Department for
Business, Energy
& Industrial Strategy

Evaluation of the Transitional Arrangements - phase 4

Final report – appendices

May 2019

Contents

Appendix 1: Introduction to DSR in the second TA and Capacity Market

Introduction to the TA and CM

The Transitional Arrangements for Demand-Side Response (TA) formed part of the Capacity Market (CM) for security of electricity supply, within the government's Electricity Market Reform (EMR) programme. The TA aimed to support BEIS's overall objectives of promoting growth and energy security, while ensuring affordability of the energy supply.

The TA aimed to encourage development of Demand-Side Response (DSR) that is increasingly needed to balance supply and demand in a decarbonised electricity grid[1]. This report uses 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. Under this definition, DSR may be achieved through any combination of onsite generation, temporary demand reduction or load-shifting. We use the term 'turn-down' DSR to refer to the last two activities.

The TA scheme involved two auctions for specific types of capacity within the CM, the first for delivery of capacity in the 2016/17 delivery year[2] and the second for delivery of capacity in 2017/18. While the first TA scheme was open to all types of DSR and also small-scale distribution-connected generation between 2 MW and 50 MW, the second TA scheme was only open to turn-down DSR and had a lower minimum threshold of 500 kW.

The TA auctions were 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 delivered capacity in 2017/18. The main CM auctions offer generation, storage and DSR capacity.

The second TA had two main objectives: to encourage turn-down DSR and to contribute to the development of flexible capacity for the future CM. In contrast to the first TA, BEIS's aim for the second TA did not include a significant contribution to security of supply in the delivery year (2017/18), because short-term system tightness had already been addressed through introduction of the Early Auction alongside the TA. The objectives of the second TA scheme were therefore:

1. To develop a stock of flexible capacity[3] that can be available for future CM auctions, thereby contributing to competitiveness and liquidity in the CM.

[1]National 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

[2] The delivery year runs from first October of one year through to 30th September of the following year.

[3] 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.

2. To encourage enterprise and develop experience, confidence and understanding so that turn-down DSR will be able to realise its potential and ultimately compete with larger generation assets in the CM.

The TA was designed to be a stepping stone to the main CM for flexible capacity[4] that might have difficulty in competing in the main CM. While the TA did not automatically lead on to future CM participation, it aimed to build capacity and confidence so that providers of DSR were better placed to compete in future CM auctions. The timeline for the second TA and other CM auctions is shown in Figure A1.1 below. Phase 3 of the evaluation focused on the second TA auction, while Phase 4 focused on delivery of obligations during the winter period of the delivery year for the second TA.

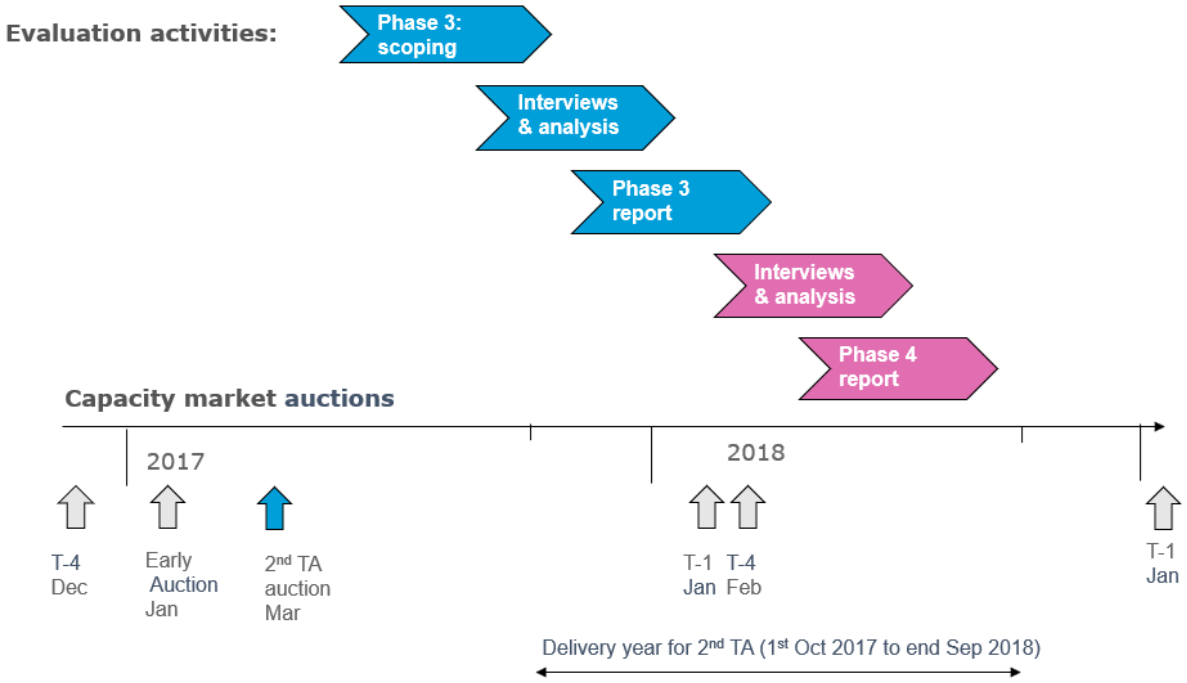


Figure A1.1: Timeline for second TA and main CM auctions, in relation to evaluation activities

Overview of the TA process

The main steps in the TA process for each 'Capacity Market Unit' (CMU) are outlined in Figure A1.2 below, with drop-out points shown in pink. The main CM auctions follow a very similar process. The grey steps were not observed during 2017/18 as there was no CM Notice or associated 'stress event' between 1st October 2017 and 30th September 2018.

[4] 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'.

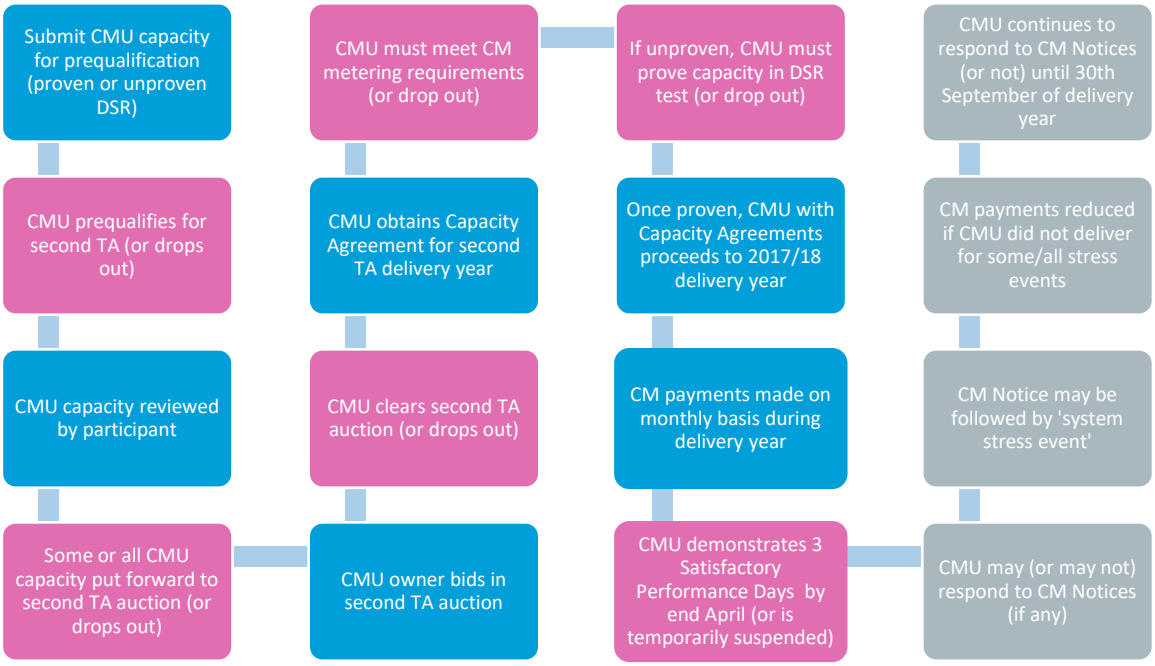


Figure A1.2: Main steps in process for second TA

The main elements of this process are described in more detail below.

Prequalification

Organisations seeking to participate in the second TA had to submit prequalification information for their CMUs to National Grid in autumn 2016. To prequalify, these CMUs had to comprise turn-down DSR within Great Britain. Their capacity had to be between 500 kW and 50 MW, but they could comprise multiple components across different sites and organisations. As in other CM auctions, the second TA was open to both direct participants, putting forward their own capacity, and aggregators, putting forward capacity on behalf of clients.

The turn-down DSR CMUs put forward for the second TA could be ‘proven’ or ‘unproven’. ‘Proven’ CMUs were known assets that had already demonstrated their capacity for the CM by passing a DSR test, as explained further below. ‘Unproven’ CMUs had not yet passed a DSR test and could consist of known assets or an ‘empty bucket’ for capacity that would be identified after the auction.

National Grid reviewed the eligibility of CMUs put forward for prequalification and published prequalification results in the initial ‘CM register’ for the second TA.

CMUs that cleared in the third T-4 (held in December 2016) or the Early Auction (held in January 2016) were excluded from participating in the second TA. There were some changes to prequalification status when the results of these auctions were known.

The TA auction process

Those participants that successfully prequalified CMUs had the option to put their CMU capacity forward to the second TA auction in March 2017. They could choose whether to put forward some or all of their prequalified capacity for each CMU into the auction.

Unproven DSR CMUs had to submit credit cover for their bidding capacity before the auction. Conditions in the TA were softer than the main CM, to encourage new entrants: the level of credit-cover collateral for unproven DSR CMUs was set at £500/MW, compared to £5,000/MW in the main CM. If a participant with a prequalified unproven DSR CMU failed to deliver 90% of the bidding capacity in DSR tests and SPDs (see below), then they lost their credit cover for that CMU. This mechanism was designed to dissuade speculative bids into the auction.

The second TA auction was held in March 2017. This was a descending clock, 'pay as clear'[5] auction in the same format as the main CM auctions. The auction price started at a price cap of £70/kW, which was specified by BEIS before the auction. The auction price was then reduced, round by round, until the volume of capacity remaining in the auction matched the demand that BEIS had specified it would buy at a given price[6]. Auction participants had the option of submitting an 'exit bid' for each of their CMUs in each round, to indicate the price below which they would withdraw a given CMU from the auction. All CMUs that remained in the auction when it cleared stood to receive the auction clearing price for their contracted capacity.

Participants successful in the auction were awarded a capacity agreement for their CMU(s) in the 2017/18 delivery year.

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 Table A1.1 and explained further below.

Table A1.1: Summary of CM 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).

[5] 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).

[6] BEIS specified a demand curve before the auction, indicating how much capacity it would buy at different prices, between the auction price cap and £0/kW.

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.[7]

About metering assessments

All sites within a CMU require 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 DSR CMUs, this involves 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.

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.

For the second TA, but not for other CM auctions, accurate metering was also required for other generating assets behind the meter. This was to ensure that capacity was delivered using turn-down DSR rather than back-up generation assets. Alternatively, participants had the option of making a declaration that they would not use back-up generation behind the meter during tests or stress events, to avoid having to install metering solely for the purposes of the second TA.

[7] The winter delivery period for the second TA is defined as 1st October 2017 to 30th April 2018. Demonstration of satisfactory performance days would not be required if there had been three stress events within this period.

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 cannot be taken for granted as industry codes do not govern these metering options.

For CMUs that know their metering arrangements in advance (e.g. proven DSR), metering assessments are submitted as part of the pre-qualification process. Unproven DSR CMUs can defer their assessments, but these need to be completed at least one month prior to the delivery year.

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 considerable evidence (presented in the main report and in the Phase 1-3 evaluation reports) that many participants found it challenging to meet these accuracy requirements, particularly for metering options (b) and (c).

Our review of Schedule 6 requirements during Phase 2 of the evaluation suggested 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)) presented 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 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 second TA that required a metering test, a test had to be completed and a certificate had to be issued by 31 August 2017 – one month prior to the delivery year for the second 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 from Phase 2 of the evaluation 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 from all Phases of the evaluation 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 method for establishing baseline demand is explained further below. 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 30mins[8] 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.
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.

In the TA, a DSR Test Certificate remained valid for so long as the components in a DSR CMU remained the same. Where they did not, the certificate was invalidated and the CMU was deemed to be an Unproven DSR CMU until such time as a new DSR Test Certificate was issued.

The requirement for CMU components to remain unchanged explained why few of the DSR CMUs that were 'proven' in the first TA were carried forward to the second TA. Only those CMUS that consisted only of turn-down DSR components could be carried forward as 'proven DSR' in the second TA: most DSR CMUs in the

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

first TA included some back-up generation components and therefore could not enter the second TA unchanged.

Similarly, very few CMUs that were 'proven' in the second TA were submitted to subsequent main CM auctions as proven DSR. Analysis of the CM register indicated that the proven DSR CMUs carried forward to subsequent CM auctions were single site or single client/organisation CMUs. Interview evidence from Phases 3 and 4 indicated that aggregators of more complex 'portfolio' CMUs, which included multiple sites and multiple clients, entered them as unproven DSR into successive CM auctions because they wanted flexibility to change components within their CMUs in response to changes in their clients' circumstances or choices. Ofgem is currently considering a rule change that would allow some changes to CMU components in the main CM, without requiring a new DSR test for the whole CMU. This may result in more proven DSR CMUs being submitted to main CM auctions.

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 in the second TA were required to provide £500/MW credit cover. If a participant with a prequalified unproven DSR CMU nominated a lower bidding capacity or failed to deliver 90% of the bidding capacity DSR tests and SPDs, then the credit cover was lost. This mechanism was designed to dissuade speculative bids into the auction.

Credit cover in the main CM is significantly higher: £5,000/MW for unproven DSR CMUs.

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.

A CM rule change introduced before the second TA auction allowed participants to retain their credit cover if they demonstrate at least 90% of their auction capacity in a joint test, while previously they had to demonstrate 100% of capacity.

Baseline methodology for DSR CMUs

Baseline methodology for DSR tests and SPDs

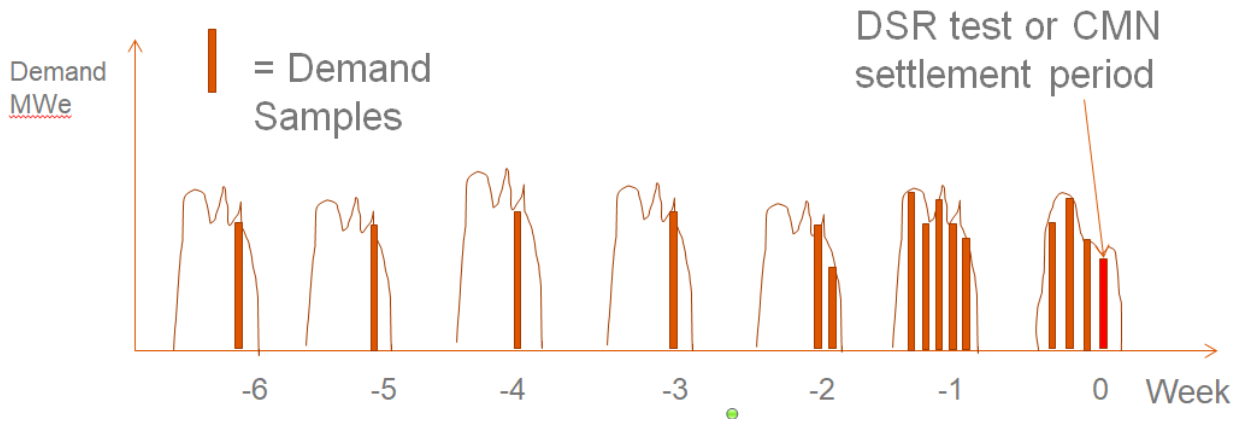
A reduction in energy demand for a DSR CMU cannot be measured directly, only estimated by comparing actual demand against what demand would have been under the same conditions (i.e. establishing a counterfactual). The baseline methodology seeks to provide a fair representation of how a DSR asset would 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 Notice (CMN) Adjustment. The Pre-CMN Adjustment is calculated as the average of the difference between the provisional baseline and the actual demand during the 6 Settlement Periods (i.e. three hours) before the CMN Settlement Period. The Pre-CMN 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:

Adjusted Baseline = Provisional Baseline Demand + the Pre-CMN 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 had to 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 were 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 and do not receive an individual despatch instruction. The Capacity Market Notice is a signal to all providers that system stress is anticipated (although it may not materialise). Four hours after the issue of the Capacity Market Notice, if a System Stress Event occurs, any participant who fails to deliver their ALFCO will be subject to penalties.

No CMNs were issued during the 2017/18 delivery year. Only two CMNs were issued during the 2016/17 delivery year, and neither developed into a 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. The penalty regime is the same as the main CM: penalties cannot exceed the original CM payment set by the auction clearing price.

Appendix 2: Candidate theory for the second TA at the start of Phase 4

The 'candidate theory' at the start of Phase 4, setting out our understanding of how the second TA was working in practice, was the revised theoretical framework from the end of Phase 3 of the evaluation. As explained in the Phase 3 evaluation report, this was based on the objectives of the second TA, as set by BEIS, together with evidence from Phase 3 and earlier phases of the evaluation.

The theoretical framework is set out in the form of realist 'Context-Mechanism-Outcome' configurations: these terms are explained in Table A2.1 below.

Table A2.1: CMO glossary

Term	Explanation
CMOs	Context-Mechanism-Outcome configurations. These are realist hypotheses about how the policy is expected to work, which are tested during the evaluation. See 'realist evaluation'
Context	The circumstances which affect whether a policy 'works' and for whom. Consideration of 'context' forms an important part of realist approaches to evaluation.
Mechanism	A change in people's reasoning, in response to the resources provided by a policy, which leads to a policy outcome. Identification of causal 'mechanisms', which operate in particular 'contexts', forms an important part of realist approaches to evaluation.
Outcome	A change in the state of the world, brought about as a result of a policy or other influences. Realist approaches to evaluation attempt to identify the 'contexts' and 'mechanisms' that lead to a particular 'outcome'.
Realist evaluation	A realist approach[9] to evaluation emphasises the importance of understanding not only whether a policy contributes to outcomes (which may be intended or unintended) but how, for whom and in what circumstances it contributes to these outcomes.

The theoretical framework consists of the following elements, set out in the figures below.

[9] Pawson and Tilley (1997), Pawson (2006)

Figure A2.7: Candidate additionality theory (H1) – TA contributes to more (and/or more competitive) flexible capacity for the CM in 2018/19 and subsequent years

While Phase 3 researched focused on testing and refining participation theory, Phase 4 research focused on testing and refining the elements of the theory relating to additionality (i.e. whether the second TA scheme really made a difference and contributed to its objectives in a way that would not have happened in the absence of the scheme). There was little direct evidence to test reliability theory during Phase 4 because there was no system stress event during the 2017/18 delivery year.

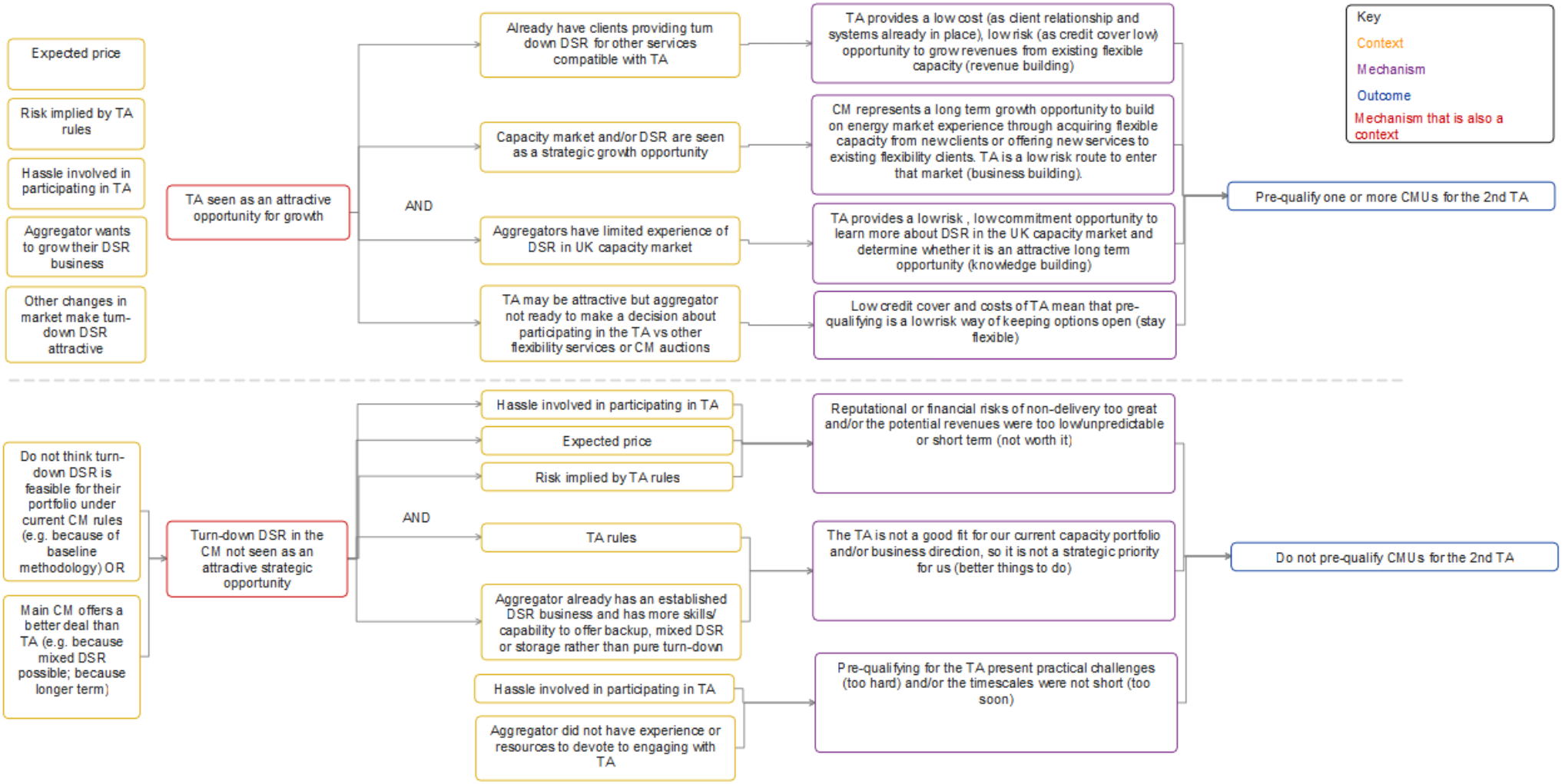


Figure A2.1: Participation theory for aggregators

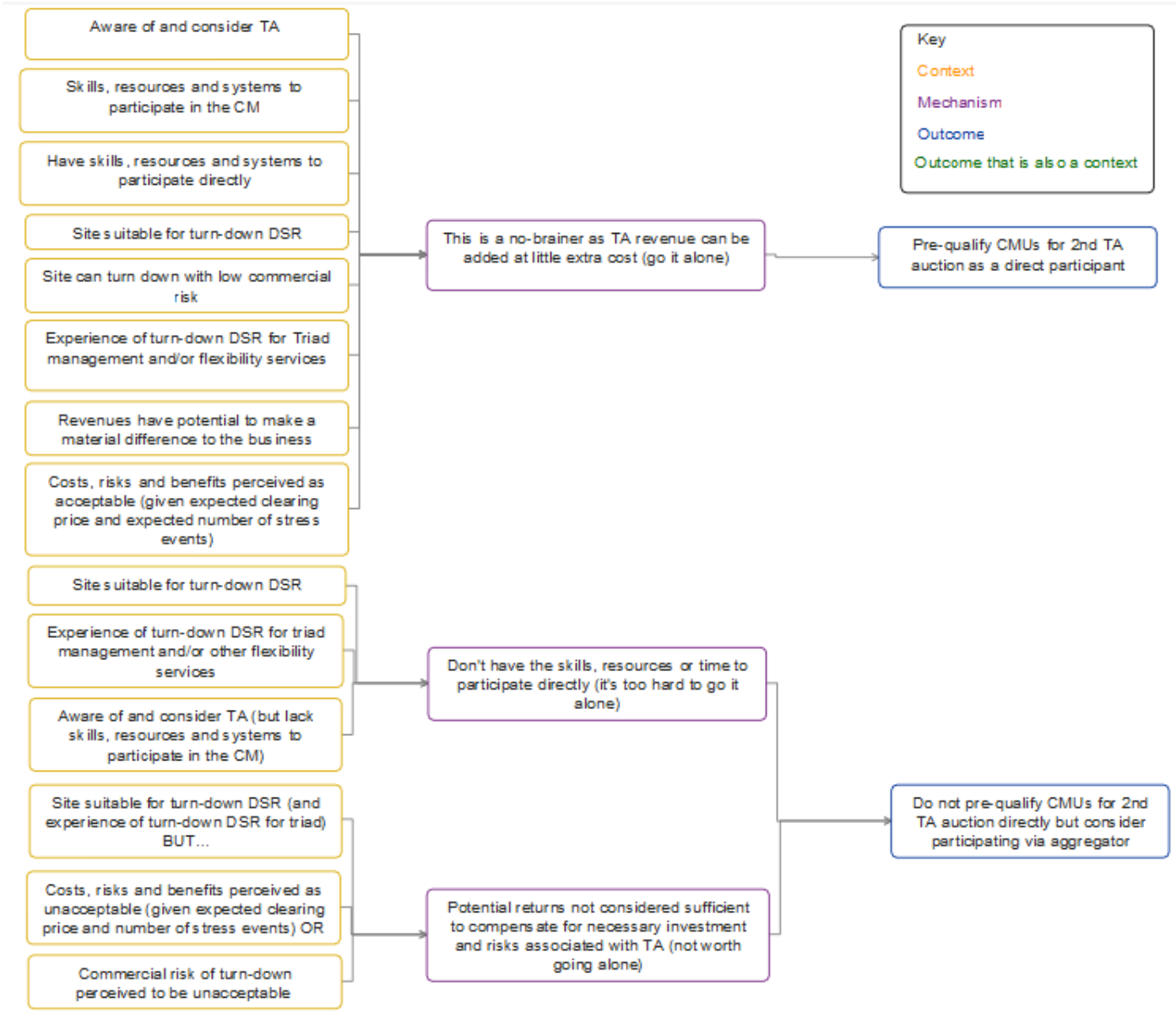


Figure A2.2: Participation theory for direct participants

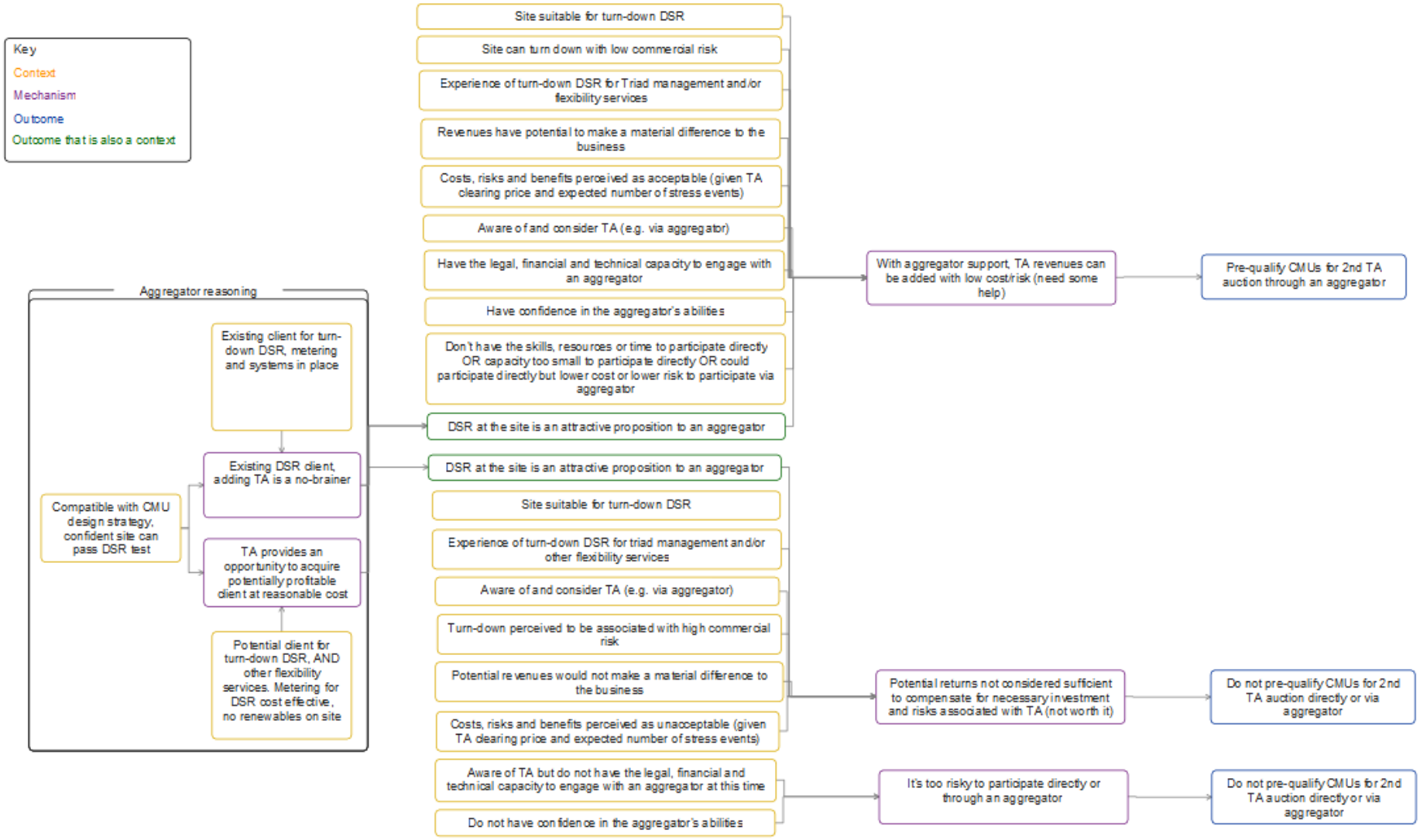


Figure A2.3: Participation theory for aggregator clients

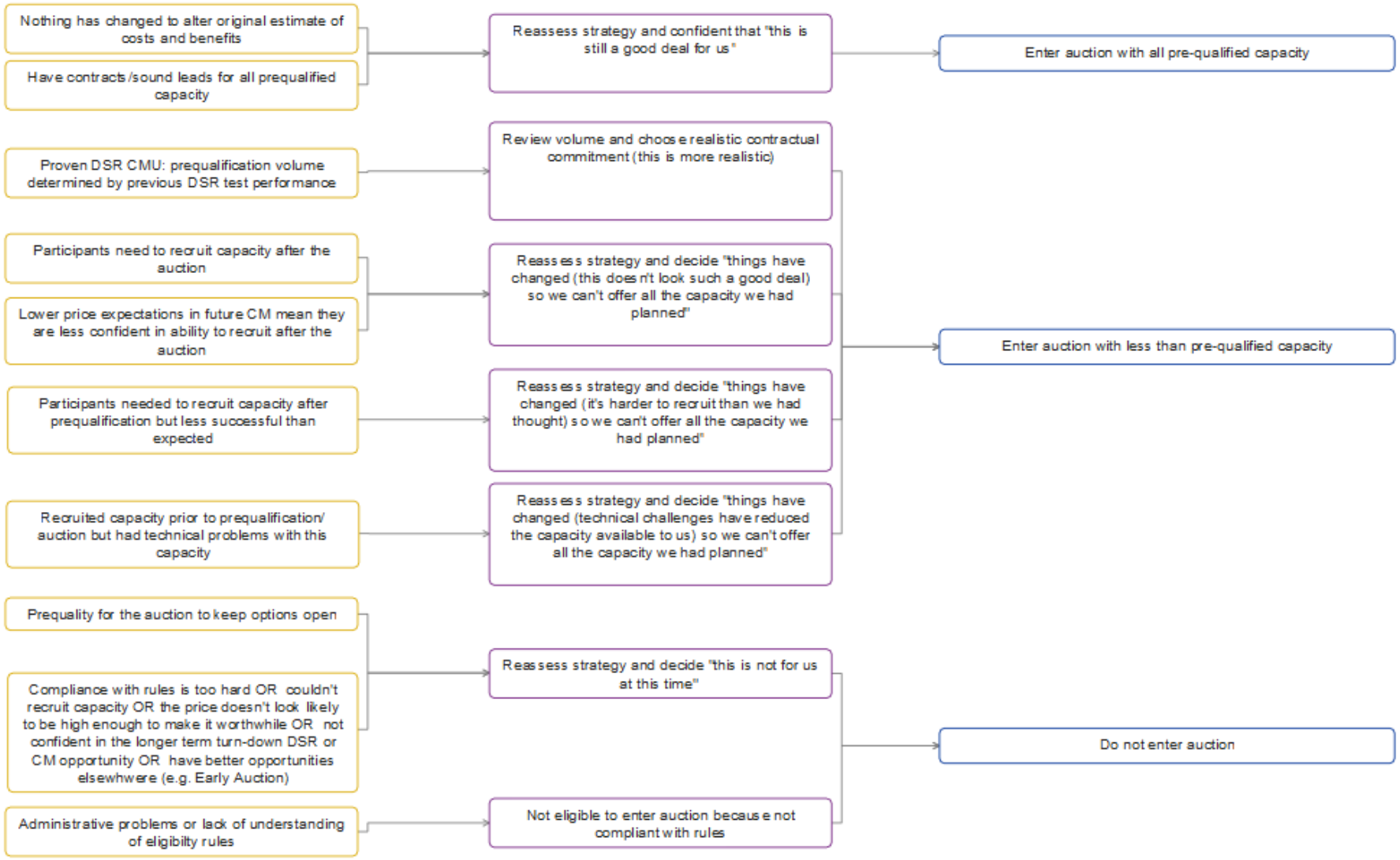


Figure A2.4: Auction participation theory (aggregators and direct participants)

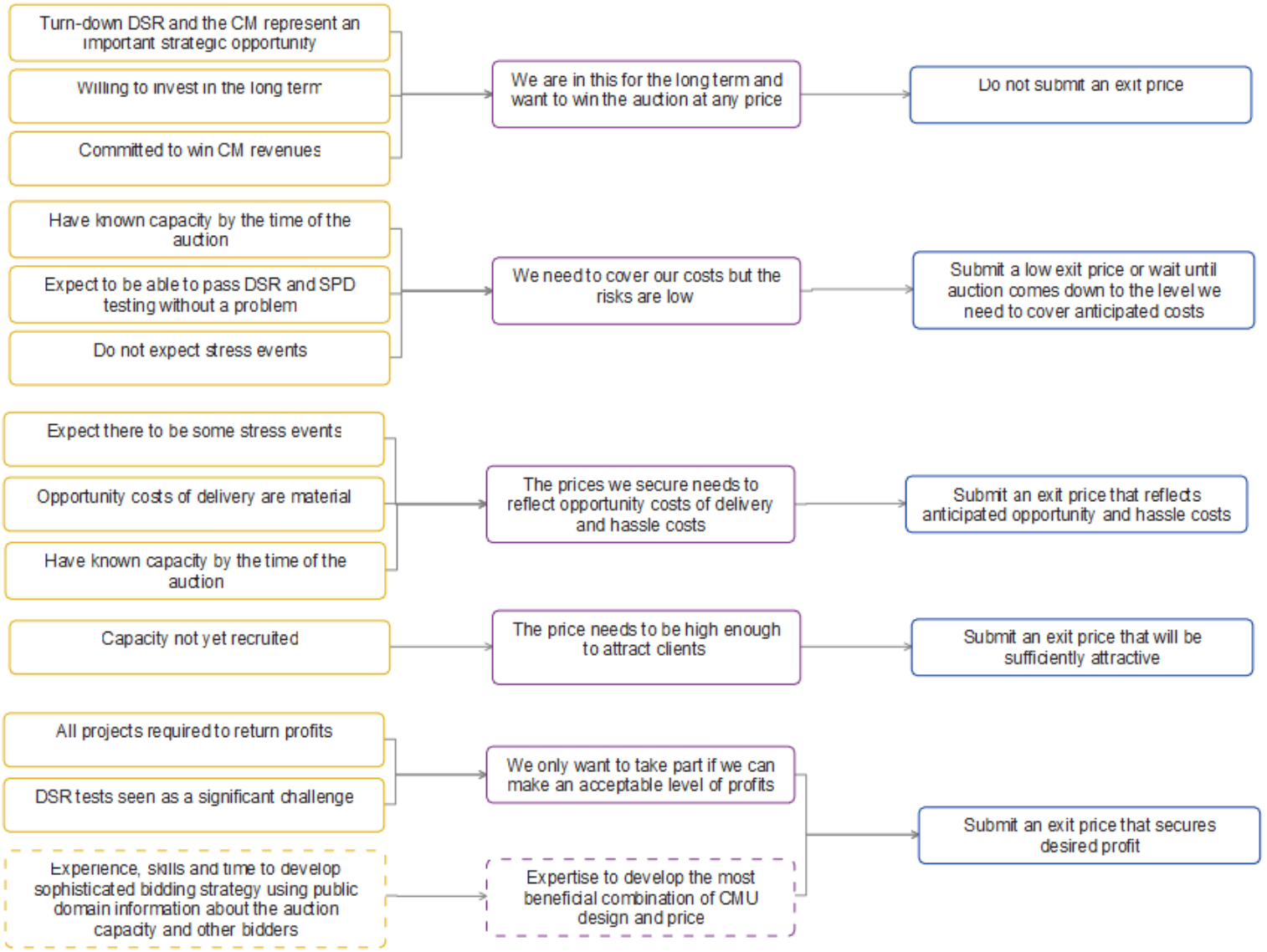


Figure A2.5: Auction bidding theory (aggregators and direct participants)

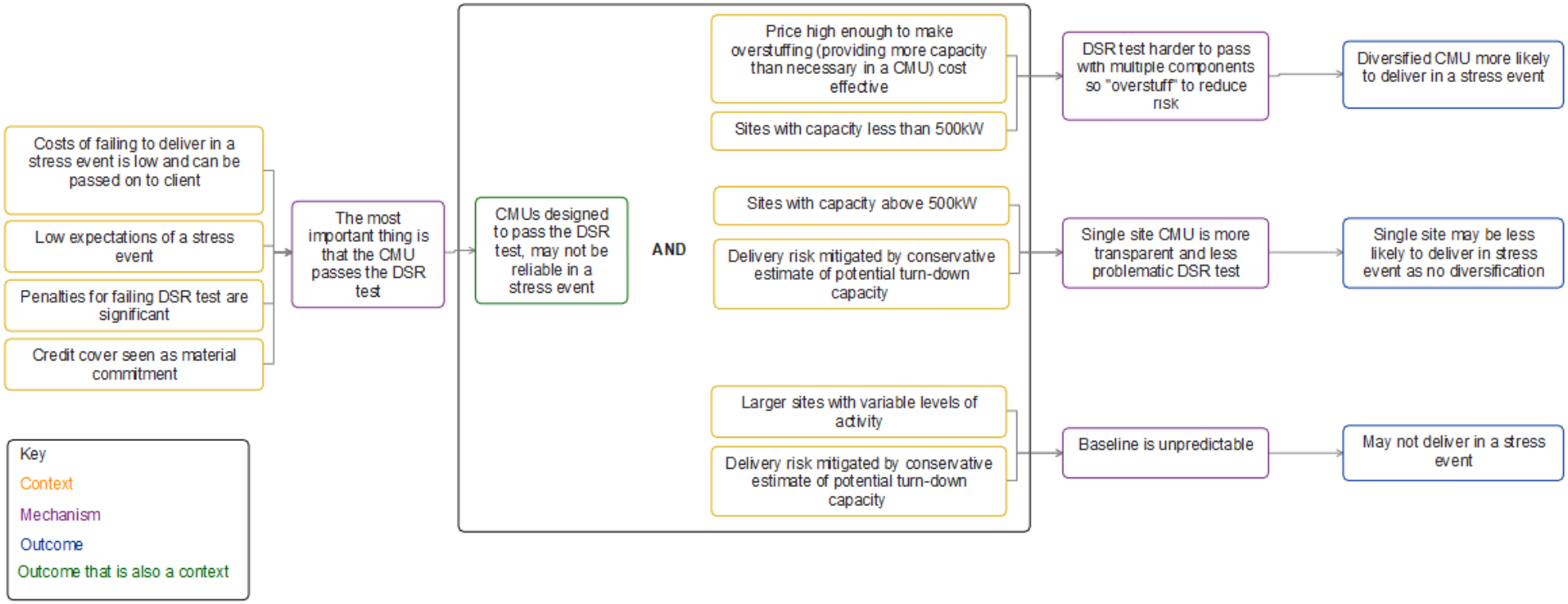


Figure A2.6: CMU design for reliability theory (aggregators and direct participants)

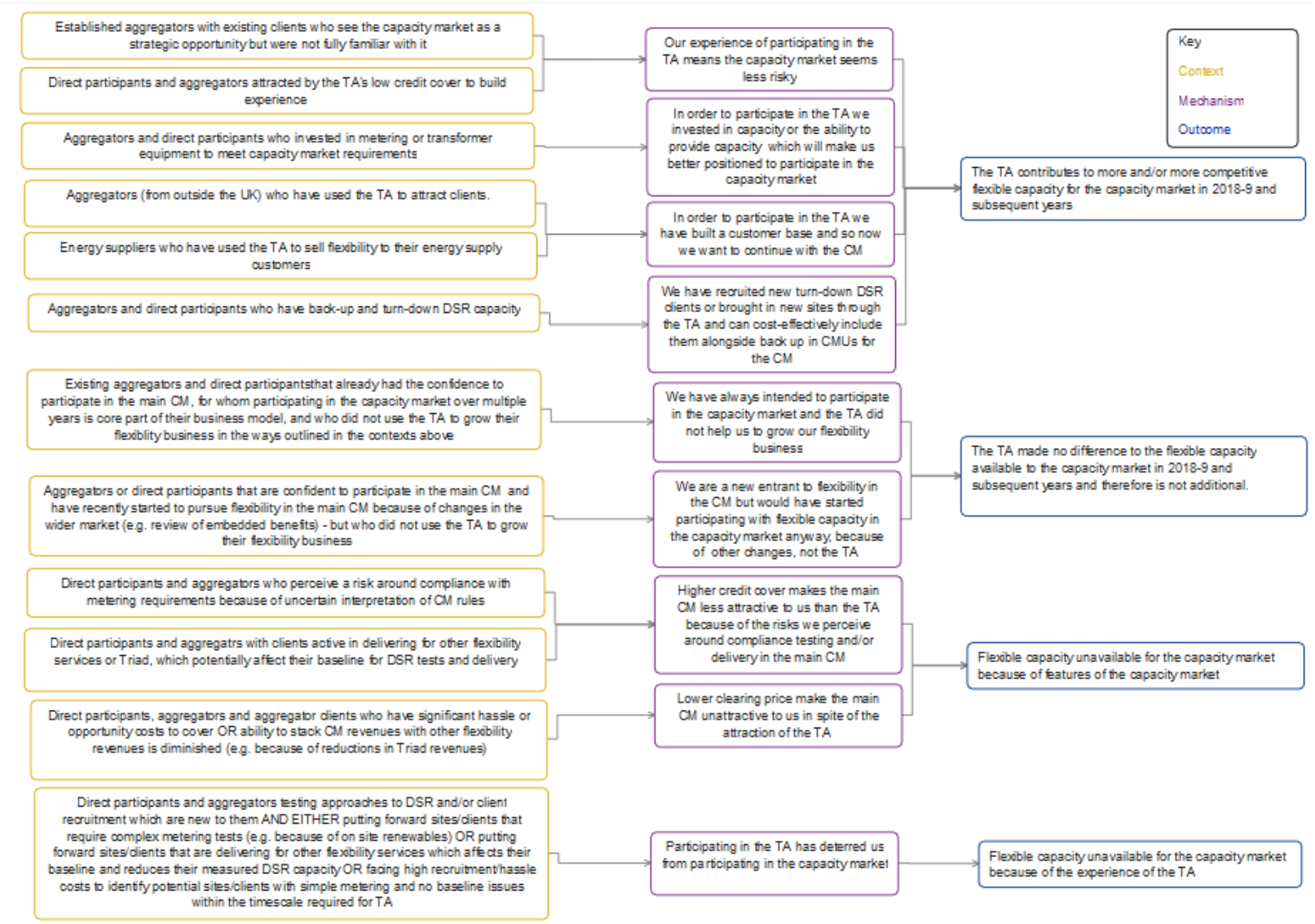


Figure A2.7: Candidate additionality theory (H1) – TA contributes to more (and/or more competitive) flexible capacity for the CM in 2018/19 and subsequent years

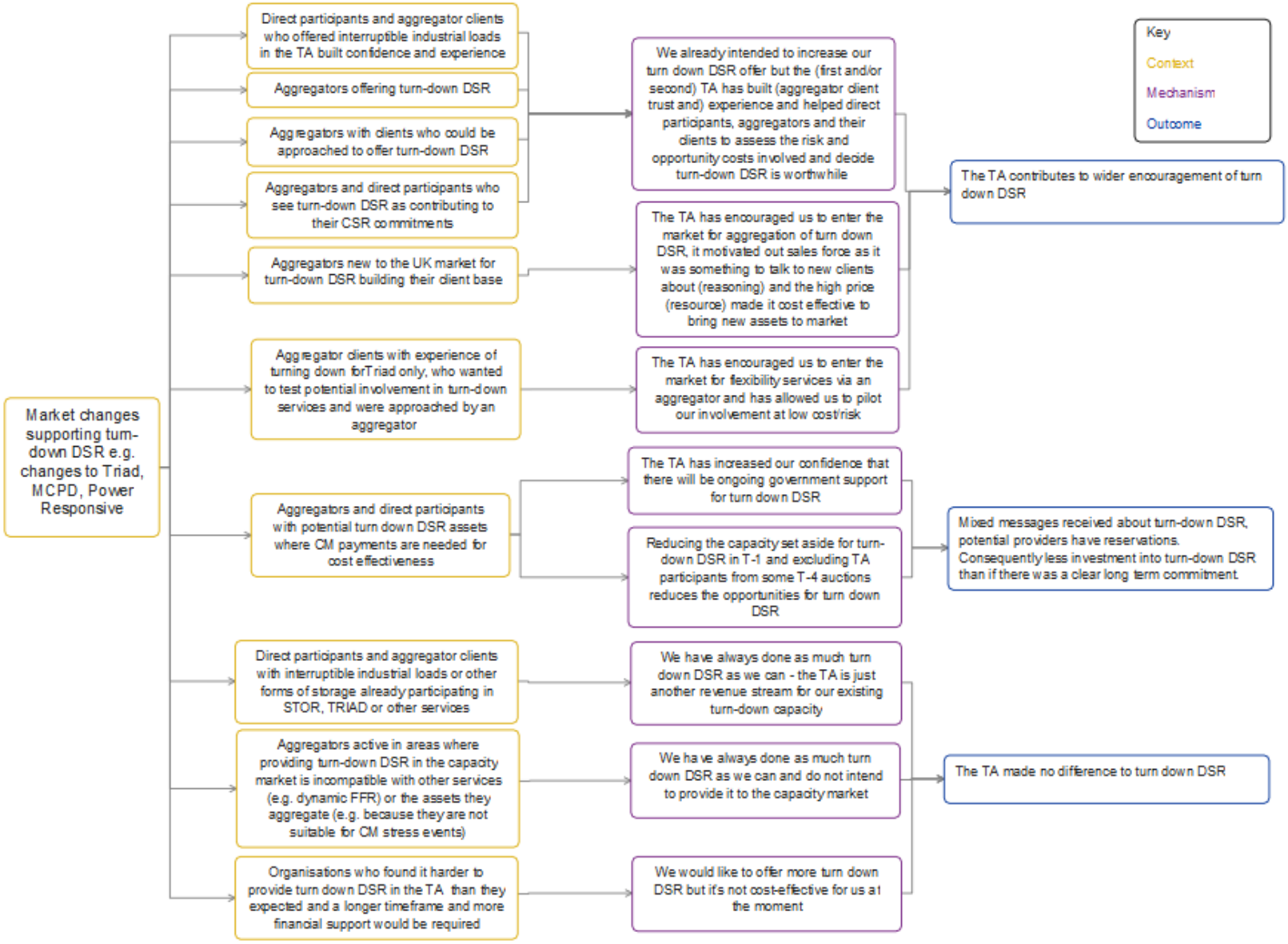


Figure A2.8: Candidate additionality theory (H2) – the TA contributes to wider encouragement of turn-down DSR

Appendix 3: Revised theoretical framework for the second TA at the end of Phase 4

Introduction

Phase 4 evidence was used to test and refine the elements of theory relating to additionality (i.e. whether the second TA scheme really made a difference and contributed to its objectives in a way that would not have happened in the absence of the scheme). This appendix presents the revised additionality theory for the two objectives of the second TA and explains how and why we refined the candidate theories for additionality. The candidate theories themselves are presented in Appendix 2.

We were not able to test the candidate reliability theory as there was no stress event during the 2017/18 delivery year. We have not therefore refined the 'CMU design for reliability theory' set out in Appendix 2.

Also, the candidate theories set out in Appendix 2 for participation, auction participation and auction bidding were not changed by evidence arising from Phase 4 and have therefore not been revised.

This appendix therefore presents revised theory relating to the two objectives of the second TA, which were framed as 'hypotheses' to be tested:

- Additionality theory (Hypothesis 1) – TA contributes to more (and/or more competitive) flexible capacity for the CM in 2018/19 and subsequent year
- Additionality theory (Hypothesis 2) – the TA contributes to wider encouragement of turn-down DSR

These two elements of revised theory are presented below, with an explanation about how and why we developed these revised theories from the candidate additionality theories presented in Appendix 2.

Additionality theory (Hypothesis 1) – TA contributes to more (and/or more competitive) flexible capacity for the CM in 2018/19 and subsequent year

The candidate theory for Hypothesis 1 is presented below in Figure A3.1, with labels for outcomes 1-4. The evidence relating to candidate CMOs for aggregators and their clients showed mixed support for different CMOs and was therefore tested using process tracing methods, as set out in Appendix 5. This showed considerable evidence to support the CMOs leading to outcome 1 and some support for

one of the CMOs leading to outcome 2, but no direct support for CMOs leading to outcomes 3 and 4. The evidence relating to candidate CMOs from the remaining group, direct participants, was clearly supportive of outcome 2, and therefore process tracing testing was not used for these cases.

These testing outcomes, together with detailed understanding of contexts and reasonings emerging from the evidence, led us to refine some aspects of the candidate theory. A diagram of the overall revised theory is presented in Figure A3.2, showing three separate sub-areas of the revised theory. The key areas of change in each of these areas are summarised in Table A3.1, while the detailed revised theory for each area is shown in Figures A3.3 to A3.5 below.

Table A3.1: Summary of changes to sub-areas of H1 theory

Sub-area of H1 theory	How the theory was revised	Why these revisions were made
A – new CMO subsequent to outcome 1	We have clarified that the additional outcomes from the second TA are dependent on what happens in the main CM, both for the 2018/19 delivery year (for which auctions have already been observed) and subsequent years (for which auctions had not been held, at the time of writing).	Chapter 3 of the main report explains that only turn-down DSR from single-site CMUs and from sites with access to frequency services revenue in the second TA cleared in the main CM auctions in 2018, but other types did not. Interviews with TA aggregators, direct participants and clients during Phase 4 provided strong evidence that the actual contribution of capacity developed in the second TA towards future CM auctions will depend on the clearing price in those auctions. While organisations may put forward turn-down DSR from the second TA into the main CM (in either proven or unproven DSR CMUs), they will only obtain capacity agreements if they are willing to accept the clearing price in a particular auction.
B – speculative theory for outcomes 3 and 4	We have condensed outcomes 3 and 4 in the candidate theory into a single 'non-additional' outcome and have marked this as 'speculative'.	We found no outright evidence of outcomes 3 and 4, which involve flexible capacity being deterred from participating in the main CM (either by negative experience of the second TA or by conditions in the main CM). This is because all the second TA participants (aggregators and direct participants) chose to prequalify DSR for the main CM auctions that were held in spring 2018. However, Phase 4 interviews identified some risks around ongoing participation in these auctions, as discussed in chapter 3 of the main report. These are captured in these 'speculative' CMOs for the new outcome 3.
C – refined theory for outcomes 1 and 2	We have refined and rationalised the CMOs for outcomes 1 and 2 as follows: a. A new CMO for aggregator clients previously doing Triad, under outcome 1.	As discussed in chapter 3, Phase 3 and 4 evidence showed that: a. Aggregator clients that were already confident about turning down for Triad did not learn about 'turn-down' per se. But the second TA brought some of them into a new aggregator contract that

Sub-area of H1 theory	How the theory was revised	Why these revisions were made
	b. Removal of direct participants from the contexts for outcome 1, c. Combined the CMOs for aggregators building a customer base and recruiting new clients, for outcome 1. d. Slightly reworded and reordered CMOs for outcome 2.	offered future opportunities for the CM participation. b. Phase 3 and 4 evidence showed that the direct participants were experienced with turn-down DSR in the CM (e.g. from the first TA) and did not find that the second TA made turn-down DSR or the CM seem less risky. c. There were strong similarities in the reasoning of new and existing aggregators that used the attractions of the second TA to build and develop their client base. d. More explicit information emerged during Phase 4 about the context and reasoning for direct participants and aggregators that preferred the main CM to the second TA.

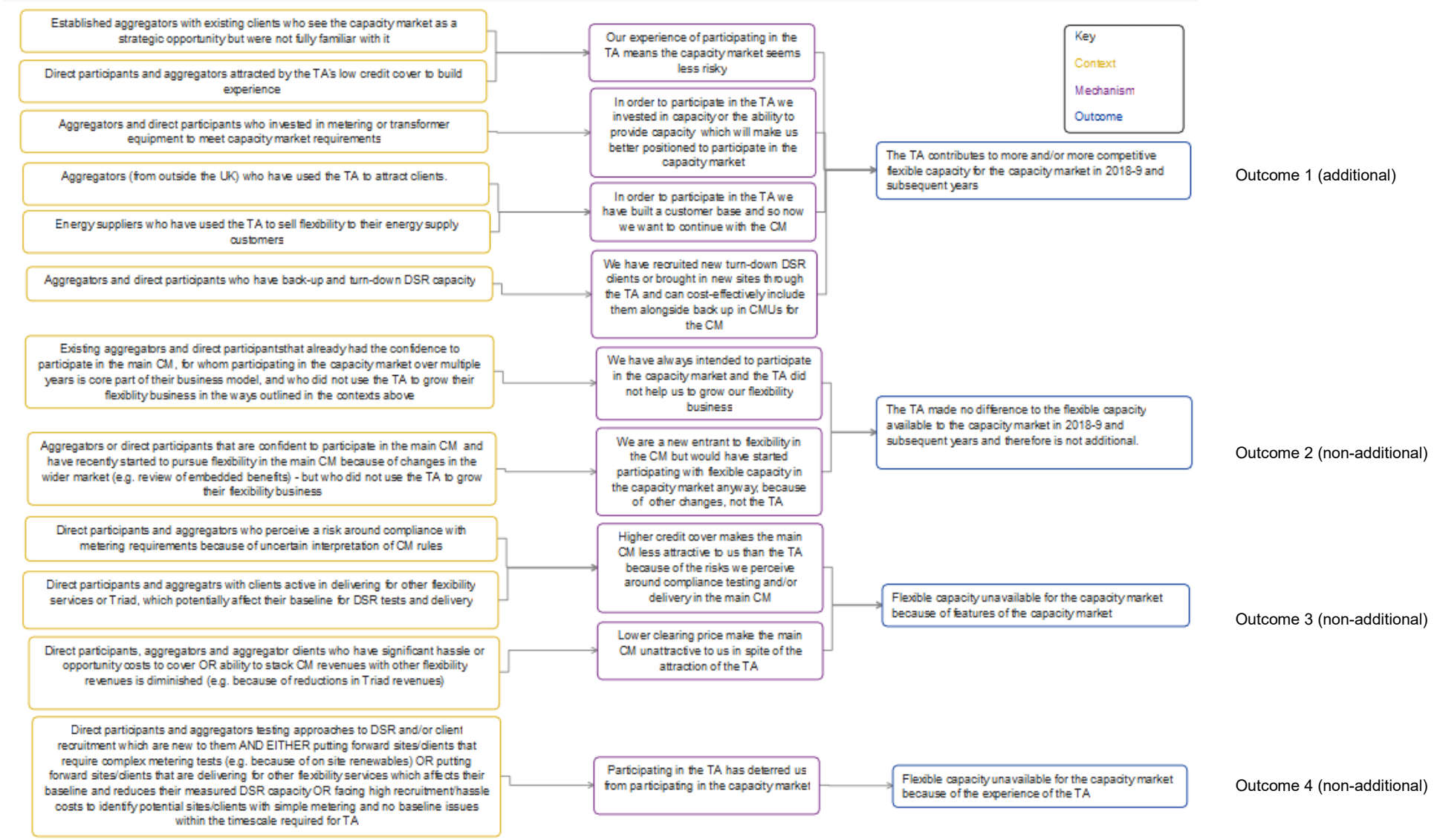


Figure A3.1: Candidate additionality theory (H1) – TA contributes to more (and/or more competitive) flexible capacity for the CM in 2018/19 and subsequent years

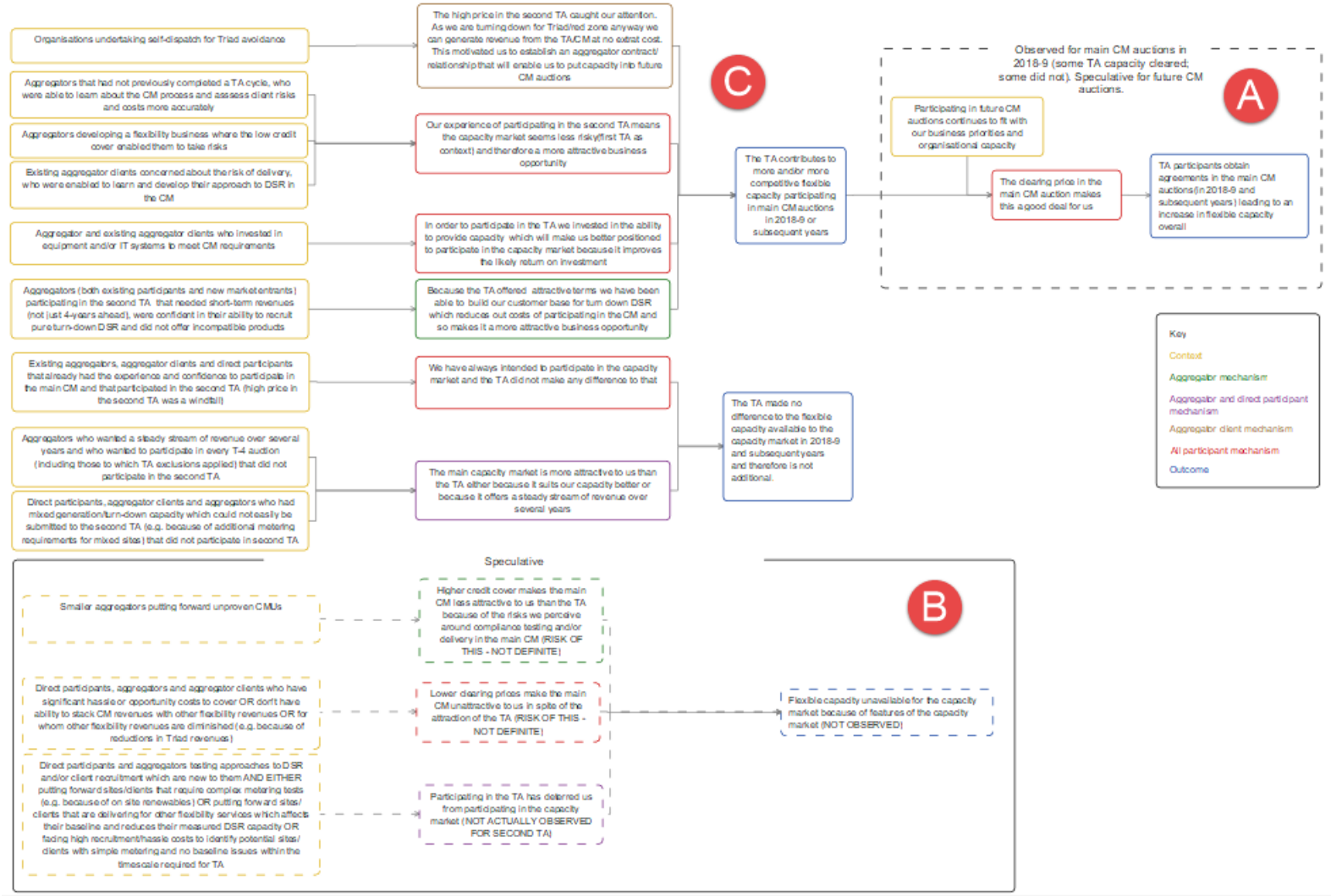


Figure A3.2: Overall map of revised additionality theory for Hypothesis 1 (see charts below for detail on areas A, B, C)

Theory area A – new CMO subsequent to outcome 1 (relating to outcomes in main CM auctions)

This new CMO has been introduced to show that – even if participants from the second TA prequalify for future CM auctions and participate in the auction process – they will only obtain capacity agreements if they think that the clearing price in those auctions offers them a good deal.

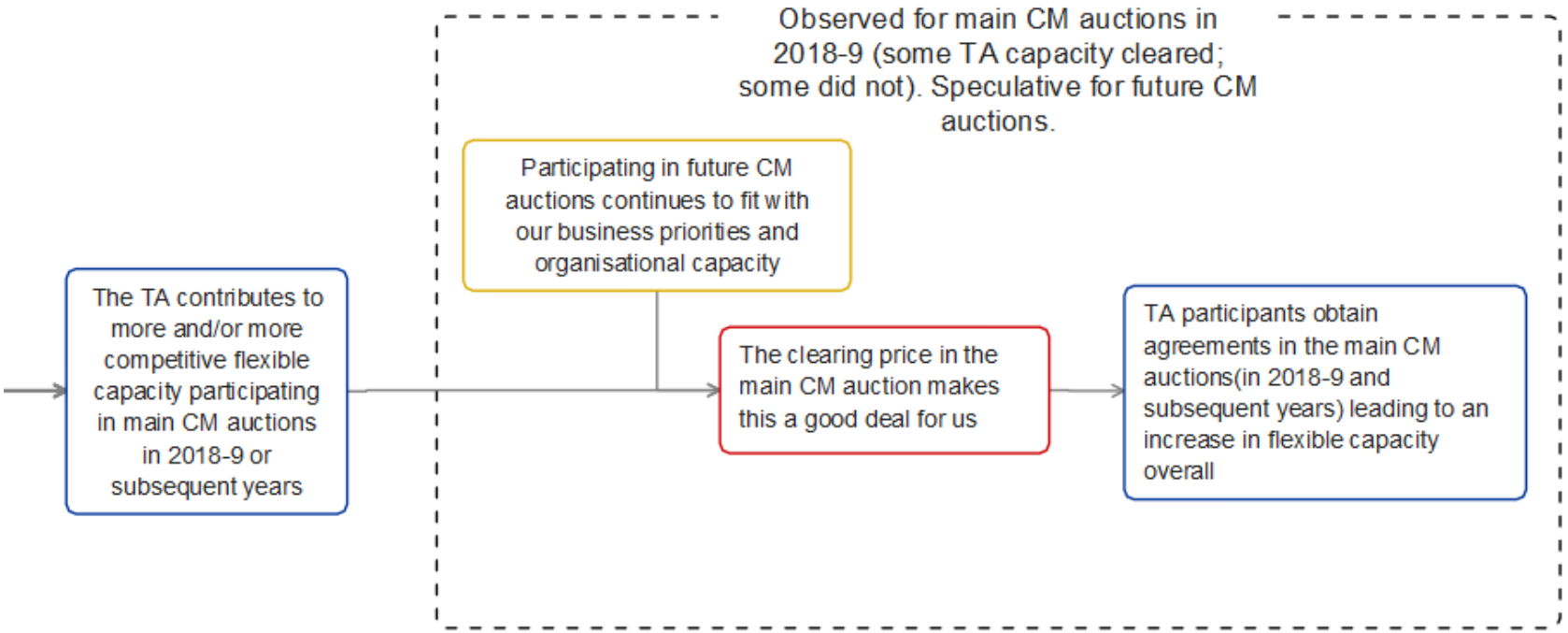


Figure A3.3: New CMO subsequent to outcome 1 in H1 theory (relating to outcomes in main CM auctions)

Theory area B – speculative CMOs for outcomes 3 and 4 in candidate theory

Outcomes 3 and 4 in the candidate theory have been combined into a single outcome (revised outcome 3) and are marked as speculative because they were observed as risks rather than as outcomes during Phase 4 of the evaluation. The evidence is summarised in the main report and in Appendix 5 on process tracing.

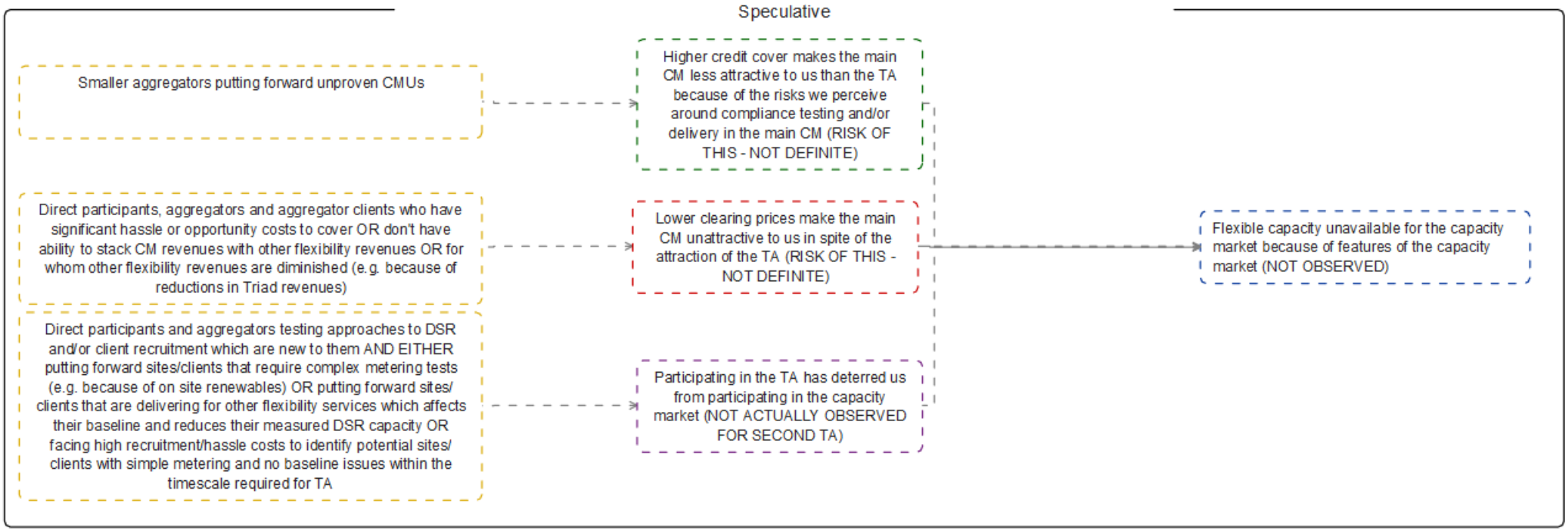


Figure A3.4: Speculative CMOs for new outcome 3 in revised additionality theory for H1

Theory area C – refined CMOs for outcomes 1 and 2 in candidate theory

Minor refinements were made to the CMOs for outcomes 1 and 2, to reflect deeper understanding of contexts and reasonings for these outcomes, as set out in Table A3.1.

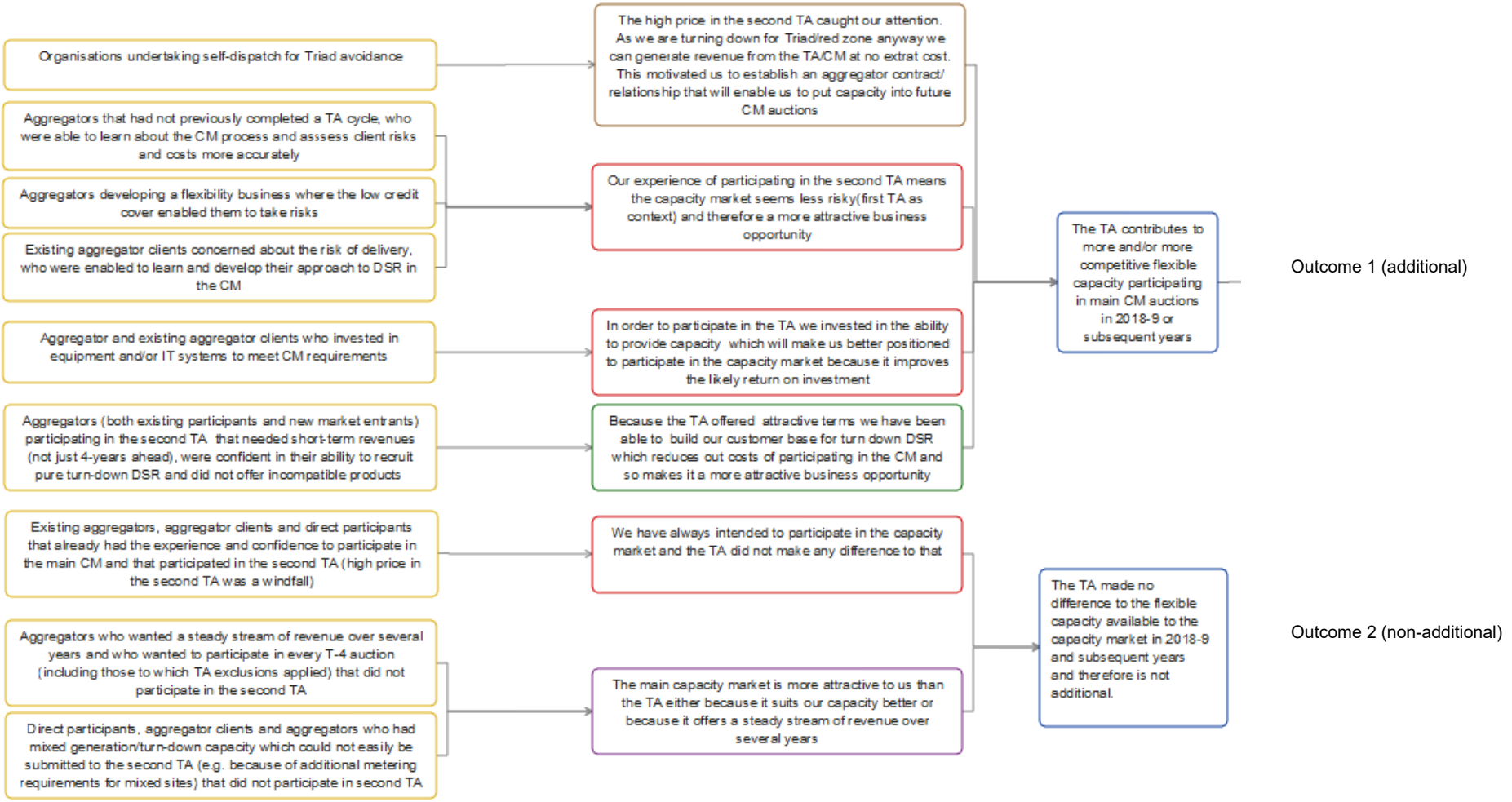


Figure A3.5: Refined CMOs for outcomes 1 and 2 in revised additionality theory for H1

Additionality theory (Hypothesis 2) – TA contributes to wider encouragement of turn-down DSR

The candidate theory for Hypothesis 2 is presented below in Figure A3.6. We tested this candidate theory against evidence from Phases 3 and 4. It was clear from interview data and observed behaviour in the CM that the turn-down DSR brought forward by direct participants was non-additional. These direct participants were already experienced providers of turn-down DSR, active in Triad management, other flexibility services and in the first TA scheme, so the second TA did not generate learning about turn-down DSR for them.

However, there was more evidence of additionality from aggregators and aggregator clients. We tested the candidate theories for the aggregators going forward to delivery in the second TA using process tracing methods, as set out in Appendix 5. The evidence tests incorporated evidence from aggregators and from clients of each aggregator. This testing process revealed considerable support for outcome 1 (additional contribution from the second TA), with some support in one case for outcome 2 (mixed additional/non-additional outcomes). There was no support for outcome 3 (non-additional outcomes for H2).

While we found evidence to support for the CMOs in the candidate theory for H2 for aggregators and some of their clients, we also found that there was considerable overlap between the theory for H1 and H2. Since the second TA was only recruiting turn-down DSR capacity, any capacity that was additional for Hypothesis 1 was (by definition) also additional for Hypothesis 2. In other words, additional turn-down DSR capacity recruited for the second TA and subsequently put forward to the main CM inherently involved an increase in supply of turn-down DSR. So the revised theory for H1 additionality already incorporated most aspects of H2 additionality. In revising H2, we therefore simplified the theory considerably and restricted the theory to outcomes involving increases in turn-down DSR for other flexibility services, outside the main CM.

Table A3.2: Summary of changes to sub-areas of H2 theory

Sub-area of H2 theory	How the theory was revised	Why these revisions were made
New additionality outcome	We introduced a new outcome, with supporting CMOs, for 'additional turn-down capacity for non-CM flexibility services'	As explained in chapter 3 of the main report, we found evidence of aggregators targeting recruitment of turn-down DSR because of the second TA, and then securing (or trying to secure) other flexibility service revenues for these clients, in addition to the main CM. Aggregators and clients less experienced with turn-down DSR also gained learning about turn-down DSR in general. We did not see evidence of additionality for direct participants in the second TA, who were already experienced with turn-down DSR.

Appendix 3: Revised theoretical framework for the second TA at the end of Phase 4

Sub-area of H2 theory	How the theory was revised	Why these revisions were made
New non-additionality outcome	We introduced a new outcome, with supporting CMOs, for 'the second TA making no difference to the provision of turn-down DSR for non-CM flexibility services by some organisations.	As explained in chapter 3 of the main report, we found evidence of aggregators, direct participants and some aggregator clients that were already active in a range of turn-down DSR services outside the CM. Where aggregators and direct participants were adding second TA revenues to existing clients or sites that already offered a fixed volume of turn-down, and already had revenue streams for this capacity from other flexibility services, this was non-additional.

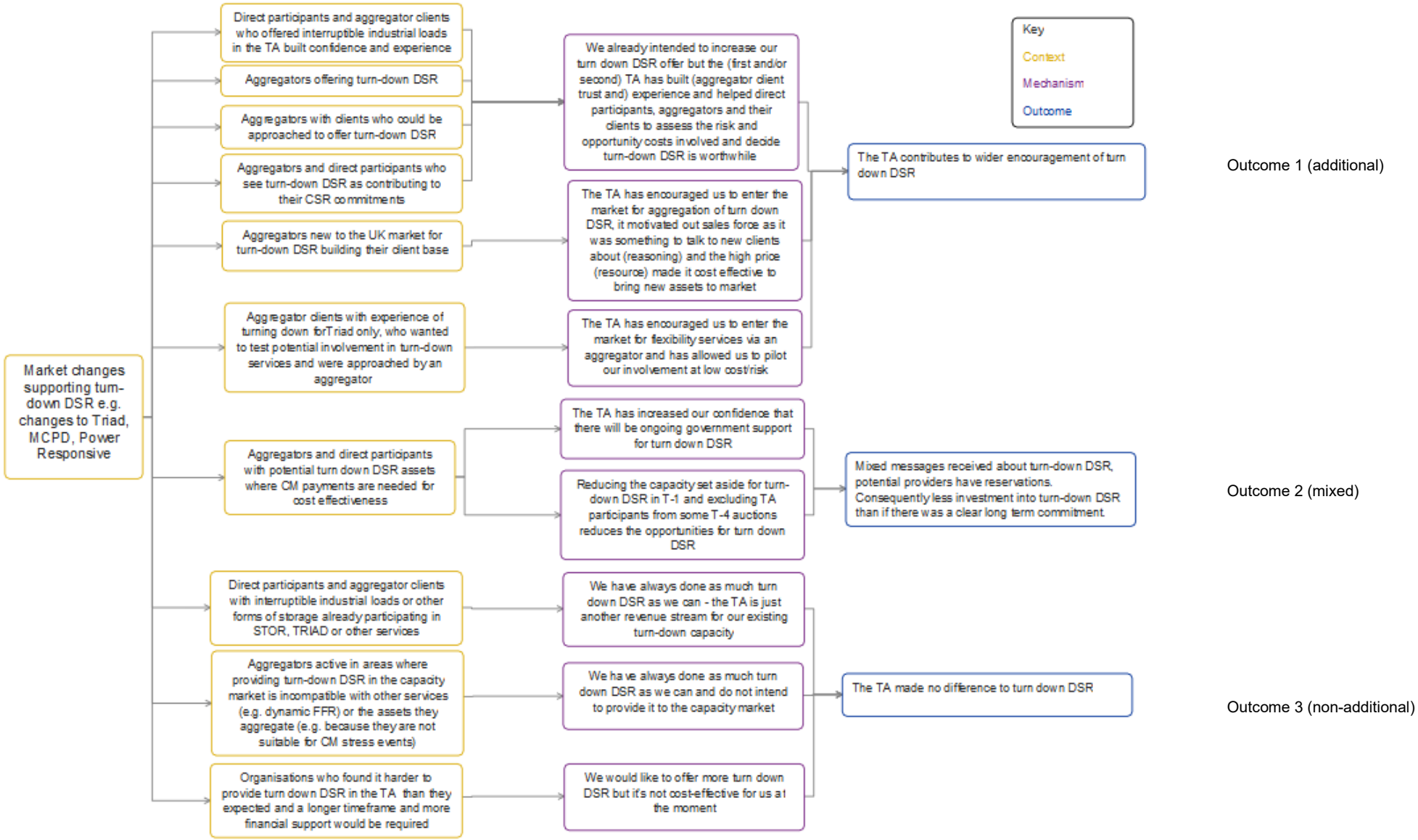


Figure A3.6: Candidate additionality theory (H2) – TA contributes to wider encouragement of turn-down DSR

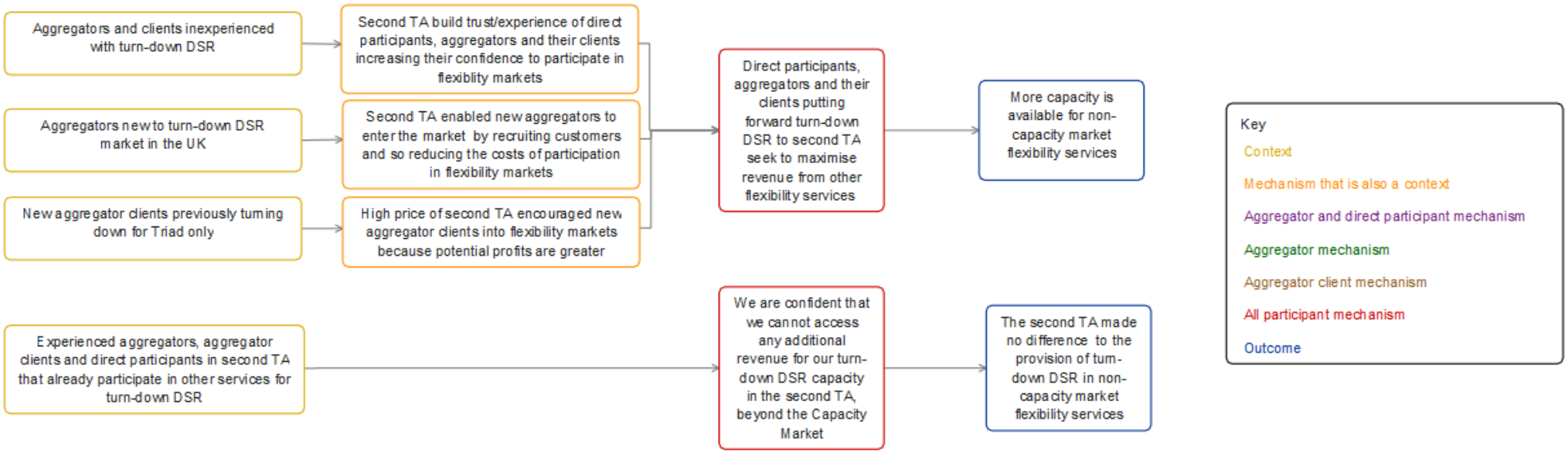


Figure A3.7: Additionality theory (H2) – TA contributes to wider encouragement of turn-down DSR

Appendix 4: Methodology for qualitative research

Introduction

The research involved in-depth telephone interviews with 18 organisations from April to May 2018. This included representatives of nine out of eleven TA participants, one non-participating aggregator that had submitted capacity via a participating aggregator and a sample of eight aggregator clients. The sampling is explained further below.

For the TA participants, the non-participating aggregator and two aggregator clients, these Phase 4 interviews extended the information already gathered through fieldwork conducted in earlier Phases of the evaluation.

Sampling and recruitment

A summary of the sampling approach and the response rates is presented in Table A4.1.

The research involved in-depth telephone interviews with representatives of:

- Nine of the eleven organisations that obtained capacity agreements in the second TA auction (taking account of the sale of one CMU to a third party, which increased the number of participants from ten to eleven).
- One non-participating aggregator that had submitted capacity via sub-contract to a participating aggregator.
- Eight aggregator clients out of an estimated population of 75 clients in the second TA.

Email surveys were used to collect additional quantitative information on the characteristics of their capacity, the costs involved in participating in the second TA and the other flexibility revenues available to these CMUs. Email surveys were completed for all of the interviewees, although some of these were incomplete owing to respondent's lack of time or commercial constraints.

Table A4.1: Summary of sample for Phase 4 qualitative research

Sample group	Description	Population	Sampling strategy	Target no. of interviews	No. of interviews completed	Email survey completed	Total responses (% target)	Response rate (% of sample contacted)
Participating aggregators	Organisations with CMUs with Capacity Agreements post-auction, including two participating aggregators that dropped out post-auction.	8	Census (8)	8	7	7	7 (87.5%)	87.5%
Direct participants	Organisations with CMUs with Capacity Agreements post-auction. Two direct participants cleared the auction, but one sold a CMU to a third direct participant.	3	Census (3)	2	2	2	2 (66.6%)	66.6%
Non-participating aggregator	Organisation that did not obtain a Capacity Agreement in the auction but submitted capacity on behalf of clients via sub-contract to a participating aggregator.	1[10]	Census (1)	1	1	1	1 (100%)	100%
Aggregator clients	Organisations submitting their own capacity via an aggregator. An additional 10 clients from this sample were interviewed during Phases 3.	75[11]	Purposive (21)	10	8	8	8 (80%)	38%

[10] There may be other such contractual arrangements of which we are unaware.
[11] The estimated population of aggregator clients is based on site address data provided by National Grid, with commercially available address databases. This has increased since Phase 3 research owing to additional data becoming available from National Grid.

Sampling for aggregator clients

National Grid provided addresses, and in most cases organisation names, for components within aggregator CMUs that were going forward to delivery in the 2nd TA. Winning Moves (then trading as Databuild) sourced contact details and business characteristics for these addresses and organisations. This enabled them to identify (for almost all sites) the main business activity, organisation name, head office telephone number and website addresses for these components. Additional data became available during Phase 4 for CMUs that were not fully defined during Phase 3, bringing the total sample of aggregator clients to 75. Screening survey information was available from Phase 3 for the initial sample but not the additional sample.

During Phase 4, attempts were made to contact 21 client organisations that had not provided in-depth interviews during Phase 3. These were drawn both from the additional data, from unused contacts from Phase 3 and – in three cases – from client organisations in the second TA that had also been in the first TA and had been interviewed during Phase 2. The sample was purposively chosen to provide representation of the main sectors (and associated asset types) providing DSR, drawing on the DSR characterisation developed during Phase 3.

Interview and email survey approach

For TA participant organisations, and other organisations interviewed in previous phases of the research (which included three aggregator clients and one non-participant aggregator), the interviews were generally undertaken with the key contact at the organisation who was involved in previous phases (i.e. the person primarily responsible for implementation of TA requirements). Respondents were encouraged to involve other individuals in their organisation if needed, to cover the range of topics under discussion. A small number of the interviews involved conference calls with more than one respondent in the organisation. There were a number of follow-up emails to obtain email survey responses.

For aggregator client organisations, we interviewed the key contact identified during the Phase 3 screening survey[12] (where available). This was generally the person responsible for liaising with the aggregator and/or the person with responsibility for energy management. Contacts for unscreened aggregator clients were identified via company switchboards, by asking for the energy manager or operations manager.

The email survey, consisting of a spreadsheet template, had already been sent to all of the participating aggregators and two out of the three direct participants. Where these organisations had not already responded during Phase 3, or where Phase 3 responses were incomplete, additional information was sought.

The email survey was also sent to the third direct participant and to the aggregator clients contacted during Phase 4. Where possible, this was completed by the interviewee. If this was not feasible, it was completed by the interviewer, using information sourced from the interview. Email surveys completed by the interviewer were sent back to the relevant interviewee so that they had the opportunity to check any inaccuracies.

[12] The methodology used for the Phase 3 screening survey is set out in Appendix 3 of the Phase 3 evaluation report: https://www.gov.uk/government/publications/evaluation-of-the-transitional-arrangements-for-demand-side-response-phase-3

Table A4.2 provides an overview of the topics covered with the different respondent types in both the Phase 4 depth interviews and the email survey.

Table A4.2: Overview of topics to be covered in Phase 4 depth interviews and email survey

Topic	Group 1: TA participants	Group 2: TA 'drop-outs' post-auction	Group 3: Aggregator clients
Introduction	✓	✓	✓
Organisational contexts	✓	✓	✓[13]
Reasons for drop-out of any CMUs	✓ (if relevant)	✓	-
Reliability and cost of turning-down	✓	-	✓
Future plans / additionality	✓	✓	✓
Email survey follow-up	✓	-	✓

The topic guides were designed to test additionality theory in detail, explicitly testing theory hypotheses with interviewees. They also gathered insights into other areas of theory, without explicitly testing the theories with interviewees. This approach was chosen to prevent interviews exceeding one hour, particularly in the light of the number of times that TA participants had already been interviewed during the evaluation. The interviews built on information already available from National Grid data, from the CM registers for the TA and other CM auctions and from any earlier interviews with the same organisations in earlier phases of the evaluation. Tailored topic guides and email surveys were prepared for each interviewee, incorporating this prior information and highlighting priority questions to be probed.

An overview of the generic topic guide for TA participants is shown below in Table A4.3. Annex A sets out a sample topic guide, while Annex B sets out a sample email survey.

Table A4.3 Overview of topic guide for TA participants (excluding drop-outs)

Topic	Sub-topics / approach	Timing (mins)
Introductions	Similar anonymity and consents to those used in Phase 3, adjusted to take into account GDPR requirements	3

[13] Only for aggregator clients not previously interviewed

Topic	Sub-topics / approach	Timing (mins)
Organisational contexts	Very brief warm-up question to check whether there have been any significant organisational changes since the last interview. Probe position/knowledge of interviewee if they were new.	5
CMU 'drop-outs'	We did not formally test theory about post-auction drop-out, because metering testing is now covered as a context within participation theory (i.e. participants select sites to avoid metering testing, wherever possible), and DSR testing is covered within reliability theory. However, we needed to elicit any learning about why CMUs dropped out post auction, so based our questions on those used in Phase 2.	5-10
Reliability and cost of turning-down	Questions covered DSR tests, SPDs, and plans/readiness for potential CMNs/stress events. The focus was on how reliable capacity is, the costs and risks involved in turning down, and interrelationships with delivery for other services/Triad. The questions were designed to help refine the 'CMU design for reliability' theory developed during Phase 3, but did not explicitly test this theory with interviewees.	15-20
Future plans and additionality	Focus on future plans, particularly in relation to the additionality evidence tests. Additionality theory was explicitly tested, with questions based on the theory.	20
Email survey follow-up	The end of the interview was used to fill gaps in, get clarifications on, and chase, email survey responses	5
Warm-down	Asked if there any other questions/comments. Highlighted BEIS publication of earlier evaluation findings.	2
Total		**50-60**

Analysis approach

We used slightly different analysis approaches for different topics, depending on whether we had formally tested the programme theory during the interviews:

- We undertook formal testing of additionality theory, as this was explicitly tested with interviewees (by testing their agreement with possible hypotheses during the interviews).

- We undertook in-depth qualitative analysis of other topics (including reasons for drop-out, testing experiences, reliability and cost of turn-down) because these spanned several areas of theory that were not explicitly tested with interviewees.

Formal testing of additionality theory

We used spreadsheets to organise and code the Phase 4 and Phase 3 interview responses against contexts, mechanisms and outcomes (CMOs) in the candidate theoretical framework for additionality with respect to Hypothesis 1 and Hypothesis 2 (see Appendix 2 for details). In these spreadsheets, we also captured additional contexts, mechanisms and outcomes that were supported by the interview evidence but not yet captured by the theory. Where appropriate, this coding made reference to pre-existing evidence from other sources (including interviews in previous phases, email survey data, public statements and CM registers for other CM auctions) where this was relevant for our assessment of additionality. We analysed the extent of support for different CMOs in the framework and for potential refined or new CMOs (see Table A4.4 for an explanation of CMOs). The coding and analysis were undertaken by a lead researcher and reviewed by a second researcher, and were cross-checked against findings from other workstreams, including data from previous phases of the evaluation. An example of how a CMO was refined from the candidate theory, through a tailored CMO reflecting the detailed evidence found, through to a final refined CMO, is given in chapter 1 of the main report.

For the six aggregators that went forward to delivery, for whom additionality issues were potentially complex, the qualitative analysis was supported by further testing of support for different additionality CMOs using process tracing techniques. Details of this process tracing analysis are presented in Appendix 5.

Table A4.4 CMO glossary

Term	Explanation
CMOs	Context-Mechanism-Outcome configurations. These are realist hypotheses about how the policy is expected to work, which are tested during the evaluation. See 'realist evaluation'
Context	The circumstances which affect whether a policy 'works' and for whom. Consideration of 'context' forms an important part of realist approaches to evaluation.
Mechanism	A change in people's reasoning, in response to the resources provided by a policy, which leads to a policy outcome. Identification of causal 'mechanisms', which operate in particular 'contexts', forms an important part of realist approaches to evaluation.

Term	Explanation
Outcome	A change in the state of the world, brought about as a result of a policy or other influences. Realist approaches to evaluation attempt to identify the 'contexts' and 'mechanisms' that lead to a particular 'outcome'.
Realist evaluation	A realist approach[14] to evaluation emphasises the importance of understanding not only whether a policy contributes to outcomes (which may be intended or unintended) but how, for whom and in what circumstances it contributes to these outcomes.

In-depth qualitative analysis on other topics

For other topics explored in Phase 4 interviews (i.e. drop-out theory, testing theory, CMU design for reliability theory and CMN response theory), we used spreadsheets to organise and code findings without explicitly organising these into CMO configurations at this stage. We applied a realist approach to this analysis, analysing what outcomes occurred for whom, in what circumstances and why, but did not attempt to formulate revised CMOs for these elements of theory at this stage. Again, the coding and analysis was undertaken by a lead researcher and reviewed by a second researcher, and was cross-checked against findings from other workstreams, including data from previous phases of the evaluation.

Limitations

The key limitation of the qualitative research findings was that we were not able to interview all TA participants. Nonetheless, the response rates for TA participants were high (66-88%) compared to other studies on non-domestic energy issues. Compared to Phase 3 research, the Phase 4 analysis of aggregator clients was more representative of the range of aggregators participating in the TA. Between Phases 3 and 4 we have interviewed clients[15] in CMUs contracted by all of the participating aggregators that have gone forward for delivery.

[14] Pawson and Tilley (1997), Pawson (2006)
[15] For one participating aggregator, we interviewed a sub-contracting aggregator that managed some of their capacity, rather than a direct client.

Annex A: Example topic guide for in-depth interview with a participating aggregator

Guidance for interviewer	Sub-topics	Prompts and probes	approx. mins
Introduction			
Aim: To introduce the research, ensure the interviewee is aware of and set the context for the proceeding discussion			3
Keep the intro as brief as possible to leave room for the interview proper	• Introduce yourself and CAG Consultants [very brief] • State that the evaluation has been commissioned by BEIS [no need to provide more detail than this] **• Introduce the study:** - Overall objective of the study is to evaluate the effectiveness of second Transitional Arrangements - Main purpose of the interview is to explore experiences of the second TA post-auction, as well as exploring TA participants' future plans regarding turn-down DSR - Findings will inform government policy development about DSR going forward **• Talk through key points about the interview:** - Length of interview [estimated 60 minutes] - Note that we would like to record the		

	interview and explain that the recording, transcription and notes will not be shared outside of the research consortium (BEIS will not have access to them) - Check that they consent to you recording the interview [if they don't, still go ahead with interview, just take notes] - Note that, as per the privacy notice in the briefing note, any views you express will be pseudonymised and our report will only contain completely anonymous data. **• Check whether they have their email survey response to hand – will be referred to during the interview if incomplete** • Ask if interviewee has any questions before you start		
Organisational contexts			
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. *Keep this section brief as specific contexts will be explored throughout the interview*			2
Only if interviewing a different person from previous interviews	Establish the interviewee's role in the organisation		

	Establish the interviewee's experience in relation to the Capacity Market		
Interviewer to refer back to the interviewee's responses from Phase 1, 2 and 3 interviews on organisational contexts	Establish if the organisation's overall approach to providing capacity in the flexibility market has evolved since they were last interviewed for this evaluation If it has changed, establish <u>how</u> it has changed, to what extent it has impacted on their overall approach to the Capacity Market (if any), and why	***Desirable but not essential*** *Probe for:* - any changes in the type of capacity they provide (e.g. between turn-down and generation capacity) - any changes in their client base or client offer evolved (for example the types of/number of clients you work with) - any changes in the organisation's business case for (and perceptions of risks associated with) providing flexible capacity changed	
Reasons for drop-out of any CMUs			
Aim: to understand the reasons why these CMUs dropped out and the contexts/factors that led to this			5
Before the interview, please: **- check how many CMUs secured Capacity Agreements and which were subsequently terminated** **- insert CMU details and add bespoke questions if necessary**			
	We understand that [x] of your CMUs in the TA was/were given (a) termination notice(s). Can you please summarise the main reason, or reasons, why this/these termination notice(s) was/were given?	Probe to understand: *- were the reasons strategic, technical, financial, TA rules-related, etc?* *- any other particular contexts that were relevant in influencing the CMU termination(s)*	

	Why were some of your CMUs terminated while others continued in the second TA? What was different about the CMUs that continued?		
Reliability and cost of turning-down			
Aim: to explore how reliable the capacity in the CMUs is, the costs and risks involved in turning down for the capacity within the CMUs, and the interrelationships with delivery for other services/Triad.			
Before the interview, please: **- review the CMU reliability theory** **- insert details about design of CMUs (e.g. number of sites, amount of 'slack' as evidenced by DSR tests or previous phases, etc)**			
Turn-down processes			5
	Can you please describe the processes involved in providing turn-down DSR	*Explore extent to which their client loads involved load shifting or load shedding* *Identify which processes are either shifted or turned down* *What does this shifting or shedding involve?*	
	What would have been the 'best case' and 'worst case' outcomes of a particular turn-down DSR request, in terms of the impact on your clients' business processes (and timing vis a vis turn-down/stoppages for other purposes)? What is your understanding of the periods of time over which different clients/CMUs can sustain turn-down?	*Explore the costs and risks associated with a request for turn-down coming at a convenient or inconvenient time, or lasting for 30 minutes, 1 hour or more, for different types of clients.* *Explore their perception of the likelihood of best/worst case occurring, for different types of clients.* *Explore their understanding of how different clients rationalised costs/risks in deciding to participate in the TA*	

	What was your understanding of your clients' overall rationale for participating in the second TA, in terms of the potential risks, costs and benefits to their businesses?	*(and whether they agreed with the client any limits about the circumstances in which they would or would not turn-down).*	
DSR test(s) ***- insert insights already shared from Phase 3 interview***			5
	For each of your CMUs, can you please describe how straightforward it was to undertake, and pass, the DSR test	*Ask respondent to explain the reason(s) for their answer*	
	For each of your CMUs, what factors influenced how straightforward the DSR tests were	Prompt: *- What was it about your CMU(s) that made the DSR test process [straightforward/not straightforward] compared with other CMUs in the TA that found the tests [not straightforward/ straightforward]* *- If you'd had a different CMU design (e.g. fewer/more sites, less/more slack), what difference would that have made?* *Probes:* *- Explore the extent to which the design of their CMU(s) influenced how straightforward the tests were [see notes in yellow row above]* *- Explore other factors that may have influenced how*	

		straightforward the test process was e.g.: *- organisational capacity and capability* *- the strategy they chose for DSR testing* *- technical issues with capacity* *- TA DSR testing rules* *- Interrelationship with other flexibility services e.g. Triad, balancing services* *- other factors*	
	Can you provide a rough estimate of the costs to your organisation associated with achieving compliance and actually doing the DSR testing?	What did it cost (for those that did) to achieve compliance with the DSR test requirements – for each CMU?: - Their understanding of client costs associated with turn-down for the DSR test - 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.	
	Were these costs above, below or the same as you expected them to be when you took part in the auction?	If they were different, by how much? And why?	
Satisfactory Performance Days			5

	For each of your CMUs, please describe your approach to demonstrating your Satisfactory Performance Days?		
	For each of your CMUs, can you please describe how straightforward it was to demonstrate your Satisfactory Performance Days?	*Ask participants to explain the reasons why demonstrating their Satisfactory Performance Days was straightforward or not*	
	For each of your CMUs, what factors influenced how straightforward it was to demonstrate their Satisfactory Performance Days?	Prompt: *- What was it about your CMU(s) that made demonstrating the Satisfactory Performance Days [straightforward/not straightforward] compared with other CMUs in the TA that found the tests [not straightforward/ straightforward]* *- If you'd had a different CMU design (e.g. fewer/more sites, less/more slack), what difference would that have made?* *Probes:* *- Explore the extent to which the design of their CMU(s) influenced how straightforward the SPDs were [see notes in yellow row above]* *- Explore other factors that may have influenced how straightforward the SPD process was e.g.:* *- organisational capacity and capability* *- the strategy they chose for DSR testing* *- technical issues with capacity* *- TA SPDs rules* *- Interrelationship with other flexibility services e.g. Triad,*	

		balancing services *- other factors*	
	Can you provide a rough estimate of the costs to your organisation associated with achieving compliance and actually doing the Satisfactory Performance Days?	What did it cost (for those that did) to achieve compliance with the SPD requirements – for each CMU?: - Their understanding of client costs associated with turn-down for the SPDs - Their own organisation's staff time required for SPDs [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.	
	Were these costs above, below or the same as you expected them to be when you took part in the auction?	If they were different, by how much? And why?	
Capacity Market Notices			5
	What system do you use to ask your clients to turn down in response to a Capacity Market Notice (and possible stress event)?	*Probe for: computerised or manual notification; degree of client control over turn-down response (e.g. is this automated? Can client opt out, and if so, how? Have you agreed/discussed a maximum turn-down period, and if so, what would this be?)*	
	What operational plans do you have in place to ensure adequate capacity will be available during a system stress event?	Probe for best/worst cases in terms of impact on business processes, as discussed above.	

	If a system stress event occurred, how confident would you be that your CMU(s) could reliably provide adequate capacity?	*Ask respondent to explain the reason(s) for their answer- e.g. what would their ability to provide capacity depend on;* *What would be the potential costs to their business of providing capacity in different circumstances (if not already covered above)*	
	How confident would be that your CMU(s) could reliably provide adequate capacity if there were: a. multiple system stress events over a short period of time b. a system stress event lasted over, say, four hours	*Ask respondent to explain the reason(s) for their answer- e.g. what would their ability to provide capacity depend on;* *Probe for differences in the lengths of time that different types of assets can typically turn-down for, given the 4 hours advance warning provided by a CMN (e.g. HVAC, cold storage, water pumping, process heating, motors & drives, other processes)*	
	Where known, what would be the potential costs to different client's businesses of providing capacity in different circumstances (best case/worst case)?	*Any insights into the costs for different types of clients of any lost production or of shifting processes to other times (e.g. overtime costs for staff; fuel costs arising from less efficient running or from having to restart production processes)*	
	What is it about the design of your CMU(s) that would help or hinder your CMU(s) to provide adequate capacity in the case of a system stress event?		
Future plans/ additionality			

Aim: to provide information to inform the additionality 'hypotheses'			
Before the interview, please: **- review the additionality theory/theories and hypotheses** **- add details of this organisation's DSR participation in the main CM.**			
Future plans with respect to DSR in the Capacity Market (covering both turn-down and back-up generation)			5
Overall strategy for DSR in CM	*(Recap from CM info)* We're aware that you submitted [..] MW of DSR for the recent [T-1/T-4] Capacity Market auctions and that you have [..] MW of DSR contracted for [T-1/T-4] Capacity Market delivery in future years. Can you explain your strategy for putting DSR into the CM? Are you planning to participate with DSR in future Capacity Market auctions as well? If not, why not?	*Probe for how much importance they attach to CM participation.* *Probe for rationale about participating with DSR in the one-year ahead vs four-year ahead auctions?*	

Factors influencing this	What are the main factors influencing your future plans in relation to DSR in the Capacity Market?	*Probe for external factors such as the Medium Combustion Plan Directive, CSR, changes to Triad charges, changes to National Grid services, Power Responsive campaign, the CM itself, TA, technology changes (e.g. battery storage, controls), aggregator access to balancing mechanism etc.* *Also probe for specific factors associated with the nature of this organisation (skills/capability/technical set-up).*	
Main CM vs TA	*(recap insights on attitudes to main CM auctions, from previous interviews)* Looking forward, does your approach to future CM auctions differ from your approach to the TA, and if so, how? Do the differences between the rules for the main Capacity Market and the first/second TA influence your plans, and if so, how?	*Probe for: different price expectations; higher credit cover; minimum CMU size; termination fees; ability to put forward back-up generation as well as turn-down DSR; baseline methodology; any proposed changes to CM rules*	
Additionality of TA with respect to DSR in CM			5

TA influence/additionality	Has your involvement in the first or second TA influenced your plans in relation to DSR in the future CM Ask participant to explain how and/or why the second TA has (or has not) influenced their plans in relation to DSR	*Probe for positive influences/additionality: systems put in place, client-base developed or increased, relationships with clients or third-parties, new sites/production processes brought in, growth in management confidence, reduced perception of risks etc.* *Probe for role of TA in helping them to recruit new clients or bringing in new sites from existing clients.* *Probe for any upfront costs associated with participation in the first and/or second TA that they won't need to cover if participating in future CM auctions.* *Also probe for negative influences of TA experience: unanticipated costs, testing processes, termination fees etc.*	

Theory testing - outcome	So would you say that overall (SELECT OUTCOME THAT SEEMS MOST APPROPRIATE OR SUMMARISE ANOTHER OUTCOME.)	*Select from:* *- the TA has contributed to you putting forward more (or more competitive) DSR to the CM in 2018/19 or future years (in at least some respects..)* *- the TA made no difference to the DSR you put forward to the CM in 2018/19 and future years* *- you do not intend to put DSR into the CM in future, because of the conditions in the main CM* *- you do not intend to put DSR into the CM in future, because of your experiences in the TA* *If none of these apply, explore what outcome does apply*	

Theory testing - reasoning	And would it be fair to say that (SELECT MECHANISMS THAT SEEM MOST APPROPRIATE OR SUMMARISE ANOTHER MECHANISM...)	*Choose most appropriate one (or more from):* ***Where the TA has contributed to more (or more competitive) DSR coming forward to the CM in 2018/19 and subsequent years:*** *- your experience of participating in the TA means the CM seems less risky* *- in order to participate in the TA you invested in capacity (or the ability to provide capacity) which has made you better positioned to participate in the CM* *- in order to participate in the TA you have built a customer base so now you want to continue with the CM* *- you have recruited new turn-down clients through the TA and can cost-effectively include them alongside back-up in CMUs for the CM* *- other reasoning (please summarise from earlier responses)* ***Where the TA has made no difference to DSR coming forward to the CM in 2018/19 and future years:*** *- you always intended to participate in the CM and the TA did not help you grow your flexibility business in any way.* *- you would have participated with DSR in the main CM anyway, but for other reasons, not because of the TA* *- other reasoning (please summarise from earlier responses)* ***Where the organisation does not intend to put DSR into the CM in future:*** *- higher credit cover makes the main CM less attractive to*	

		you than the TA because of the risks we perceive around compliance testing and/or delivery in the TA *- lower clearing prices make the main CM unattractive to you in spite of the attraction of the TA* *- participating in the TA has deterred you from participating in the CM* *- other reasoning (please summarise from earlier responses)* *If several of these mechanisms apply, please ask which is the most important.* *If none of these apply, explore what mechanism does apply*	

Theory testing - contexts	Were there any factors (internal to your organisation, or external market factors), other than those we've already discussed, that affected how the TA influenced your plans for the DSR in the main CM?		
Future plans for turn-down DSR			6
Overall strategy re turn-down DSR	More generally, how important is aggregation of **turn-down** DSR to your future plans, and why? Do you have a strategy of increasing your turn-down portfolio, and if so how and in which services? If you don't intend to increase your turn-down portfolio in future, why not?	*Probe for opportunistic vs strategic approach to turn-down* *(services may include Triad management, red zone management, FFR, STOR, other National Grid services, Capacity Market…)*	
Factors influencing this	What are the main factors influencing your future plans in relation to turn-down DSR?	*Probe for external factors such as the Medium Combustion Plan Directive, CSR, changes to Triad charges, changes to National Grid services, Power Responsive campaign, the CM, TA, technology changes (e.g. battery storage, controls), aggregator access to balancing mechanism etc.* *Also probe for specific factors associated with the nature of this organisation (skills/capability/technical set-up).*	

Viability of turn-down work in main CM?	Do you see turn-down DSR in the Capacity Market as a strategic opportunity that you want to pursue? If so, how/why? If not, why not? If not, what changes would be needed to the CM to make your organisation offer turn-down within the Capacity Market?	*Probe: do they think that turn-down DSR will be able to compete effectively in the Capacity Market in future?* *Probe for: changes to CM and/or other revenue streams (e.g. Triad, Firm Frequency Response, STOR); changes to the number/length of turn-down requests; changes to notice periods for turn-down; changes to CM rules or risk/reward ratio.*	
Viability of turn-down more generally	If you are not currently planning to offer increased turn-down DSR services in future, except on an opportunistic basis, what changes would be needed flexibility services to make your organisation offer more turn-down?	*Probe for: changes to CM and/or other revenue streams (e.g. Triad, Firm Frequency Response, STOR)*	
Additionality of TA for turn-down DSR (within and outside the CM)			5
	Has your involvement in the second TA influenced your organisation's attitudes to, or capacity to provide turn-down DSR (within or outside the CM), and if so how/why?	*Probe for positive influences: systems put in place, client base developed or increased, relationships with clients or third-party organisations, new sites/production processes brought in, growth in management confidence, reduced perception of risks etc.* *Probe for length of TA contracts - did the TA enable them to set up new client contracts for turn-down that will continue beyond the end of the TA?*	

		Also probe for negative influences: unanticipated costs, testing processes, baseline methodology etc.	
	Do you think the TA has influenced the proportion of turn-down DSR in your client portfolio going forward (within or outside the CM), and if so by how much (as % of capacity)?	*Probe for approx proportions of turn-down DSR and back-up DSR - in CM and other services.* *Probe for reasons for any increase/decrease in capacity of turn-down in main CM compared to first/second TA.*	
Theory testing - outcome	So would you say that overall (SELECT OUTCOME THAT SEEMS MOST APPROPRIATE OR SUMMARISE ANOTHER OUTCOME.)	*Select from:* *- the TA has encouraged you to pursue more turn-down DSR* *- the TA has made no difference to your provision of turn-down DSR* *- the TA has sent mixed messages, encouraging you to pursue turn-down DSR in some respects but not in others (if so, please explain these limitations)* *If none of these apply, explore what outcome does apply*	

Theory testing - reasoning	And would it be fair to say that (SELECT MECHANISMS THAT SEEM MOST APPROPRIATE OR SUMMARISE ANOTHER MECHANISM...)	*Choose most appropriate one (or more) from:* ***Where the TA has encouraged you to pursue more turn-down DSR:*** *- you already intended to increase your turn-down DSR offer but the (first and/or second) TA has built client trust and your/their experience, and has helped you and your clients assess the risk and opportunity costs involved* *- the TA encouraged you to enter the UK market for aggregation of turn-down DSR by giving you something to talk to new clients about, with the high price making it effective to bring new assets to market* *- other reasoning (please summarise from earlier responses)* ***Where the TA has made no difference to your provision of turn-down DSR:*** *- you have always done as much turn-down DSR as you can, and the TA was just another revenue stream for your existing turn-down capacity* *- you have always done as much turn-down DSR as you can but do not intend to provide it in the CM* *- you would like to offer more turn-down DSR but it's not appropriate for you at the moment (e.g. because of your situation, or because it needs more financial support)* *- other reasoning (please summarise from earlier responses)* ***Where the TA has sent mixed messages, check whether any of the reasoning above applies, and also***	

		consider whether: *- the TA increased your confidence that there would be ongoing government support for turn-down DSR* *- but reducing the capacity set aside for turn-down DSR in T-1 and excluding TA participants from some T-4 auctions reduced the opportunities for turn-down DSR* *- other reasoning (please summarise from earlier responses)* *If several of these mechanisms apply, please ask which is the most important.* *If none of these apply, explore what mechanism does apply*	
Theory testing - contexts	Were there any factors (internal to your organisation, or external market factors), other than those we've already discussed, that affected how the TA influenced your plans for turn-down DSR generally?		
Overview of additionality			2
	Overall, what do you think would be different about your ability to offer turn-down DSR and/or back-up generation in the UK in future if you had not participated in the first or second TA as an aggregator?		

Email survey follow-up			
Aim: to fill gaps in the email survey response and/or to chase a response if none has been provided yet			5
	Chase / clarify responses to email survey. If survey not responded to, either run through questions/answers at end of interview (if time) or ask them to complete asap	Ensure that we have details of any upfront costs of participation in the second TA (e.g. metering costs; control equipment costs; staff costs). Ensure we know which sites/clients were new to flexibility for the second TA. Ensure that we have a good understanding of the different types of turn-down DSR provided by their components, from the email survey - and how prevalent each of these types are. If not addressed earlier in interview, fill any gaps in understanding about the suitability of these different types of turn-down DSR for the TA/CM and other flexibility services. If there appear to be any inconsistencies or questions vis a vis their earlier answers, please probe and clarify.	
Interview close and thank you			
			2

	Would interviewee like to say anything else about the second TA		
	Thank the participant for their time. Reiterate that their anonymity will be protected in our reporting. Tell them they are welcome to contact members of the study team to ask questions at a later date if they wish		
	Tell them that the report from this phase of the evaluation should be published later on in the year. The report from the previous phase of the research should also be published within the next couple of months. All of the evaluation reports can be found on the gov.uk website		
	END INTERVIEW		
		Total indicative length	

Annex B: Example email survey used in Phase 4 research for an aggregator client

Thanks for making time to contribute to the evaluation of the Transitional Arrangements, which CAG Consultants and Databuild are undertaking on behalf of BEIS. We look forward to speaking with you soon about with the second TA scheme, which forms part of the Capacity Market. This is a brief email survey that aims to collect for some specific information on:

- the costs associated with participating in the second TA;
- the nature of the turn-down DSR in your sites.

Completing the email survey in advance will save time in the interview. If you're not sure how to respond to any of the questions, you can discuss your answers during the interview.

This survey is for organisations participating in the second TA through an aggregator. The results of the research will be used by BEIS to inform decisions regarding the future participation of demand-side response in the Capacity Market and to model the costs and benefits of turn-down DSR.

Any findings from the survey used in the research will be anonymised unless otherwise agreed with you; 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.

For your information, we will also be conducting direct research with your aggregator, as part of our research with all TA participants

We would be grateful if you could complete this email survey (PARTS A and B, one for each site/CMU) and email it back to your interviewer before the interview, or by 11th May at the latest.

PART A: STAFF TIME

We are interested to know roughly how much staff time you have put into the second TA, across the time categories below. This will help us to assess the costs of TA participation and will inform the research team's modelling of DSR costs.

Organisation name:	
Aggregator name:	

Category	Item	Approx number of staff days (FTE)	Explanatory notes
1. Initial time input by aggregator client to secure participation of CMUs and its component sites for the second TA (days)	a. staff time for setting up relationship with aggregator (if new contract)		*Please add notes to explain is included in these estimates ...*
	b. staff time for internal marketing & approvals of sites for second TA		
	c. staff time for pre-qualification & auction process (if relevant)		
	d. staff time associated with metering & testing for TA sites		
	e. other staff time for 2nd TA (please specify)		
2. Estimated annual time input by aggregator client to manage CM participation on an ongoing basis, for CMUs proven through the second TA **(excluding respond to stress events)**	a. liaison with/management of aggregator relationship (specific to 2nd TA if possible)		*Please indicate if this time also covers liaison in relation to other flexibility services:*
	b. internal liason with 2nd TA sites		
	c. adjustments to capacity offered		
	d. further testing/other compliance costs		
	e. any other ongoing time inputs (please specify)		

PART B: SITE DETAILS (IF YOUR SITES ARE IN MORE THAN 1 CMU - PLEASE COMPLETE A SEPARATE SHEET FOR EACH CMU)

This sheet asks for further information on the costs and revenues associated with different types of turn-down DSR, to inform the evaluation's modelling work. Specifically, it asks for:

1. Estimates of any upfront capital costs required to install metering or other equipment on your sites for the second TA (if any).
2. Estimates of revenue that these sites may be obtaining from other sources
3. Characterisation of the type of DSR provided by each site (using drop-down menus to facilitate this).

Our research team will use this information to develop generic cost/revenue models for different types of turn-down DSR, as well as estimates of the prevalence of these different types of turn-down in the second TA. These models and estimates will be shared with BEIS on a non-identifiable, anonymous basis.

Please enter CMU- ID:

Information on costs and revenues:

Category	Item	Approx value (£)	Explanatory notes	Please indicate whether these costs were borne by your organisation, or by your aggregator
1. Initial capital expenditure to enable participation in second TA, if any (£)	a. Capital expenture on controls		*Please add notes to explain what is included in these estimates …*	
	b. Installation of controls			
	c. Capital expenditure on metering equipment			
	d. Installation of metering equipment			
	e. Other capital expenditure (e.g. associated with metering or other aspects of participation) - please specify			
2. Other sources of flexibility revenue for this capacity	**Please list any flexibility services that you are aware that this capacity participates in during 2017/18.**	**Please enter ballpark estimate (£'k) of annual revenues if possible (£0k if none or n/k if unknown)**	***Add explanatory notes, if needed***	**Please indicate the approximate proportion of the TA capacity that participates in these services (%)**
	STOR/STOR Runway			
	Firm Frequency Response (static)			
	Firm Frequency Response (dynamic)			
	FCDM			
	Enhanced Frequency Response			
	Demand Turn-Up			
	Triad management			
	Red zone management			
	Wholesale electricity market			
	Balancing mechanism			
	Other flexibility services (please specify in notes column)			

O`

3 Please provide and or confirm the details of the turn-down DSR capacity for your sites.

Characterisation of turn-down

Type of component or site (please enter description)	Business activities (please select from list)	Notes on business activities (add if needed)	Type of equipment used to provide turn-down (please select from list)	Notes on type of equipment used (further details if needed)

(…continuation of characterisation table)

Does the TA capacity relate to processes that generally run 24/7? (yes/no/don't know)	How is delivery of your TA obligations likely to impact on your normal business? (please select)	Please add any comment to explain this	Please enter approximate kW of derated capacity offered for turn-down.	Approximately how much of this turn-down capacity was new for the 2nd TA (i.e. not previously covered by your flexibility contract(s) except for internal Triad/red zone management)?

Appendix 5: Process tracing analysis and overall synthesis

Introduction

This appendix explains how we have applied process tracing during Phase 4 of the evaluation and then explains our approach to overall synthesis of evidence, including use of 'realist' contribution analysis.

Process tracing

Process tracing involves the 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.

During Phase 4 of the evaluation, we used process tracing to test the strength of evidence for competing additionality hypotheses (treating each Context-Mechanism-Outcome configuration in the candidate theoretical framework as a separate hypothesis). We applied process tracing on a case-by-case basis, which was consistent with our realist approach to analysis and synthesis.

Process tracing is described further in the CECAN working note 2.1 by Barbara Befani on 'Testing Contribution Claims with Bayesian Updating' (December 2016).[16]

Process tracing with Bayesian updating (Phase 2)

In full process tracing with Bayesian updating, as undertaken for Phase 2 of the TA evaluation, the steps can be summarised as follows:

- Define the competing hypotheses to be tested, and the case or cases to be examined.
- Assess the 'prior' probability of each hypothesis being true (for all cases or for a particular case).
- Specify a set of independent evidence tests for each of the competing hypotheses, drawing on a range of evidence sources. These are 'clues' to be looked for in the research evidence, to help distinguish between the competing hypotheses.
- Assess the probability that a given 'clue' will be observed if the relevant hypothesis is true.
- Similarly, assess the probability that this clue will be observed if this hypothesis is false.
- Undertake research, looking for the evidence 'clues' for each case.
- Assess whether each potential 'clue' has or has not been observed for a given case (i.e. whether each evidence test has or has not been passed for this case).

[16] http://www.cecan.ac.uk/resources; and 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.

- Apply Bayesian updating[17] to update the probability that each hypothesis is true, on a case-by-case basis, taking into account whether particular clues have been observed for that case.
- The 'posterior' probability of each hypothesis being true, for a given case, provides a better assessment of the relative merits of the competing hypotheses for that case.

Our use of process tracing with Bayesian updating in Phase 2 is explained further in Appendix 5 to the Phase 2 evaluation report.[18] We called the Phase 2 method 'contribution tracing' because we were using process tracing to assess the contribution of the TA to its objectives. While the Phase 2 method used Bayesian updating, the method was only used to test three 'additional' hypotheses and three competing 'non-additional' hypotheses that reflected high-level outcomes for the first TA scheme. These six high-level hypotheses did not fully reflect the detailed CMO hypotheses in the candidate additionality theory for Phase 2.

Non-quantified approach to process tracing (Phase 4)

In Phase 4, we sought to use process tracing in a way that directly tested the CMO hypotheses in the Phase 4 candidate additionality theory. We developed and applied evidence tests for each CMO hypothesis in the candidate additionality theory, based on process tracing concepts. However, we did not quantify probabilities and did not use Bayesian updating during the Phase 4 work. There were two reasons for this:

- There were 17 CMO hypotheses in the Phase 4 candidate additionality theory. With (say) four independent tests per CMO there would have been over 50 evidence tests. We would have needed to estimate (or define ranges for) over 100 probabilities and we thought this was unmanageable.
- It was problematic to define independent evidence tests for different aspects or elements of a CMO, because they were causally related.

Our approach in Phase 4 was therefore to develop a set of evidence tests for each CMO in the Phase 4 candidate additionality theory, covering both additional and non-additionality outcomes. We categorised the tests using process tracing concepts, according to the rough likelihood of that piece of evidence being observed if the CMO was or was not true for a particular case (i.e. TA participant). We used this categorisation to assess the weight that should be attached to a particular piece of evidence when considering whether a given case (i.e. organisation) exhibits a particular CMO.

We did not apply the evidence tests to direct participant cases, because it was clear from the evidence that one of the non-additional CMOs (for outcome 2) applied to these cases. This allowed us to focus on applying the evidence tests to six cases: namely, the six aggregators that went forward to delivery. Evidence relating to the clients of these aggregators was incorporated into the evidence tests.

If none of the CMOs had fitted a given case well, we would have been prepared to refine or revise the theory and associated evidence tests, until we were confident that our refined theory was well supported by the evidence. However, in practice we found that each of the cases tested supported one or more of the candidate CMOs.

[17] 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

[18] https://www.gov.uk/government/publications/evaluation-of-the-transitional-arrangements-for-demand-side-response-phase-2

Application of process tracing during Phase 4

Defining the competing hypotheses

As explained in Appendix 2, the candidate additionality theory had two parts, reflecting the two objectives of the second TA. Each objective was used to frame a high-level additionality hypothesis as follows:

- H1: The (second) TA leads to more and/or more competitive flexible capacity for the Capacity Market in 2018-19 and subsequent years. (Note: this can be back-up or turn-down DSR)
- H2: The (second) TA leads to wider encouragement of turn-down DSR (Note: this can be within or outside the Capacity Market).

For each of these hypotheses, the candidate additionality theory presents a number of contexts and mechanisms which were expected to lead to different outcomes under these hypotheses - some of them additional, some non-additional. As explained in Appendix 2, this candidate theory had been developed during earlier phases of the evaluation. For each high-level project hypothesis, the theory defined detailed 'Context-Mechanism-Outcome' (CMO) configurations, as shown in Figures A5.1 and A5.2 below, explaining how the objectives of the second TA might or might not be achieved.

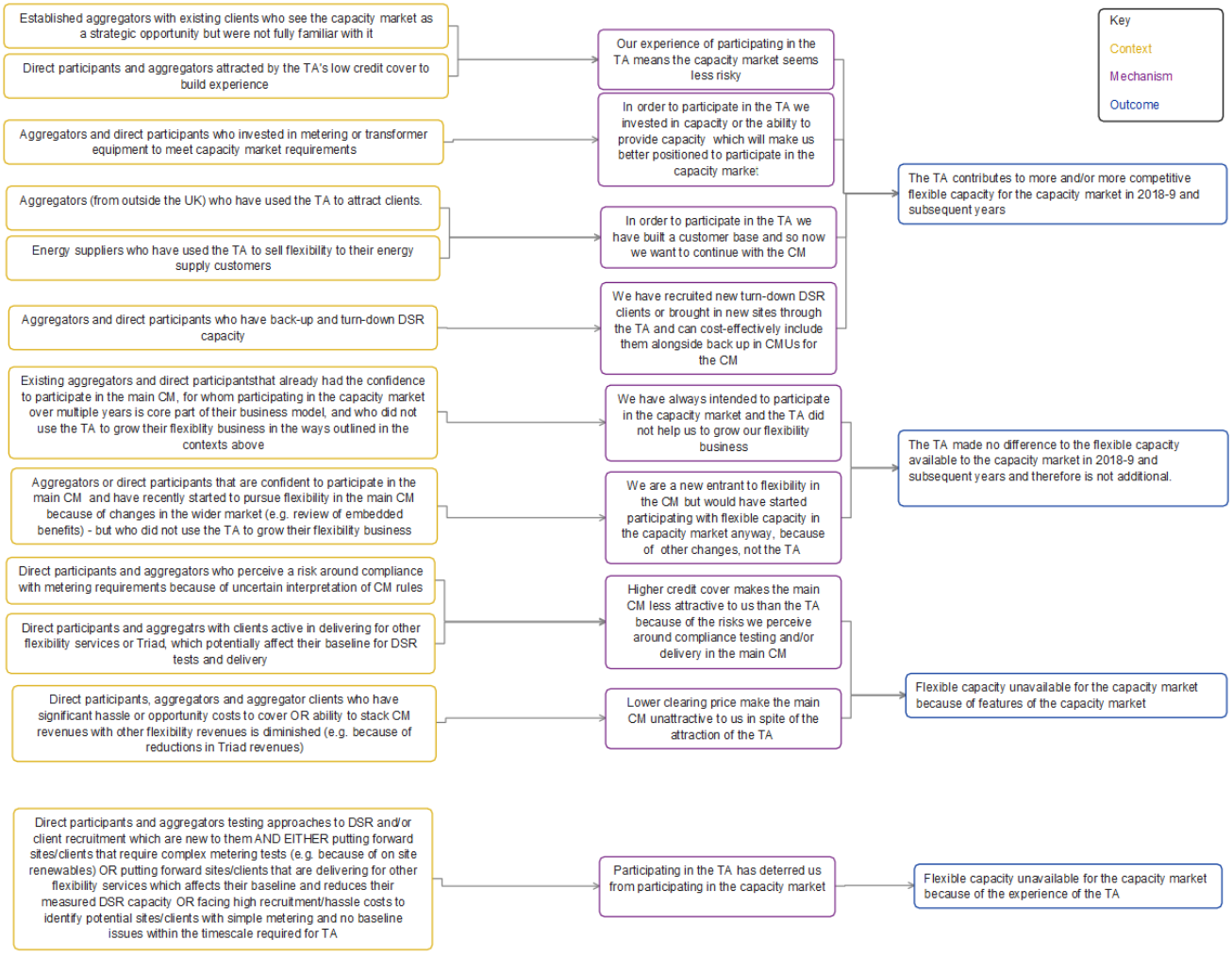


Figure A5.1: CMOs for hypothesis H1: the (second) TA contributes to more and/or more competitive flexible capacity for the Capacity market in 2018-19 and subsequent years

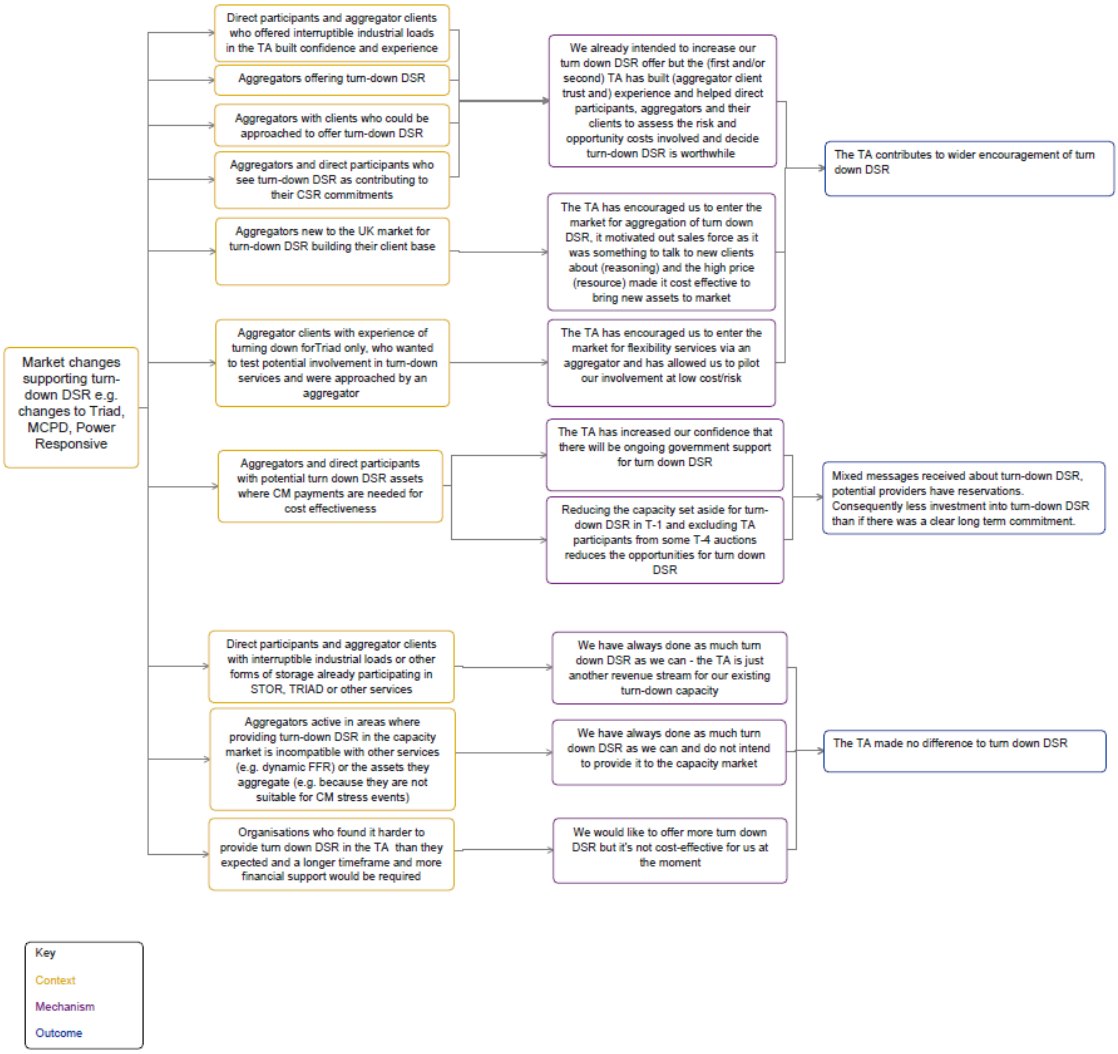


Figure A5.2: CMOs for hypothesis H2: the (second) TA leads to wider encouragement of turn-down DSR

Developing the evidence tests

We developed a set of evidence tests that specified the evidence that we would expect or like to see if each of the CMO configurations in the H1 and H2 theory was true.

We then used process tracing concepts, as set out in Barbara Befani's paper for CECAN on 'Testing Contribution Claims with Bayesian Updating'[19], to categorise the strength of each piece of evidence. Process tracing categorises evidence into four types:

- **Hoop tests** – necessary but not sufficient (reject/weaken the CMO if not found but not sufficient to confirm the CMO; these are pieces of evidence that we would expect to see if the given CMO is true)
- **Doubly-decisive** – necessary and sufficient (confirm/strengthen the CMO if observed and if not observed the CMO is rejected/weakened; these are pieces of evidence that are expected but are also confirmatory of the CMO).
- **Smoking gun** – sufficient but not necessary (confirms/strengthens the CMO if observed but does not reject/weaken the CMO if not observed; these are pieces of evidence that we would 'like to see')

[19] https://www.cecan.ac.uk/sites/default/files/2018-01/BARBARA%20v2.5.pdf

- **Straw-in-the-Wind** – neither necessary nor sufficient (not sufficient to confirm the CMO if observed or to reject the CMO if not observed, but if observed would slightly strengthen the case for the CMO to be true (and slightly weaken the case if not observed))

The tables below list the evidence tests for each CMO; Table A5.1 lists the tests for CMOs in the theory for H1 and Table A5.2 lists the tests for CMOs in the theory for H2. In each case, tests for the outcome 'O' are presented first, followed by further tests for mechanisms 'M' and associated contexts 'C'.

The colour coding in the evidence test tables is as follows:

- blue rows - evidence tests relating to outcomes 'O'
- green rows - evidence tests relating to mechanisms 'M'
- orange rows - evidence tests relating to contexts 'C'

While we tried to specify evidence tests that related specifically to the linkages between M-Os and C-Ms, these were in practice difficult to distinguish from the tests for Cs, Ms and Os.

For each CMO hypothesis, we looked at all the tests for the constituent 'Cs', 'M' and 'O'. The outcome test provided evidence that the outcome had been observed, while the context and mechanism tests provided evidence of how and why the outcome occurred. For example, for CMO1.1, we looked at test results for O1, M1.1 and C1.1.1 and C1.1.2. The numbering convention for the evidence tests emphasises linkages between the elements of a CMO (e.g. O1-H1 is hypothesised to happen as a result of M1.1-H1, which is in turn expected to be triggered by contexts C1.1.1-C1.1.2-H1).

There is considerable repetition in the evidence tests, so we assigned nicknames to the tests. The tables below indicate the source of evidence for each test, its categorisation using the four process tracing categories, the competing explanations for observing that evidence, and the rationale for classifying the test.

The evidence tests were reviewed by two peer reviewers, a technical peer reviewer with expertise in DSR and by Dr. Barbara Befani, expert in process tracing. We made some minor adjustments to the categorisation and wording of evidence tests during the testing process, to improve consistency across the tests.

Table A5.1 - Tests for H1 – outcome 1 "second TA was additional for H1"

Element of CMO	Description	Nickname for test	Evidence tests for elements and linkages	Source of evidence	Type of test	Competing explanations	Rationale for classification of test
Outcome 1 - H1	The second TA contributes to more and/or more competitive flexible capacity for the capacity market in 2018-19 and subsequent years	Capacity agreement in 2018 CM	**H1 - O1-test(a.1)** Second TA participant obtains capacity agreements for flexible capacity in T-1 or T-4 auctions in 2018	CM registers for T-1 and T-4 held in Jan and Feb 2018.	Expect to see (hoop)	Necessary for O1. Could be observed for cases supporting Outcome 2 - flexible capacity put forward in CM but not attributable to TA	Evidence that this outcome applies (although there might be some external reason why they don't bid/clear in 2018/19). Could be observed even if TA had no influence on the flexible capacity they offer in the future CM.
as above	as above	Proven DSR in 2018 CM	**H1-01-test (a.2)** Second TA participant enters DSR into T-1 or T-4 auctions in 2018, using proven DSR CMU that was proven in TA.	CM registers for T-1 and T-4 held in Jan and Feb 2018.	Like to see (smoking gun)	None - a CMU that was tested and 'proven' in the first or second TA is directly submitted to a T-1 or T-4 auction in 2018, without any changes to the underlying capacity.	Not very likely for portfolio CMUs, since current rules mean that CMU only stays proven if no components could change between auctions. May observe for single site CMUs. If observed, this would definitely mean that this participant had lower costs in main TA (since they wouldn't need to retest this capacity).
as above	as above	Prequalified for 2018 CM	**H1 -O1- test (a.3)** - Second TA participant prequalifies capacity in T-1 or T-4 auctions in 2018.	CM registers for T-1 and T-4 held in Jan and Feb 2018.	Expect to see (hoop)	Could be observed even if not attributable to TA.	Evidence that this outcome applies. A less stringent test than (a.1) since they might have prequalified but failed to clear in either auction (e.g. because the price was too low). As above, could be observed even if TA had no influence on flexible capacity offered in the future CM.
as above	as above	Flexibility in CM is strategic opportunity	**H1 - O1 - test(b)** Second TA participant states in interview that they see flexible capacity in the CM as a strategic opportunity and intend to put forward flexible capacity to future CM auctions	Phase 3 or 4 interview (or earlier interviews), or public statements (e.g. online, publications).	Expect to see (hoop)	Necessary for O1. Could be observed for cases supporting Outcome 2 - flexible capacity put forward in CM but not attributable to TA	Evidence that this outcome applies. Could be observed even if TA had no influence on the flexible capacity they offer in the future CM.

as above	as above	Sunk costs	**H1 - O1-test (c)** Second TA participant says in interview that the costs they would look to cover in future CM auctions are lower than the costs they looked to cover for equivalent capacity in the second TA (because some costs now sunk i.e. 'more competitive')	Phase 4 interview	Like to see (smoking gun)	no significant competing explanations	We did not expect to see this, in the second TA, because in the first TA research participants were generally reluctant to discuss bidding strategies. Or they said that bidding strategies were based on other factors, not costs. But we got better at asking the question in a way that the respondents were willing to answer and found this evidence in a number of cases for the second TA. It was against their interests to say that DSR capacity was becoming more competitive due to sunk costs, so the likelihood of a false positive was low.

Further tests for H1 – CMO1.1 "second TA made CM seem less risky"

Element of CMO	Description	Nickname for test	Evidence tests for elements and linkages	Source of evidence	Type of test	Competing explanations	Rationale for classification of test
Mechanism 1.1 - H1 'CM learning'	Our experience of participating in the second TA (resource) means the capacity market seems less risky (reasoning)	Learning	**H1- O1- M1.1 Test (e) Evidence of causal mechanism:** The participant says in interview that they now have more confidence in being able to meet CM rules and regulations/be competitive in other CM auctions as a result of their participation in the second TA (e.g. because they developed skills/strategies/learning)	Phase 4 interviews	Expect to see and like to see (doubly-decisive)	Necessary and sufficient for M1.1. No significant competing explanations	Likely to see this if CMO applies, although those with experience of several auctions may have difficulty in distinguishing learning from the second TA. Unlikely to see if this evidence if mechanism does not apply.
Context 1.1.1 - H1 (additional test for Mechanism 1.1)	Established aggregators with existing clients who see the CM as a strategic opportunity but were not fully familiar with it	New player	**H1 - C1.1.1 test (f) Evidence of context:** - aggregator did not obtain capacity agreement for DSR in the T-4 auction, prior to the first TA (i.e. not familiar with CM) (NB this test refers to the first TA, since second TA participants have been in both)	Historic CM registers (Phase 1)	Expect to see (hoop)	Could be observed for cases supporting Outcome 3/4 -they still aren't interested in main CM, despite the TA	If the TA was needed to increase confidence and reduced perceived risk about DSR in the CM, the participant is unlikely to have obtained a capacity agreement for DSR in T-4 auctions prior to the (first) TA. But it's possible that participants were inexperienced in the CM at the start of the TA and the TA did not build their confidence.

Context 1.1.2 - H1 (additional test for Mechanism 1.1)	Direct participants and aggregators attracted by the second TA's low credit cover to build experience	Softer conditions	**H1 -C1.1.2 - test (g) Evidence of context:** Direct participants states in interview that the second TA's lower credit cover (or other 'softer' conditions) had enabled them to build their experience so that they plan to participate in the main CM in future, when they might not otherwise have gone straight into the main CM.	Phase 3 or 4 interview	Like to see (smoking gun)	no significant competing explanations	While Phase 1/2 participants did not mention credit cover much, it was discussed more in Phase 3 interviews by aggregators as well. So we now consider it more likely that this evidence will be seen if this context applies. But they have little reason to lie about this (i.e. don't expect false positives).

Further tests for H1 – CMO 1.2 "invested in assets for second TA that could be used in main CM"

Element of CMO	Description	Nickname for test	Evidence tests for elements and linkages	Source of evidence	Type of test	Competing explanations	Rationale for classification of test
Mechanism 1.2 - H1 'invested in assets'	In order to participate in the second TA (resource) we invested in capacity or the ability to provide capacity which will make us better positioned to participate in the main CM (reasoning)	Investment (reported)	**H1 - M1.2 - test (h): Evidence of causal mechanism:** Second TA participant saying in interview that they or their clients have developed or invested in assets (e.g. controls/metering) for the second TA that reduce costs of participation in future CM	Phase 3 and 4 interviews	Expect to see (hoop)	Necessary for M1.2. Could be observed for cases supporting Outcome 3/4 -they may have invested for the TA but may not go forward in the CM	Likely to see if the second TA has positively influenced the flexible capacity they offer to the future CM, and if this mechanism applies, but may also see if controls will really be used for other flexibility services, not the CM.

Context - 1.2.1 - H1 (additional test for Mechanism 1.2)	Aggregators and direct participants who invested in metering or transformer equipment to meet CM requirements	Investment (detail)	H1 - C1.2.1 - test (i): **Evidence of context:** details of significant investment in metering or control assets (e.g. more than £1k per CMU) by aggregator, direct participant or one of the aggregator's clients (for at least one of this participant's CMUs)	Email survey responses for TA participants and clients	Straw in the wind	Could be observed for cases supporting Outcome 3/4 -they may have invested for the TA but may not go forward in the CM	Less likely to see than M1.2 because email survey responses unlikely to be complete. Specific details in email survey provide more confidence than test (j) but there's still a possibility that controls will really be used for other flexibility services, not the CM.
as above	as above	Metering tests	**H1 - C1.2.1 -test (j): Evidence of context:** metering certificate or National Grid/Elexon statements indicate that meter testing has been completed for one or more components within this participant's CMUs (except if testing was only related to metering for onsite generation that could already have participated in wider CM)	Metering certificate (plus clarification on purpose of metering from Phase 4 interview data or National Grid/Elexon)	Like to see (smoking gun)	No significant competing explanations	Unlikely to see as most participants avoided meter testing through careful site selection. Undertaking metering testing was itself an investment of time and effort. Metering testing is specific to CM so very unlikely to invest in metering unless planning future CM involvement. Stronger test than test (i).

Further tests for H1 – CMO1.3 "started to develop customer base for second TA that can be used in main CM"

Element of CMO	Description	Nickname for test	Evidence tests for elements and linkages	Source of evidence	Type of test	Competing explanations	Rationale for classification of test
Mechanism 1.3 - H1 'started to develop customer base'	In order to participate in the second TA (resource) we have built a customer base and so now we want to continue with the CM (reasoning)	Building client base	**H1 - M1.3 - (test k): Evidence of causal mechanism:** The participant saying in interview that they have developed markets (e.g. building a client base, entering the UK market) for the second TA that they plan to use in one or more main CM auctions.	Phase 4 interview, supplemented by Phase 2 or 3 interviews where relevant	Expect to see (hoop)	Necessary for M1.3 Could be observed for cases supporting Outcome 3/4 -they may have invested for the TA but may not go forward in the CM	Likely to see if the TA has positively influenced the flexible capacity they offer to the future CM, but may also see if their client base ends up contributing to other flexibility services, not the CM. Or they may put capacity into main CM auctions but not clear the auctions.
Context 1.3.1 - H1 (additional test for Mechanism 1.3)	Aggregators (from outside the UK) who have used the TA to attract clients. OR Energy suppliers who have used the TA to sell flexibility to their energy supply customers	New entrant	**H1-C1.3.1 (test l):Evidence of context:** Evidence from National Grid records confirms that this TA participant was a new entrant to the GB flexibility market for the first or second TA (EITHER because they were an aggregator operating outside GB, OR because they were an energy supplier new to flexibility OR possibly other diversification contexts e.g. previously in energy management)	Published National Grid statistics	Expect to see (hoop)	Necessary for M1.3. Could also be observed for cases supporting Outcome 3/4 -they may have started in the TA but may not go forward in the CM.	Likely to see if the TA has positively influenced the flexible capacity they offer to the future CM, but may also see if their client base ends up contributing to other flexibility services, not the CM.

Context 1.3.2 - H1 (additional test for Mechanism 1.3)	as above	Clients new to flexibility	**H1-C1.3.2 (test m): Evidence of context:** One or more clients of this 'new' aggregator states in interview that they were new to flexibility for the TA (except for Triad), that the TA provided the motivation to get involved in flexibility AND that they have offered or plan to offer capacity in other CM auctions.	Phase 4 aggregator client interviews.	For a new aggregator (i.e. one that passes test I), this is a smoking gun - like to see	Power Responsive and other external influences may have contributed to some degree, alongside TA - so need evidence of specific influence by TA	Only doing a few client interviews, so risk that we may miss some new clients. Very unlikely to see this evidence if Mechanism does not apply.

Further tests for H1 – CMO1.4 – "new clients for existing aggregators"

Element of CMO	Description	Nickname for test	Evidence tests for elements and linkages	Source of evidence	Type of test	Competing explanations	Rationale for classification of test
Mechanism 1.4 - H1 'added new capacity to existing portfolio'	We have recruited new turn-down DSR clients or brought in new sites through the second TA (resource) and can cost-effectively include them alongside back-up in CMUs for the main CM (reasoning)	Some new capacity	**H1 - M1.4 - (test n): Evidence of causal mechanism:** We have recruited new turn-down DSR clients or brought in new sites because of the second TA (resource) and can cost-effectively include them alongside our existing clients for the main CM (reasoning)	Phase 4 interview, possibly supplemented by Phase 2 or 3 interviews where relevant	Expect to see and like to see (doubly-decisive)	Necessary for M1.4. No significant competing explanations.	Likely to see if this Mechanism and Outcome apply. Test is specific to TA influence.

Context 1.4.1 - H1 (additional test for Mechanism 1.4)	Aggregators and direct participants who have back-up and turn-down DSR capacity.	Already active	**H1 - C1.4.1 - (test o): Evidence of context:** Evidence from National Grid records that this TA participant was already active in the GB flexibility market prior to the first and second TA	National Grid records about participation in CM and other flexibility services.	Expect to see (hoop)	Necessary for mechanism 1.4.	Likely to see if this Mechanism applies (instead of Mechanism 1.3 relating to building customer base for flexibility services in the UK).
Context 1.4.2 - H1 (additional test for Mechanism 1.4)	as above	Clients new to flexibility	**H1-C1.4.1 (test p): Evidence of context:** One or more clients of this existing aggregator state in interview that they were new to flexibility for the TA (except for Triad), that the TA provided the motivation to get involved in flexibility AND that they have offered or plan to offer capacity in other CM auctions.	Phase 4 aggregator client interviews.	For an existing aggregator (i.e. one that passes test o), this is a smoking gun - like to see	Power Responsive and other external influences may have contributed to some degree, alongside TA - so need evidence of specific influence by TA	Only doing a few client interviews, so risk that we may miss some new clients. Very unlikely to see this evidence if Mechanism does not apply.
as above	as above	New sites for existing clients	**H1-C1.4.1 (test q): Evidence of context:** One or more clients of this existing aggregator states in interview (or shows from site data) that they were not new to flexibility but have brought more sites into flexibility services because of the second TA AND that they have offered or plan to offer this additional capacity in other CM auctions.	Phase 4 aggregator client interviews.	For an existing aggregator (i.e. one that passes test o), this is a smoking gun - like to see	Power Responsive and other external influences may have contributed to some degree, alongside TA - so need evidence of specific influence by TA	Only doing a few client interviews, so risk that we may miss some new clients. Very unlikely to see this evidence if Mechanism does not apply.

Tests for H1 – outcome 2 "participating with flexibility in the main CM but the second TA did not make any difference"

Element of CMO	Description	Nickname for test	Evidence tests for elements and linkages	Source of evidence	Type of test	Competing explanations	Rationale for type of test
Outcome 2 - H1	The second TA made no difference to the capacity available to the CM in 2018/19 and subsequent years and therefore is not additional	Capacity agreement in 2018 CM	**H1 - O2-test(a)** Second TA participant obtains capacity agreements for DSR in T-1 or T-4 auctions in 2018	CM registers for T-1 and T-4 held in Jan and Feb 2018.	Expect to see (hoop)	Could be observed for cases supporting Outcome 1 - flexible capacity put forward in CM and attributable to TA	Necessary for O2 - evidence that this outcome applies. Could also be observed if TA did influence the flexible capacity they offer in the main CM.
as above	as above	Flexibility in CM is strategic opportunity	**H1 - O2 - test(b**) Second TA participant states in interview that they see flexible capacity in the CM as a strategic opportunity and intend to put forward flexible capacity to future CM auctions	Phase 3 or 4 interview (or earlier interviews), or public statements (e.g. online, publications).	Expect to see (hoop)	Could be observed for cases supporting Outcome 1 - flexible capacity put forward in CM and attributable to TA	Necessary for O2 - evidence that this outcome applies. Could be observed even if TA did influence the flexible capacity they offer in the future CM.
as above	as above	No sunk costs	**H1 - O2-test (c)** Second TA participant says in interview that the TA had no impact on the costs they would look to cover in future CM auctions for flexible capacity (i.e. no sunk costs)	Phase 4 interview	Expect to see (hoop)	No significant competing explanations	Necessary for O2 - expect to see this if O2 applies. Participants may say that lots of other factors affect their bidding strategies, other than sunk costs. There might also be an element of lobbying bias ('costs are just as high as ever; need to bid high to cover uncertainties…').

Further tests for H1 – CMO2.1 "existing player - second TA made no difference to our participation with flexible capacity in the main CM"

Element of CMO	Description	Nickname for test	Evidence tests for elements and linkages	Source of evidence	Type of test	Competing explanations	Rationale for classification of test
Mechanism 2.1 - H1 (test for CMO2.1)	We have always intended to participate in the CM and the TA did not help us to grow our flexibility business.	No influence on capacity	**H1- M2.1 (test e): Evidence of causal mechanism:** Existing aggregators and direct participants state in the interview that they would have invested in, or maintained, capacity for future CM auctions regardless of the TA.	Phase 4 interview (supplemented by Phase 2 and 3 interviews where relevant)	Expect to see and like to see (doubly-decisive)	no significant competing explanations	Very likely to see this for existing CM participants, if CMO applies. Unlikely to agree with this if the TA did have influence.
Context 2.1.1 - H1 (additional test for CMO 2.1)	Existing aggregators and direct participants already had the confidence to participate in the main CM, for whom participating in the CM over multiple years is core part of their business model, and who did not use the TA to grow their flexibility business in the ways outlined in the contexts for outcome 1.	Multi-year business model	**H1 - C2.1.1 (test f): Evidence of context** - participant states in interview that they have a multi-year business model for CM involvement	Phase 4 or earlier interviews.	Straw in the wind	Could be observed for cases in supporting outcome 1 (additional flexible capacity in CM, attributable to TA) and outcomes 3 or 4 (flexible capacity from TA unavailable to CM) .	Fairly likely to be observed for those participating in CM going forward - those with a long-term business model may have been less interested in the temporary TA. But a participant might have a multi-year business model and still have valued TA as an entry point to the CM, or have decided that the main CM was not attractive after all. So not a strong test.

as above	as above	Already active	**H1 - C2.1.1 - (test g): Evidence of context:** Evidence from National Grid records that this TA participant was already active in the GB flexibility market prior to the first and second TA	National Grid records about participation in CM and other flexibility services.	Expect to see (hoop)	This context may also apply to mechanism 1.4 for Outcome 1, or for Outcomes 3 or 4.	Necessary for M2.1 to apply, so that TA not used for building customer base. But this context may apply to other Ms and Os too (ie some Type 1 error).
as above	as above	No learning	**H1 -C2.1.1 - test (h) Evidence of context:** no evidence that the lower credit cover and higher price of the second TA auction enabled this participant to build their experience/skills/organisational capacity for DSR in the CM and that this learning reduced the perceived riskiness of the main CM	Phase 3 or 4 interview	Expect to see (hoop)	Could also be observed for cases supporting other mechanisms in Outcome 1, or outcomes 3 or 4.	Necessary for M2.1 - expect to see this if M2.1 applies. Could also be observed for other outcomes.
as above	as above	No investment	**H1 - C2.1.1 - test (i): Evidence of context:** no evidence found of significant investment for TA in metering or control assets that could be used in main CM	Email survey responses and metering test certificates (for TA participants and aggregator clients)	Expect to see (hoop)	Could also be observed for cases supporting Outcome 3 or outcome 4	Necessary for M2.1 - expect to see this if M2.1 applies. Could also be observed for other outcomes.

as above	as above	No clients or sites new to flexibility	**H1-C2.1.1 (test j) : Evidence of context:** no evidence found of clients of this aggregator starting flexibility services (other than Triad) because of the second TA and then proceeding to the main the CM; or of direct participants/clients bringing in new sites because of the second TA and then offered/planning to offer these sites in other CM auctions	Phase 3 and 4 aggregator client and direct participant interviews. (In theory the email surveys include information on whether clients are new to flexibility, but email survey info on this point is poor quality/limited - e.g. some aggregators have not responded, and others have just selected the same option for all their clients.)	Expect to see (hoop)	Power Responsive and other external influences may have contributed to some degree, alongside TA - so need to ask about TA's role in motivation vs other influences.	Highly necessary for M2.1 that no new CM capacity was brought forward by the TA that was attributable to the TA rather than other influences. But high risk of Type 1 error as we might miss this evidence for other reasons (e.g. because doing only a few client interviews; or because Outcomes 3 or 4 apply).

Further tests for H1 – CMO2.2 "new entrant, but second TA made no difference to our participation with flexibility in the main CM"

Element of CMO	Description	Nicname for evidence test	Evidence tests for elements and linkages	Source of evidence	Type of test	Competing explanations	Rationale for probabilities
Mechanism 2.2 - H1 (test for CMO 2.2)	We are a new entrant to flexibility in the CM but would have started participating with flexible capacity in the CM at the same level anyway,	Entered market for other reasons	**H1- M2.2 (test k): Evidence of causal mechanism:** TA participants new to flexibility in the CM state in the interview that they would have invested in, or maintained, the same level of flexible capacity for future CM auctions regardless of the TA.	Phase 4 interview (supplemented by Phase 2 and 3 interviews where relevant)	Expect to see (doubly-decisive)	no significant competing explanations	Very likely to see for new entrants, if CMO applies. Unlikely to agree with this if the TA did have influence.

	because of other changes, not the TA						
Context 2.2.1 (additional test for CMO 2.2)	Aggregators or direct participants have recently started to pursue flexibility in the main CM because of changes in the wider market (e.g. review of embedded benefits) but they were already confident to participate in the main CM and did not use the TA to grow their flexibility business	New entrant	**H1-C2.2.1 (test l):Evidence of context:** Evidence from National Grid records confirms that this TA participant was a new entrant to the GB flexibility market for the first or second TA (EITHER because they were an aggregator operating outside GB, OR because they were an energy supplier new to flexibility OR possibly other diversification contexts e.g. previously in energy management)	Published National Grid statistics	Expect to see (hoop)	Could be observed for cases supporting Outcome 3/4 - they may have started in the TA but may not go forward in the CM.	Necessary for CMO2.2 to apply. But CMO2.2 won't apply to all new entrants, to some Type 1 error.
	as above	No learning	**H1 -C2.2.1 - test (m) Evidence of context:** no evidence that the lower credit cover and higher price of the second TA auction enabled this participant to build their experience/skills/organisational capacity for DSR in the CM and that this learning reduced the perceived riskiness of the main CM	Phase 3 or 4 interview	Expect to see (hoop)	Could also be observed for cases supporting other mechanisms in Outcome 1, or outcomes 3 or 4.	Necessary for CMO2.2 - expect to see this if CMO2.2 applies. Could also be observed for other outcomes.

	as above	No investment	**H1 - C2.2.1 - test (n): Evidence of context:** no evidence found of significant investment for TA in metering or control assets that could be used in main CM	Email survey responses and metering test certificates	Expect to see (hoop)	Could also be observed for cases supporting Outcome 3 or outcome 4	Necessary for CMO2.2 - expect to see this if CMO2.2 applies. Could also be observed for other outcomes.
	as above	No clients or sites new to flexibility	**H1-C2.1.1 (test j) : Evidence of context:** no evidence found of clients of this aggregator starting flexibility services (other than Triad) because of the second TA and then proceeding to the main the CM; or of direct participants/clients bringing in new sites because of the second TA and then offered/planning to offer these sites in other CM auctions	Phase 3 and 4 aggregator client and direct participant interviews.	Expect to see (hoop)	Power Responsive and other external influences may have contributed to some degree, alongside TA - so need to ask about TA's role in motivation vs other influences.	Highly necessary for CMO2.2 that no new CM Capacity was brought forward by the TA that was attributable to the TA rather than other influences. But high risk of Type 1 error as we might miss this evidence for other reasons (e.g. because doing only a few client interviews; or because Outcomes 3 or 4 apply).

Tests for H1 – Outcome 3 "Flexible capacity unavailable for the CM because of features of the CM"

Element of CMO	Description	Nicname for evidence test	Evidence tests for elements and linkages	Source of evidence	Type of test	Competing explanations	Rationale for probabilities
Outcome 3 - H1	Flexible capacity unavailable for the CM because of features of the CM	No capacity agreement in 2018 CM	H1 - O3 (test (a)) : Former TA participants do not obtain capacity agreements for DSR in the main CM auctions in 2018.	CM registers for 2018 auctions.	Expect to see (hoop)	Changes in the organisation's circumstances, unrelated to CM.	Likely to see this if CMO applies (although participants could still participate invisibly via an (or another) aggregator). May not participate for other reasons, rather than main CM conditions/prices.
	as above	Flexibility in CM NOT a strategic opportunity	H1-O3 (test (b)) Former TA participant states in interview that they do NOT intend to put forward flexible capacity in the future CM	Phase 4 interview (+ earlier interviews)	Expect to see (hoop)	Other factors about the flexibility market, rather than main CM conditions.	Necessary for the CMO to apply (effectively a screening test for outcomes 3 or 4). But may choose not to participate for other reasons, rather than main CM conditions.

Further tests for H1 – CMO3.1 "higher credit cover makes main CM unattractive"

Element of CMO	Description	Nickname for evidence test	Evidence tests for elements and linkages	Source of evidence	Type of test	Competing explanations	Rationale for probabilities
Mechanism H1-3.1 (test for CMO 3.1)	Higher credit cover makes the main CM less attractive to us than the TA because of the risks we perceive around compliance testing and/or delivery in the main CM	Main CM too risky	H1-M3.1 (test c): TA participant states in interview that higher credit cover makes the main CM more risky and therefore unattractive to them	Phase 4 interview (+ earlier interviews)	Expect to see (hoop)	Lobbying stance	Likely to see this if Mechanism applies. Some risk of lobbying, in claiming that CM conditions are their reason for not participating in future CM (in an attempt to influence future rules).
Context H1 - 3.1.1 (additional test for CMO 3.1)	Direct participants and aggregators who perceive a risk around compliance with metering requirements because of uncertain interpretation of CM rules.	Metering risks	H1-C3.1.1 (test d): TA participant states in interview that they perceive significant risks about compliance with metering test requirements in future CM	Phase 4 interview (+ earlier interviews)	Straw in the wind	Lobbying stance	May see this detail if this Mechanism applies. Some risk of lobbying, in claiming that CM conditions are their reason for not participating in future CM (in an attempt to influence future rules).

Context H1-3.1.2 (additional test for CMO 3.1)	Direct participants and aggregators with clients active in delivering other flexibility services or Triad, which potentially affect the baseline for DSR tests and delivery.	Risks to baseline	H1-C3.1.2 (test e): TA participant states in interview that they perceive significant risks about baselining for the future CM	Phase 4 interview (+ earlier interviews)	Straw in the wind	Lobbying stance	May see this detail if this Mechanism applies. Some risk of lobbying, in claiming that CM conditions are their reason for not participating in future CM (in an attempt to influence future rules).

Further tests for H1 – CMO3.2 "lower clearing price makes main CM unattractive"

Element of CMO	Description	Nickname for test	Evidence tests for elements and linkages	Source of evidence	Type of test	Competing explanations	Rationale for classification of test
Mechanism H1 - 3.2 (test for CMO 3.2)	Lower clearing price makes the main CM unattractive to us in spite of the attraction of the TA.	Did not clear auction	H1-M3.2 (Test f): TA participant entered DSR CMUs into main CM auction in 2018 but did not clear.	CM registers for 2018 auctions.	Expect to see (doubly-decisive)	External factors affect decision to exit auction, not just rules in main CM.	This shows they seriously considered participating with DSR in main CM but judged that the risk/reward balance was not right for them. Strong evidence because observed behaviour rather than interview statements.
Context H1-3.2.1 (additional test for CMO 3.2)	Direct participants, aggregators and aggregator clients who have significant hassle or opportunity costs to cover	Significant costs to cover	H1-3.2.1 (Test g): TA participants and aggregator clients report that they need to cover significant costs/risks to participate in main CM.	Email survey data and interview data	Straw in the wind	Lobbying stance	Participants may not share details; scope for lobbying; relationship between costs and auction behaviour not necessarily direct.

Context H1-3.2.2 (additional test for CMO 3.2)	Direct participants, aggregators and aggregator clients whose ability to stack CM revenues with other flexibility is diminished (e.g. because of reductions in Traid revenues)	Reduction in non-CM revenues (reported)	H1-3.2.2 (test h): TA participant states in interview that other sources of flexibility revenue have declined, and provides examples	Phase 4 interview (+ earlier interviews)	Straw in the wind	Lobbying stance	May see this detail if this Context applies, but not necessary to support this Mechanism. Some risk of lobbying, in claiming that CM conditions are their reason for not participating in future CM (in an attempt to influence future rules).
as above	as above	Reduction in non-CM revenues (market info)	H1-3.2.2 (test i): Industry-wide evidence of reductions in non-CM revenues for flexibility	Power Responsive annual report or working group minutes.	Straw in the wind	May not actually apply to TA participants.	May see this if Context applies, although not necessary to support this Mechanism. Industry wide info will not necessarily be relevant to all participants.

Tests for H1 – outcome 4 and CMO4.1 "put off main CM because of negative experience of second TA"

Element of CMO	Description	Nickname for test	Evidence tests for elements and linkages	Source of evidence	Type of test	Competing explanations	Rationale for type of test
Outcome 4 - H1	Capacity unavailable for the CM because of the experience of the TA	No capacity agreement in 2018 CM	H1 - O4 (test (a)) : Former TA participants do not obtain capacity agreements for DSR in the main CM auctions in 2018.	CM registers for 2018 auctions.	Expect to see (hoop)	Changes in the organisation's circumstances, unrelated to CM.	Likely to see this if CMO applies (although participants could still participate invisibly via an (or another) aggregator). May not participate for other reasons, rather than main CM conditions/prices.
	as above	Flexibility in CM NOT a strategic opportunity	H1-O4 (test (b)) Former TA participant states in interview that they do NOT intend to put forward flexible	Phase 4 interview (+ earlier interviews)	Expect to see (hoop)	Other factors about the flexibility market, rather than main CM conditions.	Necessary for the CMO to apply (effectively a screening test for outcomes 3 or 4). But may choose not to participate for other reasons, rather than main CM conditions.

			capacity in the future CM				
Mechanism 4.1 - H1 (test for CMO 4.1)	Participating in the TA (resource) has deterred us from participating in the CM	TA put us off	H1-M4.1 (test c) Respondents state in the interview that their experience of the TA deterred them from participating in the main CM.	Phase 4 interview (+ earlier interviews)	Expect to see (doubly decisive)	Might conceivably ascribe decision not to pursue main CM to TA experiences, while actually the issue was really the tougher conditions in the main CM - but this is not likely. So don't really expect this to be attributable to other explanations.	Very likely to see this if CMO applies. Unlikely to agree with this if there are other reasons for not participating in the main CM.
Context 4.1.1 - H1 (test for CMO 4.1)	Direct participants and aggregators testing approaches to DSR and/or client recruitment that are new to them	Testing new approach	H1-C4.1.1 (test d) Aggregator or direct participant states in interview that they were using the TA to test a new approach	Phase 4 interview (+ earlier interviews)	Straw in the wind		Weakish test - not necessary for M4.1 to apply (may have been active in other CM auctions but still have been put off by TA). And they could have been testing a new approach but found that it went well (so not sufficient either).
Context 4.1.2 - H1 (test for CMO 4.1)	Direct participants and aggregators putting forward sites/clients that require complex metering tests (e.g. because of onsite renewables)	Complex metering	H1-C4.1.2 (test e) Aggregator or direct participant undertook a significant number of meter tests for second TA	Metering certificates from National Grid/EMRS	Like to see (smoking gun).		Hard evidence of TA hassle. Don't expect to see large number of metering tests because they are so much hassle (according to Phase 1-3 evidence). But metering tests are just one of several possible explanations for the TA putting them off.

Context 4.1.3- H1 (test for CMO 4.1)	Direct participants and aggregators putting forward sites/clients that are delivering for other flexibility services which affects their baseline and reduces their measured DSR capacity	Risks to baseline	H1-C4.1.3 (test f) Aggregator or direct participant cites baseline issues relating to delivery for Triad or other flexibility services as the reason why the TA has put them off future participation in the main CM	Phase 4 interview (+ earlier interviews)	Straw in the wind	Lobbying stance	Less strong evidence than test (e) because based on interview evidence. One of several possible explanations for the TA putting them off.
Context 4.1.4- H1 (test for CMO 4.1)	Direct participants and aggregators facing high recruitment/hassle costs to identify potential sites/clients with simple metering and no baseline issues within the timescale required for TA	Recruitment difficult	H1 - C4.1.4 (test g) Aggregator or direct participant faces high recruitment/hassle costs to identify potential sites/clients with simple metering and no baseline issues within the timescale required for TA.	Phase 4 interview (+ earlier interviews)	Straw in the wind	Lobbying stance	Less strong evidence than test (e) because based on interview evidence. One of several possible explanations for the TA putting them off. More likely to be seen for aggregators than direct participants.

Table A5.2: Tests for H2 – Outcome 1 – "second TA contributes to wider encouragement of turn-down DSR"

Element of CMO	Description	Nickname for test	Evidence tests for elements and linkages	Type of test	Competing explanations	Rationale for type of test
Outcome 1	The second TA contributes to wider encouragement of turn-down DSR	More turn-down DSR in CM	**H2 - O1 (Test a):**Greater volumes of turn-down DSR come forward for the CM in future	Straw in the wind	Other market changes supporting turn-down DSR (e.g. changes to Triad, MCPD, Power Responsive)	Difficult to observe, because turn-down DSR is not always identified separately in main CM (except in optional business case detail). There may be other reasons for any increase/decrease in turn-down DSR volumes within the CM, not just the TA.
as above	as above	More turn-down DSR for other services	**H2 - O1 (Test b):** Greater volumes of turn-down DSR made available to other flexibility services in future	Straw in the wind	Other market changes supporting turn-down DSR (e.g. changes to Triad, MCPD, Power Responsive)	Look for information in Power Responsive report. But there may be many other reasons for any increase/decrease in turn-down DSR volumes, not just the TA.
as above	as above	Second TA encouraged more turn-down DSR	**H2-O1 (Test c):** Participants in TA say in interview that the second TA has encouraged them to make more turn-down DSR available to the CM or other services in future	Expect to see (doubly-decisive)	Other market changes supporting turn-down DSR (e.g. changes to Triad, MCPD, Power Responsive)	Very likely to see if true; only a small risk of lobbying even if not true (to obtain future support for turn-down DSR).

Further test for H2 – CMO1.1 – "already doing turn-down DSR but second TA has helped"

Element of CMO	Description	Nickname for test	Evidence tests for elements and linkages	Type of test	Competing explanations	Rationale for type of test
Mechanism 1.1 (test for CMO 1.1)	We ALREADY INTENDED to increase our turn-down DSR offer but the second TA has built (aggregator client trust and) experience (resource) and helped direct participants, aggregators and their clients to assess the risk and opportunity costs involved (reasoning)	Second TA helped organisations along their 'turn-down journey'	**H2-M1.1 (test d.1):** Participants say in interview that the second TA contributed to them making more turn-down DSR available - for the TA and the wider flexibility market - by helping them, or their clients, assess the risk and opportunity costs involved in turn-down DSR. [in other words, evidence of the second TA encouraging organisations to start their 'turn-down journey'].	Expect to see (hoop)	Other market changes supporting turn-down DSR (e.g. changes to Triad, MCPD, Power Responsive)	Second TA may have been one of several factors encouraging clients to try out turn-down DSR
as above		Already active in turn-down DSR	**H1 - M1.1 (test d.2)** Evidence from National Grid records that this TA participant was already active in turn-down DSR in GB prior to the first and second TA	Expect to see (hoop)	n/a	Necessary for this CMO to apply.
Context 1.1.1 (test for CMO 1.1)	Direct participants and aggregator clients who offered interruptible industrial loads in the TA built confidence and experience	Some new turn-down capacity (reported)	**H2 - C1.1.1 (test e):** Interviewee states that the second TA helped them to bring greater volumes of turn-down DSR from their industrial loads (e.g. bringing in new sites) into the TA and future flexibility market.	Like to see (smoking gun)	Other market changes supporting turn-down DSR (e.g. changes to Triad, MCPD, Power Responsive)	Like to see more turn-down capacity if this CMO holds. Small risk that interviewee may lie about it being to the TA rather than other factors, not second TA

as above	as above	Some new turn-down capacity (detail)	**H2 - C1.1.1 (test f)**: Evidence of direct participants and aggregator clients contracting for greater volumes of turn-down DSR in the 2nd TA, compared to the first (difference from test e is that test f does not involve any evidence that this increase is attributable to the 2nd TA)	Straw in the wind	Other market changes supporting turn-down DSR (e.g. changes to Triad, MCPD, Power Responsive)	Expect to see more turn-down capacity if this CMO holds, but it may be attributable to other factors, not second TA.
Context 1.1.2 (test for CMO 1.1)	Aggregators [already] offering turn-down DSR	High price brought in new turn-down clients	**H2-C1.1.2 (test g) :** Aggregator states in interview that the financial stimulus of the second TA helped them to reach NEW clients for their turn-down DSR portfolio, for the TA and future flexibility market.	Expect to see (doubly-decisive)	Other market changes supporting turn-down DSR (e.g. changes to Triad, MCPD, Power Responsive)	Likely to see if this context applies. Unlikely to lie about this if context does not apply.
Context 1.1.3 (test for CMO 1.1)	Aggregators with clients who could be approached to offer turn-down DSR	High price encouraged existing clients to put forward additional turn-down sites	**H2 - C1.1.3 -(test h):** Aggregator (and their EXISTING clients) state in interview that the financial stimulus of the second TA helped them to start offering turn-down for the first time and/or bring additional turn-down sites into the TA and future flexibility market.	Like to see (smoking gun)	Other market changes supporting turn-down DSR (e.g. changes to Triad, MCPD, Power Responsive)	Likely to see if this context applies. Unlikely to lie as this evidence test involves quite specific statements about the role of the second TA.
Context 1.1.4 (test for CMO 1.1)	Aggregators and direct participants who see turn-down DSR as contributing to their CSR commitments	CSR motives for turn-down DSR	**H2 - C1.1.4 (test i):** CSR commitments play a role in the choice, by aggregators, direct participants and/or clients, to offer increased turn-down DSR to the TA and future flexibility market.	Straw in the wind	Other market changes supporting turn-down DSR (e.g. changes to Triad, MCPD, Power Responsive)	Might see this evidence, showing that CSR is a supportive motivation alongside financial incentive, although there was little mention of it in earlier Phases. CSR might not be enough motivation to contribute much to the outcome/mechanism.

Further tests for H2 – CMO1.2 – "second TA encouraged us to enter the market for turn-down DSR"

Element of CMO	Description	Nickname for test	Evidence tests for elements and linkages	Type of test	Competing explanations	Rationale for type of test
Mechanism 1.2 (test for CMO 1.2)	The (first and) second TA has encouraged us to ENTER the market for aggregation of turn-down DSR, it (resource) gave us something to talk to new clients about (reasoning) and the high price (resource) made it cost effective to bring new assets to market (reasoning).	High price brought in new aggregators, seeking new turn-down clients	**H2 - M1.2 (test j)** Aggregators and clients/potential clients report active marketing of turn-down DSR opportunities by aggregators NEW to the UK market, focused - at least initially - on the deal available through the second TA.	Expect to see (hoop)	Other market changes supporting turn-down DSR (e.g. changes to Triad, MCPD, Power Responsive)	Expect to see new entrants to turn-down market in GB if this mechanism applies. But market entry may be attributable to other factors, not just second TA.
Context 1.2.1 (test for CMO 1.2)	Aggregators new to the GB market for turn-down DSR building their client base.	New player for turn-down DSR in GB	**H2 - C1.2.1 test (k) Evidence of context:** - aggregator not active in GB turn-down market immediately prior to 1st and 2nd TA	Expect to see (hoop)	n/a	Necessary for this CMO to apply. But not sufficient as the aggregator might have entered the market for other reasons, not for the TA.
as above	as above	High price brought in new turn-down clients	**H2-C1.2.2 (test l) : A**ggregator states in interview that the financial stimulus of the second TA helped them to reach new clients and build their turn-down DSR portfolio, for the TA and future flexibility market.	Expect to see (doubly-decisive)	Other market changes supporting turn-down DSR (e.g. changes to Triad, MCPD, Power Responsive)	Very likely to see if this context applies. Unlikely to lie about this if context does not apply.

Further tests for H2 – CMO1.3 "second TA acted as a pilot for turn-down DSR (aggregator client)"

Element of CMO	Description	Nickname for test	Evidence tests for elements and linkages	Type of test	Competing explanations	Rationale for type of test
Mechanism 1.3 (test for CMO 1.3)	The TA (resource) has encouraged us to enter the market for flexibility services via an aggregator and has allowed us to pilot or increase our involvement at low cost/risk (reasoning)	Aggregator clients treating second TA as pilot	H2 - M1.3 (test m) Aggregator client states that they were treating the second TA as a pilot for offering turn-down services	Like to see (smoking gun)	Other market changes supporting turn-down DSR (e.g. changes to Triad, MCPD, Power Responsive)	Not particularly likely to see this evidence - as we were only interviewing a sample of clients. Not likely to observe this evidence if CMO does not apply.
as above	as above	Aggregator clients gained confidence and plan to offer turn-down DSR in future.	H2-M1.3 (test n) Aggregator client states that they have more confidence because of second TA and intend to offer turn-down DSR services in future.	Like to see (smoking gun)	Other market changes supporting turn-down DSR (e.g. changes to Triad, MCPD, Power Responsive)	Not particularly likely to see this evidence - as we were only interviewing a sample of clients. Not likely to observe this evidence if CMO does not apply.
as above	as above		H2-M1.3 (test o) Aggregator client states that second TA triggered a new relationship with an aggregator that has opened a range of new opportunities for flexibility revenues, including turn-down DSR	Like to see (smoking gun)	Other market changes supporting turn-down DSR (e.g. changes to Triad, MCPD, Power Responsive)	Not particularly likely to see this evidence - as we were only interviewing a sample of clients. Not likely to observe this evidence if CMO does not apply.
Context 1.3.1 (test for CMO 1.3)	Aggregator clients with experience of turning down for Triad only, who wanted to test potential involvement in turn-down services and	Organisation new to turn-down DSR (reported)	H2-C1.3.1 (test n) Aggregator client states that they previously turned down only for Triad (or not at all)	Expect to see (hoop)	Other market changes supporting turn-down DSR (e.g. changes to Triad, MCPD, Power Responsive)	Necessary for this context/mechanism to apply.

	participated via an aggregator					
as above	as above	Second TA seen as low risk	H2-C1.3.1 (test m) Aggregator client states that they saw the second TA as a low-risk way of offering turn-down services.	Straw in the wind	Other market changes supporting turn-down DSR (e.g. changes to Triad, MCPD, Power Responsive)	Contributory evidence that this mechanism applies - but not conclusive.

Tests for H2 – Outcome 2 and CMO2 "second TA gave mixed messages for turn-down DSR"

Element of CMO	Description	Nickname for test	Evidence tests for elements and linkages	Type of test	Competing explanations	Rationale for type of test
Outcome 2 - H2	Mixed messages received about turn-down DSR, potential providers have reservations. Consequently less investment into turn-down DSR than if there was a clear long-term commitment.	Mixed messages	**H2-O2 (test a):** Participants (and possibly non-participants) state in interview that they are committed to - and have invested in - turn-down DSR but that lack of certainty about TA/T-1/T-4 pathways has constrained this investment.	Expect to see (hoop)	Lobbying stance	If this outcome applies, we would expect them to make these interview statements. But they might lobby for more certainty, even if this outcome does not apply.
Mechanism 2.1 - H2 (test for CMO 2)	The TA (resource) has increased our confidence that there will be ongoing support for turn-down DSR (reasoning) AND	TA is sign of government commitment	**H2-M2.1 (test b):** Participants state in interview that the first and second TA has increased their confidence in government commitment to turn-down DSR.	Straw in the wind	Could have increased confidence in government commitment without having reservations about mixed messages.	Could see this evidence even without rest of CMO applying.

Mechanism 2.2 - H2 (test for CMO 2)	Reducing the capacity set aside [for turn-down DSR/other capacity] in T-1 and excluding TA participants from T-4 (resource) reduces the opportunities for turn-down DSR (reasoning)	T-1 uncertainties and/or TA/T-4 exclusion rules problematic	**H2-M2.2 (test c):** Participants state in interview that the lack of set aside capacity for T-1, and/or the exclusion rules applying to the TA/T-4, adversely affects their perception of future opportunities for turn-down DSR.	Expect to see (hoop)	Lobbying stance	If this mechanism applies, we would expect them to make these interview statements. But they might lobby for changes in set aside capacity or T-1/T-4 rules, even if this CMO does not apply.
Context 2.1.1 H2 (test for CMO 2)	Aggregators and direct participants with potential turn-down DSR assets where [ongoing] CM payments are needed for cost-effectiveness.	Uncertain about future CM revenues	H2-C2.1.1 (test d): Participants state in interview that they will not increase their provision of turn-down DSR, or may not keep available the capacity put forward for the TA, because of lack of certainty over the future availability of CM revenues	Expect to see (hoop)	Lobbying stance	If this mechanism applies, we would expect them to make these interview statements. They might lobby for more certainty in T-1, even if this mechanism does not apply.

Tests for H2 – Outcome 3 "Second TA made no difference to turn-down DSR"

Element of CMO	Description	Nickname for test	Evidence tests for elements and linkages	Type of test	Competing explanations	Rationale for type of test
Outcome 3 - H2	The TA made no difference to turn-down DSR	Second TA did NOT encourage more turn-down DSR	**H2 - O3 (test a)**: Participants (and possibly non-participants) state in interview that the TA made no difference to their commitment to, or volume of offer for, turn-down DSR.	Expect to see and like to see (doubly-decisive)	Lobbying for other support for turn-down DSR, beyond TA.	Likely to see this, if true, and participants are unlikely to claim this unless it is true.
		No increase in turn-down DSR in CM	**H2-O3 (test b):** No increase in volumes of turn-down DSR coming forward for the CM in future	Expect to see (hoop)	Turn-down DSR still not attractive, despite TA.	There may be other reasons for lack of increase, not just lack of TA influence.
		No increase in turn-down DSR in other services	**H2-O3 (test c):** No increase in volumes of turn-down DSR made available via other services.	Expect to see (hoop)	Turn-down DSR still not attractive, despite TA.	There may be other reasons for any increase/decrease in turn-down DSR volumes, not just the TA.

Further test for H2 – CMO3.1 "Second TA did not help us bring in new turn-down capacity or new turn-down clients, so made no difference"

Element of CMO	Description	Nickname for test	Evidence tests for elements and linkages	Type of test	Competing explanations	Rationale for type of test
Mechanism 3.1 - H2 (test for CMO 3.1)	We have always done as much turn-down DSR as we can.	TA was just another revenue stream for us	**H2-O3 (test d): P**articipants state in interview that they were already providing as much turn-down DSR as they could, before the TA (and the TA was just another revenue stream for their existing capacity).	Like to see (straw in the wind)	Aggregator may have used second TA as an additional revenue stream for existing clients, but still have brought in new clients using the attraction of the high price.	May see this evidence for some direct participants and aggregator clients, but there may still be H2 additionality for other participants & clients of the same aggregator
Context 3.1.1 H2 (test for CMO 3.1)	Direct participants and aggregator clients with interruptible industrial loads or other forms of storage participating in STOR and/or Triad.	Not dependent on TA revenues	**H2 - C3.1.1 (test e):** Participants state in interview that turn-down DSR projects are already cost-effective because of non-TA revenues available to them (both CM and non-CM).	Expect to see (hoop)	Other market changes supporting turn-down DSR (e.g. changes to Triad, MCPD, Power Responsive)	May see this evidence for some direct participants and aggregator clients, but still have additionality for other participants & clients (for whom other CMOs apply)
as above		No new turn-down capacity in second TA (reported)	**H2 - C3.1.1 (test f)**: Participants (and aggregator clients) state in interview that they did NOT bring sites new to turn-down into the second TA	Expect to see (hoop)	Other factors may have constrained growth in turn-down capacity.	Likely to observe this if this part of the CMO applies. But any constraint on growth in turn-down DSR volumes in the TA may arise from external factors.
as above		No new turn-down capacity in second TA (detail)	**H2 - C3.1.1 (test g):** Direct participants and aggregator clients do NOT contract for greater volumes of turn-down DSR in the 2nd TA, compared to the first.)	Expect to see (hoop)	Other factors may have constrained growth in turn-down capacity.	Likely to observe this if this part of the CMO applies. But any constraint on growth in turn-down DSR volumes in the TA may arise from external factors.

Further tests for H2 – CMO3.2 "we do turn-down DSR but not for the CM, so second TA made no difference"

Element of CMO	Description	Nickname for test	Evidence tests for elements and linkages	Type of test	Competing explanations	Rationale for type of test
Mechanism 3.2 - H2 (test for CMO 3.2)	We have always done as much turn down DSR as we can and do not intend to provide it in the CM.	We do turn-down anyway, but outside CM	**H2 - M3.2 (test h):** Participants state in interview that they were already doing as much turn-down as they can, before the TA, but that incompatible services constrain the turn-down DSR that they can do in the CM.	Like to see (smoking gun)	Might choose not to participate in CM for other reasons, not strict incompatibility.	Fairly likely to observe this, if true, and participants are unlikely to claim this unless it is true.
Context 3.2.1 H2 (test for CMO 3.2)	Aggregators active in areas where providing turn-down DSR in the CM is incompatible with other services (e.g. dynamic FFR) or the assets they aggregate (e.g. because they are not suitable for CM stress events)	Incompatible services	**H2 - C3.2.1 (test i)**: Non-participant aggregators market turn-down for other services (e.g. firm frequency response) but not for the CM.	Like to see (smoking gun)	Might just happen not to mention CM on their website and in their materials.	Might possibly see this evidence. Fairly unlikely if this CMO does not apply.
as above		Do not offer DSR in CM	**H2-C3.2.1 (test j):** Aggregator does not participate in CM with DSR. (note: test specified in terms of DSR, not turn-down, because difficult to identify turn-down in CM register)	Expect to see (hoop)	Might offer DSR in main CM, but not turn-down DSR. This test does not distinguish.	Likely to see this evidence if CMO applies.

Further tests for H2 – CMO3.3 "no interest in turn-down DSR, so second TA made no difference"

Element of CMO	Description	Nickname for test	Evidence tests for elements and linkages	Type of test	Competing explanations	Rationale for type of test
Mechanism 3.3 - H2 (test for CMO 3.3_	We did not provide turn-down DSR for the TA and have no interest in or capacity for turn-down DSR	Did not clear in 2nd TA auction	**H2 - M3.3 (test k)** This organisation did not clear any turn-down DSR capacity in the 2nd TA.	Expect to see (hoop)	Other factors may have influenced non-participation in second TA, even if interested in turn-down DSR.	If the organisation is not interested in turn-down, it's a necessary pre-condition that they haven't cleared turn-down capacity in the second TA. However, they might still be interested in turn-down but have other reasons for not participating in the 2nd TA.
as above		No interest in turn-down DSR (reported)	**H2 - M3.4 (test l)** Interview statements that the organisation currently has no interest in providing turn-down DSR.	Expect to see (doubly-decisive)	n/a	Likely to see if this part of the CMO applies; unlikely to see this if CMO does not apply.
Context 3.3.1 H2 (test for CMO 3.3)	Organisations that provide peaking or back-up generation.	No interest in turn-down DSR (detail)	**H2 - C3.3.1- (test m):** Phase 1 and 2 interview evidence indicates that this participant provided only peaking generation or back-up generation (not turn-down DSR) to the first TA.	Expect to see (doubly-decisive)	n/a	Likely to see if this part of the CMO applies; unlikely to see this if CMO does not apply.

Further tests for H2 – CMO3.4 "not appropriate for us to do turn-down DSR at the moment, so second TA made no difference"

Element of CMO	Description	Nickname for test	Evidence tests for elements and linkages	Type of test	Competing explanations	Rationale for type of test
Mechanism 3.4 - H2 (test for CMO 3.4)	We would like to offer more turn-down DSR but it's not appropriate for us at the moment	Turn-down DSR not competitive at the moment	**H2 - M3.4 (test n)**: Interview statements about the organisation being theoretically willing to offer turn-down DSR in future, but it not currently being viable.	Expect to see (hoop)	Lobbying stance	Likely to see plea for more support, possibly even if it's not needed.
Context 3.4.1 H2 (test for CMO 3.4)	Organisations that found it harder to provide turn-down DSR in the TA than they expected: a longer	Turn-down DSR requires further support, beyond TA	**H2 - C3.4.1 (test o):** Interview statements claiming that turn-down DSR needs more support or a longer timeframe to come forward than provided by the TA.	Expect to see (hoop)	Lobbying stance	Likely to see plea for more support, possibly even if it's not needed.

	timeframe and more financial support would be required.					
as above	as above	Have not cleared turn-down DSR in 2018 CM	H2 - C.3.4.1 (test p): Organisation has not cleared proven DSR from 2nd TA in the main CM (which would be identifiable as turn-down DSR).	Expect to see (hoop)	Slim chance that they may have decided not to participate in main CM for other reasons, other than non-viability.	Necessary for this mechanism to apply - if they cleared DSR in the main CM then it must be viable for them.

How we have applied the tests

We applied the evidence tests to each potential CMO as a mini hypothesis in itself. Given the number of evidence tests, and the limited resources available, we focused the testing on the most complex and important cases. The tests were therefore applied to the cases of the six aggregators that went forward to delivery in the second TA, taking into account evidence from these aggregators and any of their clients that were interviewed during Phases 3 and 4. In one case, we also took account of evidence from a sub-aggregator that had submitted capacity via one of the aggregators but was not themselves a participant in the second TA.

We did not apply the tests to direct participants in the second TA for two reasons: firstly, there was limited evidence of additionality from these participants in our main analysis of their evidence; and secondly, there were only three direct participants, so the test findings were likely to be disclosive. Similarly, we did not apply the tests to the two aggregators that dropped out of the second TA because there was little additionality in these cases.

We streamlined the process and reduced duplication by only applying tests where relevant to a particular case. For example, where evidence tests for an outcome were failed, we did not test for the supporting mechanism and context. Similarly, where evidence tests for a mechanism were failed, we did not test for supporting contexts. The tests therefore indicate those C-M-Os are well supported by the evidence. Where there are competing mechanisms for the same outcome (e.g. one additional and one not), the evidence tests show the relative support for additional and non-additional CMOs in the theory.

In applying the tests, we have synthesised evidence from a range of sources including:

- Publicly available data (e.g. Capacity Market Register, published surveys and reports, aggregator and National Grid websites)
- Interviews with these aggregators (primarily from Phase 4 of the evaluation but drawing on earlier interviews where relevant. Where there was conflicting evidence from different evaluation phases conflicts, this was noted in the detailed assessment against each test.)
- Interviews with their clients (some were interviewed during Phase 4 and some during Phase 3 and in one case we draw on an earlier interview with a client during Phase 2 of the evaluation, for information on how their situation had changed before/after the TA).
- Email survey information for aggregators and clients, including cost data and characterisation of capacity as new or existing, where available.

We used a spreadsheet to code evidence for each case against the evidence tests. The evidence summaries and coding were prepared by one researcher and reviewed by another member of the project team. The detailed evidence and coding were also reviewed by Dr. Barbara Befani. Although this evidence was anonymised it was potentially disclosive because of the small number of TA participants. So we prepared non-disclosive summaries of the results as shown in Table A5.3 and Table A5.4 below.

To present the detailed results for evidence tests we made use of the following colour coding:

Key:	Explanation
Does not support	Fails a ‘hoop’ test or ‘doubly-decisive’ test
Slightly negative	Fails a ‘straw in the wind’ or ‘smoking gun’ test
Slightly supportive	Mixed evidence for this test
Some support	Passes a ‘hoop’ or ‘straw in the wind’ test
Strongly supportive	Passes a ‘smoking gun’ or ‘doubly-decisive’ test

We then created higher level summarises which combined test results using the following synthesis rules, to indicate the combined level of support for each CMO. These rules were developed by the project team but have been peer reviewed by Dr. Barbara Befani.

Key:	Explanation	Process tracing concepts:
Strong support	Confirmatory evidence: at least one 'sufficient' or 'necessary and sufficient' test passed. No necessary tests failed. Allow failure of some tests which are 'not necessary or sufficient'.	Confirmatory evidence: at least one 'smoking gun' or 'double-decisive' test passed. No 'hoop tests' failed but allow failure of some 'straw in the wind' tests.
Some support	No necessary tests failed. Allow failure of some tests which are 'not necessary or sufficient'. No 'sufficient' or 'necessary and sufficient' tests passed.	No 'hoop tests' failed but allow failure of some 'straw in the wind' tests. No 'smoking gun' or 'double-decisive' test passed.
Mixed support	Apparently contradictory results - including at least one 'necessary' test being failed but also at least one 'sufficient' test being passed.	Mix of 'hoop' test failures and 'smoking gun' or 'double-decisive' tests being passed.
No support	At least one necessary test being failed, and no 'sufficient' tests being passed.	At least one 'hoop' test failed. No 'smoking gun' or 'double-decisive' tests passed.

Process tracing findings for hypothesis H1 – TA contributes to more flexible capacity in the main CM

The detailed results of evidence tests for the CMOs in the candidate H1 theory are given in Table A5.3 below.

Table A5.3 Detailed evidence test results for H1

		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Element of CMO	**Type of test**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	**Observed?**
Outcome 1 - H1	Expect to see (hoop)	yes	yes	yes	Yes	yes	yes
as above	Like to see (smoking gun)	Redacted to avoid disclosure – one case passed this test					
as above	Expect to see (hoop)	yes	yes	yes	Yes	yes	yes
as above	Expect to see (hoop)	yes	yes	yes	yes	yes	yes
as above	Like to see (smoking gun)	yes	yes	yes	yes	no	yes
Mechanism 1.1 - H1 'CM learning' (test CMO1.1)	Expect to see and like to see (double-decisive)	yes	yes	yes	no	yes	no
Context 1.1.1 - H1 (test for CMO 1.1)	Expect to see (hoop)	yes	yes	yes	no	yes	yes
Context 1.1.2 - H1 (test for CMO 1.1)	Like to see (smoking gun)	yes	yes	yes	no	yes	no
Mechanism 1.2 - H1 'invested in assets' (CMO1.2)	Expect to see (hoop)	yes - slightly for second TA	yes	yes	yes	no	yes
Context - 1.2.1 - H1 (test for CMO 1.2)	Straw in the wind	yes	yes	yes	yes	no	yes
as above	Like to see (smoking gun)	yes	yes	yes	no	no	no

		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Element of CMO	**Type of test**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	**Observed?**
Mechanism 1.3 - H1 'started to develop customer base' (test CMO 1.3)	Expect to see (hoop)	yes (although some concerns about clearing prices in CM going forward.)	no		yes	yes	yes
Context 1.3.1 - H1 (test for CMO 1.3)	Expect to see (hoop)	yes	no	no	yes	yes	yes
Context 1.3.2 - H1 (test for CMO 1.3)	For a new aggregator, this is a smoking gun - like to see	yes	not relevant	not relevant	yes	partial	yes
Mechanism 1.4 - H1 'added new capacity to existing portfolio' (test CMO 1.4)	Expect to see and like to see (doubly decisive)	no	yes	yes	no	no	no
Context 1.4.1 - H1 (additional test for Mechanism 1.4)	Expect to see (hoop)	no	yes	yes	no	not relevant	no
Context 1.4.2 - H1 (additional test for Mechanism 1.4)	For an existing aggregator, this is a smoking gun - like to see	not relevant	partial	no	not relevant	not relevant	not relevant
as above	For an existing aggregator, this is a smoking gun - like to see	not relevant	no	partial	not relevant	not relevant	not relevant
H1 - CMOs for outcome 2							

		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Element of CMO	**Type of test**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	**Observed?**
Element of CMO	**Type of test**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	
Outcome 2 - H1	Expect to see (hoop)	yes	yes	yes	yes	yes	yes
as above	Expect to see (hoop)	yes	yes	yes	yes	yes	yes
as above	Expect to see (hoop)	no	no	no	no	yes	no
Mechanism 2.1 - H1 (test for CMO2.1)	Expect to see and like to see (doubly-decisive)	no	no	no	no	no	no
Context 2.1.1 - H1 (test for CMO 2.1)	Straw in the wind	not relevant	partial	no	not relevant	not relevant	not relevant
as above	Expect to see (hoop)	not relevant	yes	yes	not relevant	not relevant	not relevant
as above	Expect to see (hoop)	not relevant	no	no	not relevant	not relevant	not relevant
as above	Expect to see (hoop)	not relevant	no	no	not relevant	not relevant	not relevant
as above	Expect to see (hoop)	not relevant	no	no	not relevant	not relevant	not relevant
Mechanism 2.2 - H1 (test for CMO 2.2)	Expect to see (doubly-decisive)	No	no	no	mixed	mixed	no
Context 2.2.1 (test for CMO 2.2)	Expect to see (hoop)	not relevant	not relevant	not relevant	yes	yes	not relevant
	Expect to see (hoop)	not relevant	not relevant	not relevant	yes	no	not relevant
	Expect to see (hoop)	not relevant	not relevant	not relevant	no	yes	not relevant
	Expect to see (hoop)	not relevant	not relevant	not relevant	no	no	not relevant
H1 - CMOs for outcome 3							

		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Element of CMO	**Type of test**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	**Observed?**
Element of CMO	**Type of test**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	**Observed?**
Outcome 3 - H1	Expect to see (hoop)	no	no	no	no	no	no
	Expect to see (hoop)	no	no	no	no	no	no
Mechanism H1-3.1 (test for CMO 3.1)	Expect to see (hoop)	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant
Context H1 - 3.1.1 (test for CMO 3.1)	Straw in the wind	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant
Context H1-3.1.2 (test for CMO 3.1)	Straw in the wind	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant
Mechanism H1 - 3.2 (test for CMO 3.2)	Expect to see (doubly-decisive)	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant
Context H1-3.2.1 (test for CMO 3.2)	Straw in the wind	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant
Context H1-3.2.2 (test for CMO 3.2)	Straw in the wind	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant
as above	Straw in the wind	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant
H1 - CMOs for outcome 4							
Element of CMO	**Type of test**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	**Observed?**
Outcome 4 - H1	Expect to see (hoop)	no	no	no	no	no	no
	Expect to see (hoop)	no	no	no	no	no	no
Mechanism 4.1 - H1 (test for CMO 4.1)	Expect to see (doubly-decisive)	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant
Context 4.1.1 - H1	Straw in the wind	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant

		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Element of CMO	**Type of test**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	**Observed?**
(test for CMO 4.1)							
Context 4.1.2 - H1 (test for CMO 4.1)	Like to see (smoking gun)	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant
Context 4.1.3- H1 (test for CMO 4.1)	Straw in the wind	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant
Context 4.1.4- H1 (test for CMO 4.1)	Straw in the wind	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant

The detailed test results have been combined into an overall assessment of the support for each CMO (as shown in Table A5.4), using the key explained in the previous section.

Table A5.4: Summary of support for CMOs in H1 theory

		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Outcome 1: (additional)	**The second TA contributes to more and/or more competitive flexible capacity for the capacity market in 2018-19 and subsequent years**	Strong support	Strong support	Strong support	Strong support	Some support	Strong support
CMO 1.1	Our experience of participating in the second TA (resource) means the capacity market seems less risky (reasoning)	Strong support	Strong support	Strong support	No support	Strong support	No support
CMO 1.2	In order to participate in the second TA (resource) we invested in capacity or the ability to provide capacity which will make us better positioned to participate in the main CM (reasoning)	Strong support	Strong support	Strong support	Some support	No support	Some support
CMO 1.3 (new entrants)	In order to participate in the second TA (resource) we have built a customer base and so now we want to continue with the CM (reasoning)	Strong support	No support	No support	Strong support	Some support	Strong support
CMO 1.4 (existing aggregators)	We have recruited new turn-down DSR clients or brought in new sites through the second TA (resource) and can cost-effectively include them alongside back-up in CMUs for the main CM (reasoning)	No support	Strong support	Strong support	No support	No support	No support
		Case 1	**Case 2**	**Case 3**	**Case 4**	**Case 5**	**Case 6**
Outcome 2: (non-additional)	The second TA made no difference to the capacity available to the CM in 2018/19 and subsequent years and therefore is not additional	No support	No support	No support	No support	Some support	No support
CMO 2.1 (existing aggregators)	We have always intended to participate in the CM and the TA did not help us to grow our flexibility business.	Not relevant	No support	No support	Not relevant	Not relevant	Not relevant

CMO 2.2 (new entrants)	We are a new entrant to flexibility in the CM but would have started participating with flexible capacity in the CM at the same level anyway, because of other changes, not the TA	Not relevant	Not relevant	Not relevant	Mixed support	Mixed support	Not relevant
No support for Outcomes 3 and 4, because all these aggregators are committed to future CM participation							
Outcome 3	Flexible capacity unavailable for the CM because of features of the CM	(although some comments made about aspects of CM that could be made more favourable for DSR)					
Outcome 4	Capacity unavailable for the CM because of the experience of the TA	(although some negative comments about some aspects of the second TA)					

To summarise, the findings indicate strong support for the additionality CMO for H1, with limited support for non-additional CMOs. All of the aggregators had gone ahead to participate in the main Capacity Market (CM), and all attributed some growth in their portfolios or knowledge to the second TA. But there was some support for the non-additional CMO from two aggregators who commented in interview that they would have gone straight into the main CM even without the TA (although the scale of their portfolios might have been reduced). This is discussed in chapter 3 of the main report.

Process tracing findings for hypothesis H2 – TA contributes to wider encouragement of turn-down DSR

The detailed results of evidence tests for CMOs in candidate H2 theory are given below.

Table A5.5 Detailed evidence test results for H2

		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Element of CMO	**Type of test**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	**Observed?**
Outcome 1	Straw in the wind	no	yes	no	no	yes	no
as above	Straw in the wind	partial (market-wide)	yes	partial - market wide	yes	yes	partial - market wide
as above	Expect to see and like to see (doubly-decisive)	yes	yes	yes	yes	yes	yes
Mechanism 1.1 (test for CMO 1.1)	Expect to see (hoop)	yes	yes	yes	yes	yes	yes
as above	Expect to see (hoop)	no	yes	yes	no	no	no
Context 1.1.1 (test for CMO 1.1)	Like to see (smoking gun)	not relevant	no	no	not relevant	no	not relevant
as above	Straw in the wind	not relevant	yes	yes	not relevant	no	not relevant
Context 1.1.2 (test for CMO 1.1)	Expect to see and like to see (doubly-decisive)	not relevant	yes	yes	not relevant	partial	not relevant
Context 1.1.3 (test for CMO 1.1)	Like to see (smoking gun)	not relevant	yes	yes	not relevant	no	not relevant
Context 1.1.4 (test for CMO 1.1)	Straw in the wind	not relevant	yes	yes (for aggregator and two of the three clients interviewed)	not relevant	no	not relevant
Mechanism 1.2 (test for CMO 1.2)	Expect to see (hoop)	yes	not relevant	not relevant	yes	yes	yes

Context 1.2.1 (test for CMO 1.2)	Expect to see (hoop)	yes	not relevant	not relevant	yes	yes	yes
as above	Expect to see and like to see (doubly-decisive)	yes	not relevant	not relevant	yes	yes	yes
Mechanism 1.3 (test for CMO 1.3)	Like to see (smoking gun)	no	no	no	yes	not available	yes
as above	Like to see (smoking gun)	no	no	no	yes	not available	yes
as above	Like to see (smoking gun)	no	yes	no	no	not available	no
Context 1.3.1 (test for CMO 1.3)	Expect to see (hoop)	yes	yes	no	yes	not available	yes
as above	Straw in the wind	yes	yes	no	yes	not available	yes
H2 - CMOs for outcome 2							
Element of CMO	**Type of test**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	**Observed?**
Outcome 2 - H2	Expect to see (hoop)	no	no	no	yes	no	no
Mechanism 2.1 - H2 (test for CMO 2)	Straw in the wind	not relevant	not relevant	not relevant	partial	not relevant	not relevant
Mechanism 2.2 - H2 (test for CMO 2)	Expect to see (hoop)	not relevant	not relevant	not relevant	yes	not relevant	not relevant
Context 2.1.1 H2 (test for CMO 2)	Expect to see (hoop)	not relevant	not relevant	not relevant	yes	not relevant	not relevant
H2 - CMOs for outcome 3							
Element of CMO	**Type of test**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	**Observed?**	**Observed?**
Outcome 3 - H2	Expect to see and like	no	no	no	no	no	no

	to see (doubly-decisive)						
	Expect to see (hoop)	no	no	no	no	no	no
	Expect to see (hoop)	no	no	no	no	no	no
Mechanism 3.1 - H2 (test for CMO 3.1)	Straw in the wind	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant
Context 3.1.1 H2 (test for CMO 3.1)	Expect to see (hoop)	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant
as above	Expect to see (hoop)	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant
as above	Expect to see (hoop)	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant
Mechanism 3.2 - H2 (test for CMO 3.2)	Like to see (smoking gun)	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant
Context 3.2.1 H2 (test for CMO 3.2)	Like to see (smoking gun)	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant
as above	Expect to see (hoop)	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant
Mechanism 3.3 - H2 (test for CMO 3.3_	Expect to see (hoop)	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant
as above	Expect to see and like to see (doubly-decisive)	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant
Context 3.3.1 H2 (test for CMO 3.3)	Expect to see and like to see (doubly-decisive)	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant
Mechanism 3.4 - H2	Expect to see (hoop)	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant

(test for CMO 3.4)							
Context 3.4.1 H2 (test for CMO 3.4)	Expect to see (hoop)	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant
as above	Expect to see (hoop)	not relevant	not relevant	not relevant	not relevant	not relevant	not relevant

As for H1, we have then summarised these test results to give a high-level assessment of support for each CMO in the H2 theory. These results are presented in Table A5.6 below.

Table A5.6: Summary of support for CMOs in H2 theory

		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Outcome 1: (additional)	**The second TA contributes to wider encouragement of turn-down DSR**	Strong support	Strong support	Strong support	Strong support	Strong support	Strong support
CMO 1.1 (existing aggregators)	We ALREADY INTENDED to increase our turn-down DSR offer but the second TA has built (aggregator client trust and) experience (resource) and helped direct participants, aggregators and their clients to assess the risk and opportunity costs involved (reasoning)	No support	Strong support	Strong support	No support	No support	No support
CMO 1.2 (new aggregators)	The (first and) second TA has encouraged us to ENTER the market for aggregation of turn-down DSR, it (resource) gave us something to talk to new clients about (reasoning) and the high price (resource) made it cost effective to bring new assets to market (reasoning).	Strong support	No support	No support	Strong support	Strong support	Strong support
CMO 1.3 (clients only)	The TA (resource) has encouraged us to enter the market for flexibility services via an aggregator and has allowed us to pilot or increase our involvement at low cost/risk (reasoning)	Some support	Strong support	No support	Strong support	No evidence avail-able	Strong support
		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Outcome 2: (additionality could have been greater)	**Mixed messages received about turn-down DSR, potential providers have reservations. Consequently less investment into turn-down DSR than if there was a clear long-term commitment.**	No support	No support	No support	Some support	No support	No support

CMO 2.1	Reducing the capacity set aside [for turn-down DSR/other capacity] in T-1 and excluding TA participants from T-4 (resource) reduces the opportunities for turn-down DSR (reasoning)	Not relevant	Not relevant	Not relevant	Some support	Not relevant	Not relevant

No support for Outcomes 3 - The TA made no difference to turn-down DSR.

In conclusion, these findings indicate strong support for the additionality CMOs for H2 at aggregator level, with aggregators developing their confidence and systems to aggregate turn-down DSR and/or building their portfolios of turn-down clients. There was less consistent support for the additionality CMO at client level (because some clients were already involved in turn-down DSR for Triad and other services and were simply adding the TA as another revenue stream). The apparent inconsistency between these findings is explained by the fact that some clients changed aggregator as a result of the second TA: while these were perceived as 'new' to the aggregator they were not necessarily new to flexibility. Clients previously delivering for Triad only were counted as additional.

Overall synthesis process

The final section of this appendix describes the methods that we used to synthesise findings from all the evidence sources and analysis methods involved in Phase 4, including qualitative analysis, cost analysis and process tracing.

Contribution analysis within a realist evaluation

We used contribution analysis to synthesise the evidence and analysis from Phase 4 (and Phases 1-3 where relevant) to test and review the 'contribution story' as set out in the theoretical framework. Phase 4 qualitative analysis focused largely on testing, additionality theory and reliability of delivery and preliminary indications of potential CMN responses. Testing of participation theory and auction theory was largely completed during Phase 3. The overall contribution analysis drew on the findings from process tracing set out above (in relation to additionality theory), and also drew on wider evidence from qualitative research and quantitative capacity/cost analysis (beyond the specific tests specified in process tracing) and on evidence from non-participants and wider stakeholders (primarily from Phase 3). We revised the theoretical framework at the end of Phase 4 in the light of this contribution analysis, as explained in Appendix 3. The process that we used for this 'realist' contribution analysis was similar to the approach used in earlier phases, and involved the following steps:

- We refined the theoretical framework at the end of Phase 3/start of Phase 4, on the basis of evidence gathered during Phase 3. This was the 'candidate theory' for Phase 4.

- We used the theoretical framework to inform the detailed design of topic guides for Phase 4 fieldwork, with a view to using these interviews to test and refine the candidate theory.

- As part of our descriptive analysis activities, we organised qualitative and quantitative data into readily accessible spreadsheets.

- As explained in Appendix 4, we undertook normal qualitative analysis of evidence for topics such as reasons for drop-out, testing, reliability and potential delivery in response to CMNs. We analysed in-depth interview responses alongside test results, CM register details and email survey responses for each interviewee (where available).

- For additionality theory, we undertook more formal testing of CMOs. We developed a CMO coding spreadsheet for additionality theory, with rows for every organisation for which we had in-depth interview evidence about the second TA (from Phase 3 and 4). This was organised so that evidence for each aggregator was analysed alongside evidence from their clients (where available). Again, we incorporated relevant evidence from test results, CM register details, public statements and email survey responses, not just in-depth interviews. Where relevant, we drew on contextual information from Phases 1 and 2.

- We then coded the contexts, mechanisms and outcomes for each case against the Cs, Ms and Os in the candidate additionality theory. We reviewed evidence from Phase 3 and earlier phases where there were gaps in Phase 4 evidence. Where there were inconsistencies between evidence from different phases, we endeavoured to understand the reason for any apparent inconsistencies and based our coding on the most relevant evidence. For example, where changes appeared to be attributable to lack of recall, we gave more weight to evidence closer in time to the events that were being discussed. But where the most recent qualitative evidence superseded evidence from earlier phases, we gave priority to up-do-date evidence gathered in Phase 4. Where possible, we cross-checked qualitative evidence with observed behaviour (e.g. we cross-checked information provided on costs with publicly available information on whether the respondent had obtained capacity agreements for DSR at a given clearing price in another CM auction). Where motivations were not clear, we checked participants' websites or public statements for other insights into their rationale.

- For the six aggregators going forward to delivery, we formally tested the evidence supporting additional and non-additional candidate CMOs using process tracing, as described in this Appendix. If the candidate theory did not exactly fit the observed evidence for a given case, we developed 'tailored' CMOs which were variants of the candidate theory.

- The analysis and refinement of CMOs was an iterative process: we saw patterns emerging as we analysed successive cases. The coding was undertaken by two analysts within the project team and was discussed (in non-disclosive form) with the wider team and BEIS at a policy review meeting.

- The detailed coding spreadsheets formed the basis of our findings in the main report and were used to inform our final revision of the theory at the end of Phase 4 (as described in Appendix 3).

Participatory analysis

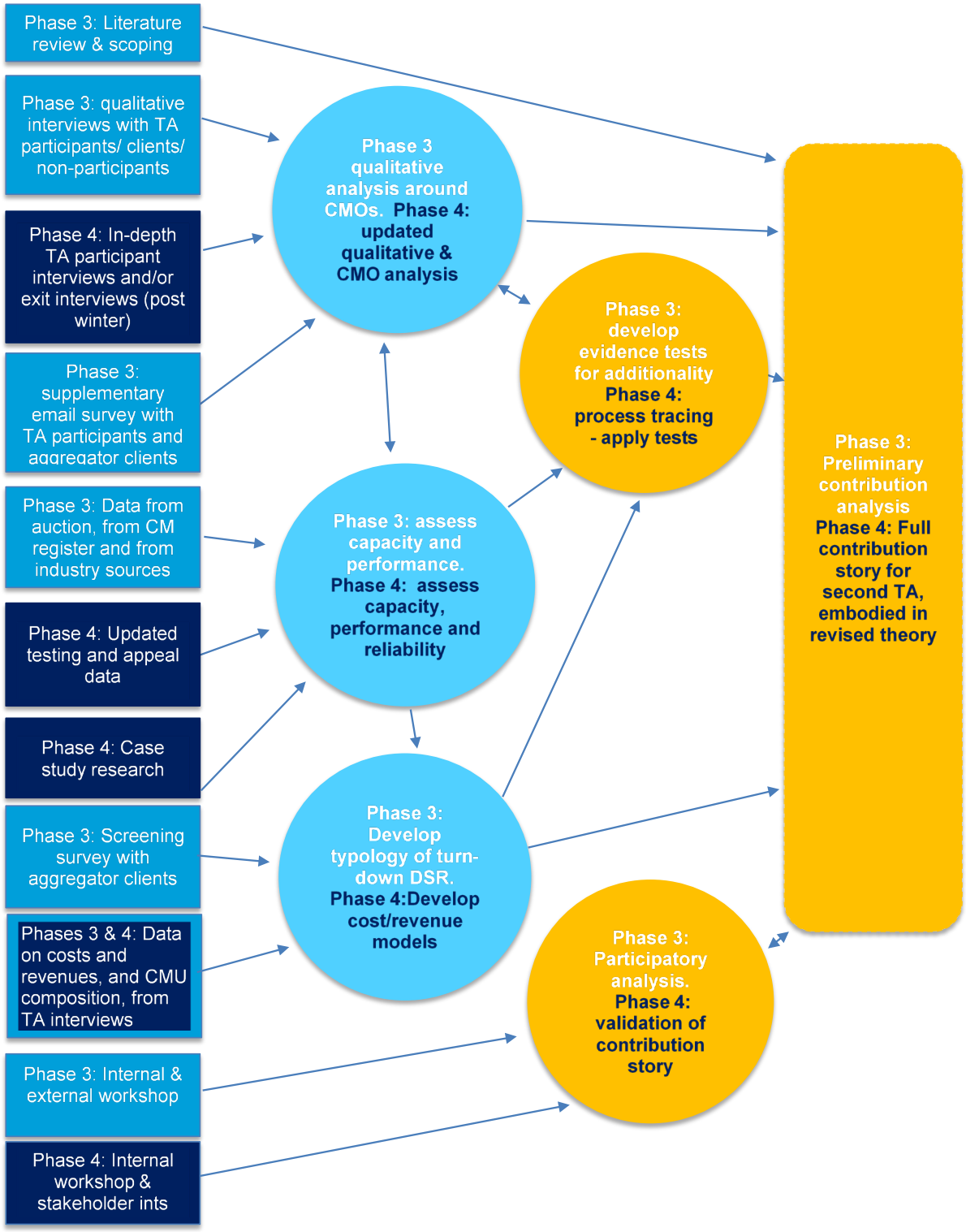
We tested the robustness of findings from Phase 4 with members of the technical and evaluation expert panels, at appropriate points in the process. We also held an internal workshop towards the end of Phase 4, with members of the project team, BEIS representatives and two peer reviewers, to test and refine the emerging contribution story.

Overall synthesis in Phase 4

The synthesis process for Phase 4 addressed all the HLQs. The process tracing particularly informed our assessment of HLQ 1 and HLQ 2 (i.e. what outcomes the TA has achieved and how), while the wider contribution analysis explained how we understand these outcomes were achieved, for whom and in what circumstance and the role played by TA design (HLQs 2 and 4). The analysis of costs and revenues for different types of turn-down DSR supported our analysis of outcomes for HLQ1 and our assessment of value for money for HLQ 3. Our assessment of HLQs 1-4 informed our assessment of implications for the future for HLQ 5.

Our synthesis report at the end of Phase 4 focused on the second TA. We have also prepared a slide-pack presenting high-level findings from all Phases of the evaluation, covering both the first and second TA.

Figure A5.1 Relationship of evidence sources, evaluation phases, descriptive analysis and generative causation methods in Phases 3 and 4



Appendix 6: Methodology for DSR characterisation and cost analysis

Introduction

This appendix explains the methodology used to derive DSR characterisation and analyse costs during Phases 3 and 4 of the TA evaluation and presents additional commentary on the findings presented in chapter 2 of the main report.

Objectives

The original aim of the cost analysis was to identify categories of turn-down DSR that share common characteristics, appear to be subject to similar decision-making processes and have broadly similar cost/revenue structures in relation to the TA/CM and other flexibility services. The overall purpose of this work was to inform BEIS's analysis of turn-down DSR.

In practice, analysis of costs by category of turn-down DSR was problematic, for reasons explained below. The aims of this analysis were therefore revised to:

- Characterise turn-down DSR within the second TA by industry group and asset type
- Analyse cost data by MW, without distinguishing between different categories of turn-down DSR.

Scope

The research focused on turn-down DSR, with data taken from the second TA. A small amount of research data from the first TA was also used to provide additional information on sites that participated in both years of the TA. Information from DSR participating in the other CM auctions was not included; indeed scant information is available for this, as turn-down DSR is difficult or impossible to distinguish from back up generation with the available information.

Terminology clarification – CMUs versus components

A CMU Component refers to an individually metered DSR asset. A DSR asset refers to the electrical system that can provide flexible demand, which can range from an individual piece of equipment such as a pump or motor, through to an industrial process comprising hundreds of power consuming elements, such as a manufacturing line.

Multiple components are aggregated together into a Capacity Market Unit (CMU) for participation in the capacity market. A CMU can also be based on a single component, provided it meets the minimum CMU size threshold.

Contents

This appendix sets out the following elements of analysis:

- Method and data sources for characterisation of DSR
- Characterising second TA components by business activity
- Characterising second TA components by DSR asset type
- Method for analysing DSR cost and revenue data
- Normalised costs by MW
- Estimates of opportunity costs
- Estimated revenues from other flexibility services

Method and data sources for characterisation of DSR

The first task was to characterise the types of DSR put forward for delivery in the second TA. The research focused on the constituent components of a CMU i.e. individual DSR measures which may be aggregated under a single CMU. The initial approach was as follows:

Phase 3:

- Review of the latest Capacity Register.
- Matching of component Meter Point Administration Numbers[20] (MPANs) against address database, to identify client sites and business activity.
- Inclusion of questions regarding DSR component characteristics, costs and revenues within the e-mail survey and in-depth interviews with TA participants and aggregator clients (see Annex A and Annex B to Appendix 4).
- The cost data was requested in three categories: the initial metering and equipment set up cost (£), the initial staff time (hrs) and ongoing staff time (hrs/yr). The staff time data has been converted into a monetary value by assuming a labour rate, as described in the cost data section below. Participants had the opportunity to respond with costs per component but chose to respond with estimated costs per CMU or for all CMUs.
- Cross referencing relevant Phase 1 and 2 data to bring additional contextual information.
- Synthesis of all data sources into a single 'Component database' file.
- Categorisation of components into DSR typologies.

Phase 4:

- Extraction of component capacities for all delivered CMUs from DSR test data (i.e. proven capacity).
- The preparation of six detailed case studies drawing on all available information and follow up calls to the DSR asset operators. This focused on context for the turn-down DSR, to investigate issues of opportunity cost, reliability and duration of response.
- Extracts from these case studies are presented in the main report. They cover all the turn-down DSR types identified in this report with the exception of Horticultural lighting. The components may be across multiple sites, but within the umbrella of a single business. Cost data is correspondingly generally provided per groups of aggregated components or by CMU.
- Updating of the phase 3 component data base file with Phase 4 interview data and additional e-mail surveys.
- Preparation of charts and tables for the Phase 4 report.

[20] Unique codes used to identify fiscal electricity consumption meters.

The e-mail survey responses contained the most directly relevant data; however they were not completed in all cases. Whilst all of the organisations completed at least part of the email survey, some did not provide all of the information requested. In particular, only seven of the nine TA participants going forward to delivery (two direct participants and five aggregators) provided cost data at participant-level and CMU-level. Eight aggregator clients provided cost data, but this was a small sample relative to the estimated population of 75 client organisations involved in the TA. Where possible, email survey data was supplemented with interview data, case study information and deductions from the business type and meter location. The small sample size limited the detail of data that can be reported without disclosing commercially sensitive data, therefore only aggregated data is presented.

Confidence in the categorisation is recorded as high, medium or low as follows:

- High confidence – direct response from participant in e-mail survey or clear information from other sources such as interview responses. In contrast to Phase 3, the majority of components now fall into this category.
- Medium confidence – derived from contextual information such as site or business activity with a good level of confidence e.g. cold stores.
- Low confidence – categorisation derived from site or business activity with a lower level of confidence e.g. broad categories of manufacturing.

We report categorisation information for all assets in the second TA below. We have included assets for which our categorisation confidence is low as these represent a relatively low proportion of the capacity put forward in the second TA and this approach enables us to provide characterisation for the whole scheme.

Characterising second TA components by business activity

From the DSR test data of the 28 CMUs in the 2nd TA we identified 333 components with a total proven capacity 493MW[21]. The business activity of the site where each CMU component is located is shown in the Figure A6.1 below. It is evident that DSR components are spread across a large number of business activities, mostly manufacturing. There are few components in the commercial sector.

The contractual capacity of the second TA was 293MW. 'Proven capacity' is determined by the DSR tests and is commonly greater than bidding and contracted capacities, as participants seek to ensure that they can fully meet their obligations and not be penalised for under performance. Furthermore, proven component capacities may (individually and in aggregate) legitimately vary between tests, so long as the contracted CMU capacity is met.

[21] The total proven capacity shown in the CM register for the same CMUs is 483MW. We believe the 10MW (2%) difference is due the existence of some meter data including renewable generation which is not discernible within the raw data provided. Cost normalisation is done using delivered capacity from the Capacity Market register in any case – see below.

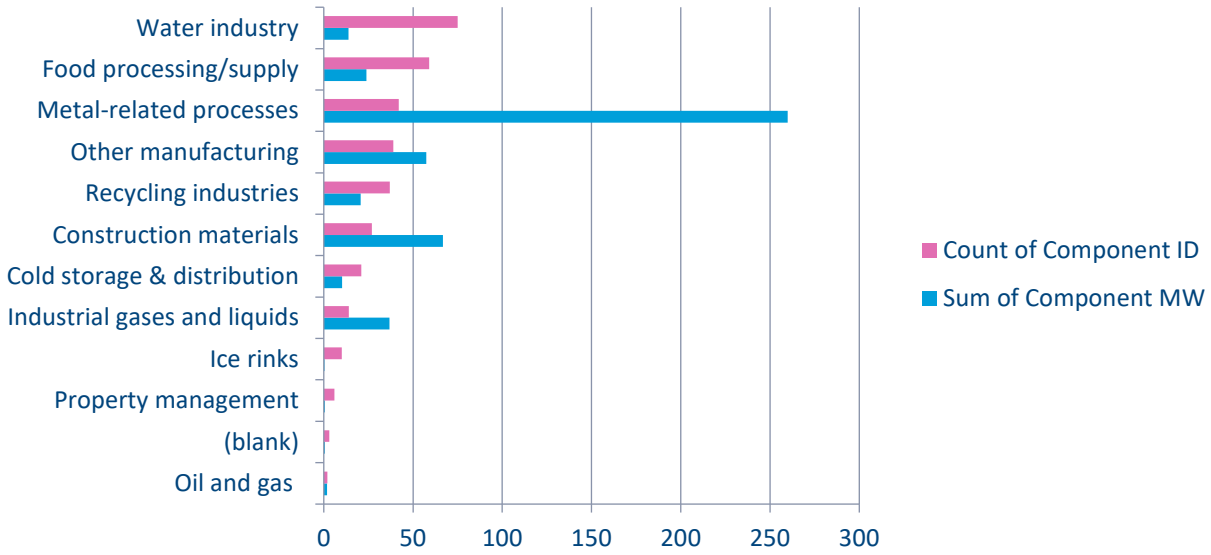


Figure A6.1: Number of components and DSR Test capacity by business activity (source: consultant's analysis of DSR test and MPAN data for sites going forward to delivery)

Characterising second TA components by DSR asset type

The second TA is exclusively for turn-down DSR, thus the 'DSR type' in this context is taken to refer to the type of electricity-consuming equipment that is being turned down. The DSR asset type was offered as drop-down options in the e-mail survey.

The categories were proposed based on the types of technologies expected to be able to provide flexible demand. An 'other' category with a text box for respondents to provide further description was also included. The proposed categories were peer reviewed by the wider research team and by external stakeholders including Power Responsive, National Grid and the Major Energy Users Council.

We cannot categorically state that other turn-down DSR types were not present in other CM auctions; however the fact that the original categorisation was based on first principles and was peer reviewed suggests that the categorisation covers the main possibilities for turn-down DSR. Building HVAC, pumps, refrigeration and chillers are sub-types of 'motors and drives' however and there was felt to be limited value in creating too many categories, hence the use of the 'Other motors and drives' category. There was only one example of lighting DSR, so the more precise label of 'Horticultural lighting' has been used, given the particular characteristics of this lighting application.

Table A6.1: Description of turn-down DSR types

DSR asset type	Further description
Building HVAC	Motors attached to air conditioning compressors in a commercial building. There is only one case of this, providing six components.
Horticultural lighting	There is only one component of this type. The lighting is for herb growing and provides both heat and illuminance.
Other motors and drives	Motors to drive conveyor belts, milling machines, crushers. The manufacturing activities associated with this category include: food, animal feed, quarrying and construction products, flour milling metals manufacture and metal recycling.
Process heating	Electrical heating used within an industrial process. The following sectors/processes are present: steel manufacture (arc furnace), glass manufacture, insulation manufacture, aluminium smelter, plastic extrusion and baked products.
Pumps	Motors attached to hydraulic or gaseous pumps.
Refrigeration and chillers	Motors attached to industrial refrigerant pumps, compressors and fans. These are found in cold stores, food processing and ice rinks.
Other	This relates to 5 components from 3 CMUs associated with 'various process related equipment' for the manufacture of paper and inorganic chemicals, plus a window manufacturer (process unknown).

Figure A6.2 and Figure A6.3 below cover all components in the 2017/18 delivery year.

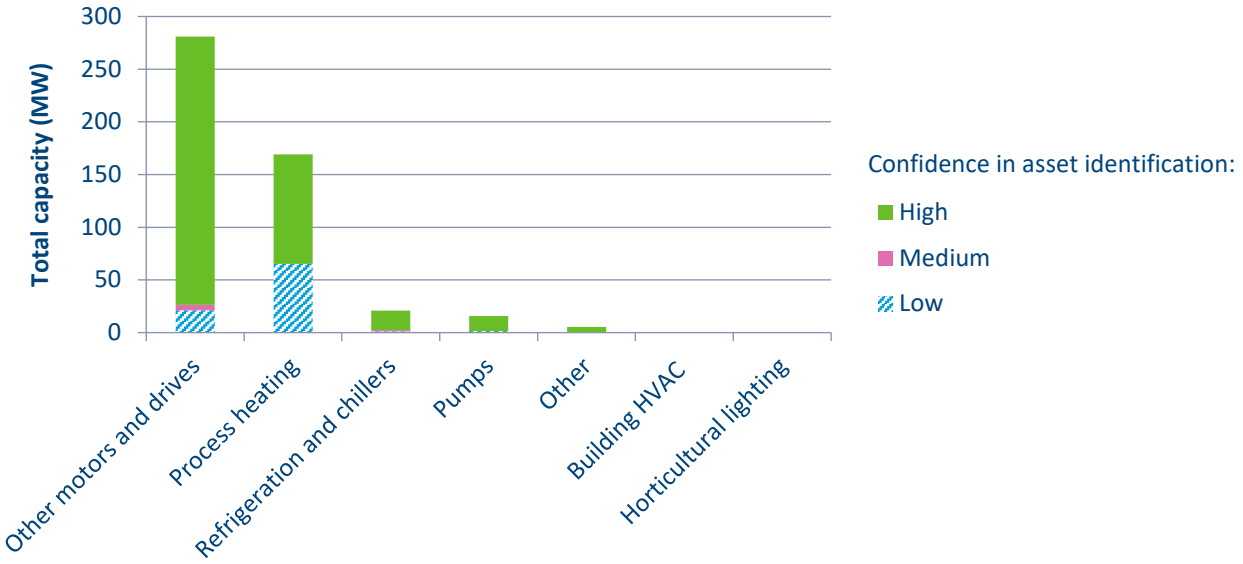


Figure A6.2: Sum of component capacities (MW) by type from DSR test data (source: consultant's analysis of Phase 3 and 4 data)

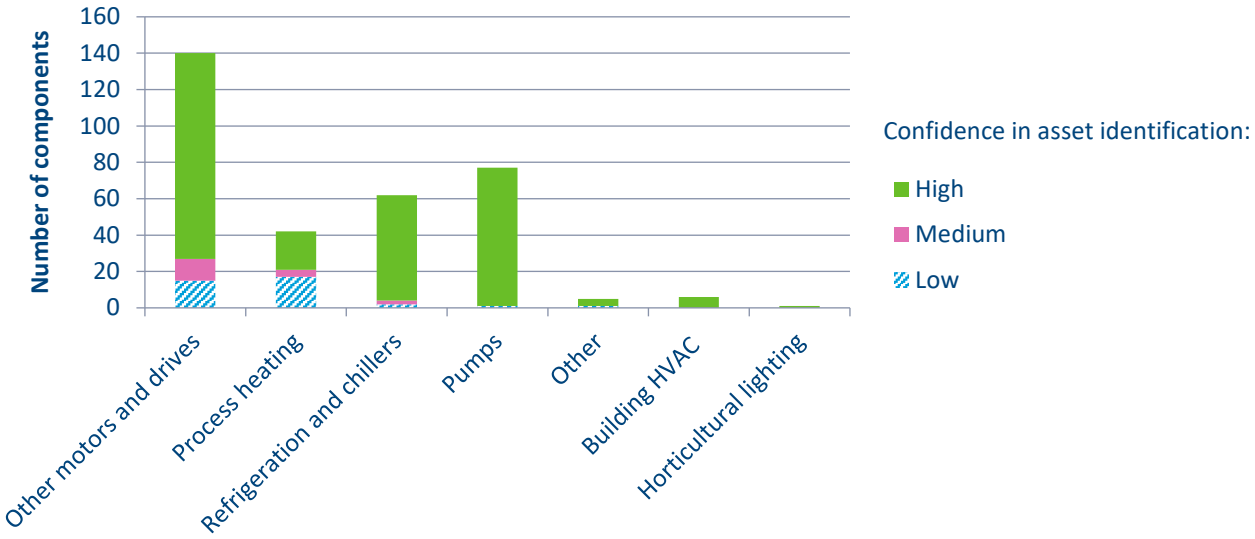


Figure A6.2: Sum of component capacities (MW) by type from DSR test data

The mean MW component capacity from the DSR Test data for the 2nd TA is 1.5MW. The median is just 200kW, reflecting the existence of some very large single components (the largest being 70MW and six others over 10MW) within the process heating and 'other motor and drives' categories.

The minimum size of CMU in the second TA was 500kW. This implies that any components beneath this size must be aggregated within a portfolio of assets to participate.

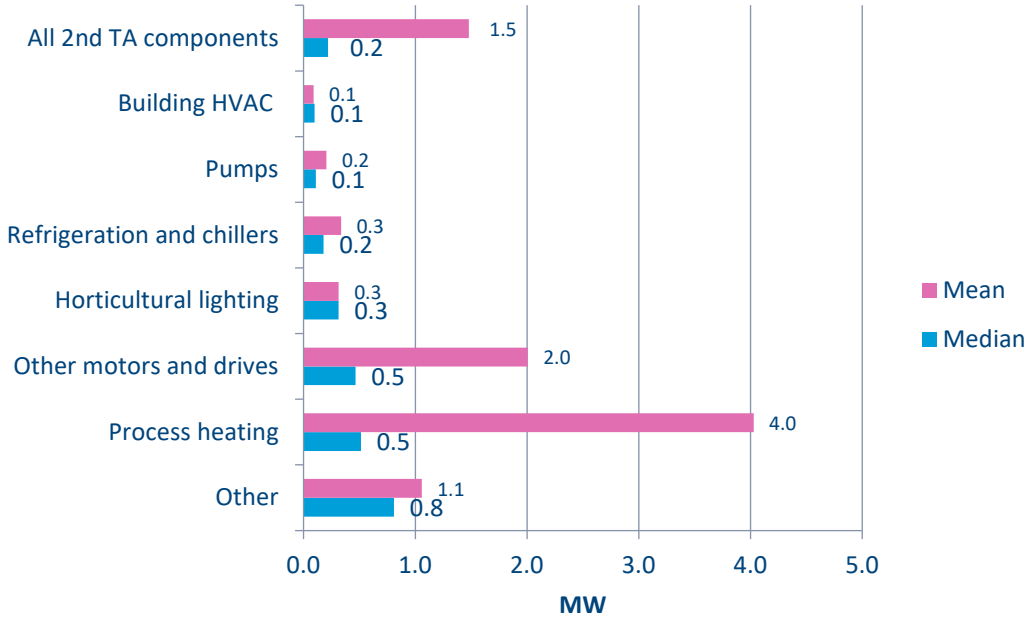


Figure A6.4: Mean and median component capacity (MW) by DSR type from DSR Test data (source: consultant's analysis of Phase 3 and 4 data)

As shown in Figure A6.4 above, refrigeration, chillers, building HVAC, pumps and horticultural lighting components have means and medians ranging between 0.1-0.3MW and thus will normally have to aggregate to achieve the 0.5MW CMU minimum size. Of the 146 components of this type in the second TA, only one, a large pump with 0.9kW proven capacity, participated in its own CMU.

Components related to industrial manufacturing (process heating, other motor and drives and 'other' industrial processes) have medians and medians of 0.5MW or more. These types are thus more likely to be present in a single component CMUs, and indeed account for the remaining five single component CMUs in the second TA.

Strategies for aggregated portfolios of DSR assets are explored further in the main report for Phase 3.

Method for analysing DSR cost data

Introduction

We collected data on the cost of capacity made available to the second TA, using data from email surveys, interviews and case studies during Phase 3 and 4 of the evaluation. To avoid over-burdening respondents, most cost data was collected at participant and CMU level, supplemented where possible with component-level data. There was limited component-level cost data because only eight out of an estimated 75 aggregator clients provided cost data during Phases 3 and 4.

Reasons for analysis approach

Analysis of costs for different types of turn-down DSR proved problematic for a number of reasons:

- Some CMUs had mixed technologies (i.e. they contained multiple components that used different types of turn-down DSR assets). So cost analysis by type of turn-down DSR required analysis of costs at component rather than CMU level.
- There was little component-level cost data and there were low numbers of components in some of the categories of DSR specified above.
- Most of the cost data was provided at TA participant or CMU level, so allocation of costs to component level involved several normalisation assumptions (effectively 'smearing' the cost data across individual components). The resulting costs per component were widely distributed and did not provide meaningful results for different technology types.

For these reasons, it was not feasible to develop meaningful costs estimates by type of turn-down DSR. An alternative approach was therefore used. This involved analysis of costs at a higher level, across all types of turn-down DSR. Because component-level cost data was very limited, the data for this high-level cost analysis was sourced from the email survey of TA participants which provided consistent data across most of the TA participants. Estimates of client costs are presented separately below, for a small sample of clients.

Methodology for high-level cost analysis

TA participants

High-level cost data was provided by seven out of the nine TA participants, comprising two direct participants and seven aggregator participants. The costs covered 19 CMUs and represented 243 MW of proven, contracted capacity (compared to the total of 28 CMUs and 293 MW of proven, contracted capacity that went forward to delivery in the second TA). The following data was collected from these participants through the email survey. Some of the surveys were completed during Phase 3 and others during Phase 4 of the research:

- Upfront metering and equipment costs (£) – data for 19 CMUs going forward to delivery, collected from TA participants at CMU-level.
- Initial staff time (hrs) – data for seven TA participants[22], provided at organisational level.

[22] There were eight responses to the email survey, but one of these was for an aggregator that provided capacity for several CMUs to a TA participant on a sub-contracting basis. The cost incurred by the sub-contracting aggregator have been included within the total costs for that TA participant.

- Ongoing staff time – annual (hrs) – data for seven TA participants, as for initial staff time.

The data was normalised by MW to make the figures more useful for BEIS policy purposes. Costs were divided by the proven, contracted capacity going forward to delivery in a given CMU, or by a given TA participant. Average costs were derived by dividing the total costs across all seven participants (or all 19 CMUs) by the total proven capacity in these CMUs (i.e. 243 MW). Proven, contracted capacity going forward to delivery was the capacity contracted post-auction for these 19 CMUs, adjusted for any reductions in proven capacity arising from DSR tests and metering tests.

Aggregator clients

High-level cost data was provided by eight out of the estimated 75 aggregator clients in the second TA in email survey responses. According to the capacity estimates provided by these aggregator clients, the cost data covered 34 MW out of the total 243 MW of capacity in the second TA. It covered 43 sites or components out of the 333 sites in the second TA. The client data was collected for the same three cost categories set out above. It was normalised using capacity estimates provided by the respondents, where available, because proven, contracted capacity was not defined at component level. Owing to the small sample size relative to the overall capacity and number of clients involved in the second TA, this data is presented separately and should be interpreted with caution.

Limitations of the cost analysis

- The costs presented below do not include the opportunity cost to businesses of turning down demand in response to stress events or tests (see later sub-section).

- As the cost data was collected from TA participants and clients by email (with some responding in autumn 2017 and others in spring 2018), it only includes costs incurred up to the time of the email survey. It does not include costs incurred during the delivery year such as the cost of meeting Satisfactory Performance Day requirements. Interview evidence suggests that there are some costs involved in the coordination of SPDs across multi-component CMUs.

- Similarly, the costs do not include any costs associated with metering requirements that arose during the delivery year (e.g. new sub-metering requirements because of renewables being installed on a particular site).

- Not all TA participants responded to the email survey – the costs for TA participants below are based on staff time data for seven TA participating organisations and capital costs for 19 CMUs, while the costs for aggregator clients are based on a small sample of eight clients. A few additional data points provided by TA participants via interview or case studies are not included because these are not available on a consistent basis.

- The costs and time estimates for TA participants do not include any capital costs or time incurred by aggregator clients, because of the lack of consistent data across different CMUs. Some evidence on the costs incurred by aggregator clients is presented separately below: the client costs should be interpreted with care as they are based on a small sample.

- The TA participant data includes a mix of direct participants and aggregators. These are not presented separately to avoid disclosure, because of the small number in each category.

- The email survey asked TA participants and aggregator clients to provide estimated costs specific to the second TA. This means that costs already incurred for participation in the first TA or other flexibility services are not included here.

- The staff time inputs include abortive development time as well as clients/sites that actually went forward to delivery. For example, they include time that aggregators spent recruiting clients for CMUs that were not contracted in the second TA auction or that were terminated post auction because of failing to pass testing requirements.

- As the data were provided in 2017/18, they do not capture later developments in the DSR market, such as the progressive trend towards DSR aggregation being offered as part of integrated energy solutions rather than as a standalone service by specialist DSR aggregators.

Normalised costs for aggregators and direct participants in second TA

Tables A6.2-A6.4 below show the average normalised costs and time inputs for participants in the second TA (both aggregators and direct participants), together with the minimum and maximum costs. The costs presented are as follows:

- 'Initial capital expenditure' is the capital cost of any metering or control equipment met by the TA participant.

- 'Initial staff time' represents staff time inputs by aggregators and direct participants, including time for marketing and for recruiting clients which varied widely between aggregators.

- 'Ongoing staff time' represents respondents' estimates of the annual costs associated with ongoing participation in the future CM. Staff time inputs have been converted to estimated costs using standard labour rates (see below).

Initial capital expenditure

Initial capital expenditure on metering equipment varied between £0/MW and £580/MW of proven, contracted capacity. The average initial capital expenditure across the dataset was £150/MW. This expenditure related solely to metering equipment and its installation. Interview evidence indicated that aggregators chose (wherever possible) to select sites that had simple metering requirements and did not require any up-front capital investment to participate in the second TA. Metering costs were only incurred for relatively large sites that could offer sufficient capacity to justify capital investment for the second TA and future CM.

Table A6.2: Initial capital expenditure specific to second TA (£/MW)

Item	Average cost (mean)	Min	Max
a. Capital expenditure on controls	0	0	0
b. Installation of controls	0	0	0
c. Capital expenditure on metering equipment	70	0	290
d. Installation of metering equipment	80	0	290
e. Other capital expenditure	0	0	0
Total initial capital expenditure	**150**	**0**	**580**

Source: email survey data from 19 CMUs, normalised by proven, contracted capacity

Caveat: The indicative costs quoted were calculated by the evaluation contractors using self-reported costs from a small sample of participants in the second Transitional Arrangements auction held in 2017. They may not be representative of all the TA DSR participants or of all the costs faced by DSR market participants more generally. They are likely to be under-estimated.

Initial staff time inputs

The staff time required to put forward capacity specifically to the second TA varied between zero and 52 days per MW of proven, contracted capacity. The average staff time inputs were 13 days per MW, equivalent to £4,600 per MW at standard labour rates. Most of this was associated with aggregators signing up clients and organising testing of client capacity. Some aggregators spent less time on client recruitment because they brought in their existing electricity supply clients or flexibility customers, but significant time inputs were still required to meet TA/CM testing processes. Time inputs by direct participants were much lower, partly because they had fewer sites and partly because they did not need to meet testing requirements, having proven their

capacity in the first TA. Although all the TA participants required some time to participate in the second TA, the minimum figures rounded down to zero when normalised by capacity.

Table A6.3: Initial staff time inputs specific to second TA (full time equivalent days per MW)

Item	Average days (mean)	Min	Max
a. staff time for marketing to clients/internal sites	1	0	10
b. staff time for signing up clients	7	0	16
c. staff time associated with testing	4	0	23
d. staff time for pre-qualification & participation in auction	1	0	11
e. Other staff time for 2nd TA	0	0	0
Total initial staff time	**13**	**0**	**52**
Estimated initial staff cost (£ per MW)[23]	**£4,800**	**£0**	**£19,300**

Source: email survey data from seven TA participants, normalised by proven, contracted capacity.

Caveat: The indicative costs quoted were calculated by the evaluation contractors using self-reported costs from a small sample of participants in the second Transitional Arrangements auction held in 2017. They may not be representative of all the TA DSR participants or of all the costs faced by DSR market participants more generally. They are likely to be under-estimated.

Ongoing staff time inputs

The predicted time inputs required to participate in the CM on an ongoing basis were lower, because they excluded client recruitment time. The average staff inputs were predicted to be 7 days per MW (equivalent to £2,600 per MW at standard labour rates). Staff inputs varied between zero and 21 days per MW. Again, all the TA participants predicted some time requirements, but the minimum figures rounded down to zero when normalised by capacity. These predicted costs were dominated by ongoing client/site engagement and by the annual auction process. Some testing requirements were still envisaged because of potential changes to the composition of multi-site CMUs, which could necessitate retesting of some sites or CMUs.

[23] Staff time has been converted into costs by using the latest available (2012) Green Book labour rate of £336/day, assuming 5 years of wage inflation at 2%, giving a day rate of £371/day. Rounded to nearest £100.

Table A6.4: Ongoing staff time inputs for CM participation (full time equivalent days per MW)

Item	Average days (mean)	Min	Max
a. client/internal site engagement	3	0	8
b. adjustments to CMU composition (if required)	0	0	1
c. further testing/other compliance costs	1	0	5
d. annual auction process (e.g. future T-4 or T-1)	3	0	13
e. any other ongoing time inputs (e.g. data flow issues)	0	0	0
Total ongoing staff time	**7**	**0**	**21**
Estimated ongoing staff cost per year (£ per MW per year) [24]	**£2,600**	**£0**	**£7,800**

Source: email survey data from seven TA participants, normalised by proven, contracted capacity.

Caveat: The indicative costs quoted were calculated by the evaluation contractors using self-reported costs from a small sample of participants in the second Transitional Arrangements auction held in 2017. They may not be representative of all the TA DSR participants or of all the costs faced by DSR market participants more generally. They are likely to be under-estimated.

Normalised costs for aggregator clients

We have analysed the costs incurred by aggregator clients separately. This data should be interpreted with care as we only have cost data for eight clients out of the estimated population of 75 clients in the second TA. These costs are additional to the TA participant costs presented in Tables A6.2-A6.4 above. Table A6.5 below shows the average cost and time inputs for this small sample of aggregator clients in the second TA, together with the minimum and maximum costs reported. Staff time inputs have been converted to estimated costs using standard

[24] Staff time has been converted into costs by using the latest available (2012) Green Book labour rate of £336/day, assuming 5 years of wage inflation at 2%, giving a day rate of £371/day. Rounded to nearest £100.

labour rates[25]. Costs have been normalised by MW of capacity, using capacity estimates provided in email survey responses (where available)[26]. The costs presented are as follows:

- 'Initial capital expenditure' is the capital cost of any metering or control equipment met by the client.
- 'Initial staff time' represents staff time inputs by aggregator clients, including set-up time with the aggregator, internal marketing and approvals, liaison with the aggregator during pre-qualification and auction processes and coordination of metering and testing processes up to the start of the delivery year.
- 'Ongoing staff time' represents aggregator clients' estimates of the annual costs associated with ongoing participation in the future CM.

Six of the eight clients reported no initial capital costs because metering and control equipment costs were met by aggregators, but two large client sites (2MW or higher) reported significant expenditure on control or metering equipment. The maximum capital cost per MW appears higher than the maximum cost presented in Table A6.2 because there was less averaging of high costs for these sites with zero costs for other sites. Average staff time inputs were 2 days per MW, lower than those incurred by TA participants. Low initial staff costs were generally reported by clients already active in the first TA, but higher staff inputs were reported by those aggregator clients with large numbers of smaller sites. Staff time inputs were dominated by the initial set-up of the aggregator contract, internal liaison and coordination of metering/testing processes.

Table A6.5: Costs reported by aggregator clients (source: email survey data, normalised by capacity)

Item	Average (mean)	Min	Max
Initial capital cost (£ per MW)	£200	£0	£1,800
Initial staff time (days per MW)	2	0[27]	17
Ongoing staff time (days per MW per year)	2	0	19
Estimated initial staff cost (£ per MW)	£740	£0	£6,300
Estimated ongoing staff cost per year (£ per MW per year)	£740	£0	£7,000

Caveat: The indicative costs quoted were calculated by the evaluation contractors using self-reported costs from a small sample of aggregator clients in the second Transitional Arrangements auction held in 2017. They may not be representative of all the TA aggregator clients or of all the costs faced by DSR market participants more generally. They are likely to be under-estimated.

[25] Staff time has been converted into costs by using the latest available (2012) Green Book labour rate of £336/day, assuming 5 years of wage inflation at 2%, giving a day rate of £371/day. Rounded to nearest £100.

[26] Capacity estimates were provided by seven out of the eight clients that provided cost data. For the remaining client that did not provide a capacity estimate, we used the average capacity per component, calculated using the proven, contracted capacity for the relevant CMU.

[27] Rounded down to zero.

Estimates of opportunity costs

Opportunity costs were explored through qualitative interviews and the case studies. Where possible the research team tried to obtain quantitative data on the opportunity cost, such as underlying staff costs, additional energy costs and production loss. Production loss can lead to damaged products, lost sales and damage to the core business, with significant potential costs for the participants.

The contexts for providing turn-down capacity are very wide, as demonstrated by the broad range of business activities present in the TA. In general participants and DSR asset owners were unable or unwilling to provide specific cost figures. As a result there is limited quantitative data.

The proportion of those that did not know, or would not disclose, their opportunity costs has not explicitly been researched. However our general impression from reviewing interview data and the case studies is that participants rarely consider opportunity costs in purely numerical terms and assessment is often based on judgement. This raises the question of how they make bidding decisions in the auction.

Direct participants had sophisticated DSR strategies and fall into the category of participants that know their opportunity costs, but do not always wish to disclose full details. There were only three participants of this type.

Aggregators do not need to understand the minutiae of their client's operations and opportunity costs, only the commercial terms that have been agreed and their judgement of the delivery risks (or an expectation of what they can sell to clients, given that the auction may proceed signing clients up). Only specific aggregator staff are likely to have a good understanding of a particular client's circumstances, hence aggregator interviews did not necessarily yield good site-level information.

Aggregator clients themselves do not participate directly in the auction. If they are in negotiation with an aggregator pre-auction, then they may agree a floor price for participation. More commonly, if the auction has taken place or there is an expectation of the likely price, then it is a 'take it or leave it' offer and in-depth interrogation of the opportunity costs is not required. The client always reserves the right to curtail or decline a turn-down instruction based on a range of operational factors and the penalties for doing so are not great.

The above provides some explanation as to why quantitative information on opportunity costs are illusive. Nonetheless it is possible to draw broad conclusions which allow a generic framework for considering opportunity costs to be proposed. The remainder of this section sets out our conclusions in this area, principally based on the Phase 4 case studies.

We can consider the opportunity costs of providing turn-down DSR in four generic stages, principally based on the duration of the turn-down event. The variation of opportunity costs by these four generic stages is shown in Table A6.6.

Table A6.6: Generic stages of opportunity cost by event duration (source: case study analysis)

Stage	Typical duration	DSR service offered	Opportunity cost	Typical cost factors
1. Standard response	< 2 hours	Yes	Negligible – Low	Management time
2. Extended / full response	2 to 4 hours	Yes, but most hope it won't happen	Negligible to £13k/MW/hr* (base on 5 CMUs)	As above + energy and staff overheads, service disruption or minor production loss.
3. Long duration	4+ hours	Some can, but for many this would be problematic	High £10+k	As above + temporary production loss / service interruption
4. Very long duration	8+ hours	No	Significant business impact	As above + business service disruption

The duration at which a DSR provider will move to the next stage is typically case specific and depends on multiple variables, such as the weather, time of day, the season, levels of product stock and business activity/production orders. These variables have an impact not only on the day of a stress event (or test) but in the preceding period that makes up the baseline. The preparation time to turn down can also have a big effect. With adequate warning, several of the case study providers could make provisions to coincide turn-down events with scheduled maintenance or to increase stock levels (water, aggregate, paper…etc.) to negate or at least mitigate the impact of a loss of production.

Reliability and opportunity costs are closely related, as DSR providers will cease to provide a turn-down response once cost and risk exceed a certain threshold. Making this judgement is usually the role of an operator, although automated thresholds may be used to trigger the end of stages 1 or 2, for example through temperature limits for thermals assets (cold stores, building HVAC, and process heating) and water levels for pumping.

One case study participant was only able to participate in the TA through an aggregator. Their opportunity costs were too unpredictable – due to uncontrollable issues such as weather and baseline issues - to provide capacity with certainty, so for the revenues available through the TA they could only participate when delivery risks were mitigated through aggregation.

Estimated revenues from other flexibility services

Cost avoidance data was provided for 9 CMUs for Triad, 2 CMUs for DUoS (Red zone) avoidance and 1 CMU for Frequency Response.

Normalisation of these revenues by MW for general comparison against TA revenues is not possible, as the revenue data is typically provided at a participant or site level. For example, the Triad revenues may refer to the cost avoidance for a company (consumption unknown), while the TA revenue is associated with DSR of individual pieces of equipment.

Comparison of revenues with other DSR services is complex. The payment structures vary between DSR options, with availability and utilisation payments for balancing services such as STOR and Firm Frequency Response (FFR), but cost avoidance for TRIAD and DUoS charges. A full analysis of this is beyond the scope of this evaluation, however £/kW values of the CM versus Triad and Dynamic FFR are shown in Table A6.7 below.

Table A6.7: £/kW value of common DSR options (source: email survey and interview data)

	£/kW		Comment
	Low	High	
Capacity Market	6.95	45	Low case is for the 2017/18 T-1 auction and high is 2nd TA
Triad cost avoidance	26	54	2017 rates as published by National Grid; varies by region.
Dynamic FFR	40	80	Indicative values only, payments based on availability and utilisation

Dynamic FFR is the most lucrative service on a £/kW basis. It is however a demanding service to deliver and requires the installation of specialist equipment and acceptance of a degree of automated / third party asset control. Triad cost avoidance generally offers a slightly greater £/kW value than the Capacity Market and is much more accessible; however it requires more frequent responses (circa 20-30 each winter season) to ensure that the maximum cost avoidance is achieved.

Stacking of CM revenues with FFR and Triad cost avoidance is common. Case studies and interviews indicated that participants' approaches to Triad in relation to CM capacity varied: some set their declared CM capacity taking account of their likely Triad response (to avoid CM baseline issues), while others (typically aggregator clients with a less sophisticated DSR strategy) did not appear to take into account the possible interactions.

This publication available from www.gov.uk/beis

Contact us if you have any enquiries about this publication, including requests for alternative formats, at:

Department for Business, Energy and Industrial Strategy
1 Victoria Street
London SW1H 0ET
Tel: 020 7215 5000

Email: enquiries@bis.gsi.gov.uk