Department for
Energy Security
& Net Zero

Reviewing energy supplier evidence on impacts of smart metering on domestic energy consumption

An Independent Review completed by the Behavioural Insights Team

Authors: Edwin Chan, Tim Hardy, Tom O'Keeffe, Toby Park, Andrew Schein
(Behavioural Insights Team)

June 2023

© Crown copyright 2023

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit nationalarchives.gov.uk/doc/open-government-licence/version/3 or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gsi.gov.uk.

Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned.

Any enquiries regarding this publication should be sent to us at: [add team email address]

Executive Summary

Background

Smart meters are replacing traditional gas and electricity meters across Great Britain, with the goal of making the energy system more efficient and flexible, helping Great Britain use more renewable energy, and delivering net zero greenhouse gas emissions by 2050. At the end of 2022, there were 31.3 million smart and advanced meters in Great Britain in homes and small businesses, representing 55% of all meters. A key principle of the roll-out is for consumers to be able to use smart meter data to gain a better understanding of their energy consumption and, where desirable, reduce it to save money and minimise carbon emissions. The Government anticipates that smart meters will average energy consumption reductions of 3.0% for electricity and 2.2% for gas (0.5% for gas prepay),[1] driven by multiple behavioural mechanisms:

1. Direct feedback on consumption via In-Home Displays (IHDs) - offered at no extra cost during the smart meter installation
2. Indirect feedback via more accurate billing
3. Engagement and advice provided before the installation, during (e.g. on how to use the IHD and how to save energy) and after (e.g. Smart Energy GB's communication campaigns and post-install communications from energy suppliers).

Since its inception in 2011, the Smart Metering Implementation Programme (SMIP) in the Department for Energy Security and Net Zero has been monitoring and evaluating its impacts on household energy consumption. During the early stage of the roll-out, SMIP focused on learning from early installations to understand how best to deliver the anticipated consumer benefits. This was the aim of the Smart Metering Early Learning Project (ELP), a programme of social research that also included statistical analysis of consumption reductions from early installations. That analysis found that compared to traditional meters, smart-type meters enabled an average annual reduction of 2.3% of domestic customers' electricity consumption and 1.5% for their gas consumption.[2] Based on these findings, international evidence on the efficacy of consumption feedback in driving energy savings, and the fact that several of the consumer engagement policies had not yet been introduced, the ELP concluded that it would be reasonable to expect average savings of 3%.

As the roll-out has progressed to the main installation stage, monitoring of these benefits has continued, mainly via analysis conducted by energy suppliers on the impacts of smart metering on their customers' energy consumption. In 2019, SMIP completed an internal review of the energy supplier analyses and found that their results aligned with the ELP's findings. However, they also concluded that a more comprehensive set of supplier analyses would help draw robust conclusions about market-wide impacts. SMIP consequently commissioned the Behavioural Insights Team (BIT)[3], to work with energy suppliers to produce tailored guidance on conducting this analysis.[4]

[1] BEIS, Smart Metering Implementation Programme: a report on progress of the realisation of smart meter consumer benefits (September 2019): https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/830668/smart-meters-benefits-realisation.pdf

[2] Department of Energy and Climate Change. (2015.) Smart Metering Early Learning Project. https://www.gov.uk/government/publications/smart-metering-early-learning-project-and-small-scale-behaviour-trials

[3] BIT is a social-purpose consultancy, now wholly owned by Nesta. It generates and applies behavioural insights and conducts evaluations to inform policy and improve products, services, and policies.

[4] The Behavioural Insights Team. (2020.) Guidance on conducting energy consumption analysis https://www.bi.team/publications/guidance-on-conducting-energy-consumption-analysis/

In January 2022, the Department for Energy Security and Net Zero commissioned us, the Behavioural Insights Team (BIT), to complete a more comprehensive evidence review, gathering the latest analyses from the ten largest domestic energy suppliers[5] to understand what conclusions could be drawn from their studies and help SMIP draw conclusions about the energy savings attributable to the roll-out. This report summarises that work.

Methodology

Our first step was to assess the rigour of the evidence submitted by energy suppliers. To do this, we (i) completed an initial review of the evidence provided by the energy suppliers to decide which studies were robust enough to warrant further exploration in methodological review meetings, (ii) conducted these methodological review meetings, (iii) completed a formal review of each study, and (iv) based on these reviews, identified the studies to include in a formal evidence synthesis.

The suppliers' studies attempt to identify the impact of smart meters on energy consumption by comparing smart-meter customers' energy consumption to that of traditional-meter customers. The general approach is to find a suitable group of customers who did not install a smart meter and compare their energy consumption to that of smart-meter customers; and/or compare smart-meter customers' *change* in energy consumption to traditional-meter customers' *change* in energy consumption over a similar time period. Suppliers looked at the impact of smart meters on energy consumption within 12 months of installation; a small number of suppliers also investigated the impact on a longer horizon.

Results

The first phase of this project involved reviewing 14 analyses from seven suppliers.[6] Each analysis estimated the effect of installing a smart meter on domestic customers' energy consumption in the first year after installation, expressed in percentage change in consumption before and after installation. Based on our reviews, we concluded that seven of these studies (from four suppliers) were sufficiently rigorous to include in an evidence synthesis: four studies from **Supplier A**, and one study each from **Supplier B**, **C**, and **D**.[7] We undertook a formal meta-analysis to synthesise the studies' energy consumption impacts. This evidence synthesis shows a reduction in energy consumption from smart meters, summarised below. Each study's square is sized in proportion to its weighting in the pooled estimate, and the orange whiskers around each study (and the pooled estimate) represent the 95% confidence intervals. We discuss these plots in more depth in Section 4.

[5] As of Autumn 2022.

[6] Beyond these 14 studies, we excluded various Supplier A analyses due to Supplier A's inability to retrieve key analysis details, particularly confidence intervals around the smart meter impact estimate. See Appendix A2 for a full list.

[7] We recommended excluding one study from Supplier A, two studies from Supplier C, the study by Supplier E, both studies from Supplier F and the study by Supplier G. This was because these studies either (i) had study designs subject to excessively severe threats to internal validity, external validity, and/or precision and rigour, or (ii) did not provide standard errors to quantify the uncertainty of their estimates. We show the included *and* excluded studies results in Appendix A3.

Figure 1: Forest plot for the impact of smart meters on electricity consumption

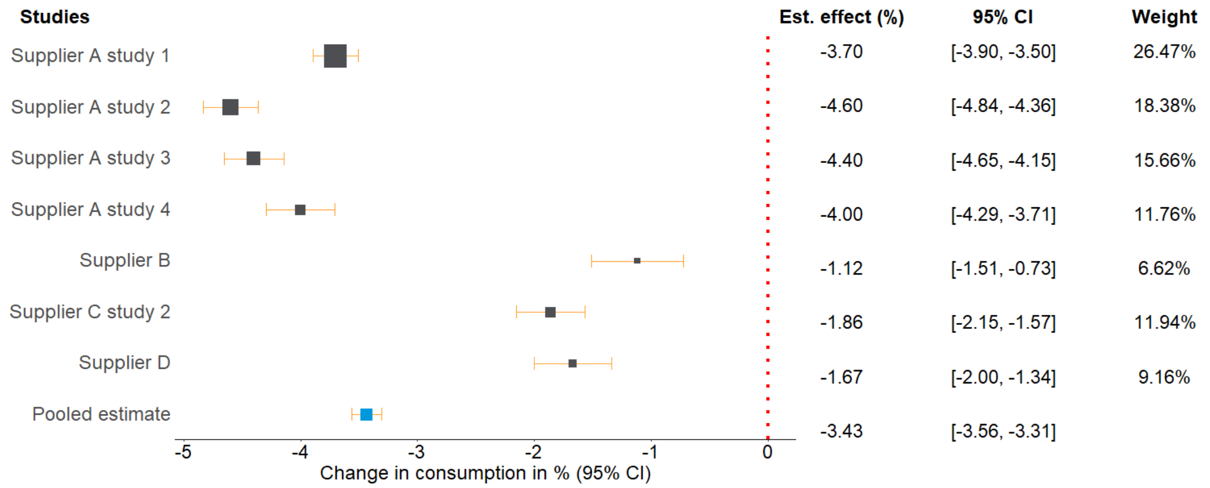
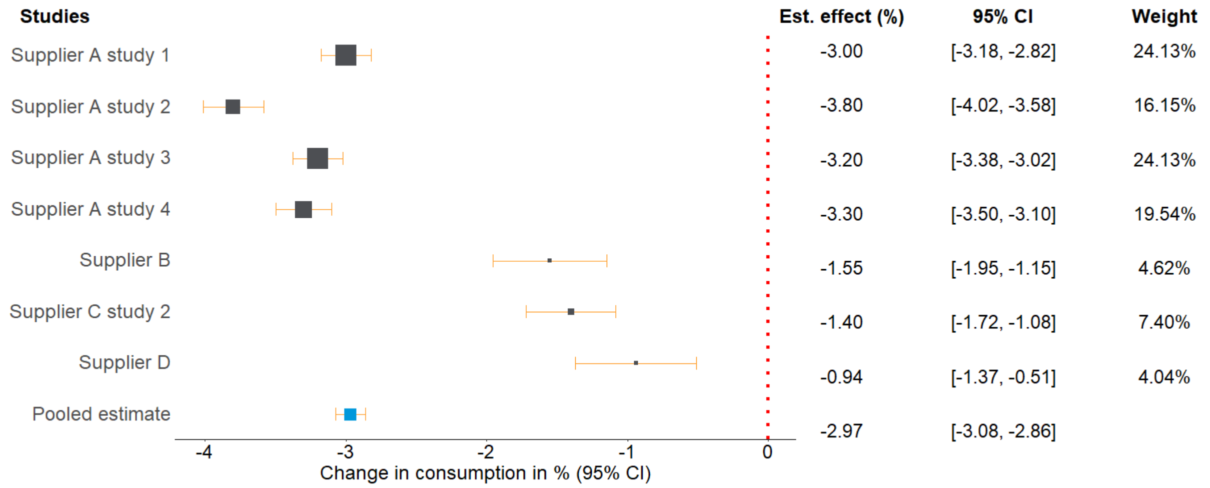


Figure 2: Forest plot for the impact of smart meters on gas consumption



The pooled estimates are -3.43% for electricity and -2.97% for gas. The estimates have narrow 95% confidence intervals of [-3.56%, -3.31%] for electricity and [-3.08%, -2.86%] for gas. Both estimates are statistically significantly different from 0 ($p<0.001$). These estimates are higher than the effects of -3.0% (electricity) and -2.2% (gas) that the Department for Energy Security and Net Zero anticipated based on the evidence available before this review.

There is variation in the estimated effects from smart meters across suppliers:

- For electricity, the studies from Supplier A (from 2017 Q2 through 2018 Q1 installation windows) show electricity consumption reductions of 3.70% to 4.40%. The Supplier B, Supplier C, and Supplier D impact estimates are lower, finding reductions between 1.12% to 1.86%.
- The picture is similar for analyses of gas consumption. The studies from Supplier A show gas consumption reductions of 3.00% to 3.80%. The Supplier B, Supplier C, and Supplier D impact studies found reductions between 0.94% to 1.55%.

It was not possible to draw strong conclusions about the reasons for this variation in estimated impacts in this review. It is possible that smart meters from Supplier A drove a greater energy consumption reduction than smart meters from other suppliers. This could be due to differences in smart-meter roll-out and installation strategies between suppliers (e.g. the comparative maturity of their roll-out at the time of their installations, the proportion of customers receiving an In-Home-Display, the quality of IHDs provided and/or the energy efficiency advice customers received during the installation). (Note that the Department for Energy Security and Net Zero *does* assume some variation in smart meter impact between suppliers.[8]) However, it is also possible that the differences in estimated effects reflect differences in the suppliers' sample for the analysis, analysis strategies, and/or customer base.

Our pre-specified primary analysis treats the Supplier A studies as separate when conducting the meta-analysis. If we treat Supplier A's studies as a single study, the estimated impact of smart meters on consumption changes from -3.43% to -2.61% for electricity and from -2.97% to -2.43% for gas. The estimates from this robustness check are both still statistically significantly different from 0 ($p<0.001$) and roughly in line with the reductions the Department for Energy Security and Net Zero anticipated.

Strength of evidence overall

All of the studies use quasi-experimental methods that rely on methods such as matching and difference-in-differences that 'build' a suitable comparison group (rather than using sources of random or quasi-random variation, such as a natural experiment). The quality of these methods depends on the availability of data. While many studies used data on historical consumption and geographic location for matching, some important factors such as environmental attitudes are missing. In their samples, customers chose to get a smart meter. These households will be different from the 'general population' of domestic energy customers in a variety of ways that may bias the estimated effects of smart meters for this group of customers. Even if estimates from the studies were perfectly accurate, they would not represent the average effects from rolling out smart meters to the 'general population'. Nevertheless, we believe that the included studies have:

- **Moderate-to-high internal validity**. Internal validity refers to whether an analysis produces an unbiased estimate of the effect of smart meters for customers in the study sample. This means that estimates only reflect the true impact of smart meters (and some random noise due to the fact that sample sizes are finite). All included studies use difference-in-differences or a similar approach. These designs reduce biases from differences between smart- and traditional-meter customers, particularly from characteristics which do not vary over time. However, there is a meaningful risk in each study that trends in characteristics differ between groups over time (e.g. the smart-meter group is becoming more environmentally-friendly on average than the traditional-meter group prior to installation), which biases the estimated impacts. We also have concerns about specific methodological choices for all the studies we have included (discussed in Appendix A3). Nevertheless, we think that they all have moderate-to-high internal validity.
- **Moderate external validity**. This refers to how well the study identifies an effect that is generalisable to other customers. The studies' samples are different to the general population in a variety of ways. For example, the requirement for reliable consumption data before and after the installation means that suppliers' study samples are less likely to include customers who have switched to a new supplier, moved to a new home, or do not regularly provide meter readings. In addition, later adopters are naturally underrepresented in most of the studies we included in the evidence synthesis. That said, it is worth noting that even though the included

[8] In its cost-benefit analysis, the Department for Energy Security and Net Zero assumes that ⅔ of customers will experience smart meter impacts of 2.8% electricity reduction (for all customers) and 2.0% gas reduction (for credit customers), but that ⅓ will experience greater reductions, due to 'more mature and sophisticated engagement approaches', leading to the average figures of 3.0% electricity reduction (credit and prepayment) and 2.2% gas reduction (for credit customers; 0.5% gas reduction for prepayment customers).
Department for Business, Energy and Industrial Strategy. (2019). Smart meter roll-out: cost-benefit analysis. https://www.gov.uk/government/publications/smart-meter-roll-out-cost-benefit-analysis-2019

studies are taken from a range of different suppliers and years, all of them find significant savings from smart meters. This suggests that the smart meters reduce energy consumption for different compositions of customers.

- **Strong precision and rigour**. Although some studies had suboptimal analysis choices (perhaps related to the limited resources that suppliers can generally invest in their smart meter impact analyses), they nevertheless have large sample sizes and straightforward, rigorous analysis strategies and outcome measures.

We find reductions in electricity consumption of 3.3% to 3.6% and gas consumption of 2.9% to 3.1% from the meta-analysis. We believe these analyses and this estimate provide useful evidence for the Department for Energy Security and Net Zero for assessing the overall impact of the smart meter roll-out on household energy consumption.

Contents

1. Policy background and research rationale

Great Britain's smart meter roll-out is expected to drive reductions in household energy consumption through multiple behavioural mechanisms:

1. Direct feedback on consumption via IHDs that are offered at no extra cost during smart meter installation;
2. Indirect feedback via more accurate billing; and
3. Engagement and advice provided before, during (e.g. on how to use the IHD and how to save energy) and after (e.g. Smart Energy GB's communications campaigns and post-install communications from energy suppliers) the installation.

Based on international and domestic evidence, the Department for Energy Security and Net Zero anticipates the roll-out will result in average savings of 3.0% in annual electricity consumption (credit and prepayment) and 2.2% (credit) / 0.5% (prepayment) in annual gas consumption for households, totalling £4.7 billion of savings to domestic consumers through to 2034.[9] With such significant investment and the importance of these savings to the roll-out's business case, the Department for Energy Security and Net Zero views it as important to update the evidence base on these savings as the roll-out progresses.

The Smart Metering Implementation Programme (SMIP) has been monitoring and evaluating the impacts of the roll-out on household energy consumption since the Programme's inception in 2011. However, some of the most important evidence for the estimated consumption reductions from smart meters is now somewhat out of date. The Energy Demand Research Project from 2006-2010 and the Early Learning Project analysis undertaken by SMIP in 2015 predate the modern smart meter roll-out – consumers were not given energy efficiency advice at installation, the IHDs were rudimentary, and it is possible the consumers themselves, as 'early adopters', were not representative of the wider population.

In 2019, BEIS reviewed analysis conducted by energy suppliers on the impacts of smart metering on their customers' energy consumption, which included more recent installations. That review supported earlier positive evidence but recognised that a more comprehensive set of supplier analyses would help draw robust conclusions about market-wide impacts. In January 2022, the Department for Energy Security and Net Zero therefore commissioned the Behavioural Insights Team (BIT)[10] to complete an evidence review, gathering the latest evidence from a larger number of energy suppliers to understand what this evidence suggests about the energy savings attributable to the roll-out.

Specifically, the aims of this project were to:

1. Synthesise evidence from energy suppliers about the impact of smart meters on household energy use and conduct a formal meta-analysis if the evidence was compatible with this approach.

[9] Department for Business, Energy and Industrial Strategy. (2019). Smart meter roll-out: cost-benefit analysis. https://www.gov.uk/government/publications/smart-meter-roll-out-cost-benefit-analysis-2019

[10] BIT is a social-purpose consultancy, now wholly owned by Nesta. It generates and applies behavioural insights and conducts evaluations to inform policy and improve products, services, and policies.

2. Identify the strengths and weaknesses of this evidence, providing recommendations for the Department for Energy Security and Net Zero on how it can be interpreted and for suppliers on how it can be improved.

2. Methodology

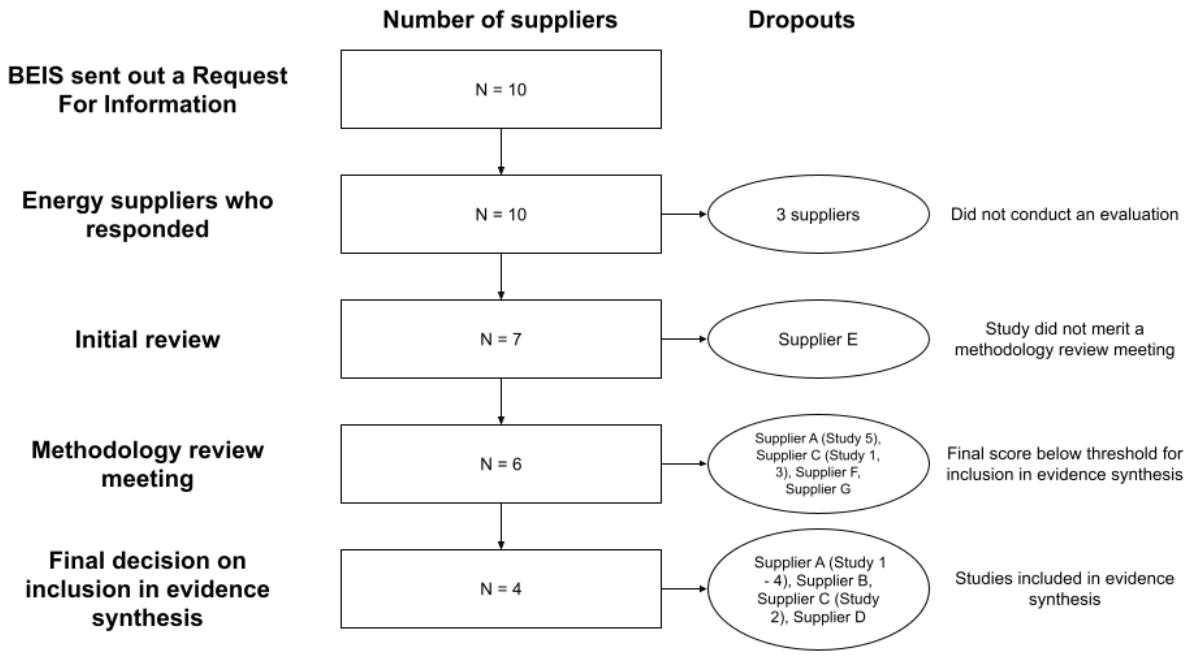
2.1 Review of supplier studies

At the start of the project, the Department for Energy Security and Net Zero requested information for this evidence review from energy suppliers, defined as any studies that suppliers had completed that quantitatively analysed the impact of smart meters on customers' energy consumption. The Department for Energy Security and Net Zero sent out requests for information to the 10 largest energy suppliers (who had a combined domestic energy market share of 99% as of Autumn 2022). All 10 suppliers responded. Seven had conducted an evaluation of the impact of smart meters on domestic energy consumption.

In advance of our initial review of suppliers' smart meter impact studies, we developed a review framework for assessing the quality of the studies. We used the same framework for each supplier, scoring separately on three categories: 1) internal validity, 2) external validity, and 3) precision and rigour, each on a 0-5 scale.[11] We attach the framework in Appendix A1.

We conducted an initial review of all studies from seven suppliers. One study was filtered out at this stage. Studies from **six** suppliers merited proceeding to the next stage, in which we held a methodological review meeting with each supplier. We clarified details of the studies' methodologies in these meetings and then updated our scores. Our final assessment concluded that **seven** studies from **four** suppliers were of sufficient quality to include in the evidence synthesis (a formal meta-analysis) in this report. Figure 3 details how many suppliers dropped out in each stage.

Figure 3: CONSORT diagram showing how suppliers' studies dropped out at each stage



[11] Internal validity refers to whether the study identifies the effect of a smart meter within the study sample in an unbiased fashion, avoiding confounding the effect of the smart meter with other factors (e.g. the characteristics of smart-meter customers). External validity refers to how well the effect the study identifies is generalisable to other customer populations. Precision and rigour capture considerations of sample size, the use of controls to reduce statistical noise, and straightforward and well-justified outcome measures.

2.2 Evidence synthesis

To formally synthesise the evidence from the included supplier studies, we conducted a meta-analysis. As explained in the Cochrane Handbook for Systematic Reviews of Interventions:[12]

> *"[A] meta-analysis is the statistical combination of results from two or more separate studies. Potential advantages of meta-analyses include an improvement in precision, the ability to answer questions not posed by individual studies, and the opportunity to settle controversies arising from conflicting claims. However, they also have the potential to mislead seriously, particularly if specific study designs, within-study biases, variation across studies, and reporting biases are not carefully considered."*

In this case, we conducted a fixed effects (FE) meta-analysis, which treats each analysis independently, as follows:

1. Compile the estimated treatment effects (standardised as percentage changes) from each study
2. Calculate the inverse-variance weighted average of the estimates:

   $$\textit{inverse-variance weighted average} = sum(estimate \times (1/SE_i^2))/sum(1/SE_i^2)$$

   where SE_i is the standard error of the estimate from energy supplier study *i*. Studies with small SEs are given relatively higher weight while studies with larger SEs are given relatively smaller weight, as SEs between studies reflect differences in the precision of estimated impacts.
3. Use the standard error of the pooled intervention effect – given by $sqrt(1/sum(1/SE_i^2))$ – to derive a 95% confidence interval. This 95% confidence interval communicates the precision of the pooled intervention effect – it contains all values of the point estimate that are not different from the point estimate at the 5% level of significance.

Implicit in the choice of the FE model (instead of a random effects model)[13] in the meta-analysis is that the single pooled estimate is of policy interest. We believe this to be a reasonable assumption because policymakers and other stakeholders tend to informally pool results – conducting the pooling formally informs and enhances this informal logic. The FE model does not represent an estimate of the smart meter impact for all customers: it represents the pooled estimate of the impact of smart meters for the customers in the suppliers' studies. Researchers and policymakers can infer from it what the impact of smart meters for other customers would be only insofar as the effects estimated in the studies are generalisable.

---

[12] Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). Cochrane handbook for systematic reviews of interventions. John Wiley & Sons.

[13] The main three choices are a common effects (CE), fixed effects (FE) or random effects (RE) model. The CE model assumes the true effect is the same across all studies, the FE model allows effects to differ between studies but assumes they make up the entire population of interest, and the RE model assumes the studies (and associated effect sizes) are drawn randomly from a distribution and attempts to estimate the true mean of that distribution. The RE model is theoretically appealing given differences between suppliers' smart meter roll-out and associated services. However, our view is that RE models are unstable when there are few studies, as characterises our situation. We suggest using the FE model since it is more realistic than the CE model (though note that in meta-analyses of a small number of studies, there is usually very little difference in results between FE and CE models). Finally, we believe that a RE model is best when it is plausible that studies are drawn randomly from the true distribution of effects; but we do not believe this assumption to be realistic in this case. Instead, they are more likely to come from larger firms that have the capacity to do these sorts of analysis. For all of these reasons, we believe the FE model to be most appropriate.

3. Quality assurance of suppliers' studies

In this section, we summarise our quality assessment for each energy supplier's study. Full details are in Appendix A3.

3.1 Quality assurance criteria

We scored studies that evaluated the impact of smart meters on domestic energy consumption on three categories: 1) internal validity, 2) external validity, and 3) precision and rigour, each on a 0-5 scale.[14]

A study proceeded to a methodological review meeting if it satisfied the following conditions:

- Minimum internal validity score of 2
- Minimum total score of 6

Scores were updated based on the methodological review meetings. A study was then used in the evidence synthesis if it satisfied the following:

- Minimum internal validity score of 3
- Minimum external validity of 2
- Minimum total score of 7
- Must have standard errors (as these are necessary to construct confidence intervals)

3.2 Study inclusion/exclusion decisions

In this section, we explain our rationale on the inclusion and exclusion decision of each supplier's smart meter impact evaluation study. See full details in Appendix A3.

Supplier A

Supplier A compared changes in (weather-adjusted) energy consumption, before and after installing a smart meter, between smart-meter customers and a matched group of traditional-meter customers[15] – a 'matched difference-in-differences analysis'. Specifically, the supplier performed 1:1 matching without replacement, matching on previous consumption and property size (bedroom count) and geographical location (Electricity DNO geographical region). Each of their studies used installations during one specific quarter to construct the smart-meter group.

---

[14] As noted above:

- Internal validity refers to whether the study identifies the effect of a smart meter within the study sample in an unbiased fashion, avoiding mixing up the effect of the smart meter with other factors (e.g. the characteristics of smart-meter customers).
- External validity refers to how well the effect the study identifies is generalisable to other customer populations.
- Precision and rigour capture considerations of sample size, the use of controls to reduce statistical noise, and straightforward and well-justified outcome measures.

The full framework is in Appendix A1.

[15] Pre-installation and post-installation periods are defined by a predefined 3-month installation window – for both the smart-meter customers (whose meter was installed during that window) and the matched group of traditional-meter customers (who, of course, had no meter installation during that window).

Findings

Supplier Smart-meter customers' installation dates	Sample size[16]	Electricity smart meter impact estimate (95% confidence interval)	Gas smart meter impact estimate (95% confidence interval)
Study 1 (2017 Q2)	Electricity: **61.6k** Gas: **69.4k**	**-3.70%** (-3.90%, -3.50%)	**-3.00%** (-3.18%, -2.82%)
Study 2 (2017 Q3)	Electricity: **58k** Gas: **64k**	**-4.60%** (-4.84%, -4.37%)	**-3.80%** (-4.02%, -3.58%)
Study 3 (2017 Q4)	Electricity: **48k** Gas: **90k**	**-4.40%** (-4.66%, -4.15%)	**-3.20%** (-3.38%, -3.02%)
Study 4 (2018 Q1)	Electricity: **36.2k** Gas: **68.6k**	**-4.00%** (-4.29%, -3.71%)	**-3.30%** (-3.50%, -3.10%)
Study 5 (2020 Q3)	Electricity: **3.2k** Gas: **5k**	**-2.20%** (-3.12%, -1.28%)	**-1.90%** (-2.80%, -1.00%)

Strengths

By using matching, the supplier ensured both groups were similar in various dimensions – chiefly in terms of their pre-installation consumption and geographical location. While Supplier A's methodology is very unlikely to give perfectly unbiased estimates (since results are biased by characteristics whose trends over time differ by smart-meter and traditional-meter customers), it removes bias from characteristics which are constant over time and may differ between the two groups and accounts for differences in pre-installation energy consumption.

They also conducted the analysis multiple times using very similar analytical strategies each time, and were transparent in the sample selection criteria. This gives us confidence in the integrity of the studies.

Weaknesses

Although most sample selection criteria seem reasonable, ultimately a majority (50% to 90%) of the potential sample was excluded. In addition, for Study 5, we are concerned that the installation window and post-installation-window includes COVID-19 affected months. Our concern is that the effect of smart meters may be different during COVID-19 months compared with non-COVID-19 months – but the magnitude and even direction of the COVID-19 influences are difficult to predict without other (and ideally larger) studies.

Decisions

Supplier A - Studies 1-4		
Internal validity 4/5	**External validity** 4/5	**Rigour** 4.5/5

[16] The sample size is split 50/50 between smart- and traditional-meter customers (because they use 1:1 matching without replacement).

Decision
Include Studies 1-4 in the evidence synthesis

Supplier A - Study 5		
Internal validity 4/5	**External validity** 1/5	**Rigour** 2/5
Decision Exclude Study 5 from the evidence synthesis		

Supplier B

Supplier B also performed a matched difference-in-differences analysis, using installations between January 2016 and February 2018 to form the group of smart-meter customers. They divided customers into three bands - low, medium and high - defined by Ofgem's Typical Domestic Consumption Values[17]. They matched the smart-meter customers to a random sample of traditional-meter customers with the same region / consumption band.

Findings

Smart-meter customers' installation dates	Sample size (smart- and traditional-meter customers)	Electricity smart meter impact estimate (95% confidence interval)	Gas smart meter impact estimate (95% confidence interval)
Jan 2016 - Feb 2018	Electricity: **25.8k** (13.0k smart, 12.9k traditional) Gas: **21.6k** (12.9k smart, 8.7k traditional)	**-1.12%** (-1.51%, -0.73%)	**-1.55%** (-1.95%, -1.15%)

Strengths

Supplier B's study design and sample restriction criteria are similar to Supplier A's. While the matched difference-in-differences does not completely eliminate bias, it guards against biases in estimating the smart meter impact from differences in some customer characteristics (specifically, ones that do not vary over time). Even though Supplier B uses a smaller sample than Supplier A, we believe the Study is conducted using a sufficiently large sample size.

[17] Ofgem https://www.ofgem.gov.uk/publications/review-typical-domestic-consumption-values-2019 defines:
- the 'low' (25th percentile) thresholds as: 8,000 kWh for gas, 1,800 kWh for electricity;
- the 'medium' (50th percentile) thresholds as: 12,000 kWh for gas, and 2,900 kWh for electricity; and
- the 'high' (75th percentile) thresholds as: 17,000 kWh for gas, and 4,300 kWh for electricity.

We are not certain exactly how Supplier B turned these three thresholds (which imply four groupings) into three groupings. Based on conversations with Supplier B, we believe the Ofgem figures may have been used as midpoints for the three groups, with two thresholds implied by these midpoints.

Weaknesses

As opposed to Supplier A, Supplier B conducted matching coarsely (using very wide consumption bands and different installation dates), resulting in a less closely matched and comparable sample.[18]

Decision

Supplier B		
Internal validity 3.5/5	**External validity** 3.5/5	**Rigour** 4/5
Decision Include the study in the evidence synthesis		

Supplier C

Supplier C conducted a difference-in-differences-style analysis (without matching), comparing the difference in consumption 1-3 months before the installation window and one year later (10-12 months after the installation window) between customers who had a traditional meter and customers who had a smart meter. In their two separate studies, they defined smart-meter customers as those who installed a smart meter between (i) June 2018 - August 2018 and (ii) June 2019 - August 2019.

Findings

Smart-meter customers' installation dates	Sample size (smart- and traditional-meter customers)	Electricity smart meter impact estimate (95% confidence interval)	Gas smart meter impact estimate (95% confidence interval)
Study 1 (Jun 2017 - Aug 2017)	Unknown	**-2.40%** (no confidence intervals available)	**-1.60%** (no confidence intervals available)
Study 2 (Jun 2018 - Aug 2018)	Electricity: **403.2k** (37.4k smart) Gas: **317.7k** (28.5k smart)	**-1.86%** (-2.15%, -1.57%)	**-1.40%** (-1.72%, -1.08%)
Study 3 (Jun 2019 - Aug 2019)	Electricity: **367.4k** (33.9k smart) Gas: **286.1k** (25.7k smart)	**1.79%** (no confidence intervals available)	**3.58%** (no confidence intervals available)

Strengths

The analysis uses a very large sample size. In addition, the exclusion/inclusion criteria are sensible; many of them mirror suggestions we made in our energy consumption analysis guidance.[19] Finally,

[18] What we mean by 'different installation dates': Supplier A's analysts showed that the 'dummy' installation dates for traditional-meter customers were, on average, earlier than the smart meter customers' actual installation dates. We consider this a moderate threat to internal validity.

[19] https://www.bi.team/publications/guidance-on-conducting-energy-consumption-analysis/

smart- and traditional-meter customers have similar trends in their prior to smart-meter installation in Study 2, which suggests that the parallel trends assumption is more likely to (approximately) hold.

Weaknesses

As opposed to Supplier A and B's method, the lack of matching makes the smart-meter customers and traditional-meter customers less comparable. They also used a transformed measure of energy usage that is more subject to measurement errors.

Furthermore, as we examined the data provided by Supplier C's Study 3, we found evidence that smart- and traditional-meter customers had diverging trends in their consumption prior to smart-meter installation. This suggests that the parallel trends assumption is likely to be violated, which would bias difference-in-differences estimates.[20] This raised concerns that the estimated impact may be biased, leading us to exclude Study 3 from the evidence synthesis. Lastly, since Supplier C was unable to provide information about the sample size for Study 1's analysis and the relevant standard errors, we also excluded the study from the evidence synthesis.

Decisions

Supplier C - Study 1		
Internal validity 3/5	**External validity** 3/5	**Rigour** 2/5
Decision Exclude Study 1 from the evidence synthesis (no standard errors or sample sizes reported)		

Supplier C - Study 2		
Internal validity 3/5	**External validity** 3/5	**Rigour** 3/5
Decision Include Study 2 in the evidence synthesis		

Supplier C - Study 3		
Internal validity 2/5	**External validity** 2/5	**Rigour** 3/5
Decision Exclude Study 3 from the evidence synthesis		

[20] Difference-in-differences analyses assume parallel trends in the absence of an intervention. That is, the difference-in-differences will correctly identify the impact of smart meters even where smart- and traditional-meter customers' historic *levels* of electricity/gas consumption differ, but it will fail to do so where these groups' energy consumption *trends* differ. We found a diverging energy consumption pre-installation trend between traditional-metered and smart-metered customers in Supplier C's Study 3.

Supplier D

Supplier D performed a matched difference-in-differences analysis using a similar method to supplier A. Their smart-meter group was defined by installations between December 2015 and July 2016. They use coarsened exact matching on total combined 2015 gas and electricity consumption (in 200kWh/month consumption bands), outward postcode, property type and property age.

Findings

Smart-meter customers' installation dates	Sample size[21]	Electricity smart meter impact estimate (95% confidence interval)	Gas smart meter impact estimate (95% confidence interval)
Dec 2015 - Jul 2016	Electricity: **41.7k** Gas: **22.9k**	**-1.67%** (-2.00%, -1.34%)	**-0.94%** (-1.37%, -0.51%)

Strengths

Supplier D's study design and sample restriction criteria are very similar to Supplier A's. While the matched difference-in-differences does not completely eliminate bias, it guards against biases in estimating the smart meter impact from differences in some customer characteristics (specifically, ones that do not vary over time). Even though Supplier D uses a smaller sample than Supplier A, we believe the study is conducted using a sufficiently large sample size.

Furthermore, Supplier D provided an anonymised copy of the dataset underpinning the analysis. We were able to conduct various verification checks using the data, increasing our confidence in the analysis.

Weaknesses

While our verification checks increased our confidence in the internal validity of the analysis, we did find high sensitivity of some of the results to model specifications. Furthermore, the analysis is fairly representative of Supplier D's customer base in 2015 through 2017, but the focus is on early adopters and has more electricity-only customers than is representative of the overall customer base.

Decision

Supplier D		
Internal validity 4/5	**External validity** 3/5	**Rigour** 3/5
Decision Include the study in the evidence synthesis		

Supplier E

Supplier E calculated the average annual electricity consumption of all Profile Class 1 smart-meter customers in its customer base with installations up to December 2020, and the average annual gas

[21] The sample is split almost exactly 50/50 between smart- and traditional-meter customers

consumption of all smart-meter customers in its customer base (though note that they also included some traditional-meter customers in the calculation of smart-meter customers' consumption, due to the way they grouped customers). They compared these averages to Ofgem's 'medium' Typical Domestic Consumption Values (TDCV) in order to estimate the impact of smart meters.

Findings

Smart-meter customers' installation dates	Sample size (smart- and traditional-meter customers)	Electricity smart meter impact estimate	Gas smart meter impact estimate
Until Dec 2020	Most of Supplier E's customer base	Electricity consumption among Supplier E smart meter customers was 17% **higher** than Ofgem's medium TDCV	Gas consumption among Supplier E smart meter customers was 17% **lower** than Ofgem's medium TDCV

Weaknesses

We believe this is not a valid study design in this context. Supplier E is different in meaningful ways from other suppliers, so its customers are unlikely to be representative of other UK customers; thus, the difference in consumption between Supplier E customers and Ofgem's 'medium' Typical Domestic Consumption Value could be driven by differences between their customers and other customers, rather than by the installation of a smart meter.

Note that the supplier had not claimed to have done a robust study, but submitted this to comply with the Department for Energy Security and Net Zero's request because the study technically fell within its scope.

Decision

Supplier E		
Internal validity 0.5/5	**External validity** 2/5	**Rigour** 1/5
Decision Exclude the study from the evidence synthesis		

Supplier F

Supplier F's study design is similar to a traditional difference-in-differences analysis. They compared the change in consumption (before vs after smart meter installation) among smart-meter households. They normalise each customer's consumption using the demand coefficients estimated by Elexon (electricity) and Xoserve (gas) to estimate what each smart meter customer would have used in the absence of a smart meter.[22] In their two separate studies, they defined smart-meter customers as those who installed a smart meter between (i) June 2012 - June 2017 and (ii) June 2012 - March 2018.

Findings

Smart-meter customers' installation dates	Sample size (smart- and traditional-meter customers)	Electricity smart meter impact estimate (95% confidence interval)	Gas smart meter impact estimate (95% confidence interval)
Study 1 (Jun 2012 - Jun 2017)	Electricity: **17.6k** Gas: **1.5k**	**-2.45%** (-2.67%, -2.22%)	**-2.76%** (-3.52%, -1.99%)
Study 2 (Jun 2012 - Mar 2018)	Electricity: **82k - 83k** (80.5k smart) Gas: **6.1k** (4.6k smart)	**-2.20%**[23] (-2.31%, -2.08%)	**-3.15%** (-3.68%, -2.61%)

Strengths

Supplier F's study design is akin to a traditional difference-in-differences analysis. This approach reduces bias, in particular from factors that do not vary over time. Unlike most suppliers' analyses, which relied on some amount of estimation (interpolation between meter reads) that may be subject to measurement errors, Supplier F used an outcome measure that used *only actual meter reads*.

Weaknesses

While the exclusion restrictions seem reasonable, they ultimately caused a high exclusion rate in Supplier F's electricity analysis (60% of the original sample, which is similar to Supplier A's exclusion rate) and very high exclusion rate in their gas analysis (94% of the original sample).

In addition, the sample used to estimate the demand coefficients tends to be small relative to the samples required for this kind of analysis (typically less than 2,000 households). While the study design is otherwise sufficiently robust, we believe the precision of the analysis could be overestimated if the analysis does not consider the sampling variation from the households used to create the demand coefficients. We therefore excluded Supplier F's analysis from the evidence synthesis.

Decision

Supplier F		
Internal validity 3/5	**External validity** 3/5	**Rigour** 2/5
Decision Exclude both studies from the evidence synthesis		

[22] Elexon uses a sample of customers with special AMR meters to generate the coefficients suppliers use for demand estimation. The total number of customers in this sample varies over time (between 2,000 and 3,000). After households with invalid data are excluded, there are approximately 600-700 Profile-Class-1 customers across the 14 GSPs in a given year. Similar to Elexon, Xoserve uses special AMR meters to generate coefficients. As of April 2022, there were ~1,500 meter points in this sample with valid data.

[23] Supplier F only estimated the *median* change rather than the *mean* change for this study.

Supplier G

Supplier G used a traditional difference-in-differences study design (without matching), comparing the difference in consumption for smart-meter customers before and after installation to the difference in consumption for traditional-meter customers. It is our understanding that they used installations between 2017 and 2019.

Findings

Smart-meter customers' installation dates	Sample size (smart- and traditional-meter customers)	Electricity smart meter impact estimate (95% confidence interval)	Gas smart meter impact estimate (95% confidence interval)
est. 2017-2019	83% of Supplier G's customer base[24]	**-1.31%** (no confidence intervals available)	**0.04%** (no confidence intervals available)

Strengths

As with other difference-in-differences analyses, the methodology removes bias from factors that are constant over time and differ between smart- and traditional-meter customers, but assumes that underlying consumption trends are the same in the two groups (parallel trends). The analysis was also conducted with a large sample, based on our interpretation of the study.

Weaknesses

Although the methodology described seems similar to valid methods used by other suppliers, there was insufficient information available to be confident in its internal validity. This also meant it was not possible to calculate the standard errors of their estimates.

Decision

Supplier G		
Internal validity 2/5	**External validity** 4/5	**Rigour** 2/5
Decision Exclude the study from the analysis		

[24] This is based on our interpretation of the study. Supplier G did not explicitly state the sample size of the study.

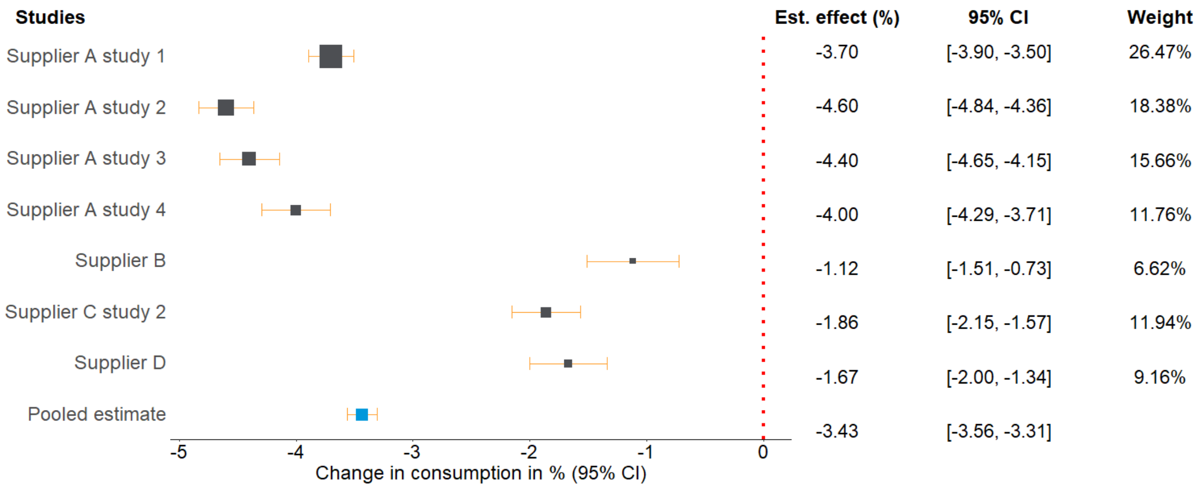
4. Results

4.1 Primary findings

To formally synthesise the evidence from the included supplier studies, we conducted a fixed effects (FE) meta-analysis that treated each analysis as independent, including the numerous Supplier A studies. Our main justification for this decision about the Supplier A analyses is that the samples in each should not overlap much and therefore can be seen as discrete studies.[25] We present our main findings for electricity and gas as forest plots in Figures 4 and 5 respectively.

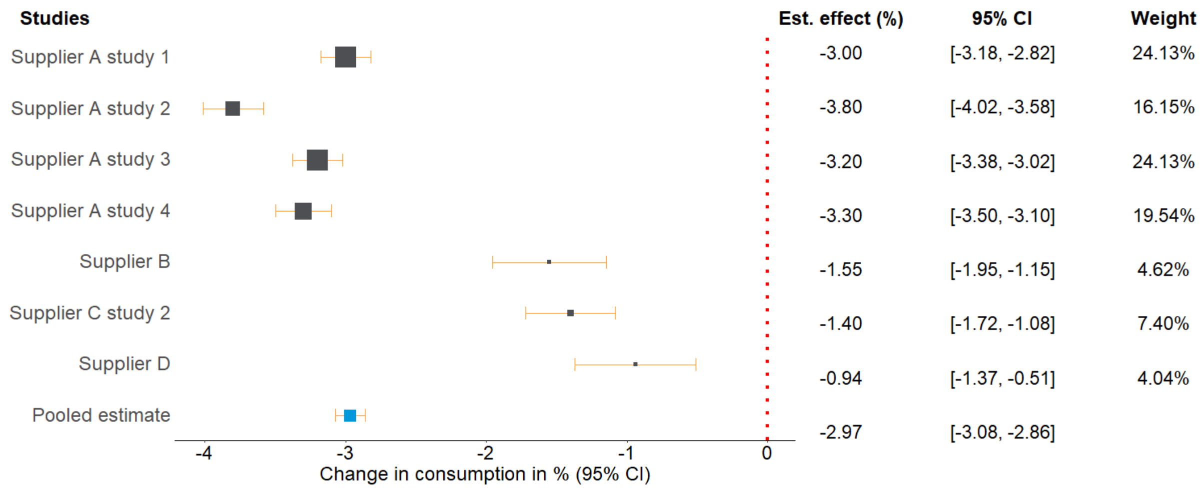
In the following forest plots, each study's square is sized in proportion to its weighting in the pooled estimate. The orange whiskers around each study represent the 95% confidence intervals. The lowest square (coloured blue) is the pooled estimate, and it has its own 95% confidence interval. We mark zero (no effect) with a red dotted line. If a study's 95% confidence interval crosses zero, we do not consider its point estimate to be statistically significantly different from zero; if the study's confidence interval is completely to the left or to the right of zero, we do consider its point estimate statistically significantly different from zero.

Figure 4: Forest plot for the impact of smart meters on electricity consumption



[25] They were conducted at different times, so it is possible for a customer who obtained a smart meter later to be in the traditional-meter group for an earlier study. It is also possible for a customer to be in the traditional-meter group for more than one study. However, the population of traditional-meter customers is extremely large, and as Supplier A used 1:1 matching, we expect it is rare for a traditional-meter in an earlier study to be re-used in a later study (as either a traditional-meter comparator again or as a smart-meter customer).

Figure 5: Forest plot for the impact of smart meters on gas consumption



The pooled estimates are -3.43% for electricity and -2.97% for gas. In other words, the meta-analysis suggests that smart meters reduce electricity consumption by 3.43% and gas consumption by 2.97% on average. The estimates have narrow 95% confidence intervals of [-3.56%, -3.31%] for electricity and [-3.08%, -2.86%] for gas. (Both estimates are statistically significantly different from 0, $p<0.001$.)

These pooled estimates are roughly in line with the 3.0% and 2.2% reductions in electricity and gas consumption (respectively) that the Department for Energy Security and Net Zero anticipated based on the existing international and domestic evidence.

4.2 Interpretation of primary results

4.2.1 Between-supplier heterogeneity

There is variation in the estimated effects from smart meters across suppliers:

- For electricity, the studies from Supplier A show electricity consumption reductions of 3.70% to 4.40%. The Supplier B, Supplier C, and Supplier D impact estimates are lower, indicating reductions of 1.12% to 1.86%.
- For gas, the studies from Supplier A show consumption reductions of 3.00% to 3.80%. The Supplier B, Supplier C, and Supplier D impact estimates indicate reductions of 0.94% to 1.55%.

It was not possible to draw strong conclusions about the reasons for this variation in estimated impacts in this review. It is possible that installations by different suppliers drove different consumption impacts. This could be due to differences in smart-meter roll-out and installation strategies between them – for example, differences in the in-home display (IHD) or advice received during installation or the proportion of customers receiving an IHD and/or energy efficiency advice during the installation. It could also reflect differences in the maturity of each supplier's roll-out at the time of the installations in their analysis. However, it is also possible that the differences in estimated effects reflect inherent differences in their customer bases which result in different smart meter impacts, or even just differences in the suppliers' sampling and analysis strategies. We therefore advise against overinterpretation of these differences.

4.2.2 Quality of evidence

We judged the studies included in the meta-analysis to be sufficiently robust – specifically, they needed to have an internal validity score of at least 3/5 and an external validity score of at least 2/5. However, none of the studies received full marks and there are some consistent threats to internal and external validity across studies. These threats influence how we should interpret the results from the evidence synthesis.

Internal validity - the following threats may systematically bias the included studies' estimates for the samples they examine:

- **Non-parallel underlying trends:** Every included study relied on a difference-in-differences style analysis. This method allows for historic levels of energy consumption to differ between smart- and traditional-meter customers, but assumes that the underlying trends (i.e. the difference in consumption over time in the absence of the intervention) would be the same in both groups. This is known as the parallel trends assumption. Because the allocation of smart meters is not random, smart meter uptake could be correlated with changes in customer attributes that influence energy consumption (e.g. engagement with energy-saving behaviours). If this is true, the parallel trends assumption would be violated, which would bias estimates. We have limited evidence about whether trends are parallel; few of the suppliers' studies checked whether trends were parallel in the past.
- **Differential attrition:** In almost every analysis, suppliers excluded customers without sufficient actual meter reads. We view this as a reasonable choice to guard against measurement error from incorrectly estimated consumption. However, this exclusion bears more heavily on traditional-meter customers (whose meters must be read manually) than on smart-meter customers (whose meter is read automatically after the smart meter installation). In a difference-in-differences analysis, this is a problem if the resulting differences in characteristics between groups vary over time. It is possible that some characteristics that affect both energy consumption and meter reading frequency do indeed vary with time, which would bias estimates. For example, as people become more attentive to their energy bills, they may submit meter readings more frequently and at the same time reduce their energy consumption. Our view is that this issue could lead to a difference-in-differences analysis to overestimate *or* underestimate smart meters' impact on energy consumption.
- **Differences in outcome measurement by meter type:** Suppliers measure energy consumption in different ways depending on meter type. For smart meter customers, consumption is automatically transmitted to suppliers after installation (but not before). On the other hand, consumption of traditional meter customers is calculated from manual reads only. To the extent that meter type affects measured consumption (without affecting 'true' consumption), this leads to bias in difference-in-differences estimates.[26]

External validity - the following threats may mean the effects of smart meters in the included studies are not representative of the effects on the UK general population:

---

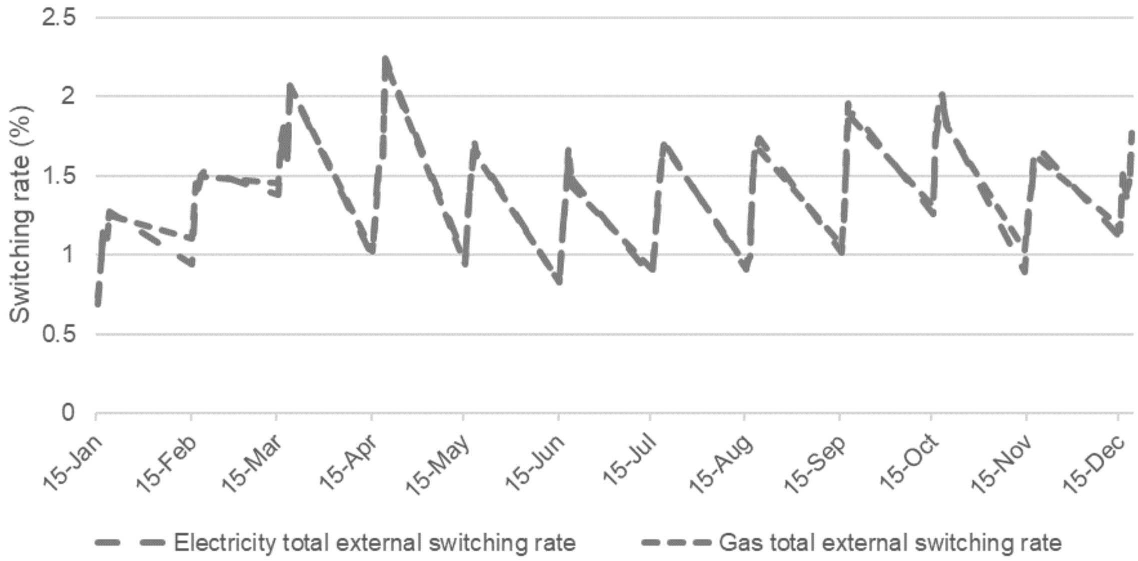
[26] The supplier studies varied in the extent to which they cleanly identified 'pre' and 'post' consumption in the traditional-meter group. Some suppliers estimated daily consumption by interpolating between two meter reads (and then adjusting for weather and seasonality using daily demand coefficients). As discussed in Appendix A3:

- Supplier B addresses this directly by using an actual meter read as the 'dummy installation date' for traditional-meter customers, leaving no need to interpolate consumption between the pre and post periods.
- Supplier C's design – by drawing on a measure of consumption seven to nine months after installation – also mostly avoids this issue.
- Supplier A protects against this issue by requiring a certain amount of actual meter reads in the pre and post periods to avoid over-interpolation. However, their method does leave room for some interpolation between periods for the traditional-meter group, and it leads to differential attrition (as discussed above).
- Supplier D's design is the most vulnerable to this issue.

Note that this issue does not affect the smart-meter group because suppliers almost always receive a 'closing read' from the traditional meter on the day of smart meter installation.

- **Exclusion of switchers and home-movers:** Almost all supplier studies excluded customers who moved home or switched suppliers during the study period from the analysis. As mentioned, customers were also excluded due to insufficient meter readings in most studies. This means that customers who were more likely to move home, read their meters infrequently, and/or switch energy suppliers were underrepresented in suppliers' analyses. Additionally, during the installation periods used in these studies, nearly all suppliers had roll-out strategies that focused on particular customer segments (e.g. avoiding certain regions such as London).

Figure 6: Monthly switching rates in the residential gas and electricity sector (six largest suppliers only)



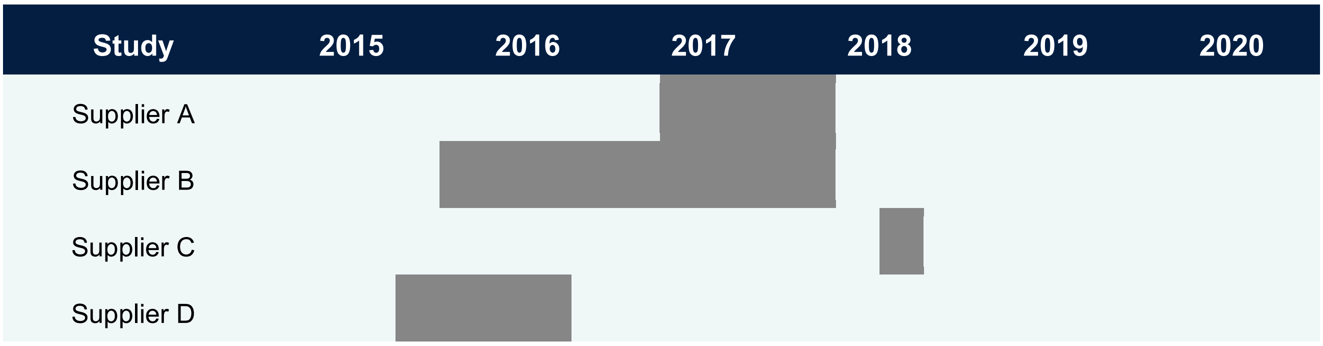
Source: https://www.ofgem.gov.uk/energy-data-and-research/data-portal/retail-market-indicators (April 2020)

As depicted in Figure 6, switchers represent a non-negligible share of the customers for electricity and gas. We hypothesise that switchers may be more attentive to their energy consumption, so the impact of smart meters may be different for this group.

The exclusion of home-movers from studies is also important to keep in mind. The estimates will be more applicable to older home-owners, as renters and younger people move at higher rates.

- **Late vs early adopters:** Early and late adopters may differ in meaningful ways. Early adopters may have been more energy-engaged – meaning that the impact of smart meters may be greater (given their interest in energy) or lower (because engaged customers may have less to 'learn' from their smart meter's more frequent and salient feedback) among this group than among later adopters. In Table 1, we visually summarise the dates of smart-meter installations (for the studies included in the evidence synthesis). Overall, the evidence base has moderate intertemporal diversity, but it is slightly skewed towards earlier-adopting smart-meter customers (23% of meters were smart meters as of mid-2018).

Table 1: Cohorts of smart-meter installations in studies that we include in the evidence synthesis, by year and quarter



4.3 Robustness checks

We perform robustness checks to examine whether our results are sensitive to the exact analytical strategy used. Specifically, we perform two robustness checks:

- **Re-weight the Supplier A studies at 1/4 the weight they have in the main analysis, before re-weighting all studies so that weights sum to 100%:** The Supplier A studies have a combined weight of more than 70% for the main electricity and gas analyses. Supplier A's estimates arguably all reflect the same type of sampling variation (due to the similarity of their methodology and all customers being from the same supplier). As a result, it is valuable to check if our results are robust when the Supplier A studies are essentially treated as one study.
- **Include Supplier A's 2020 Q3 analysis:** This analysis was excluded from the main synthesis because it examines the impact of smart meters during the COVID-19 period.

Figures 7 and 8 present the results of the first robustness check (re-weighting the four Supplier A studies used in the main synthesis). The pooled estimates change from -3.43% to -2.61% for electricity and from -2.97% to -2.43% for gas. These are both still statistically significantly different from 0 ($p<0.001$).

Figure 7: Forest plot for the impact of smart meters on electricity consumption, re-weighting Supplier A's estimates

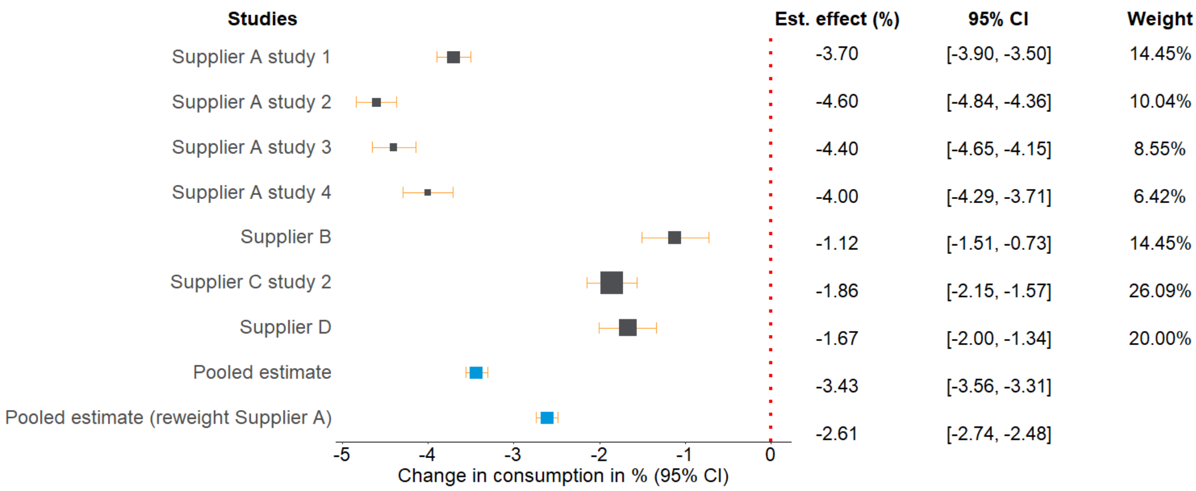
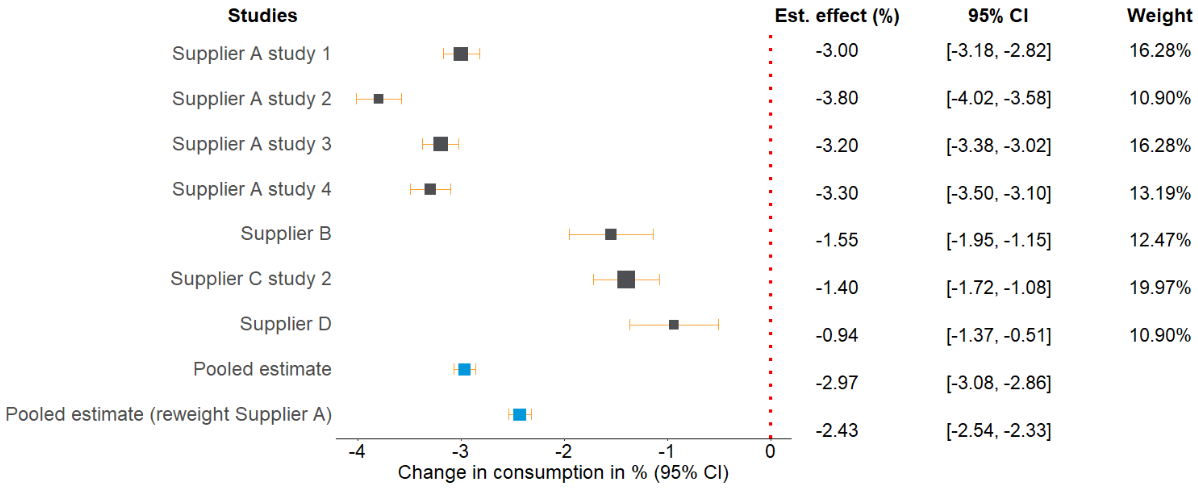


Figure 8: Forest plot for the impact of smart meters on gas consumption, re-weighting Supplier A's estimates



Figures 9 and 10 present the results of the second robustness check (including the Supplier A 2020 Q3 estimate). The pooled estimates do not change much at all: this is because the 2020 Q3 analysis was conducted using relatively small samples, resulting in very small weights in the creation of the pooled estimates.

Figure 9: Forest plot for the impact of smart meters on electricity consumption, including Supplier A's 2020 Q3 study

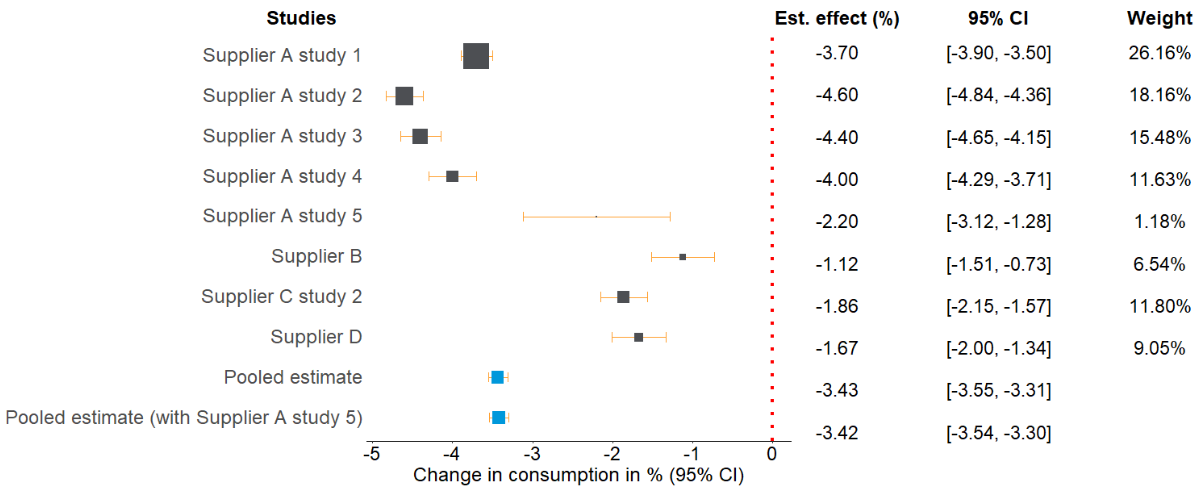
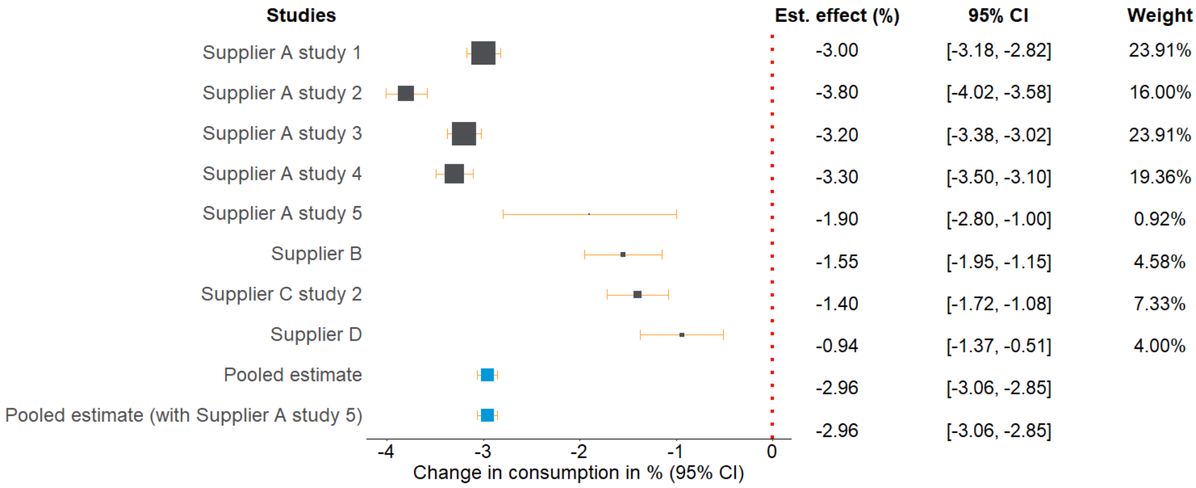


Figure 10: Forest plot for the impact of smart meters on gas consumption, including Supplier A's 2020 Q3 study



4.4 Exploratory analyses (long-term and COVID-19 specific effects)

Long-term impact of smart meters on energy consumption

Supplier A conducted analysis to understand the longer-run effect of smart meters on energy consumption.[27] The long-term impact of smart meters on energy consumption may differ from their one-year impact, as customers may learn over time and form new habits as they receive more insights on their energy consumption from their in-home-displays. Conversely, smart meters may have a novelty effect on consumption which wears off over time.

Since energy suppliers are unable to track customers' energy consumption if they switch to another supplier or move home, there is an additional sample selection issue with these analyses (as evidenced by the decreasing sample size as the duration increases). These analyses may be focusing on customers who are happier with their supplier, more loyal, and/or less engaged with their energy consumption, and are also more likely to be earlier adopters (by definition, because the analysis requires a longer time series).

We provide a summary of the findings in Table 2. Supplier A's long-term impact analysis indicates a persistent effect of smart meters on electricity consumption and a possible small diminishing trend in the effect on gas consumption, but the reduction impact remains above 3% across both electricity and gas consumption for both the two- and three-year analyses. Supplier A did not provide standard errors on these estimates, so we cannot conclude whether these changes in impact are statistically significant. Moreover, with only one supplier conducting analysis on the long-term impact of smart meters and the additional source of selection bias discussed above, we do not believe there is sufficient evidence in the energy supplier studies to draw firm conclusions on the long-term smart meter effect.

[27] Supplier B also estimated the impact of smart meters over 1.5 and 2 years but we do not report these results in this report because they are based on very small sample sizes.

Table 2: The long-term effect of smart meters on energy consumption

Supplier	Installation time	Duration	Electricity (sample size)	Gas (sample size)
Supplier A	2015 Q2	1 year	-4.1% (n = 22.0k)	-3.9% (n = 22.4k)
		2 years	-4.8% (n = 16.3k)	-3.6% (n = 19k)
		3 years	-4.3% (n = 7.4k)	-2.7% (n = 8k)
	2016 Q2	1 year	-3.5% (n = 34.7k)	-3.4% (n = 42.9k)
		2 years	-3.2% (n = 21.3k)	-3.3% (n = 28.5k)
		3 years	-3.1% (n = 10.5k)	-3.0% (n = 14.6k)

COVID-19 specific smart meter impact

Two suppliers investigated the effect of smart meters on customers' energy consumption during COVID-19 affected months. It should also be noted that the smart meters' effect in COVID-19 months may differ from non-COVID-19 months, as customers were more likely to be at home during the lockdown, so (i) there may have been more (or fewer) opportunities for energy reduction as customers were likely consuming more energy[28], and (ii) customers may have been paying more attention to their in-home-displays as they were staying at home.

However, we do not think the evidence is sufficiently robust to present in this report. We excluded Supplier A's 2020 Q3 analysis partly because of concerns that the post-installation-window included COVID-19 affected months. This may affect external validity – the extent to which the measured effects generalise to non-COVID-19 affected months. It may also reduce internal validity, if newly working-from-home customers were also more likely to accept smart meters during the 2020 Q3 installation window. In addition, we noted that the exclusion rate for that Supplier A analysis was particularly high – up to 90% of potential smart meter customers were excluded. We excluded Supplier C's analysis of installations from 06/2019 - 08/2019 due to concerns about non-parallel pre-trends we observed.

[28] Our uncertainty on the direction of the effect comes from our uncertainty about the ease of saving energy while working from home – it may be harder *or* easier to reduce discretionary or wasteful energy consumption than when occupants do *not* work from home.

5. Conclusions

The first phase of the project involved reviewing 14 analyses from seven suppliers.[29] Based on our reviews, we recommended using seven studies from four suppliers in an evidence synthesis: four studies from **Supplier A**, and one study each from **Suppliers B**, **C**, and **D**. We recommended excluding one study from Supplier A, two studies from Supplier C, Supplier E's study, two studies from Supplier F, and Supplier G's study. The studies included in the evidence synthesis have moderate-to-high internal validity, moderate external validity, and strong precision and rigour.

Results

Our evidence synthesis of the seven included studies finds a reduction in energy consumption from smart meters.

The pooled estimates are -3.43% for electricity and -2.97% for gas. The estimates have narrow 95% confidence intervals of [-3.56%, -3.31%] and [-3.08%, -2.86%] for electricity and gas respectively. These estimates are roughly in line with (indeed slightly higher than) the effects of -3.0% (electricity) and -2.2% (gas) that the Department for Energy Security and Net Zero anticipated based on the evidence available before this review.

There is variation in the estimated effects from smart meters across suppliers:

- For electricity, the studies from Supplier A (from 2017 Q2 through 2018 Q1 installation windows) show electricity consumption reductions of 3.70% to 4.40%. The Supplier B, Supplier C, and Supplier D impact estimates are lower, indicating reductions of 1.12% to 1.86%.
- The picture is similar for analyses of gas consumption. The studies from Supplier A show gas consumption reductions of 3.00% to 3.80%. The Supplier B, Supplier C, and Supplier D impact estimates indicate reductions of 0.94% to 1.55%.

It is possible that installations by different suppliers drove different consumption impacts. This could be due to differences in smart-meter roll-out and installation strategies between them – for example, differences in the in-home display (IHD) or the quality of the advice received during installation or the proportion of customers receiving an IHD and/or energy efficiency advice during the installation. However, it is also possible that the differences in estimated effects reflect inherent differences in their customer bases which result in different smart meter impacts, or even just differences in the suppliers' sampling and analysis strategies. We therefore advise against overinterpretation of these differences.

We conducted a robustness check re-weighting the Supplier A studies such that they are essentially treated as one study. This is because Supplier A contributes four of the seven studies. The new estimated impact of smart meters is slightly smaller when we make this analytical change: the pooled estimates change from -3.43% to -2.61% for electricity and from -2.97% to -2.43% for gas. These new results indicate that the magnitude of the pooled estimates is sensitive to the choice of study weighting. However, the pooled electricity and gas estimates are both still statistically significantly different from 0 ($p<0.001$) and roughly in line with the reductions the Department for Energy Security and Net Zero anticipated.

[29] Beyond these 14 studies, we excluded various Supplier A analyses due to Supplier A's inability to retrieve key analysis details, particularly confidence intervals around the smart meter impact estimate. See Appendix A2 for a full list.

Strength of evidence overall

All of the studies use quasi-experimental methods and so the estimates from each individual study are likely to be biased to some degree, and do not represent the average effect of smart meters on the general population's energy consumption. However, we believe the included studies have:

- **Moderate-to-high internal validity**. All included studies use a difference-in-differences or equivalent approach. These designs remove statistical biases from differences in characteristics which do not vary over time in the data (e.g. geographical region) between smart- and traditional-meter groups. However, there is a meaningful risk in each study that trends in characteristics differ between groups over time (e.g. customers' environmental attitudes), which biases estimated impacts. We also felt that the analyses sometimes made suboptimal or unconventional methodological choices. For example, in some analyses involving matching, not every smart-meter customer had a matching traditional-meter customer, which is an unconventional way to carry out matching in econometric analyses. Another example is that certain analyses exhibited moderate biases in the dates of the smart-meter versus traditional-meter group, leading to the possibility of weather effects muddying the identification of the smart meter's impact. (See Appendix A3 for further details.) Nevertheless, we think that the included studies all have moderate-to-high internal validity.
- **Moderate external validity**. The studies' samples are different to the general population in a variety of ways. For example, customers in the suppliers' study samples are:
  - less likely to switch supplier;
  - less likely to move home; and
  - more likely to consistently provide manual readings for their traditional meter before receiving the smart meter.

  In addition, late adopters are underrepresented in most of the studies we included in the evidence synthesis.
- **Strong precision and rigour**. The studies have large sample sizes and straightforward analysis strategies and outcome measures. However, some involved analytical decisions that we thought were suboptimal (perhaps related to the limited resources that suppliers can generally invest in their smart meter impact analyses).

The meta-analysis suggests that smart meters reduce electricity consumption by 3.3% to 3.6% and gas consumption by 2.9% to 3.1%. We believe this estimate provides useful evidence for the Department for Energy Security and Net Zero for assessing the overall impact of the smart meter roll-out on household energy consumption.

A1. Study inclusion/exclusion decisions methodology

To decide on inclusion/exclusion at each stage of the work, we rated studies separately on three categories: 1) internal validity, 2) external validity, and 3) precision and rigour. We will rate each category on a 0-5 scale, similar to the Education Endowment Foundation's 'five padlocks' system[30] to assess research.

To proceed to a methodological review meeting, we required:

- Minimum score of 6 overall
- Minimum internal validity score of 2.

To proceed to the evidence synthesis, we required:

- Minimum internal validity score of 3
- Minimum external validity of 2
- Minimum total score of 7
- The study to have standard errors on its smart meter impact estimate.

Internal validity

Score	Description
5	● The study is based on a randomised controlled trial (including randomised encouragement designs) or convincing regression discontinuity design ● Accurate use of confidence intervals and standard errors ● No concurrent intervention was implemented on the sample (or any intervention was orthogonal to smart meter installation) ● Low level of contamination and/or spillover in the treatment and control groups ● Low attrition and missing data
4	● Appropriate quasi-experimental method to recover a causal estimate that considers how some type of selection on unobservable characteristics may be avoided/minimised (examples include: convincing difference-in-differences, matching + difference-in-differences, synthetic control) ● Sound justification for identification assumptions, depending on the quasi-experimental design used, and supports the justification with empirical evidence[31] ● Accurate use of confidence intervals and standard errors ● Low attrition and missing data

[30] Educational Endowment Foundation. (2019.) Classification of the security of findings from EEF evaluations Version 2.0 – July 2019. https://educationendowmentfoundation.org.uk/public/files/Evaluation/Carrying_out_a_Peer_Review/Classifying_the_security_of_EEF_findings_2019.pdf

[31] Such as balance checks, parallel pre-trend inspection for difference-in-differences, cut-off manipulation check for regression discontinuity, overlap in the propensity score distribution for matching, and/or a strong first stage for instrumental-variable approach.

3	• Appropriate approach to recover a causal estimate that considers how some type of selection on unobservable characteristics may be avoided/minimised (examples include: matching) • Some justification for identification assumptions, depending on the quasi-experimental design used, but with no empirical tests of the assumptions. • Incorrect use of standard errors • Moderate attrition and missing data, with mitigation strategies such as imputation
2	• Design for comparison that considers selection only on some relevant observable characteristics • Incorrect use of standard errors • Moderate attrition and missing data, without any mitigation strategy
1	• Inappropriate approach to recover a causal estimate • No discussion on attrition and missing data/high attrition
0	• Methodology does not meet the minimum standards when considering any of the following: selection into treatment, sample selection, attrition, and the use of standard errors

External validity

Score	Description
5	• The sample for the study is representative of the supplier's customer base • The supplier's customer base is representative of the general population • Formal representativeness checks of the sample • Sufficient consideration of qualitative threats to external validity • Study conducted through a sufficiently long period (at least one year of consumption data), to avoid results being sensitive to aberrant short-term or season-specific impacts
4	• The sample for the study is representative of the suppliers' customer base • Sufficient consideration of qualitative threats to external validity, with mitigation strategies • Study conducted through a sufficiently long period (at least one year of consumption data)
3	• The sample for the study focuses on a particular set of customers who are not representative of the suppliers' customer base • Some consideration of qualitative threats to external validity • Study conducted through a moderate length of time (6-12 months of consumption data)
2	• The sample for the study focuses on a niche and unrepresentative set of customers who are not representative of the suppliers' customer base (e.g. time-of-use tariff customers) • Little consideration of qualitative threats to external validity • Study conducted through a moderate length of time (6-12 months of

	consumption data)
1	• No way to determine the representativeness of the sample. • Study conducted over a limited period of time (fewer than 6 months of consumption data)
0	• No discussion of the representativeness of the sample, despite serious concerns about extreme non-representativeness of sample

Precision and rigour

Score	Description
5	• The study has a pre-specified analysis plan • Sound justification of the outcome measure(s) • Sufficiently large sample size (ideally 10,000 customers, though fewer customers may be sufficient to earn a '5' if the supplier uses other methods to improve precision) • Sufficient and convincing robustness checks to address the possibility that results are sensitive to model selection • Appropriate consideration of the impact of COVID-19 and/or other unusual events on study logistics or analysis results • Appropriate use of covariates, potentially including past consumption data, weather, household characteristics
4	• Sound justification of the outcome measure(s) • Sufficiently large sample (at least 5,000 customers, though see note above that we accept other ways to increase precision) • Appropriate consideration of the impact of COVID-19 and/or other unusual events on study logistics or analysis results • Some appropriate, but incomplete, use of covariates
3	• Some justification of the outcome measure(s) • Moderate sample size (2,500 to 5,000 customers) • Some consideration of the impact of COVID-19 and/or other unusual events on study logistics or analysis results • Covariates to reduce standard errors not used
2	• Some, but unconvincing, justification of the outcome measure(s) • Moderate sample size (2,500 to 5,000 customers) • No consideration of the impact of COVID-19 and/or other unusual events on study logistics or analysis results
1	• Small sample size (fewer than 500 customers, with no mitigating boosts to precision) • Lacks robustness checks • No consideration of the impact of COVID-19 and/or other unusual events on study logistics or analysis results
0	• Severely underpowered study (fewer than 250 customers, with no mitigating

	boosts to precision) ● Other severe threats to precision and rigour, such as nonsensical outcome measure(s)

A2. Study inclusion/exclusion decisions

Table 3: Summary of scores on smart meter impact studies that we *included* in the evidence synthesis

Supplier (smart-meter customers' installation dates)	Sample size[32] (smart- and traditional- meter customers)	Internal validity score	External validity score	Precision and rigour score
Supplier A Study 1 (2017 Q2)	Electricity: **61.6k** Gas: **69.4k**	4	4	4
Supplier A Study 2 (2017 Q3)	Electricity: **58k** Gas: **64k**	4	4	4
Supplier A Study 3 (2017 Q4)	Electricity: **48k** Gas: **90k**	4	4	4
Supplier A Study 4 (2018 Q1)	Electricity: **36.2k** Gas: **68.6k**	4	4	4
Supplier B (Jan 2016 - Feb 2018)	Electricity: **25.8k** (13.0k smart, 12.9k traditional) Gas: **21.6k** (12.9k smart, 8.7k traditional)	3.5	3.5	4
Supplier C Study 2 (Jun 2018 - Aug 2018)	Electricity: **403.2k** (37.4k smart) Gas: **317.7k** (28.5k smart)	3	3	3
Supplier D (Dec 2015 - Jul 2016)	Electricity: **41.7k** Gas: **22.9k**	4	3	3

32 For the Supplier A studies, the sample size is split 50/50 between smart- and traditional-meter customers (because they use 1:1 matching without replacement). Supplier C and Supplier B's study sample sizes by smart/traditional meter are specified in the table. The Supplier D sample is split almost exactly 50/50 between smart- and traditional-meter customers, as is Supplier B's electricity analysis.

Table 4: Summary of the scores on smart meter impact studies that we *excluded* from the evidence synthesis

Supplier (smart-meter customers' installation dates)	Sample size (smart- and traditional-meter customers)	Internal validity score	External validity score	Precision and rigour score
Supplier A Study 5 (2020 Q3)	Electricity: **3.2k** Gas: **5k**	4	1	2
Supplier C Study 1 (Jun 2017 - Aug 2017)	Unknown	3	3	2 (no standard errors)
Supplier C Study 3 (Jun 2019 - Aug 2019)	Electricity: **367.4k** (33.9k smart) Gas: **286.1k** (25.7k smart)	2	2	3 (no standard errors)
Supplier E (until Dec 2020)	Most of Supplier E's customer base	0.5	2	1
Supplier F (Jun 2012 - Jun 2017)	Electricity: **17.6k** Gas: **1.5k**	3	3	2 (incorrect standard errors)
Supplier F (Jun 2012 - Mar 2018)	Electricity: **82k** (80.5k smart) Gas: **6.1k** (4.6k smart)	3	3	2 (incorrect standard errors)
Supplier G (est. 2017-2019)	83% of Supplier G's customer base[33]	2	4	2

[33] This is based on our interpretation of the study. Supplier G did not explicitly state the sample size of the study.

A3. Summary of energy suppliers' studies

Supplier studies: results summary

Table 5 summarises the impact estimates of the studies we included in the evidence synthesis. Table 6 summarises the studies we excluded.

Table 5: Summary of smart meter impact estimates we included in the evidence synthesis (95% confidence intervals in brackets)

Supplier (smart-meter customers' installation dates)	Sample size[34] (smart- and traditional-meter customers)	Electricity smart meter impact estimate (95% confidence interval)	Gas smart meter impact estimate (95% confidence interval)
Supplier A (2017 Q2)	Electricity: **61.6k** Gas: **69.4k**	**-3.70%** (-3.90%, -3.50%)	**-3.00%** (-3.18%, -2.82%)
Supplier A (2017 Q3)	Electricity: **58k** Gas: **64k**	**-4.60%** (-4.84%, -4.37%)	**-3.80%** (-4.02%, -3.58%)
Supplier A (2017 Q4)	Electricity: **48k** Gas: **90k**	**-4.40%** (-4.66%, -4.15%)	**-3.20%** (-3.38%, -3.02%)
Supplier A (2018 Q1)	Electricity: **36.2k** Gas: **68.6k**	**-4.00%** (-4.29%, -3.71%)	**-3.30%** (-3.50%, -3.10%)
Supplier B (Jan 2016 - Feb 2018)	Electricity: **25.8k** (13.0k smart, 12.9k trad.) Gas: **21.6k** (12.9k smart, 8.7k trad.)	**-1.12%** (-1.51%, -0.73%)	**-1.55%** (-1.95%, -1.15%)
Supplier C (Jun 2018 - Aug 2018)	Electricity: **403.2k** (37.4k smart) Gas: **317.7k** (28.5k smart)	**-1.86%** (-2.15%, -1.57%)	**-1.40%** (-1.72%, -1.08%)
Supplier D (Dec 2015 - Jul 2016)	Electricity: **41.7k** Gas: **22.9k**	**-1.67%** (-2.00%, -1.34%)	**-0.94%** (-1.37%, -0.51%)

[34] For the Supplier A's studies, the sample size is split 50/50 between smart- and traditional-meter customers (because they use 1:1 matching without replacement). Supplier D's sample is split almost exactly 50/50 between smart- and traditional-meter customers, as is Supplier B's electricity analysis. Supplier B and C's study sample sizes by smart/traditional meter are specified in the table.

Table 6: Summary of smart meter impact estimates we *excluded* from the evidence synthesis (95% confidence intervals in brackets)

Supplier (smart-meter customers' installation dates)	Sample size (smart- and traditional-meter customers)	Electricity smart meter impact estimate (95% confidence interval)	Gas smart meter impact estimate (95% confidence interval)
Supplier A (2020 Q3)	Electricity: **3.2k** Gas: **5k**	**-2.20%** (-3.12%, -1.28%)	**-1.90%** (-2.80%, -1.00%)
Supplier C (Jun 2017 - Aug 2017)	Unknown	**-2.40%** (no confidence intervals available)	**-1.60%** (no confidence intervals available)
Supplier C (Jun 2019 - Aug 2019)	Electricity: **367.4k** (33.9k smart) Gas: **286.1k** (25.7k smart)	**1.79%** (no confidence intervals available)	**3.58%** (no confidence intervals available)
Supplier E (until Dec 2020)	Between 500,000 and 1m customers[35]	Electricity consumption among Supplier E smart meter customers was 17% **higher** than Ofgem's medium TDCV	Gas consumption among Supplier E smart meter customers was 17% **lower** than Ofgem's medium TDCV
Supplier F (Jun 2012 - Jun 2017)	Electricity: **17.6k** Gas: **1.5k**	**-2.45%** (-2.67%, -2.22%)	**-2.76%** (-3.52%, -1.99%)
Supplier F (Jun 2012 - Mar 2018)	Electricity: **82k - 83k** (80.5k smart) Gas: **6.1k** (4.6k smart)	**-2.20%**[36] (-2.31%, -2.08%)	**-3.15%** (-3.68%, -2.61%)
Supplier G (est 2017-2019)	Approximately 700,000[37]	**-1.31%** (no confidence intervals available)	**0.04%** (no confidence intervals available)

Supplier A

Internal validity: 4/5 for all five of Supplier A's studies

Supplier A compared changes in energy consumption, before and after installing a smart meter, between smart-meter customers and a matched group of traditional-meter customers.[38] By using matching, they ensured both groups were similar in various dimensions – chiefly in terms of their pre-installation consumption and geographical location. While Supplier A's methodology does not completely eliminate statistical bias, it guards against biases in estimating the smart meter impact from differences in some customer characteristics (specifically, ones that do not vary over time). We did not give a '5' on internal validity to any study, as none conducted a randomised controlled trial nor exploited

[35] This is based on our interpretation of the study. Supplier E did not explicitly state the sample size of the study.

[36] Supplier F only estimated the *median* change rather than the *mean* change for this study.

[37] This is based on our interpretation of the study. Supplier G did not explicitly state the sample size of the study.

[38] Pre-installation and post-installation periods are defined by a predefined 3-month installation window – for both the smart-meter customers (whose meter was installed during that window) and the matched group of traditional-meter customers (who, of course, had no meter installation during that window).

a natural experiment that would ensure smart- and traditional-meter groups were well-matched on both observable *and* unobservable characteristics.[39]

External validity: 4/5 for Studies 1-4; 1/5 for Study 5

- Studies 1-4 (installation windows 1) 2017 Q2, 2) 2017 Q3, 3) 2017 Q4, and 4) 2018 Q1): In our initial review, we noted that Supplier A's exclusion criteria were imposed as a safeguard to internal validity threats. We still believe this choice was reasonable, but note that these exclusion criteria (e.g. excluding home moves, supplier switches, and customers who do provide sufficiently frequent actual meter reads) do end up excluding more than 50% of the original sample.[40]
- Study 5 (2020 Q3 installation window). We are concerned that the installation window and post-installation-window includes COVID-19 affected months. Our concern is that the effect of smart meters may be different during COVID-19 months compared with non-COVID-19 months – but the magnitude and even direction of the COVID-19 influences are difficult to predict without other (and ideally larger) studies. For instance, households may be more attentive to their in-home display units as they are more likely to be at home. In addition, the exclusion rate is particularly high – up to 90% of potential smart meter customers are excluded for various reasons in this analysis.

Rigour and precision: 4/5 for Studies 1-4; 2/5 for Study 5

- Studies 1-4: Supplier A conducted almost the same consumption analysis each time, so we view their studies as akin to pre-specified studies. The analyses were conducted using large sample sizes.
- Study 5: We give a lower score to Study 5 due to a smaller sample size.

Our decision: Our decision was to include in the evidence synthesis Studies 1-4 (installation windows 2017 Q2, 2017 Q3, 2017 Q4, and 2018 Q1) but exclude Study 5 (installation window 2020 Q3).

Supplier B

Internal validity: 3.5/5

Similar to Supplier A's method, Supplier B's analysis compares changes in energy consumption before and after installing a smart meter between smart-meter customers and a 'matched' traditional-meter customer. Matching ensures both groups are similar in various dimensions such as pre-installation consumption and geographical location. However, Supplier B conducted matching more coarsely (using very wide consumption bands and different installation dates), which results in a less closely matched sample.

External validity: 3.5/5

We believe Supplier B's sample exclusion criteria are reasonable, and guard against threats to internal validity and rigour. Overall, the description leads us to assume that the study sample is similar to

[39] We also have some concerns about the 2020 Q3 study's internal validity, given that people who were at home to accept a smart meter during Q3 of 2020 may differ in important unobservable ways from traditional-meter customers with whom they are close matches on observable characteristics – for example, they may have recently started working from home, which would affect their post-installation energy consumption.

[40] Note that this issue is not specific to Supplier A – we believe the general issue affects every other supplier's analysis as well. Also note that the threat is problematic insofar as home-movers, supplier-switchers, and customers with inconsistent actual meter reads may have systematically different energy consumption patterns from those included in the sample.

Supplier B's customer base – albeit with exclusions that are similar to most of the other supplier studies, such as the exclusion of frequent switchers.

Rigour and precision: 4/5

Outcome measures are straightforward and logical. The sample size is large (approximately 10k smart- and 10k traditional-meter customers in the 12-month analysis). While there is no pre-analysis plan, the analysis is sufficiently straightforward that we are not concerned about manipulation of analysis parameters.

Supplier B conducted analyses examining the smart meter impact on consumption 6, 12, 18, and 24 months after installation. Based on conversations with Supplier B, we believe the 12-month analysis is the most robust. This is for a few reasons. First, it was the main focus of their analysis, it features a full year of post-installation consumption (and therefore the same months as pre-installation consumption – whereas the 6- and 18-month analyses feature different seasons pre- versus post-installation, as we discuss below). In addition, it features a robust sample size (in contrast to Supplier B's smaller-sample analyses focusing on smart meter impacts over longer periods of time). For this reason, we make the 12-month analysis the main focus of our review here.

Our decision: Our decision was to include Supplier B's 12-month estimates in the evidence synthesis.

Supplier C

Internal validity: 3/5 for Studies 1 and 2 (2017 and 2018 installation windows); 2/5 for Study 3 (2019 window)

- Studies 1 and 2: We believe that Supplier C's guards against statistical biases in estimating the smart meter impact from differences in some customer characteristics (specifically, ones that do not vary over time). However, the lack of matching makes the smart-meter customers and traditional-meter customers less comparable. Moreover, the outcome measure is more prone to measurement errors that reduce the magnitude of the estimated impact of smart meters.
- Study 3: As we examined the data provided by Supplier C, we found evidence that smart-meter customers and traditional-meter customers had diverging trends in their consumption prior to smart-meter installation. This suggests that the parallel trends assumption is likely to be violated, which would bias estimates from a difference-in-differences approach.

External validity: 3/5 for Studies 1 and 2; 2/5 for Study 3

- Studies 1 and 2: We believe Supplier C's sample exclusion/inclusion criteria are sensible; many of them mirror suggestions we made in our energy consumption analysis guidance. They have sampled customers from all PES regions, forming a sample that is geographically representative. To guard against internal validity issues due to their outcome measure (which smooths consumption estimation based on previous consumption), the analysis was conducted with just three months of pre-installation data and three months of outcome data. (Specifically, the supplier used the three months directly preceding installation and a 3-month period 7-9 months post-installation.)
- Study 3: We are concerned that the post-installation-window includes COVID-19 affected months, and that the effect of smart meters may be different during these months compared with non-COVID-19 months.

Rigour and precision: 2/5 for Study 1, 3/5 for Studies 2 and 3

- Study 1. Supplier C could not provide information about the sample size for the analysis.

- Studies 2 and 3: Supplier C's analysis involved a large sample size, enabling high precision.

Our decision: Our decision was to include Supplier C's 2018 analysis (Study 2) in the evidence synthesis, but exclude its 2017 (Study 1) and 2019 (Study 3) analyses.

Supplier D

Internal validity: 4/5.

Similar to Supplier A's method, Supplier D's design compares changes in energy consumption before and after installing a smart meter between smart-meter customers and a 'matched' set of traditional-meter customers. By using matching, Supplier D ensured both groups were similar in various dimensions, such as pre-installation consumption and geographical location. While their methodology does not completely eliminate statistical bias, it guards against biases in estimating the smart meter impact from differences in some customer characteristics (specifically, ones that do not vary over time). With that said, the matching was not perfect, meaning the groups were not balanced even on all of the matching criteria (whereas in most matching approaches, they would be).

External validity: 3/5.

Supplier D's analysis is fairly representative of their customer base in 2015 through 2017, but the focus is on early adopters with an overrepresentation of electricity-only customers.

Rigour and precision: 3/5.

Supplier D's analysis involves a large sample size, enabling high precision. We have some concerns about overall rigour. In particular, there are potentially blurry definitions of 'pre' (before installing a smart meter) and 'post' (after installing a smart meter) in the analysis – we found instances of post-installation consumption counting as pre-installation for some customers, and vice versa, which may introduce noise into the analysis. However, our replication of Supplier D's analysis removed this blurriness. We also find high sensitivity of the results to model specification.

Our decision: Our decision was to include Supplier D's analysis in the evidence synthesis.

Supplier E

Internal validity: 0.5/5.

We believe that the cross-sectional comparison between Supplier E's whole customer base's average energy consumption to Ofgem's 'medium' Typical Domestic Consumption Value is not a valid study design in this context. Supplier E is different in meaningful ways from other suppliers, so its customers are unlikely to be representative of other UK customers. The difference in consumption between Supplier E customers and Ofgem's 'medium' Typical Domestic Consumption Value could be driven by differences between their customers and other customers, rather than by the installation of a smart meter.

External validity: 2/5.

The study used the whole customer base to estimate the impact of smart meters. Therefore, the smart-meter group, by design, is representative of the customer base. However, given Supplier E's differences from other suppliers, we caution that its customers are not representative of UK domestic customers in general.

Precision and rigour: 1/5.

Supplier E themselves noted in their study that they did not conduct detailed analysis, but submitted this to comply with the Department for Energy Security and Net Zero's request because the study technically fell within its scope. We agree that, to understand the impact of smart meters on energy consumption, it is not rigorous to compare their smart-meter customers' consumption and Ofgem's 'medium' Typical Domestic Consumption Value for electricity and gas.

Our decision: We decided to exclude Supplier E's estimate from the evidence synthesis.

Supplier F

Internal validity: 3/5.

Supplier F's study design (akin to a difference-in-differences) reduces statistical biases from differences between customers, in particular differences that do not change over time. The approach controls well for bias that could be introduced by changes in weather over time and regional effects, but the lack of a retailer specific matched group of traditional-meter customers for comparison carries more risk of bias from differences between customers that could emerge over time. The outcome measure does not rely on estimated reads, making it less prone to measurement error. This enhances the data quality of the study.

External validity: 3/5.

While the sample exclusion restrictions seem reasonable, they ultimately cause a high exclusion in Supplier F's electricity analysis (exclusion of 60% of the original sample, which is similar to Supplier A) and very high exclusion in their gas analysis (exclusion of 94% of the original sample).

Rigour and precision: 2/5.

Supplier F's analysis uses a large sample of electricity smart meter customers but the demand coefficients used to estimate expected consumption in the absence of a smart meter are based on a relatively small sample customers (typically less than 2,000). We believe the precision of the analysis may be overestimated, potentially quite significantly.

Our decision: Our decision was to exclude Supplier F's analysis from the evidence synthesis. While the study design is sufficiently robust, the standard errors (a measure of accuracy of the estimated impact) are calculated without considering the sample size of the traditional-meter customers, and we do not have sufficient information to correct them ourselves.

Supplier G

Supplier G's study did not contain certain details about the analysis performed. Consequently, we have scored their study conservatively in areas where we are uncertain about their exact analysis approach.

Internal validity: 2/5.

The design of Supplier G's study guards against biases in estimating the smart meter impact from differences in some customer characteristics (specifically, ones that do not vary over time). However, the lack of matching makes the smart-meter customers and traditional-meter customers less comparable.

External validity: 4/5.

The study seemingly examines all Supplier G customers at the time, implying high external validity.

Precision and rigour: 2/5.

The sample size is theoretically large (though note it is never specified), but the outcome measure is unclear and there are no standard errors provided.

Our decision: Our decision was to exclude Supplier G’s analysis from the evidence synthesis because we cannot construct standard errors with the available information.

A4. Constructing standard errors

In some cases, suppliers had estimated the impact of their smart meter by using difference-in-differences but had not constructed its standard error.

Suppliers B and C provided the standard deviation of the **percentage change** in consumption and the sample sizes for the smart-meter and traditional-meter groups. With this information, we can calculate the pooled standard error with no further assumptions as:

$$\sqrt{\frac{((N_{smart} - 1) * sd.perc_{smart}^{2}) + ((N_{trad} - 1) * sd.perc_{trad}^{2})}{N_{smart} + N_{trad} - 2}}$$

Here:

- N_{smart} (N_{trad}) is the sample size for the smart-meter (traditional-meter) group
- $sd.perc_{smart}$ ($sd.perc_{smart}$) is the standard deviation of the percentage change in consumption and the sample sizes for the smart-meter (traditional-meter) group

This is equal to the standard error that you would get from a simple t-test (assuming equal variances) which compares the mean of the percentage change in consumption for the smart-meter group to the mean for the traditional-meter group.

This publication is available from: www.gov.uk/government/publications/impacts-of-smart-metering-roll-out-on-household-energy-use

If you need a version of this document in a more accessible format, please email alt.formats@beis.gov.uk. Please tell us what format you need. It will help us if you say what assistive technology you use.