d | Participants in TA say in interview that they are implementing turn down DSR for the first time | 4 | 11 | 0 | Not generally observed because most participants doing turn-down were already active in this area – with a few exceptions. |
| H3e | Participants in TA say in interview that they are considering bidding turn down DSR projects into the CM or other services in future | 12 | 3 | 0 | Generally observed |
| A3a | Respondents state in the interview that the first TA made no difference to their commitment to turn down DSR | 7 | 7 | 1 | Mixed evidence – some already committed; some not committed |
| A3b | Respondents say in interview that turn down DSR projects are considered able to compete effectively in the CM. | 0 | 14 | 1 | No evidence to support this |
| A3c | Respondents say in interview that turn down DSR projects are considered cost effective because of the existing non-CM revenues available to them | 2 | 13 | 0 | Not generally observed – most respondents see turn-down DSR as requiring multiple sources of revenue, of which the CM is one. |
| A3d | Respondents say in interview that turn down DSR projects that exploit smart grids and other new developments are considered likely to be commercially viable in the long run. | 4 | 10 | 1 | Not generally observed – a few references to battery technology changing the market for turn-down DSR (either positively or negatively) |
| A3e | Marketing material and websites for TA aggregators indicate that they are actively marketing other services for turn-down DSR in preference to the TA ²¹ (where those other services are apparently not compatible with the TA – e.g. frequency-related services). | 1 | 11 | 3 | Not generally observed. We particularly sought this information where aggregators appeared to favour other services over CM for turn-down DSR. |

The next section presents the findings from contribution tracing, drawing on the evidence for each case in relation to each hypothesis.

²¹ This test effectively applies to both TA auctions, since current marketing materials are likely to relate to the second rather than first TA.

5. Summary of contribution tracing test findings

Summary of findings for hypothesis 1

The contribution tracing in Phase 2 provided mixed evidence for attribution hypothesis 1 and for the competing alternative hypothesis. Nearly half of the TA capacity was assessed as 'very likely' to support H1, while a similar amount was assessed as 'very likely' to support A1. But support for hypothesis 1 was slightly stronger than in the Phase 1 analysis. This was largely because of increased evidence about participants investing in the TA (e.g. building up their client base; or investing considerable time in getting to grips with CM rules), as well as evidence of compliance with tests in the delivery phase.

As explained above, we have applied these tests at CMU level so that it is possible to estimate the proportion of capacity covered by different outcomes. The figures below have been derived by multiplying the posterior likelihood of the hypothesis for a given CMU by the capacity of that CMU. This is a crude estimate of additionality: for example a CMU would be coded 'Yes' against the test for keeping capacity available, even if just 20% of the CMU was influenced in this way. It was not possible to undertake the contribution tracing at component level, below CMU level, because of lack of data on components. The overall assessment of additionality for H1 depends on all the tests, not just that particular test, so takes account of other evidence.

To avoid a spurious semblance of accuracy from precise probability findings, we present our results using banded categories of probability. To avoid distortion, we have used bands of equal width.

- Very likely (posterior probability > 80%)
- Likely (60% > posterior probability >= 80%)
- Neither likely nor unlikely (40% > posterior probability >= 60%)
- Unlikely (20% > posterior probability >= 40%)
- Very unlikely (posterior probability < 20%)

While these bands are used for presentation purposes in Figure A5.12, more precise numbers have been used to estimate the capacity (weighted by probability) within in each band.

Figure A5.12 Estimated capacity of CMUs providing different levels of support for H1 and A1

H1: TA contributes to 2016/17 security

| Posterior likelihood of H1 | Sum of Delivery capacity (MW) |
|-----------------------------|-------------------------------|
| Very likely | 273 |
| Likely | 0 |
| Neither likely nor unlikely | 104 |
| Unlikely | 60 |
| Very unlikely | 184 |
| Grand Total | 620 |

A1: Capacity would have been available to system anyway in 2016/17

| Posterior probability of A1 | Sum of Delivery capacity (MW) |
|-----------------------------|-------------------------------|
| Very likely | 276 |
| Likely | 41 |
| Neither likely nor unlikely | 147 |
| Unlikely | 0 |
| Very unlikely | 155 |
| Grand Total | 620 |

Further detail of the outcomes across all the CMUs are provided below, in anonymous format. The cases at the top of the table offer low support for H1: many of these are existing generation CMUs. Those at the bottom of the table offer strong support for H1: many of these are unproven DSR CMUs. Support for H1 is strengthened in more than half of the cases, but there are some unproven DSR CMUs which nevertheless provide stronger support for A1 than H1. In most cases, strong support for H1 is accompanied by weak support for A1, and vice versa. We have commented on a few cases which appear anomalous, owing to unusual combinations of circumstances (e.g. existing baseload generation that was new to the flexibility market; or DSR that was not new to the flexibility market but where the participant was attracted by the specific conditions offered by the TA). The qualitative analysis provides more in-depth analysis of the contexts and mechanisms associated with different outcomes.

Figure A5.13 Case by case findings for hypothesis H1 and A1

| Type of CMU | H1 prior (30%) | H1 post | Change in H1 prob | H1 likelihood | A1 prior (70%) | A1 post | Change in A1 prob | A1 likelihood | Commentary |
|--------------------------|----------------|---------|---------------------|-----------------------------|----------------|---------|---------------------|-----------------------------|--|
| Existing Generating CMU | | 1% | weakens | Very unlikely | | 98% | strengthens | Very likely | |
| Existing Generating CMU | | 3% | weakens | Very unlikely | | 100% | strengthens | Very likely | |
| Existing Generating CMU | | 3% | weakens | Very unlikely | | 98% | strengthens | Very likely | |
| Existing Generating CMU | | 3% | weakens | Very unlikely | | 98% | strengthens | Very likely | |
| Existing Generating CMU | | 3% | weakens | Very unlikely | | 98% | strengthens | Very likely | |
| Existing Generating CMU | | 3% | weakens | Very unlikely | | 100% | strengthens | Very likely | |
| New Build Generating CMU | | 5% | weakens | Very unlikely | | 98% | strengthens | Very likely | |
| Existing Generating CMU | | 6% | weakens | Very unlikely | | 100% | strengthens | Very likely | |
| Existing Generating CMU | | 6% | weakens | Very unlikely | | 100% | strengthens | Very likely | |
| Existing Generating CMU | | 6% | weakens | Very unlikely | | 100% | strengthens | Very likely | |
| Existing Generating CMU | | 11% | weakens | Very unlikely | | 98% | strengthens | Very likely | |
| Existing Generating CMU | | 11% | weakens | Very unlikely | | 98% | strengthens | Very likely | |
| Unproven DSR CMU | | 11% | weakens | Very unlikely | | 92% | strengthens | Very likely | |
| New Build Generating CMU | | 22% | little or no change | Unlikely | | 95% | strengthens | Very likely | |
| Existing Generating CMU | | 22% | little or no change | Unlikely | | 100% | strengthens | Very likely | |
| Existing Generating CMU | | 22% | little or no change | Unlikely | | 100% | strengthens | Very likely | |
| Unproven DSR CMU | | 34% | little or no change | Unlikely | | 59% | weakens | Neither likely nor unlikely | |
| Unproven DSR CMU | | 34% | little or no change | Unlikely | | 59% | weakens | Neither likely nor unlikely | |
| Unproven DSR CMU | | 42% | strengthens | Neither likely nor unlikely | | 85% | strengthens | Very likely | |
| Unproven DSR CMU | | 53% | strengthens | Neither likely nor unlikely | | 59% | weakens | Neither likely nor unlikely | |
| Unproven DSR CMU | | 53% | strengthens | Neither likely nor unlikely | | 59% | weakens | Neither likely nor unlikely | |
| Unproven DSR CMU | | 53% | strengthens | Neither likely nor unlikely | | 59% | weakens | Neither likely nor unlikely | |
| Unproven DSR CMU | | 53% | strengthens | Neither likely nor unlikely | | 98% | strengthens | Very likely | |
| Unproven DSR CMU | | 53% | strengthens | Neither likely nor unlikely | | 98% | strengthens | Very likely | |
| Unproven DSR CMU | | 53% | strengthens | Neither likely nor unlikely | | 98% | strengthens | Very likely | |
| Unproven DSR CMU | | 53% | strengthens | Neither likely nor unlikely | | 98% | strengthens | Very likely | |
| Unproven DSR CMU | | 53% | strengthens | Neither likely nor unlikely | | 98% | strengthens | Very likely | |
| Unproven DSR CMU | | 53% | strengthens | Neither likely nor unlikely | | 99% | strengthens | Very likely | |
| Existing Generating CMU | | 82% | strengthens | Very likely | | 76% | little or no change | Likely | Baseload generation, new to flexibility, wanted 1 year ahead revenue |
| Existing Generating CMU | | 82% | strengthens | Very likely | | 59% | weakens | Neither likely nor unlikely | Baseload generation, new to flexibility, wanted 1 year ahead revenue |
| Unproven DSR CMU | | 82% | strengthens | Very likely | | 6% | weakens | Very unlikely | |
| Unproven DSR CMU | | 82% | strengthens | Very likely | | 6% | weakens | Very unlikely | |
| Unproven DSR CMU | | 82% | strengthens | Very likely | | 6% | weakens | Very unlikely | |
| Unproven DSR CMU | | 82% | strengthens | Very likely | | 6% | weakens | Very unlikely | |
| Unproven DSR CMU | | 82% | strengthens | Very likely | | 6% | weakens | Very unlikely | |
| Unproven DSR CMU | | 82% | strengthens | Very likely | | 98% | strengthens | Very likely | Existing DSR capacity but helped into CM by TA conditions. |
| Existing Generating CMU | | 91% | strengthens | Very likely | | 59% | weakens | Neither likely nor unlikely | |
| Existing Generating CMU | | 91% | strengthens | Very likely | | 59% | weakens | Neither likely nor unlikely | |
| Unproven DSR CMU | | 91% | strengthens | Very likely | | 59% | weakens | Neither likely nor unlikely | |
| Unproven DSR CMU | | 91% | strengthens | Very likely | | 59% | weakens | Neither likely nor unlikely | |
| Unproven DSR CMU | | 91% | strengthens | Very likely | | 59% | weakens | Neither likely nor unlikely | |
| Unproven DSR CMU | | 91% | strengthens | Very likely | | 59% | weakens | Neither likely nor unlikely | |
| Unproven DSR CMU | | 95% | strengthens | Very likely | | 6% | weakens | Very unlikely | |
| Unproven DSR CMU | | 98% | strengthens | Very likely | | 6% | weakens | Very unlikely | |
| Unproven DSR CMU | | 98% | strengthens | Very likely | | 6% | weakens | Very unlikely | |
| Existing Generating CMU | | 99% | strengthens | Very likely | | 59% | weakens | Neither likely nor unlikely | |

Summary of findings for hypothesis 2

The contribution tracing in Phase 2 provides quite strong support for attribution hypothesis 2 and mixed support for the competing alternative hypothesis. The probability of H2 being true was strengthened in all but a few cases. This is consistent with strong support for H2 in the Phase 1 research. Interpretation of evidence about the competitiveness of the main CM was complicated by a number of other factors. For example, some respondents only responded in terms of the influence of the 2nd TA, in spite of prompting by the interviewer; we did not have access to exit price data for any auction apart from the 1st TA; no interviewees volunteered information on their exit prices in later auctions in response to questions about competitiveness; and some interpreted the question as being about the outturn of the other auction (e.g. the overall clearing price, which was influenced by many non-TA factors) rather than about DSR participant volumes and bid prices in these other auctions.

For this hypothesis, we analysed evidence at organizational level. The figures below present the ‘count’ of TA participants for which evidence seemed to provide different levels of support for the competing hypotheses. Most of the cases showed that H2 was ‘likely’ or ‘very likely’ to be true for this participant. But some cases showed support for A2 as well: there was an even split between those cases showing strong support for A2 and this hypothesis as ‘neither likely nor unlikely’.

Figure A5.14 Estimated number of TA participants providing different levels of support for H2 and A2

H2: TA will make CM more competitive

| Posterior probability of H2 | Count of Cases |
|-----------------------------|----------------|
| Very likely | 11 |
| Likely | 7 |
| Unlikely | 1 |
| Very unlikely | 1 |
| Neither likely nor unlikely | 3 |
| Grand Total | 23 |

| Posterior probability of A2 | Count of Cases |
|-----------------------------|----------------|
| Very likely | 12 |
| Likely | 0 |
| Neither likely nor unlikely | 11 |
| Unlikely | 0 |

Appendix 5. Contribution tracing paper

| | |
|--------------------|-----------|
| Very unlikely | 0 |
| Grand Total | 23 |

Further detail of the outcomes across all the TA participants are provided below, in anonymous format. The cases at the top of the table offer low support for H2, while those at the bottom of the table offer strong support for H2. Nearly all the cases show that the evidence strengthens the estimated probability of H2 being true. Support for A2 is more mixed, with some strengthening of support but some cases largely unchanged. There appears to be no particular pattern between aggregators and direct participants. In most cases, strong support for H2 is accompanied by weak support for A2, and vice versa. But we have commented on a few cases which appear anomalous, offering support for both or neither hypothesis. Those that support both H2 and A2 are cases where there is evidence that the participant has gained learning from their TA experience, but little evidence that this will actually affect the competitiveness of their offer in the main CM. Those that support neither H2 nor A2 are cases in which the participant has some reservations about bidding into the main CM in future, in some cases based on problems encountered during the TA. The qualitative analysis provides more in-depth analysis of the contexts and mechanisms associated with different outcomes.

Figure A5.15 Case by case findings for hypothesis H2 and A2

| | Makes CM more competitive | | | | Participants would have bid into future CM at similar prices anyway. | | | | |
|------------|---------------------------|---------|---------------------|-----------------------------|--|---------|---------------------|-----------------------------|--|
| | 60% | | | | 40% | | | | |
| Category | H2 prior | H2 post | H2 change | H2 likelihood | A2 prior | A2 post | A2 change | A2 likelihood | Commentary |
| Aggregator | | 13% | weakens | Very unlikely | | 100% | strengthens | Very likely | |
| Aggregator | | 33% | weakens | Unlikely | | 41% | little or no change | Neither likely nor unlikely | Negative view of CM following TA experience |
| Direct | | 40% | weakens | Neither likely nor unlikely | | 41% | little or no change | Neither likely nor unlikely | Negative view of CM (limited evidence) |
| Direct | | 44% | weakens | Neither likely nor unlikely | | 41% | little or no change | Neither likely nor unlikely | Cautious about bidding into main CM |
| Aggregator | | 57% | little or no change | Neither likely nor unlikely | | 93% | strengthens | Very likely | |
| Direct | | 67% | little or no change | Likely | | 96% | strengthens | Very likely | |
| Direct | | 75% | strengthens | Likely | | 92% | strengthens | Very likely | |
| Aggregator | | 75% | strengthens | Likely | | 41% | little or no change | Neither likely nor unlikely | |
| Aggregator | | 75% | strengthens | Likely | | 92% | strengthens | Very likely | |
| Aggregator | | 75% | strengthens | Likely | | 100% | strengthens | Very likely | |
| Aggregator | | 75% | strengthens | Likely | | 92% | strengthens | Very likely | |
| Aggregator | | 75% | strengthens | Likely | | 41% | little or no change | Neither likely nor unlikely | |
| Direct | | 89% | strengthens | Very likely | | 84% | strengthens | Very likely | little evidence on A2; evidence for test A2c only |
| Direct | | 90% | strengthens | Very likely | | 41% | little or no change | Neither likely nor unlikely | |
| Aggregator | | 92% | strengthens | Very likely | | 41% | little or no change | Neither likely nor unlikely | |
| Aggregator | | 92% | strengthens | Very likely | | 41% | little or no change | Neither likely nor unlikely | |
| Direct | | 97% | strengthens | Very likely | | 41% | little or no change | Neither likely nor unlikely | |
| Direct | | 97% | strengthens | Very likely | | 93% | strengthens | Very likely | Learning from TA but it does not affect their offer. |
| Direct | | 97% | strengthens | Very likely | | 41% | little or no change | Neither likely nor unlikely | |
| Direct | | 97% | strengthens | Very likely | | 100% | strengthens | Very likely | Learning from TA but it does not affect their offer. |
| Aggregator | | 97% | strengthens | Very likely | | 41% | little or no change | Neither likely nor unlikely | |
| Aggregator | | 97% | strengthens | Very likely | | 100% | strengthens | Very likely | Learning from TA but see alternatives to CM. |
| Aggregator | | 97% | strengthens | Very likely | | 93% | strengthens | Very likely | Learning from TA but it does not affect their offer. |

Summary of findings for hypothesis 3

The contribution tracing findings for hypothesis 3 are more difficult to interpret. Overall, there is weak support for both hypotheses, although support for H3 is slightly stronger than for its alternative A3. But quite a few cases offer strong support or weak support for both the attribution hypothesis and its alternative.

As for hypothesis 2, we analysed evidence for H3 and its alternative A3 at organisational level. There are fewer cases because we screened out organisations which had no interest in turn-down DSR at all (e.g. because their business was solely generation). The figures below present the ‘count’ of TA participants for which evidence seemed to provide different levels of support for the competing hypotheses. Evidence for H3 was mixed: a few cases showed that H3 was very likely, but more showed that it was very unlikely. But evidence for A3 was rather weaker, with most cases showing this was neither likely nor unlikely, or very unlikely.

Figure A5.16 Estimated number of TA participants providing different levels of support for H3 and A3

H3: 1st TA supports growth in turn-down DSR

| Posterior probability of H3 | Count of Cases |
|-----------------------------|----------------|
| Very likely | 5 |
| Likely | 1 |
| Neither likely nor unlikely | 1 |
| Unlikely | 0 |
| Very unlikely | 8 |
| Grand Total | 15 |

| Posterior probability of A3 | Count of Cases |
|-----------------------------|----------------|
| Very likely | 2 |
| Likely | 0 |
| Neither likely nor unlikely | 8 |
| Unlikely | 1 |
| Very unlikely | 4 |
| Grand Total | 15 |

Further detail of the outcomes across all the TA participants are provided below, in anonymous format. The cases at the top of the table offer low support for H3, while those at the bottom of the table offer strong support for H3. Slightly under half of the cases showed strengthened support for H3, while just over half showed weakened support for this hypothesis. Of those showing little support for H3, only two show A3 as very likely while most show A3 as neither likely nor unlikely. The evidence indicates that the latter group consists of organisations already committed to providing turn-down DSR, for whom the 1st TA did not influence their commitment, but who think that turn-down DSR still has difficulty competing with generation in the Capacity Market. The level of support for A3 was weak except in two cases. The qualitative analysis presented in the main report provides more in-depth analysis of the contexts and mechanisms associated with different outcomes.

Figure A5.17 Case by case findings for hypothesis H3 and A3

| | TA supports growth in turn-down DSR. | | | Turn-down DSR seen as a long-term opportunity, even if not currently cost-effective, and TA not needed to encourage this. | | | | | |
|------------|--------------------------------------|---------|---------------------|---|----------|---------|---------------------|-----------------------------|--|
| | 55% | | | 45% | | | Commentary | | |
| Category | H3 prior | H3 post | H3 change | H3 likelihood | A3 prior | A3 post | A3 change | A3 likelihood | |
| Aggregator | | 11% | weakens | Very unlikely | | 8% | weakens | Very unlikely | Cautious about turn-down generally, particularly in CM |
| Aggregator | | 11% | weakens | Very unlikely | | 99% | strengthens | Very likely | |
| Aggregator | | 11% | weakens | Very unlikely | | 56% | strengthens | Neither likely nor unlikely | Cautious about turn-down generally, particularly in CM |
| Direct | | 16% | weakens | Very unlikely | | 56% | strengthens | Neither likely nor unlikely | |
| Aggregator | | 16% | weakens | Very unlikely | | 56% | strengthens | Neither likely nor unlikely | Already committed to turn-down; it needs more support |
| Aggregator | | 16% | weakens | Very unlikely | | 56% | strengthens | Neither likely nor unlikely | Already committed to turn-down; it needs more support |
| Aggregator | | 16% | weakens | Very unlikely | | 56% | strengthens | Neither likely nor unlikely | Already committed to turn-down; it needs more support |
| Aggregator | | 16% | weakens | Very unlikely | | 91% | strengthens | Very likely | |
| Aggregator | | 54% | little or no change | Neither likely nor unlikely | | 27% | weakens | Unlikely | |
| Direct | | 72% | strengthens | Likely | | 56% | strengthens | Neither likely nor unlikely | |
| Aggregator | | 82% | strengthens | Very likely | | 8% | weakens | Very unlikely | |
| Aggregator | | 83% | strengthens | Very likely | | 8% | weakens | Very unlikely | |
| Aggregator | | 83% | strengthens | Very likely | | 43% | little or no change | Neither likely nor unlikely | |
| Direct | | 98% | strengthens | Very likely | | 8% | weakens | Very unlikely | |
| Aggregator | | 98% | strengthens | Very likely | | 43% | little or no change | Neither likely nor unlikely | |

6. Sensitivity of results to probability assumptions and coding

Given the number of assumptions involved in these tests, we have checked the sensitivity of results to probability assumptions and coding decisions. We have particularly focused on H3 and A3, because of the relevance of turn-down DSR to the second TA auction.

Sensitivity to slight variations in probability assumptions

We tested the impact on results of amending the prior probability of H3 from 55% to 60% (a variant suggested by the peer reviewer at the beginning of Phase 2), with the prior probability amended for A3 amended to 40%. This had no impact on H3 results but slightly reduced support for A3. A larger change in the prior probabilities would have had more of an impact, but are less realistic, given that the chosen probabilities reflect current understanding and were peer reviewed. The slight change in some of the results justifies the percentage ranges used, rather than relying on point estimates.

Figure A5.18 Sensitivity to prior probability for H3 and A3 – in favour of H3

H3: 1st TA supports growth in turn-down DSR

| Posterior probability of H3 | Basecase | Sensitivity |
|-----------------------------|-----------|-------------|
| Very likely | 5 | 5 |
| Likely | 1 | 1 |
| Neither likely nor unlikely | 1 | 1 |
| Unlikely | 0 | 0 |
| Very unlikely | 8 | 8 |
| Grand Total | 15 | 15 |

| Posterior probability of A3 | Basecase | Sensitivity |
|-----------------------------|----------|-------------|
| Very likely | 2 | 2 |
| Likely | 0 | 0 |
| Neither likely nor unlikely | 8 | 6 |
| Unlikely | 1 | 3 |

Appendix 5. Contribution tracing paper

| | | |
|--------------------|-----------|-----------|
| Very unlikely | 4 | 4 |
| Grand Total | 15 | 15 |

Similarly, we tested the impact on results of changing the Type 1 probability for test H3c from 0.01 to 0.1, and for test A3d from 0.05 to 0.1. The rationale for amending these particular probabilities is that these probabilities were defined to two decimal places, rather than one, which may imply an unrealistic level of accuracy. This influenced the results slightly, reduced the likelihood of H3 and also slightly reduced the likelihood of A3, but did not change the overall picture. It is possible that a larger change in these probabilities, or changes in the probabilities for other tests, might have had more impact.

Figure A5.19 Sensitivity to probabilities for evidence tests H3c and A3d

H3: 1st TA supports growth in turn-down DSR

| Posterior probability of H3 | Basecase | Sensitivity |
|-----------------------------|-----------|-------------|
| Very likely | 5 | 3 |
| Likely | 1 | 1 |
| Neither likely nor unlikely | 1 | 1 |
| Unlikely | 0 | 8 |
| Very unlikely | 8 | 2 |
| Grand Total | 15 | 15 |

| Posterior probability of A3 | Basecase | Sensitivity |
|-----------------------------|----------|-------------|
| Very likely | 2 | 2 |
| Likely | 0 | 0 |
| Neither likely nor unlikely | 8 | 6 |
| Unlikely | 1 | 3 |
| Very unlikely | 4 | 4 |

| | | |
|--------------------|-----------|-----------|
| Grand Total | 15 | 15 |
|--------------------|-----------|-----------|

Sensitivity to coding decisions

We tested the impact on results of amending coding decisions that we had identified as marginal. This involved changing three separate pieces of evidence for H3 and four separate pieces of evidence for A3 (where a piece of evidence is one code for one evidence test). In the first sensitivity, we coded all this marginal evidence in favour of H3 and against A3. This affected the high-level results of one case for H3 and three cases for A3, but did not affect the overall shape of the results.

Figure A5.20 Marginal coding decisions in favour of H3 and against A3

H3: 1st TA supports growth in turn-down DSR

| Posterior probability of H3 | Basecase | Sensitivity |
|-----------------------------|-----------|-------------|
| Very likely | 5 | 5 |
| Likely | 1 | 2 |
| Neither likely nor unlikely | 1 | 1 |
| Unlikely | 0 | 0 |
| Very unlikely | 8 | 7 |
| Grand Total | 15 | 15 |

| Posterior probability of A3 | Basecase | Sensitivity |
|-----------------------------|-----------|-------------|
| Very likely | 2 | 2 |
| Likely | 0 | 0 |
| Neither likely nor unlikely | 8 | 5 |
| Unlikely | 1 | 1 |
| Very unlikely | 4 | 7 |
| Grand Total | 15 | 15 |

In the second test, we coded these same marginal pieces of evidence in the opposite way: against H3 and in favour of A3. The high-level results for H3 were unchanged from the basecase, but this affected two cases for A3. Again, this did not have a significant effect on the overall results.

Figure A5.21 Marginal coding decisions against H3 and in favour of A3

H3: 1st TA supports growth in turn-down DSR

| Posterior probability of H3 | Basecase | Sensitivity |
|-----------------------------|-----------|-------------|
| Very likely | 5 | 5 |
| Likely | 1 | 1 |
| Neither likely nor unlikely | 1 | 1 |
| Unlikely | 0 | 0 |
| Very unlikely | 8 | 8 |
| Grand Total | 15 | 15 |

| Posterior probability of A3 | Basecase | Sensitivity |
|-----------------------------|-----------|-------------|
| Very likely | 2 | 3 |
| Likely | 0 | 0 |
| Neither likely nor unlikely | 8 | 8 |
| Unlikely | 1 | 1 |
| Very unlikely | 4 | 3 |
| Grand Total | 15 | 15 |

Sensitivity to likelihood bands

In our analysis, we chose to use likelihood bands of even width, on the grounds that this would avoid distorting the presentation of results. However, we tested how different the high-level results would look if we changed the banding used to present results. As a sensitivity, we used the probability bands suggested in Barbara Befani's CECAN paper (op cit) for the terms we used. These revised bands, which are based on the probability bands used by the Intergovernmental Panel on Climate Change (IPCC), can be found on the right hand side of Figure A5.22, which compares the bands used in sensitivity analysis with the bands used in the original analysis:

Figure A5.22 Probability bands used in the original analysis and the sensitivity analysis

| Probability bands used in original analysis (based on bands of equal width) | Probability bands used in sensitivity analysis (based on bands used in Barbara Befani's CECAN paper, which in turn are based on bands used by the IPCC, but with bands for 'Virtually certain' and 'Extremely unlikely' omitted) |
|--|---|
| Very likely (posterior probability > 80%) | Very likely (posterior probability > 90%) |
| Likely (posterior probability > 60% but <= 80%) | Likely (posterior probability > 66% but <= 90%) |
| Neither likely nor unlikely (posterior probability > 40% but <= 60%) | Neither likely nor unlikely (posterior probability > 33% but <= 66%) |
| Unlikely (posterior probability > 20% but <= 40%) | Unlikely (posterior probability > 10% but <= 33%) |
| Very unlikely (posterior probability < 20%) | Very unlikely (posterior probability < 10%) |

We omitted the banding that CECAN recommend for 'virtually certain' (99-100%) and 'extremely unlikely' (0-1%), as this would have been less comparable to the Phase 1 results. The revised banding allocates a broader range of probabilities to the central 'neither likely nor unlikely' band compared to the used in the original analysis. Compared to the banding used elsewhere in this paper, CECAN recommends slightly broader bands for 'likely' and 'unlikely', and narrower bands for 'very likely' and 'very unlikely'.

The effect of the revised banding was to reclassify three cases which had been 'very likely' to support H3 to be simply 'likely' to support H3. There was no change in the high-level results for A3.

Figure A5.23 Revised banding for ‘likelihood’ categories.

H3: 1st TA supports growth in turn-down DSR

| Posterior probability of H3 | Basecase | Sensitivity |
|-----------------------------|-----------|-------------|
| Very likely | 5 | 2 |
| Likely | 1 | 4 |
| Neither likely nor unlikely | 1 | 1 |
| Unlikely | 0 | 0 |
| Very unlikely | 8 | 8 |
| Grand Total | 15 | 15 |

| Posterior probability of A3 | Basecase | Sensitivity |
|-----------------------------|-----------|-------------|
| Very likely | 2 | 2 |
| Likely | 0 | 0 |
| Neither likely nor unlikely | 8 | 8 |
| Unlikely | 1 | 1 |
| Very unlikely | 4 | 4 |
| Grand Total | 15 | 15 |

Conclusions from sensitivity tests

The overall results for H3 and A3 were fairly robust to the sensitivity tests applied here. This contrasts with the higher sensitivity found during Phase 1 contribution tracing as the sensitivity of results for a given hypothesis depends on the particular combination of evidence tests and evidence observed.

7. How the contribution tracing findings have been used

These results have been used on a case by case basis to support the formulation of context-mechanism-outcome configurations during the synthesis process. The analysis presented here has been cross-checked against the qualitative analysis and coding decisions have been reviewed to ensure consistency.

The contexts and reasoning mechanisms for TA participants, lying behind these outcome findings, are explored in more detail in the main report, drawing on findings from qualitative research.

Appendix 6. Methodology for email survey of TA participants in autumn 2016

Purpose

An email survey of all TA participants was undertaken in autumn 2016 to provide early findings for BEIS on testing outcomes for the first TA. Early findings were used to inform policy development.

The aim was to fill gaps in information available from National Grid data on the final characteristics of individual CMUs (e.g. size, technology used, % turndown vs back-up DSR in unproven DSR CMUs, costs and revenue streams), to ask for feedback on participants' experience of the metering assessment/testing and DSR testing processes (to inform theory testing), and to ask for indications of their future plans for the TA and other CM auctions (to help inform BEIS thinking on the design of the second TA auction). An email survey was chosen because it was more cost-effective for collection of quantitative information than another wave of in-depth interviews, and because it imposed less burden on respondents.

The survey design was informed by a series of scoping interviews to explore the issues involved in TA testing and delivery. Five semi-structured scoping interviews were conducted with representatives from key organisations involved in overseeing, delivering and regulating the TA: BEIS, National Grid, Ofgem, Elexon and Electricity Market Reform Settlements Limited (EMRS). All but one of these took place in August and early September 2016. The scoping interview with National Grid took place at the beginning of November 2016 owing to staff capacity constraints.

Email survey

An email survey was sent out to all 24 organisations that held capacity agreements in August 2016. The 'relationship manager'²² for each participant sent out and chased responses to this email survey, as we believed this would produce a better response than the chasing being done by an administrator. The invites were sent to the contacts interviewed in Phase 1 of the evaluation. The survey invitations were sent out in the week commencing 26th August 2016. The deadline for responses was 15th September 2016. The survey included:

- Nine questions (closed and short open-ended questions) about participants' experiences of the TA following the auction (covering metering and DSR testing, reasons for CMUs dropping out of the TA, and levels of preparedness for the delivery period).

²² The relationship manager for a TA participant is the social researcher who led the Phase 1 interview with this organisation and who will build up a rapport with them during successive Phases.

- A request for data on each of the participating CMUs (e.g. in relation to the type of DSR or generation involved, existing revenues received for this DSR or generation, etc.).
- A request for information about the clients involved in aggregators' Unproven DSR CMUs, to help identify clients for interview in spring 2016.
- A question about the organisations' plans for taking forward its capacity in the future.

A copy of the email survey is presented in Appendix 8.

Email survey findings

Responses were received from 17 of the 24 TA participants²³. This included responses from ten of the 13 aggregators and seven of the 11 direct participants. It also included responses from all three of the organisations who exited the TA after their CMUs received termination notices. Responses varied in depth. Some respondents provided relatively detailed responses, whilst others were very limited. The email survey findings provided useful insights for BEIS, to inform their response to the CM consultation in autumn 2016. However, the email survey findings have generally been superseded and amplified by subsequent in-depth interviews with TA participants in Spring 2017, as explained in Appendix 7, except in occurrences where full interviews could not be undertaken, in which case the email survey responses were used.

²³ Efforts were made to boost response rates. The initial invite was signed by BEIS, to help increase buy-in and the email survey was designed to be short to complete. The survey questions were contained in a word-processed attachment to the email, so that respondents could see all the questions clearly and collate their responses rather than having to enter all the data at one sitting. Contact with the TA participants' was carried out by their 'relationship managers', to build on the relationships established during the Phase 1 research. TA participants not responding to our initial requests to complete the survey were chased by email and telephone several times.

Appendix 7. Methodology for in-depth interviews with TA participants and aggregator clients in spring 2017

Introduction

In-depth interviews were conducted during April and May 2017 with representatives of nearly all TA participants and with a sample of aggregator clients. Sampling is explained further below.

For TA participants, these Phase 2 interviews extended the information already gathered through Phase 1 interviews (undertaken in March/April 2016) and through the Phase 2 email survey in September 2016.

For aggregator clients, the Phase 2 interviews represented our first source of in-depth insights into the involvement of these organisations in the TA.

Sampling and recruitment

A summary of the sampling strategy is presented in Table A7.1.

The research involved in-depth telephone interviews with representatives of 19 of the 24 organisations that obtained capacity agreements in the first TA auction, together with more limited email or partial responses from a further four organisations. For TA participant organisations, the interviews were generally undertaken with the person primarily responsible for implementation of TA requirements in each organisation (this was generally the contact at the organisation who was involved in Phase 1 research). In total, we obtained some form of response from 23 out of 24 TA participant organisations during Phase 2, adding to the findings already collected from a census of participants in Phase 1 of the evaluation. Only one TA participant organisation, an aggregator, did not provide a response to the Phase 2 qualitative research questions, despite extensive attempts to contact them.

Out of the 23 interviews undertaken with TA participants, three were 'exit interviews' involving organisations that had received capacity agreements in the first TA auction but had subsequently dropped out and had no capacity going forward to delivery. These took place in October and November 2016. The remainder were undertaken at the end of the 2016/17 winter period, in April and May 2017, to allow coverage of delivery issues arising during the winter period (the period in which Capacity Market stress events are most likely to happen). In some cases, the TA participants going forward to delivery had reduced capacity or fewer Capacity Market Units (CMUs) compared to the capacity for which they obtained agreements following the auction, so the reasons for this were probed during the interviews.

Alongside these interviews with TA participants, we undertook in-depth telephone interviews with representatives of seven clients of aggregators that obtained capacity agreements in the first TA auction,

from a pool of 12 such clients. The client organisations were identified by current TA aggregators, in response to a data request from the evaluation team²⁴. Other methods of identifying client organisations, via National Grid scheme data or via the DSR working group of the Power Responsive campaign, were unsuccessful²⁵. Four interviews with aggregator clients had previously been conducted during March and April 2016 as part of Phase 1 of the evaluation, identified via aggregator websites and pre-qualification data for the first TA. Taken together, we have now interviewed 11 aggregator clients, covering six of the 13 aggregators²⁶ that obtained capacity agreements in the first TA auction. These clients came from the following sectors:

- NHS trusts
- Universities
- Energy-intensive industry
- Other manufacturing industry
- Water companies

The limitations of sampling for aggregator clients are discussed in the limitations section further below.

²⁴ Only 3 aggregators shared details of their clients with us. Others were reluctant to share these details owing to confidentiality concerns or concern about the research burden on their clients.

²⁵ National Grid do not hold information on the identity of organisations participating in aggregator CMUs, although they do hold MPAN numbers for all sites in the TA. Matching MPAN numbers to organisation identities would be a major task which has not been attempted during Phase 2. A request for volunteers from the DSR working group of the Power Responsive campaign generated one offer from an organisation who is participating in the 2nd TA, but not the first TA, who will be included in the sample for client research during later phases of the evaluation.

²⁶ The 13 aggregators comprise the 9 aggregators identified as such by National Grid, together with 4 further aggregators or potential aggregators identified on the basis of Phase 1 evaluation evidence.

Table A7.1 Summary of sample for Phase 2 qualitative research

| Sample group | Population | Sampling strategy | Target number of interviews | Number of interviews completed | Email/short call response only | Total responses (% of target) | Response rate (% of sample) |
|---|--|--|-----------------------------|--------------------------------|--------------------------------|-------------------------------|-----------------------------|
| TA participants | | | | | | | |
| Direct participants (proceeded to delivery) | 9 | Census (9) | 9 | 6 | 3 | 9 (100%) | 100% |
| Direct participants (exited TA prior to delivery) | 1 | Census (1) | 1 | 1 | n/a | 1 (100%) | 100% |
| Aggregator participants (proceeded to delivery) | 11 | Census (11) | 11 | 9 | 1 | 10 (91%) ²⁷ | 91% |
| Aggregator participants (exited TA prior to delivery) | 2 | Census (2) | 2 | 2 | n/a | 2 (100%) | 100% |
| Total participants | 24 (excluding clients) | 36 | 34 | 26 | 4 | 30 (88%) | 30 (83%) |
| Clients of TA aggregators | | | | | | | |
| Clients of TA aggregators | Not known - possibly in the range 50-100 clients | Clients identified via aggregators; purposively sampled to provide mix of sectors/aggregators (12) | 10 | 7 | 0 | 7 (70%) | 58% |

Interview approach

For TA participant organisations, the interviews were generally undertaken with the person primarily responsible for the implementation of TA requirements in each organisation (this was generally the contact at the organisation who was involved in Phase 1 research). Respondents were encouraged to involve other individuals in their organisation if needed, to cover the range of topics under discussion. A few of the

²⁷ The non-respondent was identified as a direct participant by National Grid but, on the basis of Phase 1 evaluation evidence, we have categorised them as a potential aggregator.

interviews involved conference calls with more than one respondent in the organisation (e.g. the person responsible for the commercial side of the TA together with someone more involved in implementation of metering requirements). A few others involved follow-up responses by email on more technical points.

For aggregator client organisations, we interviewed the key contact provided by that client's aggregator. Again, this was generally the person responsible for liaising with the aggregator and coordinating delivery of the organisation's TA obligations, including both testing and delivery.

Interview length depended on the number of topics relevant to a particular interviewee. The longest interviews (up to 2 hours) were with aggregators providing unproven DSR or from a varied portfolio of clients and a large number of sites, while the shortest (up to 30 minutes) were with aggregator clients or direct participants providing capacity from existing generation on a single site.

The interviews revisited the topics covered in Phase 1 research to fill gaps and take account of any changes, and also covered new topics relating to testing and fulfilment of obligations. The realist theory was not explicitly discussed with respondents, but interviewers were briefed on the theory and encouraged to probe as required to test the theoretical framework for their interviewees. This was due to the large number of topics to be covered in the interviews, to meet the data requirements of supply curve analysis and contribution tracing.

The broad topics covered in the TA participant interviews are listed below. They were designed to help test the initial theoretical framework (see Appendix 2). Topics were omitted where irrelevant to a particular interviewee (e.g. CMU details were not discussed where CMU composition and capacity were fully known at the time of Phase 1 research and had not changed; DSR testing did not need to be discussed for existing generation CMUs; and metering tests details were only discussed for those using metering options that required testing).

1. **Revisiting organisational contexts:** understanding whether anything important has changed for the organisation since the Phase 1 interview.
2. **Capacity Market Unit (CMU) changes:** understanding why CMUs were terminated or had their capacity reduced [*only applied to organisations that had CMUs terminated or CMUs with reduced capacity*].
3. **CMU composition and capacity:** gaining additional detail on the composition and capacity of CMUs, and costs relating to making this capacity available to the TA.
4. **Metering assessment and test process:** understanding of organisational experiences of the metering assessment and (where applicable) test requirements²⁸ and the costs of meeting these requirements.
5. **DSR test process:** understanding of organisational experiences of the DSR test requirements²⁹ (where applicable) and the costs of meeting these requirements.

²⁸ While all participants had to complete a metering assessment, only sites using 'bespoke metering' and 'balancing services metering' had to undertake metering tests. Sites using supplier settlement metering were exempt from metering tests.

6. **Experience of Satisfactory Performance Days:** understand the organisation's approach to, and experiences of, demonstrating Satisfactory Performance Days, and the costs of meeting these requirements.
7. **Fulfilment: delivery choices during Capacity Market Notice periods:** understand the organisation's approach and response to Capacity Market Notices and potential system stress events, and – where possible – the cost of delivering in response to a CMN.
8. **Views towards the future:** understanding the organisation's attitude to, and plans for, providing capacity in the future, and the influence of their participation in the first TA on this.

For aggregator clients, we did not ask about CMU changes and composition, since these were not relevant at client level. In place of topics (2) and (3), we asked about the nature of the capacity they provide for the TA and their rationale for participating in the TA via an aggregator.

A master topic guide is provided in Appendix 8. The topic guide was designed to test the realist Theoretical Framework, as detailed in Appendix 2. As far as possible within an interview of reasonable length, the topic guide was designed to explore not only what happened, but why and how, including the respondent's reasoning and the factors affecting this reasoning. The realist theory was not explicitly discussed with respondents, but interviewers were briefed on the theory and encouraged to probe as required to test the theoretical framework for their interviewees. In future phases of the evaluation, we will aim to focus the interviews on key aspects of theory that require testing and refining, to allow more time for explicit discussion of theory.

The interviews were undertaken by social researchers from CAG Consultants and Databuild. For almost all TA participants, Phase 2 interviews were undertaken by the same researcher who led the Phase 1 interview, to maximise in-depth understanding of the organisation's situation. In all cases, researchers reviewed existing information on the participant before the interview. This comprised information gathered through Phase 1 interviews, email responses provided to the email survey earlier in Phase 2 (where available), and publicly available information on their participation in the first and second TA and other Capacity Market auctions.

All interviews were recorded and transcribed, except in a couple of cases where the interview was a follow-up to earlier conversations and was very short. Interview findings were written-up in spreadsheet grids prepared by the interviewers. The transcribed recordings were used to finalise the write-ups and to add direct quotes to the write-up spreadsheets.

Analysis approach

We used spreadsheets to code the Phase 2 interview responses against contexts, mechanisms and outcomes in the Theoretical Framework, and to capture additional contexts, mechanisms and outcomes that were supported by the interview evidence but not yet captured by the theory. We analysed the extent of support for different CMOs in the framework and for potential refined or new CMOs. The coding was

²⁹ DSR tests only needed to be completed by unproven DSR CMUs, not existing generation CMUs. There were no proven DSR CMUs in the first TA scheme.

undertaken by two researchers and was cross-checked against findings from other workstreams (e.g. the contribution tracing workstream and the review of final capacity).

Realist analysis during Phase 2 also drew on qualitative evidence from Phase 1. This varied according to the relevance of Phase 1 to the topic in question:

- For the ‘testing’ aspects of delivery theory (equivalent to Step 4 in the Phase 1 theoretical framework), the coding and analysis was based on Phase 2 interview data and email survey data, combined with information from National Grid and the Capacity Market Register on testing outcomes.
- For ‘fulfilment’ aspects of delivery theory (equivalent to Step 5 in the Phase 1 theoretical framework), the coding and analysis was based primarily on reported responses to Capacity Market Notices in Phase 2 interview data. National Grid have no hard data on fulfilment of TA obligations as no stress event occurred in the 2016/17 delivery year.
- For the overarching ‘participation theory’ and ‘non-participation theory’ (covering Steps 1-5 of the Phase 1 theoretical framework), the coding and analysis drew on Phase 1 in-depth interviews, the Phase 2 email survey and Phase 2 in-depth interviews.
- For the additionality theory relating to the TA’s three objectives, the coding and analysis presented here drew on Phase 2 evidence, supplemented by existing evidence from Phase 1 where appropriate. Findings were cross-checked on a case by case basis with findings from contribution tracing, and adjustments made where appropriate to provide a fuller assessment based on all available evidence.

The process for testing theory against evidence for each case is described further in Appendix 4 on Generative Causation Approaches, and is summarised in Figure A4.1.

Limitations

Key limitations of the qualitative research are that:

- Respondent fatigue meant that a few TA participants submitted only partial responses to key questions by email, which did not allow full probing of their reasoning as required to test realist theory. One TA participant did not respond at all. Most of those giving limited or no answers were direct participants offering existing generation, which were lower priority in terms of testing the Theoretical Framework. Those giving limited feedback tended to be those with least capacity to engage with the TA (e.g. where a single individual within an organisation was dealing with TA requirements, as part of a wider role). However one was a potential aggregator that – during Phase 1 – had expressed interest in potentially aggregating existing generation owned by clients. Given BEIS’s particular interest in the impact of the TA in encouraging DSR, and particularly turn-down DSR, we do not feel that these gaps in evidence on existing generation pose a major limitation.
- The aggregator client sample was limited, owing to the reluctance of many aggregators to identify their clients to the evaluation team. There is a risk of bias in that the aggregators choosing to put forward their clients may have had particularly positive or negative experiences of the TA. However, when combined with the Phase 1 sample, we now have 11 interviews, giving reasonable

coverage of sectors and some representation across six of the 13 aggregators and potential aggregators. Our plans for Phases 3 and 4 will include proposals on how to improve the sampling of aggregator clients.

- A further limitation on these findings is that there was a tension between taking a fully realist approach to the interviews (i.e. testing theory explicitly, and probing the rationale and factors underlying an organisation's choices) and keeping interviews to a reasonable length. Interviews with some aggregators extended to 2 hours, which is well beyond the normal limit of 45-50 minutes for telephone interviews. We will address this during Phases 3 and 4 by focusing interviews on key gaps in our understanding of theory, using a combination of email responses and telephone interviews.

Appendix 8. Research instruments for the email survey and in-depth interviews

This appendix presents the email survey administered to TA participants in autumn 2016 and the master topic guide that was used for in-depth interviews in spring 2017. The guide was tailored for particular interviewees (e.g. the DSR testing topic was only covered for interviewees with at least one DSR CMU). The instruments presented here are:

- The email survey
- Invitation email for in-depth interview
- Pre-interview preparation
- Introductory script
- The master topic guide for in-depth interviews

A slightly different topic guide was used for organisations exiting the TA, focusing on experiences of the TA testing process and reasons for leaving rather than on delivery issues.

Email survey

Evaluation of the Transitional Arrangements, Phase 2: survey of all participants

This survey is for all organisations that were successful at auction. The results of the research will be used by BEIS to inform decisions regarding the next Transitional Arrangements auction and the future participation of demand-side response and small-scale generation in the Capacity Market. So this is an opportunity for your organisation to feed into BEIS' decision making.

Please note that we are also planning to conduct follow-up telephone interviews with all TA participants, which provide you with an opportunity to give more in-depth feedback about your experiences to date, what worked well, challenges faced and what can be improved.

Any findings from the survey used in the research will be anonymised; neither you nor your organisation will be named in any published outputs. We will keep any information that you share with us confidential and store it securely, in accordance with the Data Protection Act.

The survey deadline is 15th September 2015.

| | |
|-------------------|--|
| Your name | |
| Organisation name | |

A. Experiences of the TA post-auction

We would like you to provide us with some brief feedback on your experiences of the TA since we interviewed you last. Note that there will be an opportunity to discuss these experiences in more depth through a follow-up telephone interview later on in the research.

1. How straightforward did your organisation find the metering assessment and/or test process?

1 = not at all straightforward, and 5 = very straightforward

Highlight or embolden one number only.

1 2 3 4 5 not applicable

2. What worked well about the metering assessment and/or test process?

Briefly list up to three key elements that worked well

| | |
|---|--|
| a | |
| b | |
| c | |

3. What challenges did you face in undertaking the metering assessment and/or test process?

Briefly list up to three key challenges you faced

| | |
|---|--|
| a | |
| b | |
| c | |

Appendix 8. Research instruments for the email survey and in-depth interviews

4. How straightforward did your organisation find the DSR test process?

1 = not at all straightforward, and 5 = very straightforward

Highlight or embolden one number only.

1

2

3

4

5

not applicable

5. What worked well about the DSR test process?

Briefly list up to three key elements that worked well

| | |
|---|--|
| a | |
| b | |
| c | |

6. What challenges did you face in undertaking the DSR test process?

Briefly list up to three key challenges you faced

| | |
|---|--|
| a | |
| b | |
| c | |

7. If some or all of your CMUs are not going forward to delivery, can you briefly explain why this is?

8. For CMUs going forward to delivery, do you anticipate any issues associated with a stress event warning, delivering your load following obligation or any other aspect of delivery? If yes, please briefly list the key issues.

9. For CMUs going forward to delivery, how prepared do you feel for undertaking the satisfactory performance days?
1 = not very prepared, and 5 = very prepared
Highlight or embolden one number only.

1 2 3 4 5

B. CMU data

10. We have replicated CMU data that you have previously provided below. If there are any changes or additional information, please could you indicate what these are?

Where CMUs have exited the TA, we are still interested in data, especially if it is relevant to your leaving.

We are particularly interested in the opportunity costs and benefits of load reduction / shifting, where we have very little data to-date. We very much appreciate your time in looking at this, as it will help to improve our estimates of the costs of TA participation and any cost barriers and incentives for doing so.

| | Previous answer | Changes / Additional information |
|--|---|----------------------------------|
| Number of sites per CMU | [interviewer to insert from previous interview write-ups] | |
| What assumptions have you made about how often and how long your clients will be able to respond to a stress event? | | |
| Can you talk us through any equipment your clients have installed (or will need to install) to facilitate participation in the Transitional Arrangements? | | |
| For Unproven DSR CMUs | | |
| Status of CMUs (i.e. kW's secured) | | |
| Customer's description (i.e. industrial sector, public sector, office etc.) | | |
| Percentage (of total kW) back-up generation versus load reduction | | |
| Details of any back-up generation (fuel, technology, main purpose, usual running hours etc.) | | |
| Details of load reduction (what is being interrupted, how is it being interrupted, is it a permanent load reduction or load shifting, what are the OPPORTUNITY COSTS and BENEFITS? | | |
| If load shifting, from when to when? | | |
| Have clients incurred any capital and operational expenditure attributable to | | |

Appendix 8. Research instruments for the email survey and in-depth interviews

| | | |
|---|--|--|
| TA? | | |
| What revenues to your clients receive from CMUs (in addition to TA?) | | |
| For existing generation CMUs | | |
| What type of generation is involved? | | |
| Details of costs and revenues for the CMU. | | |
| Annual running hours and if you expect this to change in the future | | |
| For this year, can you provide an approximate breakdown of running hours for each of wholesale market; ancillary services, triad, DNO red zone and other. Do you expect this to change in the future? | | |

Invitation email for in-depth interview

Dear xxxxxx,

You have previously participated in our research for the evaluation of the first round of the Transitional Arrangements. So far our research has focused on the application, auction and testing processes of the first Transitional Arrangements. We are now undertaking a round of more in-depth research with survey participants about testing, delivery and future plans, and as such would like to request your participation in a telephone interview.

The interview will be an opportunity to discuss in more depth the issues you raised in your email survey response, and explore more broadly your experiences of the first Transitional Arrangements, including the final composition of your CMU(s), your experience of metering assessment and DSR testing processes, any response you made to the Capacity Market Notices in the autumn, how your organisation made decisions about the TA (both strategic and operational) and how you view the future regarding DSR, embedded generation, the Capacity Market and flexibility services more generally.

We are conducting these interviews between 3rd and 28th April and we anticipate that the interview will take [no longer than an hour/about 90 minutes]. [Please let us know whether you would prefer to cover the issues in one call, or two separate calls]. We would welcome the involvement of more than one representative from your organisation in the interview, to provide both a strategic and an operational perspective.

I am available from [times]. **I would be grateful if you could suggest a convenient time or times for the interview.**

We would really value your input into this research. In this second phase of the evaluation, we are conducting telephone interviews with all organisations that are contracted for the 2016-17 delivery period. The results of this research will be used by BEIS to inform decisions regarding the second Transitional Arrangements testing and delivery processes, as well as the future participation of demand-side response and small-scale generation in the Capacity Market. So this is an opportunity for your organisation to feed into BEIS' decision-making.

Any findings from the interviews used in the research will be anonymised; neither you nor your organisation will be named in any published outputs. We will keep any information that you share with us confidential and store it securely, in accordance with the Data Protection Act.

If you have any specific questions about the interview or the research evaluation more broadly, please do contact me on [insert email address and contact telephone number].

Kind regards

[insert interviewer name]

Pre-interview preparation

Interviewer to:

- Review the Phase 1 qualitative interview notes and CMU details
- Review the email survey response
- Project analyst to add specific questions/probes about gaps in CMU information, where required, for each participant
- Read through the 'points to probe' document, which highlights specific issues we'd like to follow-up on from the phase 1 interview and email survey response

Introductory script

Good morning/afternoon. My name is [XXXX] and I am calling from [CAG Consultants or Databuild] on behalf of the Department for Business, Energy and Industrial Strategy. Thank you for making the time to talk to me today.

As I explained when arranging the interview, we are currently carrying out the second phase of our evaluation of the first Transitional Arrangements auction. As such, this interview will build on the interview we did with [you/your organisation] in Spring 2016, and the email survey we undertook in September 2016.

This current round of in-depth interviews is with organisations who are participating in the first delivery year of Transitional Arrangements.

The results will be used by BEIS to inform future Government policy in relation to the future of DSR and small-scale generation in the Capacity Market (and particularly the second Transitional Arrangements auction and delivery), so this is an opportunity for your organisation to feed into that decision making process.

Before I begin with the questions, there are a few important points to note:

- We are interested in your organisation's views in relation to the questions we cover and so it would be helpful if you could highlight where you are answering questions from a personal viewpoint only.

Appendix 8. Research instruments for the email survey and in-depth interviews

- We would like you to be as open and honest as possible during the interview. This will help improve our understanding of the first Transitional Arrangements. Neither you nor your organisation will be identifiable in our report to BEIS, unless otherwise agreed with you. We will ensure any data used from the interview is anonymised before publication. We will keep any information that you share with us confidential and store it securely, in accordance with the Data Protection Act.

Before we begin, are you happy for me to record this interview? The interview recording will only be used by the research consortium and will not be shared by the consortium without your prior consent; BEIS will not have access to the interview recording or transcript.

Phase 2 master topic guide – TA 2016-17 delivery year participants

| | Main question | Probes, sub-questions and guidance |
|--|--|---|
| <p>Revisiting organisational contexts</p> <p><i>The aim of these opening questions is to establish whether the organisational contexts we identified in the first phase of research may have changed for this organisation. We want to understand whether there might be changes to the organisation's contexts which may have an impact on its decisions (i.e. mechanisms) about its participation in the TA and the wider flexibility market.</i></p> | | |
| <p><i>Only if interviewing a different person than last time</i></p> | <p>Can you please briefly explain your role in the organisation?</p> | <p>Where do you sit within the organisation?</p> <p>How much influence would you say you have in relation to (a) strategic decision-making (for example in relation to decisions about whether to continue to participate in the Transitional Arrangements) and (b) the operational side of the TA (e.g. meeting meter-testing and DRS-testing requirements)?</p> |
| <p><i>Interviewer to refer back to the interviewee's responses from Phase 1 interview on Organisational Contexts</i></p> | <p>Has there been any significant change to your overall business situation since last year that has affected your approach or ability to provide flexible capacity? If so, please explain why?</p> | <p>For example, has your organisation merged, has its business situation changed significantly, or have there been personnel changes?</p> <p>If so, what difference has this made to your attitude and ability to provide flexible capacity, and why?</p> |
| <p>CMU changes: drop-outs and capacity reductions</p> <p><i>Only for organisations who had CMUs terminated or who had CMUs whose capacity was reduced</i></p> <p><i>A number of organisations participating in the TA had CMUs which were terminated or whose capacity were reduced. If either or both of these apply to your interviewee, these questions seek to explore the reasons and contexts behind why this happened</i></p> <p><i>We want to explore the exit theory for those CMUs that left the TA.</i></p> | | |
| <p><i>Only for organisations who had CMU(s) that were given termination notice(s) for reasons beyond their control (e.g. failing tests etc. despite trying to complete them - as</i></p> | <p>We understand that [x] of your CMUs in the TA was/were given (a) termination notice(s).</p> <p>Can you please summarise the main reason, or reasons, why this/these termination notice(s) was/were given?</p> <p>Why did some of your CMUs pass and</p> | |

Appendix 8. Research instruments for the email survey and in-depth interviews

| | | |
|---|--|--|
| <i>opposed to choosing to withdraw one or more CMUs, see next question)</i> | others fail? What was different about the CMUs that passed? | |
| <i>If they actively decided to withdraw their CMU(s) from the TA (rather than being terminated – e.g. for failing test)</i> | Can you explain the reasons why you decided to withdraw that/those CMU(s)? Why did you retain some CMUs in the TA but withdraw others? What was different about the CMUs that you retained in the TA? | |
| <i>Only for organisations who had CMU(s) whose capacity was/were reduced</i> | We understand that the capacity for [x] of your CMUs was reduced prior to the delivery year commencing. Can you please explain the reasons for this. If not all CMUs were reduced: What were the differences between CMUs that were reduced and CMUs that were not? | |
| <i>All</i> | During the research in spring 2016, you said that your rationale for taking part in the TA was [summarise from Phase 1 interview].. Has this rationale changed since, and if so how/why? | Can you explain any changes in the situation that made continued participation in the first TA unattractive, for some of your CMUs? (e.g. were costs/benefits/risks different from your expectations) Can you explain any change in your overall rationale for participating in the first TA scheme? (e.g. new strategies/perspectives) |
| <p>CMU composition and capacity</p> <p><i>We want to gain some additional detail on CMU composition and capacity. Find answers for each CMU the interviewee has (including any that exited the TA prior to the delivery period. It may be helpful to resend them their email survey responses beforehand, which include a summary of CMU and</i></p> | | |

Appendix 8. Research instruments for the email survey and in-depth interviews

| | | |
|---|--|---|
| <i>cost data previously provided, with a column for any changes.</i> | | |
| <i>For aggregators only</i> | <p>Can you explain how you went about the process of signing up, contracting and managing your clients for the TA?</p> <p>Did you incur any costs in signing up and contracting clients for the 1st TA? If so, can you give an indication of the costs typically incurred in signing up a new client?</p> <p>What types of clients/sites made good prospects for the TA, and why?</p> | <p>What went well? What worked well in attracting clients to the TA? Why? What were the challenges? Why?</p> <p>Costs would include the aggregator's time inputs. The costs might include time spent marketing the TA (or wider flexibility/energy services) to organisations which did not actually sign-up.</p> <p>Roughly what proportion of the clients' capacity you lined up were (a) your existing clients (b) new to your company and (c) new to DSR? Why was this? Roughly what proportion were offering turn-down DSR vs back-up? (refer to email survey response, if any). Why was this?</p> |
| <i>For direct participants only</i> | <p>What proportion of the capacity in your CMUs is new to the flexibility market? (i.e. not already participating in flexibility services)</p> | <p>To what extent would you have invested in or maintained this capacity regardless of the TA?</p> |
| | <p>Project analyst to provide interviewer with any further clarifications/info we would like on this organisation's CMU data, including operating costs for different types of CMUs.</p> <p>Project manager to add probes to test any other aspects of participation/additionality theory for this organisation.</p> | |
| <p>Metering assessment and test process</p> <p><i>Cover all organisations for metering assessment and only cover meter testing where applicable.</i></p> <p><i>We want to gain an understanding of organisational experiences of the metering test requirements and the costs of meeting these requirements. Again, it may be helpful to refer them to their previous response – if they responded to the email survey.</i></p> | | |

Appendix 8. Research instruments for the email survey and in-depth interviews

| | | |
|---|---|--|
| <p><i>Interviewer to have responses to email survey to hand</i></p> <p><i>In next section, ensure participants are referring to the initial metering assessment (which was a screening process that applied to all CMUs) rather than the metering statement or metering test (which was only required for some metering types).</i></p> | | |
| <p><i>All interviewees</i></p> | <p>I'd like to talk next about your experiences of the metering assessment...</p> <p>[if they responded to the email survey] You might like to refer to your responses on metering testing in the email survey (as shared with you earlier).</p> | |
| <p><i>All interviewees</i></p> | <p>Who within the organisation completed the metering assessment?</p> | <p>What technical expertise did they have in relation to metering? Was it the same person responsible for metering testing?</p> |
| <p><i>All interviewees</i></p> | <p>What worked well about the metering assessment process, and why? What worked less well about the metering assessment process, and why?</p> | <p><i>Probe for factors that helped or hindered (and how they helped/hindered) in terms of:</i></p> <ul style="list-style-type: none"> • Organisational capacity and capability (e.g. experience/knowledge of staff, time and resources put it into, preparedness)? • The organisation's technical set-up (e.g. were there particular issues around their metering, etc. Were they aware they might have these issues? If not, why?) • The metering assessment itself (e.g. what rules were an issue and why (and how aware were they of the requirements when the applied for the TA?), were timescales an issue – if so why (and how aware of these timescales)? <p><i>Were there any other factors that helped/hindered the process, and if so how?</i></p> |

Appendix 8. Research instruments for the email survey and in-depth interviews

| | | |
|---|---|--|
| <p><i>All interviewees</i></p> | <p>Did you make any adjustment to your TA strategy in the light of information about metering testing? If so, please explain what this involved and why you did it?</p> | <p><i>(e.g. did they change their choice of metering option? Or did they choose to leave out some sites? If so why, and what difference did this make..?)</i></p> |
| <p><i>All interviewees whose CMUs required a metering test but did not complete one</i></p> | <p>And moving on to the metering test and metering statement process... I understand that one [or more] CMUs required a metering test but you didn't complete the test. Can you explain why?</p> | <p><i>Probe for factors that helped or hindered (and how they helped/hindered) in terms of:</i></p> <ul style="list-style-type: none"> • Organisational capacity and capability • <i>The organisation's technical set-up</i> • The testing/statement process itself <p><i>Were there any other factors that helped/hindered the process, and if so how?</i></p> |
| <p><i>All interviewees who attempted/completed the metering test process</i></p> | <p>What elements of the metering test and metering statement process worked well, and why?</p> | <p><i>Probe for factors that helped (and how they helped) in terms of:</i></p> <ul style="list-style-type: none"> • Organisational capacity and capability • <i>The organisation's technical set-up</i> • The testing/statement process itself <p><i>Were there any other factors that helped the process, and if so how?</i></p> |
| <p><i>All interviewees who attempted/completed the metering test process</i></p> | <p>What elements of the metering test and metering statement process worked less well or was challenging, and why?</p> <p>If you found the metering test/statement challenging, what was the motivation for you to continue your participation in the TA?</p> | <p><i>Probe for factors that hindered (and why they hindered) in terms of:</i></p> <ul style="list-style-type: none"> • Organisational capacity and capability (e.g. experience/knowledge of staff, time and resources put it into, preparedness)? • <i>The organisation's technical set-up</i> (e.g. were there particular issues around their metering, etc. Were they aware they might have these issues? If not, why?) • The testing/statement process itself (e.g. what rules were an issue and why (and how aware were they of the requirements when the applied for the TA?), were timescales an issue – if so why (and how aware of these timescales)?) <p><i>Were there any other factors that hindered the metering test/statement</i></p> |

Appendix 8. Research instruments for the email survey and in-depth interviews

| | | |
|---|--|---|
| | | <i>process?</i> |
| <i>All interviewees</i> | Can you provide a rough estimate of the costs associated with (a) the metering assessment and (b) the meter testing (where applicable) | <p><i>Their own organisation’s staff time required for (a) metering assessment (b) metering testing [and whether this was a consequence of the number of sites, number of clients or complexity of particular sites]</i></p> <p><i>Can they identify any requirements that were particularly time/resource intensive, and explain why?</i></p> <p><i>If they didn’t complete the metering test: what it would have cost (or did cost) to achieve compliance with the metering test – for each CMU?</i></p> <p><i>Was, or would there have been, any need to shut down plant for the meter testing and, if so, what were, or would have been, the implications [and cost – if known] for their (or their clients’) core business? How much notice was needed to a site/client/DNO of a proposed shut down?</i></p> |
| <i>All interviewees</i> | Were these costs above, below or the same as you expected them to be when you took part in the auction? | <p><i>If they were different, by how much? And why?</i></p> <p><i>At what stage in the process did you realise what the true costs would be?</i></p> |
| <i>All interviewees</i> | If you were starting the TA process again now, is there anything you would do differently in relation to metering assessment and/or testing – and why? | <p><i>i.e. what is their organisation learning (if any) from the process?</i></p> <p><i>What difference would this change have made to them?</i></p> |
| <p>DSR test process</p> <p><i>Only for organisations required to do DSR testing</i></p> <p><i>We want to gain an understanding of organisational experiences of the DSR test requirements and the costs of meeting these requirements. Again, it may be helpful to refer them to their previous response – if they responded to the email survey.</i></p> | | |
| <p><i>Interviewer to have responses to email survey to hand</i></p> | | |
| <i>Interviewees who had CMUs that required a DSR test but didn’t complete one</i> | I understand that one [or more] of your CMU(s) required a DSR test but you didn’t complete the test. Can you explain why? | <p><i>If they started with the process, how far did they get?</i></p> <p><i>Probe for factors that helped or hindered in terms of:</i></p> <ul style="list-style-type: none"> • Organisational capacity and capability • The organisation’s technical set-up • The testing process itself <p><i>Were there any other factors that helped/hindered, and if so how?</i></p> |

Appendix 8. Research instruments for the email survey and in-depth interviews

| | | |
|--|--|--|
| <p><i>All interviewees who required a DSR test</i></p> | <p>What elements of the DSR testing process worked well, and why?</p> | <p><i>Probe for factors that helped (and how they helped) in terms of:</i></p> <ul style="list-style-type: none"> • Organisational capacity and capability • <i>The organisation's technical set-up</i> • The DSR testing process itself <p><i>Were there any other factors that helped the process, and if so how?</i></p> |
| <p><i>All interviewees who required a DSR test</i></p> | <p>What elements of the metering test and metering statement process worked less well or was challenging, and why?</p> <p>If you found the DSR test challenging, what was the motivation for you to continue your participation in the TA?</p> | <p><i>Probe for factors that hindered (and why they hindered) in terms of:</i></p> <ul style="list-style-type: none"> • Organisational capacity and capability (e.g. experience/knowledge of staff, time and resources put it into, preparedness)? • <i>The organisation's technical set-up</i> (e.g. were there particular issues around their metering, etc. Were they aware they might have these issues? If not, why?) • The testing/statement process itself (e.g. what rules were an issue and why (and how aware were they of the requirements when the applied for the TA?), were timescales an issue – if so why (and how aware of these timescales)?) <p><i>Were there any other factors that hindered the metering test/statement process?</i></p> |
| <p><i>All interviewees who required a DSR test</i></p> | <p>Can you provide a rough estimate of the costs to your organisation associated with achieving compliance and actually doing the DSR testing?</p> | <p><i>What would it have cost (for CMUs that did not complete the process) or what did it cost (for those that did) to achieve compliance with the DSR test requirements– for each CMU?:</i></p> <p><i>Costs of equipment/kit</i></p> <p><i>Their own organisation's staff time required for DSR testing [and whether this depended on the number of sites, number of CMUs, number of clients or complexity of particular sites] – and a rough estimate of cost in terms of client staff time, where relevant.</i></p> <p><i>Did they have any choices about how to organise DSR testing, and – if so – why did they choose the approach they did? (e.g. joint testing across CMUs vs separate testing of different CMUs)</i></p> <p><i>What were the implications [and cost – if known] of actually doing the DSR</i></p> |

Appendix 8. Research instruments for the email survey and in-depth interviews

| | | |
|--|---|--|
| | | <i>testing for their (or their clients') core businesses?</i> |
| <i>All interviewees who required a test</i> | Were these costs above, below or the same as you expected them to be when you took part in the auction? | <i>If they were different, by how much? And why? At what stage in the process did you realise what the true costs would be?</i> |
| <i>All interviewees who required a test</i> | If you were starting the TA process again now, with a better understanding of the processes and costs involved in participation, is there anything you would do differently in relation to DSR testing? | <i>i.e. what is their organisation learning (if any) from the process? What difference would this change have made to how you approached the TA?</i> |
| <p>Experience of Satisfactory Performance Days</p> <p><i>All CMUs must demonstrate that they have delivered their full capacity obligation for one settlement period on three separate days over winter (1st October - 30th April). Each settlement period used to evidence the CMU has met their obligation is called a Satisfactory Performance Day.</i></p> <p><i>The purpose of these questions is to understand the participants' approach to demonstrating their Satisfactory Performance Days. We want to understand the reasoning for this approach and identify the contexts that may have affected their approach.</i></p> | | |
| <i>For Unproven DSR CMUs</i> | Have you undertaken any of your Satisfactory Performance Days yet? | If so, how many, and when? |
| <i>For Generation CMUs</i> | Have you identified the settlement periods which that you will use to evidence your Satisfactory Performance Days? | If so, have you notified the Delivery Body of these? If no, by when do you intend to notify the Delivery Body about them? <i>[note that they have to deliver by 30th April]</i> |
| <i>For all CMU types</i> | Can you talk me through your approach to, and experience of, demonstrating your Satisfactory Performance Days? | |
| | What is your rationale for this approach? | <p><i>Probe for reasoning behind response:</i></p> <p><i>(e.g. attitude to risk, reputational concerns, fit with operational requirements, commercial reasoning, strategic importance of meeting obligations, concern about penalties, choosing periods when already delivering for Triad or Balancing Services)</i></p> <p><i>(For aggregators: issues re coordination of responses by clients, while</i></p> |

Appendix 8. Research instruments for the email survey and in-depth interviews

| | | |
|--|---|---|
| | | <p>scheduling to fit clients' business needs.)</p> <p>(For direct participants: possible scheduling of SPD periods to fit business needs (e.g. production down-time (for turn-down); testing of generators (for standby))</p> <p>INTRODUCE ANY PROMPTS WITH A LIGHT TOUCH – WE WANT RESPONDENT'S OWN ACCOUNT AND DON'T WANT TO FORCE ANSWERS</p> |
| | What factors or circumstances have informed this approach? | <ul style="list-style-type: none"> • Organisational capacity and capability (e.g. ability to provide/ coordinate full capacity for SPDs, able to respond for duration of SPD period, technical set-up, understanding and awareness of the SPD process, systems to manage clients if aggregator?) • The delivery process itself (e.g. how straightforward were the rules and systems for demonstrating Satisfactory Performance Days?) |
| <i>For all CMU types who HAVE NOT fully delivered their SPDs</i> | Do you envisage there will be any issues in demonstrating your three Satisfactory Performance Days by 30 th April? If so, please explain what/why. | <p><i>If any concerns about meeting SPD requirements, what would need to change to enable them to meet these requirements?</i></p> <p><i>What are the implications of not meeting SPD requirements, and how would this affect them?</i></p> |
| <i>For all CMU types</i> | Were there any costs for your organisation [or your clients] associated with delivering Satisfactory Performance Days? If so, please explain what these were. | <i>Cross-check information provided on costs to earlier answers about how far SPD periods were chosen to fit operational requirements (e.g. periods when load might already be turned down or generators might already be running).</i> |
| <p>Fulfilment: delivery choices during Capacity Market Notice periods</p> <p><i>When a shortage of generation is anticipated, a Capacity Market Notice will be issued by the System Operator. Capacity providers will not be 'called upon' to deliver capacity or receive an individual despatch instruction. The Capacity Market Notice is a signal to all providers that system stress is anticipated (although may not materialise).</i></p> <p><i>We want to understand participants' approach and response to Capacity Market Notices and potential system stress events. We want to understand how/whether they responded to Capacity Market Notices, why they responded as they did, and what contextual factors influenced their response.</i></p> | | |
| <i>For all respondents</i> | | |

Appendix 8. Research instruments for the email survey and in-depth interviews

| | | |
|--|--|---|
| | <p>What operational plans do you have in place to ensure adequate capacity will be available during a system stress event?</p> | <p>Have you made any attempt to pre-judge system stress events?</p> <p>How have you been monitoring Capacity Market Notices? (<i>e.g. by signing up to notifications on the National Grid's website, or setting up more sophisticated monitoring systems</i>)</p> <p>How do you decide whether or not to respond to a particular CMN? Who is involved in this decision?</p> |
| | <p>What factors influenced these plans?</p> | <p>Probe factors like:</p> <ul style="list-style-type: none"> - initial expectations about how many stress events there would be or how long they would last; - expectations on the proportion of your (or clients') capacity would be likely to be available - participation in other services (<i>e.g. National Grid balancing services</i>) |
| | <p>To date, there have been two Capacity Market Notices:</p> | <p>For each one – and for each of your CMUs - how much capacity did you respond with (if any)?</p> <p><i>Want to understand if they responded fully, partially or not at all</i></p> |

Appendix 8. Research instruments for the email survey and in-depth interviews

| | | |
|--|---|--|
| | <ul style="list-style-type: none"> - One issued on 31st October at 12.06pm, which was live from 4.30pm to 7.00pm. The cancellation notice was issued at 6.53pm. - One issued on 7th November at 12.06pm. This was due to go live at 4.30pm but was cancelled at 3.07pm, so never went live. - <p>How did you respond to these?</p> | <p>How did you make the decision for each CMN? (e.g. who was involved, within their own or client organisations)</p> <p>In responding to the CMN(s), did you adjust your response in line with system load at the time?</p> <p><i>Want to understand if they understood and executed their Adjusted Load Following Capacity Obligation (did they respond in full or did they estimate how much capacity they needed to provide)</i></p> |
| | <p>Why did you respond to the Capacity Market Warning(s) in the way you did?</p> | <p><i>Probe for reasoning behind response to each CMN.</i></p> <p><i>(e.g. attitude to risk, reputational concerns, fit with operational requirements, believed CMN would lead (or would not lead) to system stress event, commercial reasoning, overlap with Balancing Services or Triad activity, strategic importance of meeting obligations, concern about penalties, system failures)</i></p> <p><i>If they believed that the CMN would not lead to a system stress event, please probe what information they were basing this assessment on (e.g. other National Grid websites or system information).</i></p> <p>INTRODUCE ANY PROMPTS WITH A LIGHT TOUCH – WE WANT RESPONDENT’S OWN ACCOUNT AND DON’T WANT TO FORCE ANSWERS</p> |
| | <p>What factors influenced your decision about how to respond?</p> | <p><i>Probe for factors that influenced reasoning in terms of:</i></p> <ul style="list-style-type: none"> • Organisational capacity and capability (e.g. ability to respond in time, able to respond for duration of event, technical set-up, understanding and awareness of the CMN process, systems to manage clients if aggregator?) • The delivery process itself (e.g. were communications from National Grid clear, were rules about Adjusted Load Following Capacity Obligation clear? Was the relationship between TA obligations and any Balancing Services requirements clear?) |

Appendix 8. Research instruments for the email survey and in-depth interviews

| | | |
|---|--|---|
| <p><i>For all CMU types</i></p> | <p>Did your organisation [or your clients] incur any costs associated with responding to these CMNs?</p> | <p><i>Relate information on costs to earlier answers about how far the CMNs affected operational requirements (e.g. could loads be turned down without affecting business requirements, and if not what were typical costs of turn down? would generators already be running for other purposes, and if not what were typical costs of running generators for the CMN?).</i></p> <p><i>Probe any differences between costs of responding to CMNs and cost of delivering SPDs – and why?</i></p> <p><i>For aggregators, probe whether their payments to clients are related to whether clients respond when asked or whether they have arrangements to pass penalties on to clients in some way. (Rationale - payments to clients may be perceived as a 'cost' by aggregators).</i></p> |
| <p>Views towards the future</p> <p><i>We want to understand participants' views about their attitude to and plans for providing capacity in the future, as a result of their participation in the first TA. These questions are important for the testing the attribution hypotheses in our contribution tracing tests..</i></p> | | |
| | <p>Knowing what you know now, what would your organisation have done differently if you were starting the TA process again (and why)?</p> | <p>What are the key lessons learned for your organisation?</p> <p>What difference would it have made if you knew then what you know now?</p> <p>How have these lessons impacted on your approach to the second TA auction?</p> |
| | <p><i>Interviewer to note whether they participated in TA auction 2</i></p> <p>Can you explain how you made a decision about participating in the second TA auction (e.g. who was involved and what did they consider)?</p> <p>Why did you choose [not] to participate in the second TA auction?</p> | <p>Please probe for reasons for decision, including :</p> <ul style="list-style-type: none"> - A change in their circumstances (what and how did it affect the decision) - Or a change in the rules (what and how did it affect their decision) - Or a change in their business case (what and how did it affect the decision) <p>Probe for any influence of the first TA learning/experience on choices relating to 2nd TA.</p> |
| | <p><i>Interviewer to note whether they participated in other recent CM auctions</i></p> <p>What about previous Capacity Market</p> | <p>Please explain your answer</p> <p>Probe for any influence of first TA learning/experience on choices relating to this year's T-4 and early auction.</p> <p>(e.g. understanding of rules, perception of risk, confidence in ability to</p> |

Appendix 8. Research instruments for the email survey and in-depth interviews

| | | |
|--|---|--|
| | <p>auctions (T-4, early auction)? What was your reasoning behind whether you participated in these?</p> | <p>comply, kit/customers in place)</p> |
| | <p>What difference has the first TA made to your preparedness for future Capacity Market? How and why has / hasn't it made a difference?</p> | <p>What implications does learning from the first TA have for your future participation in the future Capacity Market (i.e. future T-4 or T-1 auctions), and why? (e.g. bidding at lower price, understanding of rules, perception of risk, kit/customers in place) What impact has your experience of the first TA had on how confident you feel about participating in the Capacity Market, and why? And what influence has it had on your confidence about providing flexible capacity via other services (e.g. STOR, frequency-related services, via DNOs etc.), and why? How would your situation be different in relation to the future Capacity Market if you hadn't participated in the first TA, and why?</p> |
| | <p>Have you developed or invested in assets (e.g. controls), markets (e.g. building a client base, or skills (e.g. knowledge of the capacity market rules and procedures) for the first TA that will help your organisation in the Capacity Market in the future?</p> | <p>To what extent would you have made these investments regardless of the first TA?</p> |
| | <p>Without TA payments in the current delivery year, has there been sufficient funding from other flexibility services (from STOR etc.) to justify investing in or keeping the capacity in your CMUs available for future CM auctions?</p> | <p><i>Probe implications of DSBR not being tendered for winter 2016/17, contrary to expectations at the time of the first TA auction.</i></p> |
| | <p>Looking ahead, to what extent does your organisation have a long term strategic commitment to turn down DSR? What role does turn down play in</p> | <p>What impact, if any, has the first TA had on this commitment (positive or negative)? And what impact has the first TA had your clients' commitment to turn down</p> |

Appendix 8. Research instruments for the email survey and in-depth interviews

| | | |
|----------------|--|--|
| | your business model? | DSR (positive or negative)? |
| | Do you intend to offer flexible capacity via frequency services or other balancing services? If so, please explain which services are you pursuing, or planning to pursue, and how important these opportunities are in relation to the Capacity Market? | Please explain your answer. |
| | Looking across the industry, what implications has the first TA had on the competitiveness of recent CM auctions and the ability of DSR to compete in these and in future CM auctions? | <i>We're asking here about their view of the CM and DSR industry generally, not just their own situation.</i> <i>Or are there other factors that have been more important in encouraging DSR?</i> |
| | Are there any new technologies, approaches or market developments that are likely to make DSR more viable in the long run? | If so, please explain your answer Probe for developments that might make DSR viable without Government support, in the long run. |
| Closing | | |
| | Before we finish, do you have any observations or comments about the first TA? | |
| | Would you be happy for us to get in touch again, if required, as part of this research? | |
| | Specifically, would you be willing to take part in a case study about your CMN response (if selected)? | <i>This would involve a further telephone interview about your choices relating to the CMN.</i> |
| | Did you have any questions before we finish the interview? | |
| | Thank you very much, we really appreciate you taking the time to be interviewed | |

Appendix 9. Methodology for DSR and SPD case studies

Purpose

The original purpose of the case studies was to research the reasoning, costs and outcomes of participants in delivering their obligations during the delivery year, particularly in response to any stress event (as defined by National Grid), to any Capacity Market Notices³⁰ (CMN) and to Satisfactory Performance Days (SPD).

In the absence of a stress event within the delivery year to date or data on the mock stress event³¹, this task focused on selecting five CMUs and using these to examine in-depth the reliability and cost of turn-down DSR, responses to the SPDs and DSR tests and some wider issues about implementation of the delivery year.

Approach

Our approach to this analysis was as follows.

Sampling and selecting clients

- The list of CMUs was reviewed to identify those with turn-down DSR components. The number of components in each CMU was also reviewed.
- During the in-depth interviews, participants were asked if they were willing to be contacted for follow-up research. The sample population was composed of those with at least one turn-down DSR component that had expressed a willingness to be re-contacted. CMUs were then purposively selected in order to highlight a diversity of experiences and issues.
- A sample of five CMUs were selected for analysis, using the criteria above. The five CMUs were from four organisations. These comprised three aggregators and one direct participant.

Analysis of case study CMUs

- The baselines were analysed at the component level. The DSR tests already had this breakdown in sheets provided by the National Grid. For the SPDs, only raw half-hourly data was available, meaning that baseline calculations needed to be replicated.

³⁰ One CMN was issued on 31st October and went live for about 2.5 hours. A second CMN was issued but cancelled before going live on 7th November 2016. Neither CMN developed into a stress event.

³¹ A mock stress event was called by National Grid primarily to test the data flows that would be needed following a stress event. TA participants were not obliged to deliver their obligations for the mock stress event.

- Statistical analysis was performed to review the dispersion of baseline readings around the mean result.
- The DSR test and SPD results of each sample CMU were compared.
- Specific questions about performance of these CMUs in the tests were asked to case study participants by email and phone.

Data sources

The following data sources were used in the case study analysis:

- From the delivery bodies, for a sample of CMUs:
 - DSR test baselines
 - SPD dates
 - SPD raw half hourly data
 - SPD baseline extraction macro
 - Mock stress event and CMN response information was not available.
- From the evaluation team
 - Delivery interviews – April 2016
 - Scoping interviews – August 2016
 - Email surveys – September 2016
 - Delivery interviews – Spring 2017
 - Phone calls and email consultations – late Spring 2017

Case study findings

Owing to potential disclosure, the case studies have not been reported separately. But the findings from case study analysis have formed part of the evidence base for realist analysis and synthesis. Key findings are summarised below.

Approaches to DSR tests and Satisfactory Performance Day (SPD) delivery

In all cases the DSR test result was close to the contractual capacity to be delivered, whereas the SPD results exceeded it by a comfortable margin.

From the feedback of participants, this is because the SPD data can be picked retrospectively to show the required demand reduction and it is usually possible to find a half-hour period when demand is comfortably reduced compared to demand in the baseline period. Nonetheless, some participants did take a proactive approach to the SPDs, particularly where they had many separate components in their CMU(s).

In contrast, the DSR test is arranged in advance requiring the DSR assets to be in a particular operational state at a defined point in time. Participants are only remunerated on availability not utilisation. To minimise cost and disruption, participants are therefore incentivised to only do enough to meet their capacity obligation while managing the risk of test failure.

Variability between tests

Those CMUs with multiple components showed that the contribution of components sometimes varied widely and was not necessarily consistent between tests i.e. one component could contribute significantly to one test but then not contribute at all to another. From the five CMUs reviewed this is most prevalent amongst aggregator CMUs with many components.

Variability of baselines

The baseline methodology did not include any provision for verifying a demand reduction compared to the variability (or dispersion) of the Demand Sample points. To explore this potential issue, the uncertainty associated with the mean MW baseline for a sample of DSR and SPD test meter readings was estimated by considering the combined error of:

- The standard error of the sample mean. This is calculated from the standard deviation of the half hourly baseline data points.
- The precision of meters. It was assumed that the relative meter accuracy was +/- 1.2%, which is consistent with the accuracy of fiscal meters typically used to obtain the half hourly data.

The standard error was then used to express a confidence interval of the mean result. The test result was then checked to determine whether it was statistically different from the baseline readings.

At the CMU level, the test result of all CMUs reviewed was well outside of the 95% confidence interval, meaning that there is a >97.5% probability that the test result is beneath the baseline average, based on the 16 Demand Sample points and the t-test distribution. The CMU level test result can be concluded to be statistically different from the baseline and hence the demand reduction is verifiable.

A minority of components within some CMUs reviewed showed test results which were not statistically different from the baseline readings. This implies that – during the baseline period - they exhibited variability of demand which was of the same order as the demand reduction offered during the test. However, this variability between components is not a concern for the reliability of demand reduction, since capacity obligations apply at the CMU not component level.

Conclusion

The case study analysis confirmed the reliability of demand reduction provided in DSR and SPD tests by CMUs with turn-down DSR components. The analysis generated insights into the contexts and reasoning affecting delivery of the tests which has been incorporated into the realist analysis presented in the main report.