Department for Energy Security & Net Zero

Domestic NEED 2024 Annex D: Methodology Note

27 June 2024 Accredited Official Statistics

Domestic NEED links together various existing data sources (administrative and commercial) to report on the electricity and gas consumption of households in Great Britain. This data framework, which holds information on individual properties, also demonstrates the real-world impact of installation of energy efficiency measures on electricity and gas consumption. Domestic NEED includes data on:

- annual electricity and gas consumption;
- characteristics of the property (floor area, property age, property type, etc.);
- characteristics of the household (household income, tenure, number of adults etc.); and
- information about the area in which the property is located (local authority, index of multiple deprivation etc.).

The data framework provides the largest source of data available for analysis of gas and electricity consumption and the impacts of energy efficiency measures. Previously, the Department for Energy Security and Net Zero (DESNZ) relied on evidence from surveys and small technical monitoring trials. The first results from the framework were published in June 2011.

This methodology note provides details of how the estimates of domestic electricity and gas consumption by property attributes and household characteristics are produced. It also sets out how the estimates of the impact of energy efficiency measures are derived.

Published alongside this are the following tables and documents:

- **NEED report: Summary of analysis 2024** – analysis of key trends in energy consumption and the impact of installing different energy efficiency measures.
- **Consumption data tables** – gas and electricity consumption estimates for different property attributes and household characteristics.
- **Impact of measures data tables** – estimated consumption savings arising from installation of different energy efficiency measures.
- **Annex A: What is Domestic NEED?** – an introductory overview of the NEED framework.
- **Annex B: Overview of data tables** – a list of all the published tables and their contents.
- **Annex C: Comparisons with other Sources** – a summary of comparisons of NEED outputs with other data sources for quality assurance purposes.

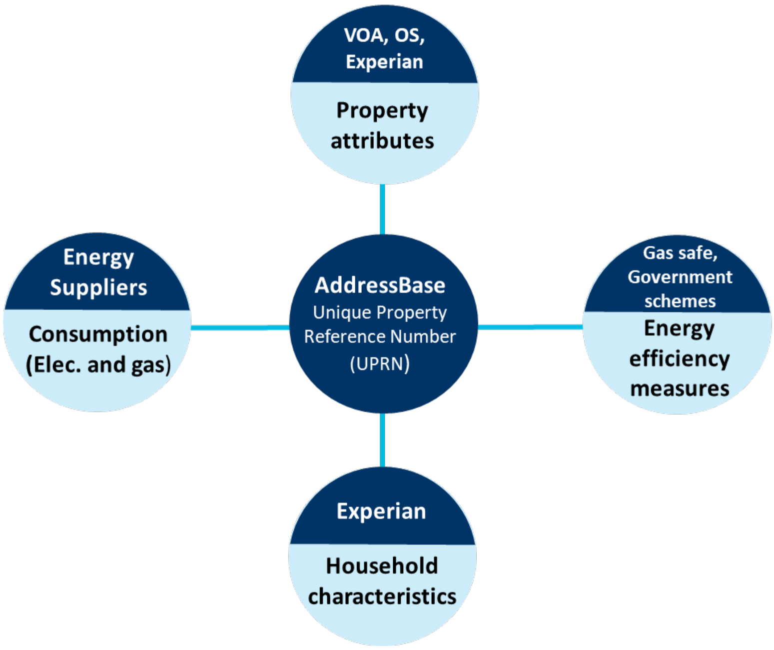
Contents

1. Overview of Domestic NEED ____ 3
    Annual gas and electricity consumption ____ 4
2. Consumption estimates methodology ____ 6
    The properties included in the analysis ____ 6
3. Impact of measures methodology ____ 9
    Overview of Difference in Difference Approach ____ 9
    Step 1: Determining years ____ 12
    Step 2: Excluding properties ____ 12
    Step 3: Intervention and control groups formed ____ 13
    Step 4: Weighting the intervention group ____ 13
    Step 5: Matching intervention properties to control properties ____ 14
    Step 6: Savings estimates for each intervention property ____ 14
    Step 7: Removing savings outliers ____ 15
    Step 8: Savings summary statistics ____ 16
    The final impact of measures estimates ____ 16
    Variations in estimated savings between years ____ 17

1. Overview of Domestic NEED

Domestic NEED is formed from linking various existing datasets (both administrative and commercial) that each provide different information on individual residential properties/ households. How these datasets are linked to each other is illustrated in Figure 2.1 below.

Figure 2.1 – The overall structure of Domestic NEED



The addresses corresponding to each record in a given dataset are matched to addresses in AddressBase (Ordnance Survey's database of all addresses in Great Britain). This is to retrieve the unique property reference number (UPRN) that Ordnance Survey have assigned to the address. This process is referred to as "address matching". Once the records in each dataset have been address-matched, the various datasets can be matched to each other using the UPRN. Note that there are some addresses which cannot be matched confidently to AddressBase, and a proportion of matches that are made are likely to be incorrect. The address matching algorithm being used is estimated to have an error rate of around 1% for domestic addresses.

Valuation Office Agency (VOA) data on domestic properties in England and Wales and Ordnance Survey AddressBase Plus (OSAB) data on domestic properties (UPRNs classed as residential dwellings) in Scotland together form the population of domestic properties in Great Britain, onto which all other data sources used in NEED (including the gas and electricity consumption data) are mapped.

Previously a Scottish Assessors Association (SAA) dataset of domestic properties has been used as the Scotland database of domestic properties. This dataset was provided 10 years ago, and it is therefore likely that some of the attribute information is out of date, and it also doesn't include domestic properties built in Scotland since the dataset was received. This year

(NEED 2024) a database of domestic properties in Scotland has been constructed using Ordnance Survey AddressBase Plus (OSAB). This means SAA data is no longer being used as part of the Scotland NEED estimates.

Since Domestic NEED 2019 (used to produce the 2017 consumption estimates) DESNZ has received a full updated dataset of all domestic properties in England and Wales. Prior to Domestic NEED 2019, a stratified random sample of approximately one in five records was selected from the complete dataset held by VOA. This is why the consumption estimates for England and Wales up to and including 2016 are only based on a small fraction of all domestic properties in England and Wales.

More information on what data is contained in Domestic NEED, how it is built and how it is used, is contained in Annex A: What is Domestic NEED?.

Annual gas and electricity consumption

The precise time periods covered by different years of data differ for gas and electricity consumption. The years according to which the gas meter point level data have been provided to DESNZ are as follows:

- Prior to 2015: October – September (same period as 2015)
- 2015: October 2014 – September 2015
- 2016: mid-July 2016 – mid-July 2017
- 2017: mid-June 2017 – mid-June 2018
- 2018: mid-May 2018 – mid-May 2019
- 2019: mid-May 2019 – mid-May 2020
- 2020: mid-May 2020 – mid-May 2021
- 2021: mid-May 2021 – mid-May 2022
- 2022: mid-May 2022 – mid-May 2023

For this report and all accompanying annexes and tables, when years are mentioned with reference to gas consumption (or savings - with the exception of solar PV), these relate to the “gas years” set out above, rather than calendar years. For electricity consumption, the years cover the months February to January. For example, the 2022 electricity consumption refers to the period 31 January 2022 to 30 January 2023.

Note that annual gas consumption data which are heavily influenced by the weather (as gas is predominantly used for space heating), are weather corrected to remove the effect of year-on-year changes in weather conditions. This is to enable more like for like comparisons of average domestic gas consumption over time. However, the weather correction process used assumes a brief lag between changes in weather and the average consumer’s change in heating behaviour. This is why it does not adjust adequately for rapid and extreme changes in the weather.

With the 2016 consumption figures, Xoserve introduced a new data collection system. Due to this, a large proportion of meters which had not reported for some time had their annual consumption figures updated in the 2017 gas consumption figures. This large update led to an increase in the total gas consumption reported in 2017. With the majority of gas meters now providing timely meter readings, the figures from 2017 onwards are a more accurate reflection of average gas consumption.

More information on meter point level gas and electricity consumption data can be found in the Subnational methodology and guidance note.

2. Consumption estimates methodology

The Domestic NEED publication includes breakdowns of average (mean and median) domestic gas and electricity consumption by various property attributes, household characteristics and information about the area in which the property is located. These estimates are based on the data held in Domestic NEED.

England and Wales estimates are based on data from the Valuation Office Agency (VOA), whereas for Scotland a different dataset is used, Ordnance Survey AddressBase Plus (OSAB). We produce separate estimates for England and Wales (combined), and for Scotland, due to important differences between the VOA and OSAB datasets:

- There are inconsistent categories used in the two datasets. For example, while the VOA data distinguishes between purpose-built flats and converted flats, the OSAB data has a single 'flats' category. The VOA data also contains more detailed categories for property type than the OSAB data.
- For England and Wales, the VOA data is used for all property attributes. For Scotland, OSAB data is used for property type, and modelled data from Experian is used for property age and number of bedrooms.

The properties included in the analysis

The latest consumption estimates (for the year 2022) provided in the Domestic NEED publication are based on one of these four samples:

- The sample for household gas consumption in England and Wales (18.7 million properties)
- The sample for household electricity consumption in England and Wales (23.2 million properties)
- The sample for household gas consumption in Scotland (1.6 million properties)
- The sample for household electricity consumption in Scotland (2.2 million properties)

The steps in the creation of these 4 samples used for the 2022 domestic consumption estimates are as follows:

Step 1 – **All domestic properties:** Start with the dataset used to capture domestic properties (VOA data for England and Wales and OSAB data for Scotland).

Step 2 – Matching energy meters: Matching the energy meters for the given fuel type (gas or electricity) to domestic properties. Up to two meters of the given fuel can be matched to a single domestic property. Properties with no meters or more than 2 meters of the given fuel type are discarded.

Steps 1 and 2 are completed as part of creating Domestic NEED 2024. See Annex A: What Domestic NEED? (Section 2) for more details of how Domestic NEED is created.

Step 3 – Removing invalid consumption figures: Properties are discarded from the sample if any of the following conditions hold:

- No consumption is recorded (or the consumption is missing) for the given fuel type.
- The 2022 consumption figure for the given property appears to be imputed, as it is identical to either of the previous two years.
- The 2022 consumption figure is either implausibly high or implausibly low. The plausible range adopted here for annual gas consumption is 100 to 50,000 kWh. For electricity, the plausible range adopted is 100 to 25,000 kWh.

Step 4 – Removing cases with missing attributes: Properties are removed if the information is unknown for any of the following attributes:

- property type
- property age
- floor area
- the number of bedrooms

This is only done for England and Wales, and not for Scotland. This is because modelled data is used for property attributes in Scotland which means a larger proportion of properties have missing attribute information and we do not want to exclude all of these from the sample.

Step 5 – Removing uncommon property types: Rarer properties such as caravans and annexes are removed from the sample. The property types that remain are:

- For England and Wales: Detached, demi-detached, end-terrace, mid-terrace, bungalows, converted flats and purpose-built flats.
- For Scotland: detached, semi-detached, terrace, flats and “unknown” properties.

Tables 2.1 and 2.2 below set out how the final number of properties in each of the four samples (two for England and Wales, and two for Scotland) is arrived at:

Table 2.1: The formation of the samples of properties for the 2022 average domestic gas consumption estimates

	Number of Properties (millions)		Percentage of starting population	
	England and Wales	Scotland	England and Wales	Scotland
Step 1 - All domestic properties	26.9	2.8	100%	100%
Step 2 - Matching gas meters	20.6	1.8	76%	64%
Step 3 - Removing invalid gas consumption	19.0	1.7	70%	59%
Step 4 - Removing cases with missing attributes	18.8	N/A	70%	n/a
Step 5 - Removing uncommon property types	18.7	1.6	69%	59%
Final sample	**18.7**	**1.6**	**69%**	**59%**

Table 2.2: The formation of the samples of properties for the 2022 average domestic electricity consumption estimates

	Number of Properties (millions)		Percentage of starting population	
	England and Wales	Scotland	England and Wales	Scotland
Step 1 - All domestic properties	26.9	2.8	100%	100%
Step 2 - Matching electric meters	24.8	2.4	92%	86%
Step 3 - Removing invalid electric consumption	23.6	2.2	88%	81%
Step 4 - Removing cases with missing attributes	23.3	N/A	87%	n/a
Step 5 - Removing uncommon property types	23.2	2.2	86%	80%
Final sample	**23.2**	**2.2**	**86%**	**80%**

3. Impact of measures methodology

A difference in difference approach is used to estimate the impact of installing energy efficiency measures. This approach has been used, with refinements, since the first publication of the NEED impact of measures analysis in 2011. This presents the methodology for the impact of measures estimates published as a part of Domestic NEED 2024.

Overview of Difference in Difference Approach

A difference in difference approach for impact of measures works by comparing the gas consumption in properties before and after an energy efficiency measure has been installed (or electricity consumption before and after Solar PV has been installed). The change in consumption for a property which has had the given measure installed (a property in the intervention group) is compared with a similar property which had had no measures installed (a property in the control group).

To do this, intervention and control groups are created for each energy efficiency measure being considered. The **intervention group** contains properties which have received the energy efficiency measure, and no other measure (including smart meters) in the year before, during or in the year after the installation, as far as recorded in Domestic NEED.

The intervention group cannot be analysed in isolation since a change in gas consumption could be due to other factors, such as a change in energy prices. Therefore, a **control group** is created containing properties with no record of an energy efficiency measure installed in the year before, during or in the year after the installation. This is to control for some of the other factors which may have influenced electricity and gas consumption.

A savings estimate is produced for each individual property in the intervention group. This is done by matching each property in the intervention group to a similar property in the control group and looking at the difference in the change in gas consumption between the two properties (as illustrated in Figure 3.1). The saving estimates can then be pooled across all properties in the intervention group to estimate the average impact of the measure being considered.

Figure 3.1: Difference in difference approach

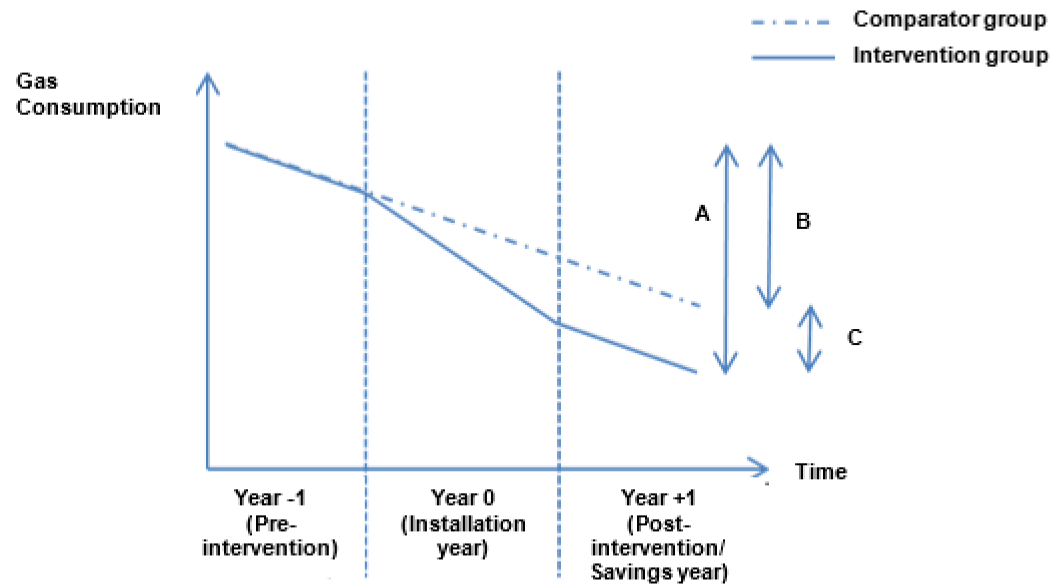
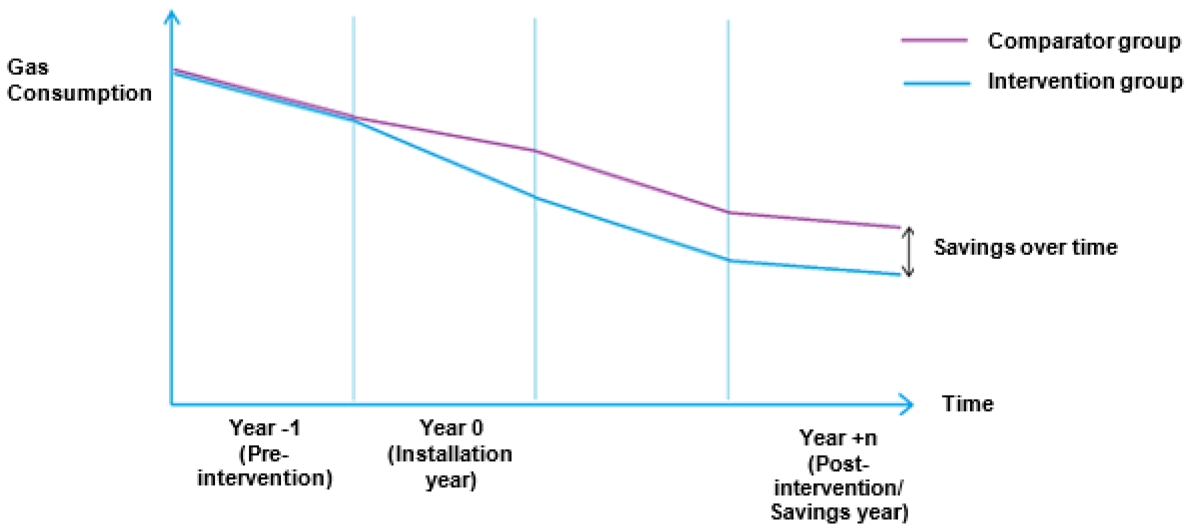


Figure 3.2 below shows how the difference in difference approach can be generalised for estimating savings for multiple years after installation ("Savings in years following installation").

Figure 3.2: Difference in difference approach for savings in years following installation

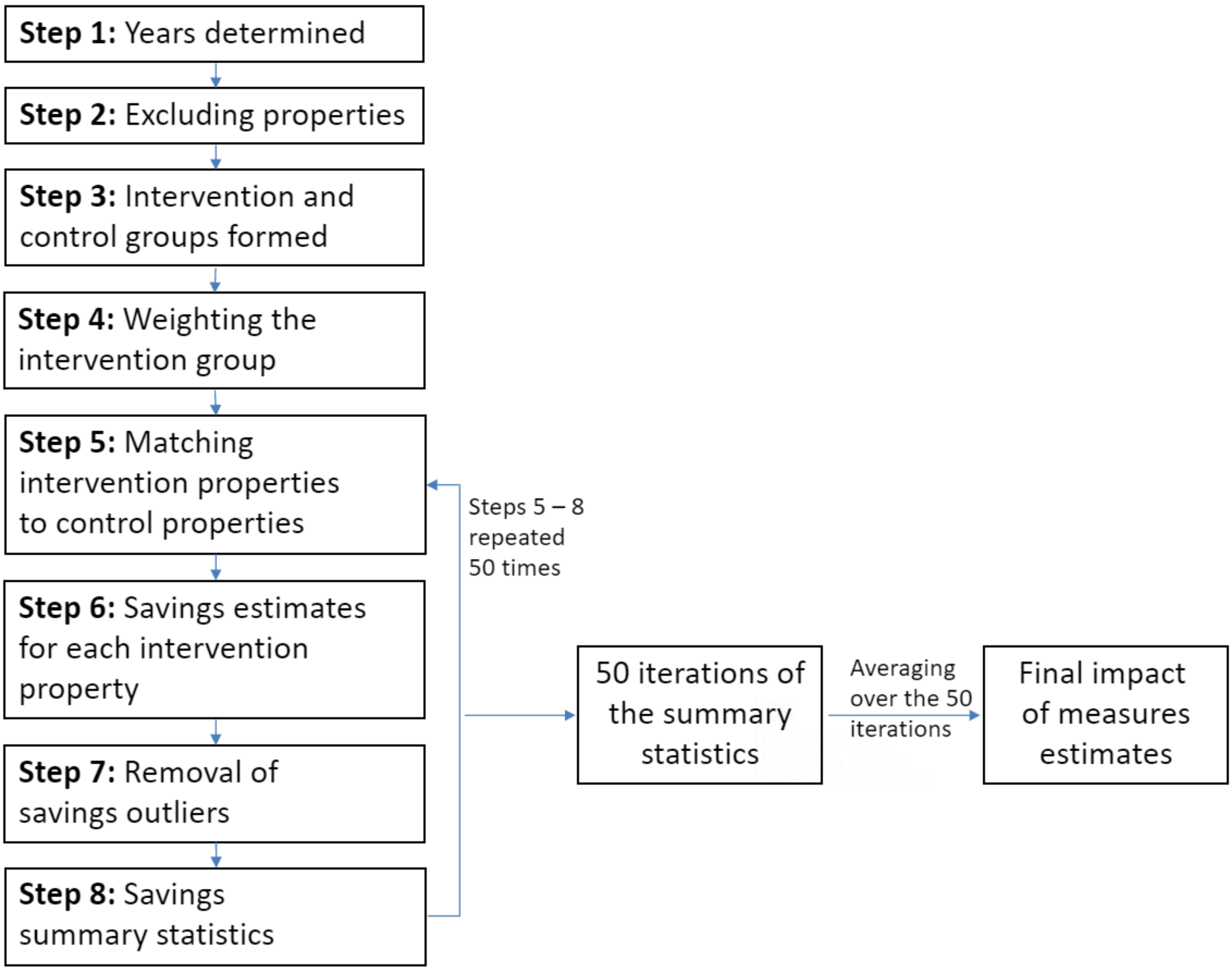


For calculating impact of measures estimates, certain parameters need to be considered, to define which savings are to be estimated:

- The measure type (e.g. loft insulation) and country (England and Wales or Scotland).
- The fuel source for which we want estimate savings. Within the Domestic NEED 2024 publication, gas savings are estimated for all measures except solar PV (for which electricity savings are estimated).
- The installation year of interest (Year 0).
- The savings year (e.g. the first year after installation (Year 0 + 1) or the 5th year after installation (Year 0 + 5)).

Figure 3.3 sets out the steps in producing impact of measures estimates and the remainder of this section provides a more detailed explanation of each step.

Figure 3.3: Difference in difference approach for savings in years following installation



Step 1: Determining years

As an example, suppose savings were being estimated for loft insulation measures installed in 2021, for the first year (2022), then:

- The pre-intervention year (Year 0 - 1) would be the gas year 2020 (mid-May 20 – mid-May 21).
- The intervention year (Year 0) would then run from just after the end of Year -1 to just before the start of Year 0 +1 (mid-May 21 – mid-May 22).
- The 1st year post-installation (Year 0 + 1) would be the gas year 2022 (mid-May 22 – mid-May 23). This would be the savings year.

Step 2: Excluding properties

The impact of measures estimates make use of the consumption figures for both the pre-intervention year (Year 0 - 1) and the savings year (Year 0 + n, when the savings are being estimated for the n^{th} year after installation). We therefore exclude properties from the analysis if any of the following conditions hold for the consumption figures for either the pre-installation year or the savings year:

- The consumption estimates are not within a plausible range. For gas the plausible range adopted is 2,500 – 50,000 kWh. For electricity, the range adopted is 500 – 25,000 kWh.
- The consumption figure is an extreme change on the previous years figure at the same property (more than an 80 per cent increase or more than a 50 per cent decrease).
- The consumption figure is suspected to be imputed. This is the case if either of these conditions hold:
  - The figure is identical to the figures for either of the previous two years at the property.
  - The figure corresponds to a spike in the distribution of the consumption values for the year in question. Such spike values in the distribution are identified by rounding the consumption values for all properties to the nearest kWh, counting the number of properties by each kWh and ordering the counts by the rounded kWh. Values which are more than 300 per cent higher than the two values either side are considered to be spike values.

The following properties are also excluded from the impact of measures analysis:

- Flats (owing to lower address matching rates among these properties) and rare types of property (caravans, annexes and cluster houses)
- Any properties which have missing information for any of the following attributes:
  - Property type
  - Property age
  - Number of bedrooms
  - Number of adults
  - Household income
  - Tenure

In addition to this, properties are also excluded from the analysis if they have had a smart meter installed anytime from the start of the pre-intervention year to the end of the savings year, as the introduction of the smart meter is expected to also reduce energy consumption.

Step 3: Intervention and control groups formed

From the remaining properties, intervention and control groups are formed for the energy efficiency measure in question.

The criteria for inclusion in the intervention group are:

- The property had the given measure installed during the intervention year.
- The property had no other installations recorded between the start of the pre-intervention year and the end of the savings year. This is to try to isolate the effect of the energy efficiency measure in question.
- If the measure is loft insulation, cavity wall insulation or solid wall insulation then properties built from 2000 onwards are excluded. This is because post-1999 properties have specific installation standards regarding loft and wall insulation.

The properties included in the control group are those which have no installations of measures recorded between the start in the pre-intervention year and the end of the savings year.

Step 4: Weighting the intervention group

The type of properties having a particular measure installed may vary between different types of measure and from one installation year to the next. The intervention group is therefore weighted to reflect the total housing stock. This is to enable more like-for-like comparisons between different types of measure and between years.

Table 3.1: Level of attributes used for weighting the intervention group, and for matching control properties to intervention properties

Level	Attributes used					
Level 1	Cons band	Property type				
Level 2	Cons band	Property type	Property age group			
Level 3	Cons band	Property type	Property age group	No. of adults		
Level 4	Cons band	Property type	Property age group	No. of adults	EPC rating	
Level 5	Cons band	Property type	Property age group	No. of adults	EPC rating	Region

Ideally, weighting should be as precise as possible, breaking down the housing stock by as many attributes as possible. However, using too many attributes in the weighting will mean that some subsections of the housing stock are not represented by any properties in the intervention group. So, we adopt the highest level of weighting (see Table 3.1), which allows at least 99.9 per cent of the housing stock to be represented by 1 or more properties in the intervention group. This process is known as elastic weighting.

The weighting factor weights each record in the intervention group based on its frequency – relative to how often it should appear if this group had the same makeup as the total housing stock. For example, if a property type is more common in the intervention group than the housing stock as a whole, then the weighting factor acts to reduce its overall contribution, while if the property type is less frequent in the group than is the case in the housing stock, then the weighting factor acts to increase its contribution.

The weighting factors are calculated using the following formula:

$$\text{Weighting factor for record n } (w_n) = \frac{Housing\ stock\ _g}{Total\ housing\ stock} \times \frac{1}{Sample\ stockg} = A \times \frac{1}{B}$$

Where:

n	*is the record number in the sample*
g	*is the group number (e.g. group 1 = 6th gas consumption decile, semi-detached, built 1965-82, 1 adult, EPC rating D)*
Housing stock$_g$	*is the number in the total housing stock (VOA) in group g*
Sample stock$_g$	*is the number of properties in the intervention group in group g*
A	is the proportion of the total housing stock that are in group g;
B	is the number of properties in the intervention group which are in group g.

Step 5: Matching intervention properties to control properties

Each property in the intervention group is matched to a property in the control group which is as similar as possible. There are several levels of increasingly more stringent criteria that are chosen from for the matching as detailed in Table 3.1.

Ideally, the intervention properties should be matched to control properties which are as similar to them as possible (matching on as many attributes as possible). However, matching on too many attributes can mean that for some properties in the intervention group there may not be a property in the control group meeting the required criteria. So, the highest level of matching (see Table 3.1) is adopted, which allows at least 99.9 per cent of intervention properties to be matched to a control property.

Control properties meeting the required criteria are picked at random for each intervention property. So rerunning Step 5 results in different control properties being paired with the intervention properties.

Step 6: Savings estimates for each intervention property

After each property in the intervention group is paired with a property from the control group, the savings resulting from the given energy efficiency measure are calculated as detailed below:

$$Estimated\ savings\ (kWh) = -(\ Change\ in\ consumption_{intervention\ property} - Change\ in\ consumption_{comparator\ property}\ )$$

$$Estimated\ percentage\ savings = -(\ Percentage\ change\ in\ consumption_{intervention\ property} - Percentage\ change\ in\ consumption_{comparator\ property}\ )$$

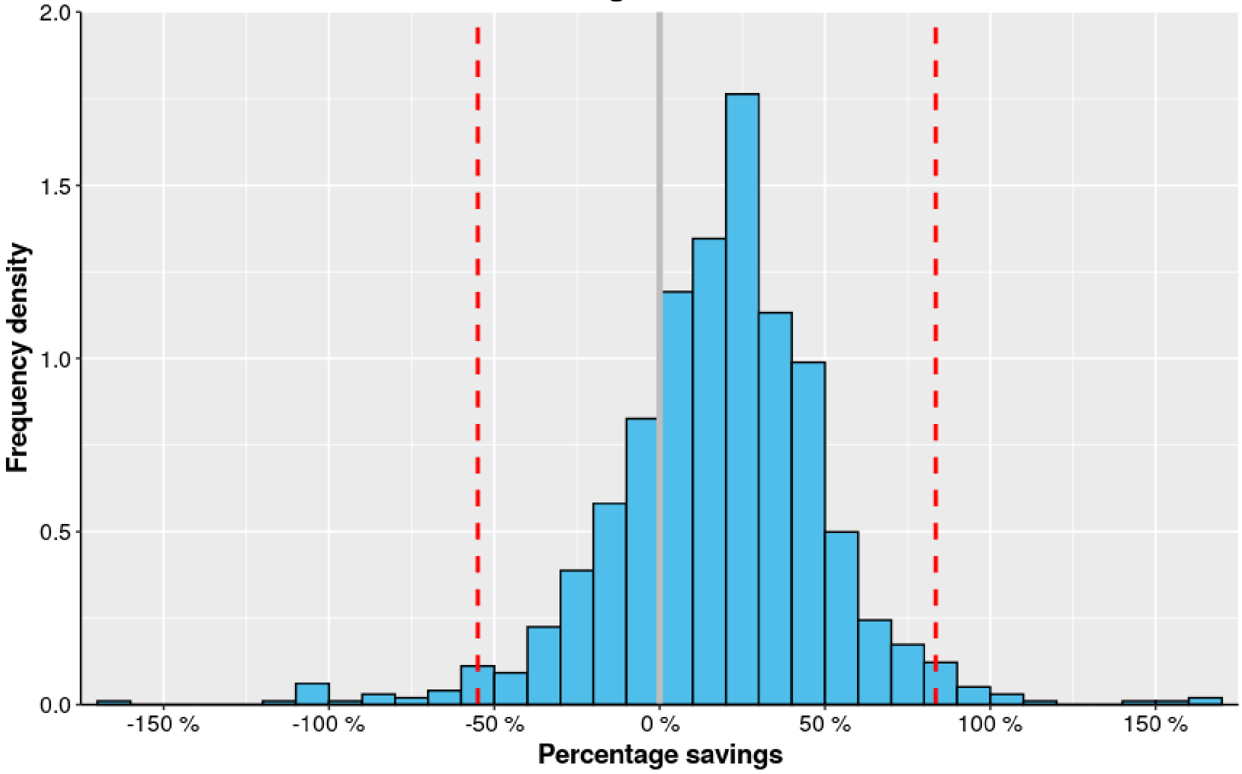
Change here refers to change between the pre-intervention year (Year 0) and the savings year (Year 0 + n).

An example of the variation in the resulting percentage savings estimates for each property in the intervention group is shown in Figure 3.4. The variation probably reflects events that take place and remain unaccounted for (e.g. a change of occupants), as well as uncertainty associated with the pairing of intervention and control properties.

Step 7: Removing savings outliers

The cases corresponding to the top and bottom 2.5 per cent of percentage savings estimates are discarded. To illustrate this Figure 3.4 below shows the distribution of savings for solid wall insulation installed in 2018, where the filter values are obtained using the percentiles at 2.5 and 97.5 per cent.

Figure 3.4: The 95 per cent filtering boundary for the gas percentage savings in 2019 from solid wall insulation installed in England and Wales in 2018



Once the pairs with outlying percentage saving estimates are removed, the weights for the remaining pairs are renormalised (scaled up to ensure that they continue to add up to 1).

Step 8: Savings summary statistics

Summary statistics for the savings are calculated from the savings estimates for each of the remaining intervention/control pairs. For both savings (kWh) and percentage savings the following summary statistics are calculated:

- Weighted mean
- Weighted median – this is the savings value for which the combined weight of all properties with a lower savings figure is 50 per cent and the combined weight of all properties with a higher saving figure is 50 per cent.
- Other weighted quantiles (5 per cent, 25 per cent, 75 per cent, 95 per cent)

The final impact of measures estimates

Steps 5-8 are repeated 50 times. Each re-run of Step 5 results in a different pairing of intervention and control properties, ultimately resulting in different savings summary statistics at Step 8. Averaging the savings summary statistics across the 50 iterations, gives the final impact of measures estimates for the measure in question. The resulting savings figures retain statistical uncertainty, in particular due to the limited population size of the intervention groups for some measures.

Variations in estimated savings between years

The methodology used for the impact of measures analysis has varied slightly from one year to the next, and a full assessment of the uncertainties in the estimates has not yet been made. The savings estimates should therefore be considered as indicative rather than precise. There are a number of factors that are likely to contribute to variations in the estimated savings:

Methodology and data

- Whilst the fundamental methodological approach used for the impact of measures estimates has remained consistent since the creation of NEED, refinements have been made over time. The sensitivity of the estimates to these changes has not been fully assessed and therefore variation seen in estimates may in part be a result of methodological changes. Comparisons between the results published in different years should therefore be treated with caution.

- Measures installed outside of government schemes are mostly “hidden” from this methodology. Properties in our comparator group having energy measures installed which are not known about can lead to saving being underestimated.

Unknown information about the installations or property

- The quality or size of installations may vary between years. For example, trends in the size and quality of solar panels being installed will impact the estimated savings.
- The attributes of the property may vary between years. For example, property extensions will likely increase consumption and could be made alongside installation of energy efficiency measures, masking the savings benefit of those measures.
- The performance of a measure can vary by the brand or subtype of measure. For example, while cavity wall insulation is considered to be a single class of intervention, there are several types of cavity fill (PDF, 162KB), notably bead and mineral wool, which may have different impacts.

Unknown information about the household

- The results may be different for early adopters of novel measures because this self-selecting treated population may have a different energy consumption pattern to other consumers.

- Any variation between the treated populations which is not available in the data cannot be controlled for, for example, age of residents and the number of children in the home.

- Changes in energy consumption behaviour which follow the installation of an energy efficiency measure and may also vary over time and between different types of household. An example is when a household chooses to heat their home to a higher temperature following installation of a measure; this is a known phenomenon called *comfort taking*.

This publication is available from: www.gov.uk/government/collections/national-energy-efficiency-data-need-framework

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