

Testing the strength of impact evidence – applying contribution tracing within a realist evaluation of Transitional Arrangements for demand-side response

Mary Anderson, CAG

Tajbee Ahmed, BEIS

Barbara Befani, University of Surrey

Charles Michaelis, SDS

What were the Transitional Arrangements for demand-side response?

A pilot scheme (2016-2018) to encourage Demand-Side Response (DSR) to participate in the Capacity Market for electricity

First TA auction held January 2016, awarding contracts for 803 MW (approx. £22 million).

Second TA auction held March 2017, awarding contracts for 313 MW (approx. £14 million).

Successful bidders obtained CM obligations for the following delivery year (2016/17 or 2017/18)

First TA: 24 participants, both DSR and small-scale generation

Second TA: 11 participants, turn-down DSR only – only 9 proceeded to delivery (subset of first TA participants)

Participants could put forward their own capacity (direct participants) or aggregate capacity from others (aggregators)

For further details on the evaluation as a whole, see published report:

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

We tested two hypotheses about the impact of the second TA scheme to its objectives:

Hypothesis 1: The second TA contributes to more (competitive) **flexible capacity for the Capacity Market** in 2018/19 and subsequent years

Hypothesis 2: The second TA contributes to wider encouragement of **turn-down demand-side response**



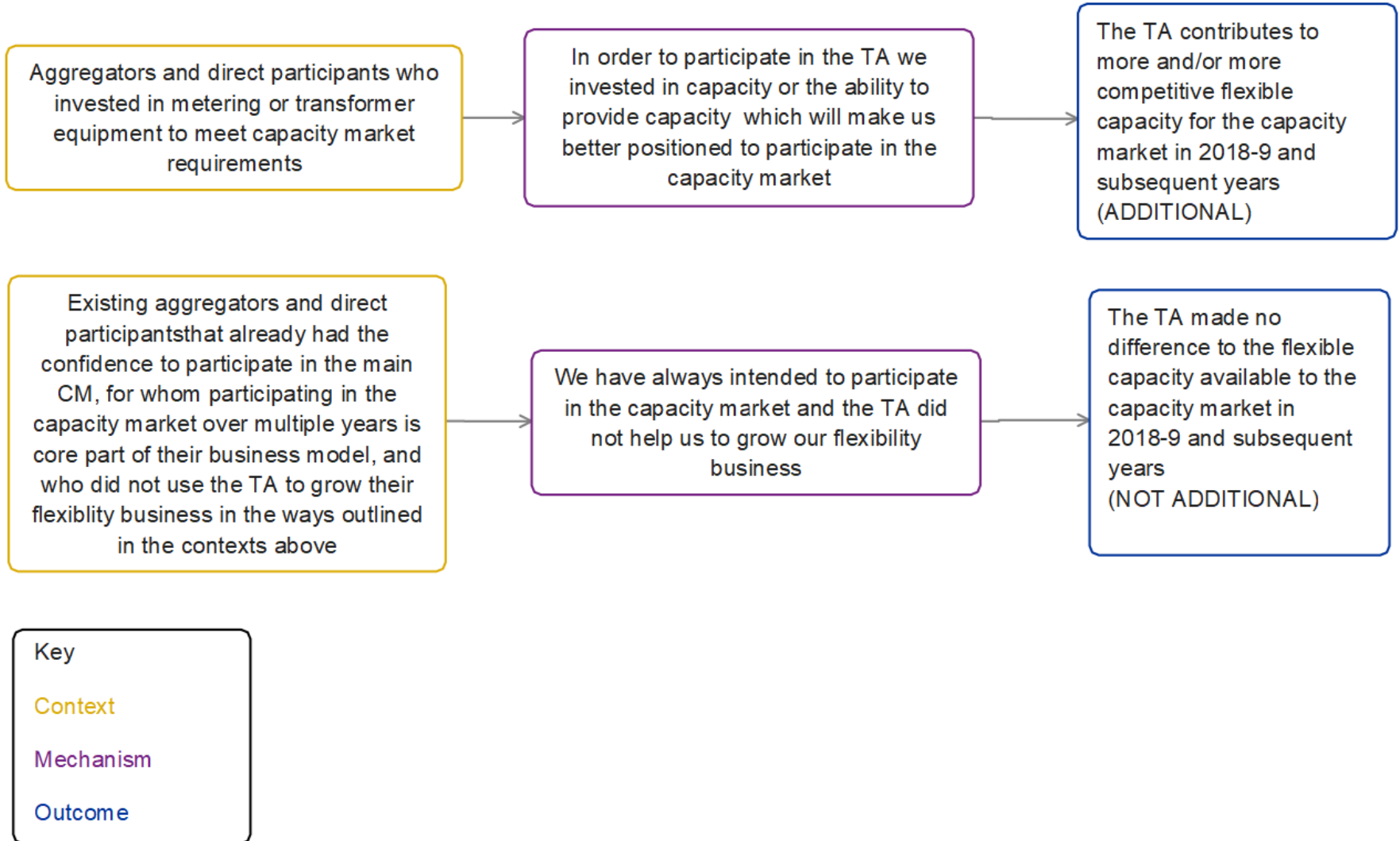
Why did we use process/contribution tracing in our assessment of the impact of the second TA scheme?

Complex area of energy policy managed by BEIS, UK

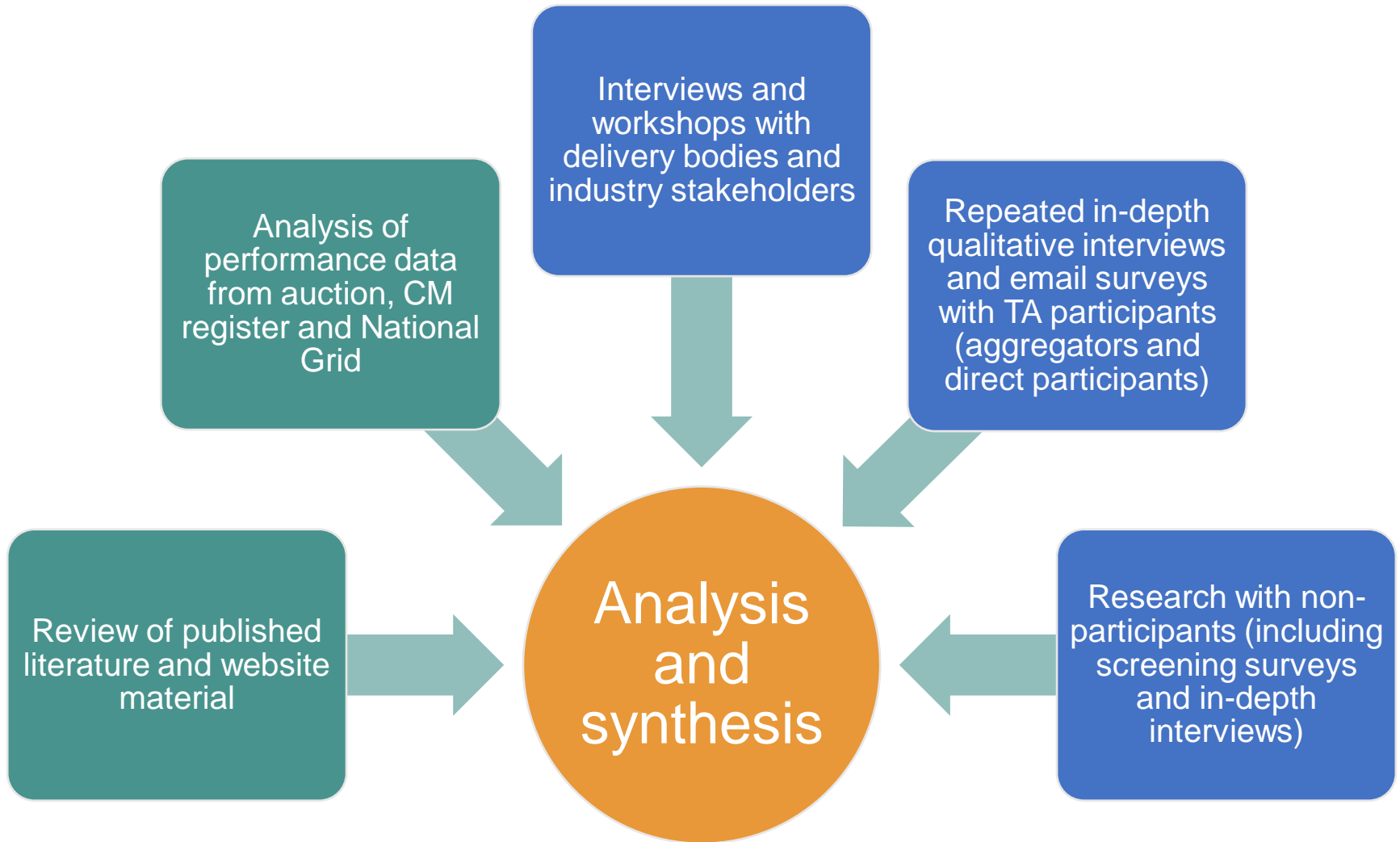
A theory-based realist evaluation

Small cohort of TA participants and scope for lobbying bias

Based on our findings from evaluation of the first TA, we developed plausible C-M-Os for our additionality hypotheses (H1 and H2). For example:



To assess the hypotheses, we gathered and then analysed primary and secondary evidence:



We then used process/contribution tracing to test the evidence in a robust way, using four types of evidence tests:

- **Hoop tests** – reject the hypothesis if this evidence is not found but not sufficient to confirm the hypothesis if found; these are pieces of evidence that we would *expect to see* if the given hypothesis is true (i.e. necessary but not sufficient)
- **Smoking gun** – *confirms* the hypothesis if observed but does not reject the hypothesis if not observed; these are pieces of evidence that we would *like to see* (i.e. sufficient but not necessary)
- **Doubly-decisive** – confirms the hypothesis if observed and if not observed the hypothesis is rejected; these are pieces of evidence that are *expected* but are also *confirmatory* of the hypothesis (i.e. both necessary and sufficient)
- **Straw-in-the-Wind** – not sufficient to confirm the hypothesis if observed nor to reject the hypothesis if not observed (i.e. neither necessary nor sufficient in itself)

Reference: Befani, D'Errico, Booker, Giuliani. (IIED Briefing, April 2016) 'Clearing the fog: new tools for improving of impact claims'

We developed evidence tests for each of the C-M-Os in the H1 and H2 theory. Example from H1 theory:

Evidence tests for elements and linkages	Source of evidence	Type of test	Competing explanations	Rationale for classification of test
H1 – OUTCOME 1-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.
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	In-depth 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.
H1 – CONTEXT 1.2.1 - test (i): details of significant investment in metering or control assets 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	Like to see (straw in 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	Specific details in email survey provide more confidence than test (h) but there's still a possibility that controls will really be used for other flexibility services, not the CM.
H1 – CONTEXT 1.2.1 -test (j): 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).

We analysed the evidence for each TA participant and combined the test results to assess the strength of support for each C-M-O in that 'case' using the following system:

KEY:	How we combined the underlying test results
Strong support	EITHER all tests passed, irrespective of strength. OR At least one 'smoking gun' or 'double-decisive' test passed. No 'hoop tests' failed, but allow failure of some 'straw in the wind' tests.
Support	No 'hoop tests' failed, but allow failure of some 'straw in the wind' tests. No 'smoking gun' or 'double-decisive' test passed.
Mixed support	Mix of 'hoop' test failures and 'smoking gun' or 'double-decisive' tests being passed.
No support	At least one 'hoop' test failed. No 'smoking gun' or 'double-decisive' tests passed.

Support for additional outcomes (H1):

		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	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	Support	No support	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	Strong 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

Findings on H1: the second TA contributed to more (competitive) flexible capacity for the Capacity Market in 2018/19 and subsequent years

- **Strong support** for additionality of the second TA amongst aggregators (tested using process tracing)
 - Aggregators attributed growth in their portfolios or CM knowledge to the second TA (e.g. because TA made CM seem less risky, or clients invested time/money that would make them better positioned to participate in the main CM)
 - For new aggregators, the second TA helped them to build a customer base for DSR
 - For existing aggregators, the second TA enabled them to bring new turn-down clients on board
- **But no support** for additionality of second TA amongst direct participants – clear cut from evidence, without need for sophisticated process tracing

Findings on H2: the second TA contributed to encouragement of more turn-down DSR

- **Strong support** for additionality outcomes amongst aggregators
 - Existing aggregators already intended to increase turn-down DSR offer but the second TA helped aggregators and their client to assess the risks and opportunity costs involved
 - New aggregators were encouraged by the first and second TA to enter the market for aggregation of turn-down DSR. The TA gave them something to talk to new clients about and the high price made it cost-effective to bring new assets to market.
- **Less consistent support** for additionality amongst aggregator clients – some were encouraged to enter the market for flexibility services via an aggregator and pilot their involvement at low risk
- **No evidence of additionality** for direct participants offering turn-down DSR – clear cut from evidence, without need for sophisticated process tracing

What value did process tracing add to this evaluation?

Process tracing concepts were useful for assessing and weighing evidence according to its likely reliability (e.g. taking into account potential lobbying bias - “they would say that, wouldn’t they?”)

The analysis process reminded us to consider alternative explanations for observed evidence

Evidence test findings lent themselves to presentation in visual form and allowed non-disclosive presentation of findings about the strength of evidence

Process tracing approach facilitated combination of evidence from different sources (e.g. objective and subjective, qualitative and quantitative)

But the development of evidence tests in collaboration with policy/technical experts took considerable time and budget

Thank you

Contact: ma@cagconsult.co.uk



www.cagconsultants.co.uk