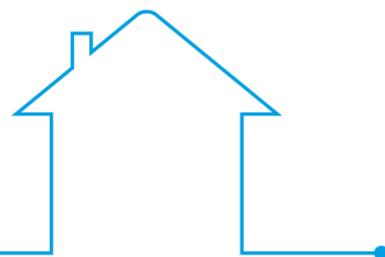




Department
of Energy &
Climate Change

ANNEX A

Quality Assurance



June 2016

Contents

Introduction	3
Matched sample	5
Consumption data	7
Gas consumption data	7
Electricity consumption data	11
Valuations Office Agency Data	16
Introduction	16
Coverage	17
Summary of data and comparison with other sources	18
Conclusion	20
Experian data	21
Introduction	21
Coverage and comparison with other sources	21
Household income	21
Tenure	23
Number of adult occupants	24
Homes Energy Efficiency Database	27
Introduction	27
Coverage	27
Data in HEED	28
Central Feed-in Tariff Register	30
Introduction	30
Coverage and comparison with other sources	30

Conclusion _____ 30

Introduction

The National Energy Efficiency Data-Framework (NEED) is based on data from various sources which are linked together using the Unique Property Reference Number (UPRN). This annex provides information on the quality assurance of data used in the production of analysis using NEED. More information on NEED, including a domestic NEED methodology note and outputs from NEED are available at the following link:

<https://www.gov.uk/government/collections/national-energy-efficiency-data-need-framework>

Outputs from NEED are based on a sample of records selected in order to be representative of the housing stock in England and Wales. Figure A1 shows how the distribution of properties in the 2013 and 2014 NEED sample compares with the Department for Communities and Local Government (DCLG) estimates of the dwelling stock in English regions and Wales in 2011.

Figure A1: Distribution of 2013 NEED sample compared with DCLG dwelling stock estimates



Footnotes:

1. The DCLG dwelling stock data can be obtained from: <https://www.gov.uk/government/statistical-data-sets/live-tables-on-dwelling-stock-including-vacants>.

Table A1 summarises the strengths and weaknesses of each of the main data sources used for the May 2016 NEED publication. The quality and coverage of the data are good but the below strengths and weaknesses should be borne in mind when interpreting the results.

Table A1: Strengths and weaknesses of data in NEED

Data source	Coverage	Strengths	Weaknesses
Consumption data	<ul style="list-style-type: none"> Covers Great Britain 	<ul style="list-style-type: none"> Good coverage of almost all properties (post matching) Data provided by energy suppliers Gas data are weather corrected 	<ul style="list-style-type: none"> Based on billing data (sometimes estimated) Gas and electricity years don't cover calendar year (or the same period as each other) Domestic/non-domestic split (by profile class, or consumption threshold)
Valuation Office Agency (VOA)	<ul style="list-style-type: none"> Covers every property in England and Wales 	<ul style="list-style-type: none"> Excellent coverage—more than 99 per cent of properties in the NEED sample for all variables 	<ul style="list-style-type: none"> No data for Scotland Some data may not be up to date
Experian	<ul style="list-style-type: none"> Data available for each household in the UK 	<ul style="list-style-type: none"> Best source of data at property level on household characteristics 	<ul style="list-style-type: none"> Modelled data with varying accuracy at property level
DECC/Ofgem/HEED energy efficiency datasets (including ECO, ECO excess and Green Deal measures)	<ul style="list-style-type: none"> Covers households in the UK 	<ul style="list-style-type: none"> These datasets contain data for measures installed in homes in the UK including the date of installation 	<ul style="list-style-type: none"> Only covers measures installed through Government schemes; no information on measures installed by households themselves or installed when the property is built Matching of (converted) flats not reliable ECO excess dataset has inconsistent address information and almost a third of the dataset could not be matched to AddressBase
Central Feed-in Tariff Register (FiTs)	<ul style="list-style-type: none"> Covers every property in Great Britain. 	<ul style="list-style-type: none"> Excellent coverage—contains detailed property information on all microgeneration installations receiving FiTs 	<ul style="list-style-type: none"> Excludes microgeneration installations that are not registered for FiTs
Land Registry Price Paid data	<ul style="list-style-type: none"> Covers households in England and Wales 	<ul style="list-style-type: none"> Data available for all residential property sales in England and Wales that are sold for market value. 	<ul style="list-style-type: none"> Excludes: transactions at a discount, including repossession sales, transfers between parties on a divorce, and giving as a gift No personal details of occupants—change in ownership may not correspond to change in occupancy

Matched sample

In order to help increase processing speed, reduce cost and ensure that DECC is not processing more data than necessary, most of the analysis is carried out on a sample of data.

To create the matched sample address information in each dataset has been matched to Ordnance Survey's AddressBase Unique Property Reference Number (UPRN). Table A2 shows the proportion of records on each dataset which could be matched to AddressBase.

Table A2: Match rates¹ (sub-building² match rates in brackets)

Data source	Match rate to AddressBase (%)
Electricity consumption	100 (87)
Gas consumption	97 (93)
VOA property attribute data	100
Experian	98
Central Feed-in Tariff Register	98
ECO measures	97
Land Registry Price Paid data	97
HEED	94
Green Deal measures	92
ECO excess measures	66

The electricity and gas consumption figures cover domestic and non-domestic properties in Great Britain. The core of the NEED data are based on valid meter reads from the electricity suppliers, and therefore every record within the sample also has valid electricity readings and the match rate is 100 per cent. Additionally, the analysis sample was selected from records in the VOA dataset which had a valid UPRN; therefore the match rate for VOA was also 100 per cent. All other match rates were high (between 92 and 98 per cent), with the exception of ECO excess measures which had a match rate of 66 per cent due to inconsistent address information.

¹ Further information about ECO excess can be found on page 2 of the following Ofgem publication: <https://www.ofgem.gov.uk/ofgem-publications/88474/superseded-energycompaniesobligationeco-excessactions-pdf>.

² A sub-building is a separate property within the same building, such as a flat within a converted property or an individual shop within a shopping centre.

A random sample of records was selected from the VOA data. To ensure the sample was representative of properties in England and Wales the sample was stratified by local authority, property age, property type and number of bedrooms³.

The sample selected was originally 19 per cent (approximately one in five records) of the complete property attribute dataset held by VOA. This results in a sample containing approximately 4 million records.

The loss of records through matching to other sources was not evenly distributed. There were more records lost for flats (as these are hard to match to addresses) and consequently proportionately more records were lost in London than other areas of England and Wales. Once the data had been matched to other sources some further records were lost as a result of invalid or missing consumption values (see Section 3 of this annex for details). For 2013 data, 82 per cent had a valid gas consumption value—to be expected as not all properties have a gas meter⁴. As an example, the impact of the loss of these records on the distribution of dwellings in the 2012 sample can be seen later in this annex (page 17 onwards).

³ Property age was recorded as pre-1919, 1919-44, 1945-64, 1965-82, 1983-92, 1993-99 and Post 1999. Property type was divided into detached, semi-detached, end terrace, mid terrace, bungalow, purpose built flat and converted flat. Categories for number of bedrooms were 1, 2, 3, 4, and 5 or more.

⁴ In 2014, it is estimated that 9 per cent of properties in England and 15 per cent of properties in Wales were not connected to the gas network. Source: DECC sub-national estimates of households not connected to the gas network: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/497510/Sub-national_estimates_of_households_not_connected_to_the_gas_network_2014_published_Jan_2016.xlsx.

Consumption data

UK Government has collected and published energy consumption data in the *Digest of UK Energy Statistics* since 1948⁵. A time series on how energy has been used, including data back to 1970, is also published in *Energy Consumption in the UK*⁶. Data at individual meter point level (which makes up the consumption part of NEED) were first obtained in 2004 in order to produce local area estimates of consumption—this work was awarded a Royal Statistical Society Award for innovation in 2010. These meter point consumption data covers both gas and electricity consumption for all homes and businesses within England, Scotland and Wales. Property-level data are not available for other heating fuels such as oil or coal. The electricity and gas data are from energy suppliers' administrative systems and cover around 30 million electricity meters and 25 million gas meters. Consumption data based on these meter level readings are published by DECC down to Lower Layer Super Output Area (LSOA)⁷. This section provides more detail on the gas and electricity consumption data used in NEED.

Gas consumption data

Data collection

DECC obtains annualised consumption estimates for all gas meters in Great Britain. The majority come from Xoserve, the company responsible for the collation and aggregation of gas consumption, with a further (approximately) one million provided by independent gas transporters. DECC is provided with annualised estimates of consumption for all the Meter Point Reference Numbers (MPRNs) in Great Britain based on an Annual Quantity (AQ). The latter is an estimate of annualised consumption using consumption recorded between two meter readings at least six months apart. The estimate is then adjusted to reflect a

⁵ DUKES can be accessed here: <https://www.gov.uk/government/collections/digest-of-uk-energy-statistics-dukes>.

⁶ ECUK can be accessed here: <https://www.gov.uk/government/collections/energy-consumption-in-the-uk>.

⁷ LSOAs are areas containing 400 to 1,200 homes. Further information on sub-national energy consumption statistics can be found in the methodology and guidance booklet: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/298335/Sub-national_methodology_and_guidance_booklet.pdf.

weather correction factor. The AQ for each MPRN represents consumption relating to the gas year—the period covering 1 October through to the following 30 September⁸.

The data are provided with permission from the owners of the Local Distribution Zones (LDZ) network (i.e. the four major gas transporters in Great Britain – National Grid, Scotia, Wales and West Utilities and Northern Gas Networks) and by agreement by the gas suppliers.

There is currently no reliable way to distinguish between gas used by domestic customers and that used by industry/commerce. The gas industry uses a cut off of 73,200 kWh, with customers using less than this assumed to be domestic. This cut off is therefore also used in DECC's published sub-national consumption publication. It means that there are a significant number of businesses (estimated to be around 2 million) misallocated in the sub-national estimates. DECC is looking to resolve this issue but it does not affect NEED data. NEED uses the allocation of property for council tax and non-domestic rates purposes to define which customers are domestic and which are non-domestic.

Coverage

The gas data excludes properties in Northern Ireland due to the differences in market structure⁹. In addition, the data exclude a considerable amount of consumption relating to power stations and very large industrial consumers.

The data represent gas transported through the national distribution system and gas that passes through the National Transmission System into other independently owned local distribution systems. The data do include the 2,500 gas consumers whose consumptions are recorded on a daily basis and who are known as Daily Metered (DM) customers. However, the data exclude any gas passing through other transmission and distribution systems such as those owned by North Sea producers. They also exclude large loads fed directly from the National Transmission System such as those loads used by certain power stations and large industrial consumers. These two exclusions only affect the non-domestic sector.

Data validation

Consistent with the approach taken for sub-national statistics publications, the NEED analysis started by excluding any records with consumption greater than 73,200 kWh as it was assumed they were non-domestic. However, further cleansing and validation was undertaken because of the nature of the analysis within NEED. This means that

⁸ The 2014 gas year runs from 1 October 2013 to 30 September 2014.

⁹ Please see Chapters 5 and 6 of the sub-national methodology and guidance booklet for further information about the differences in market structure in Northern Ireland:

<https://www.gov.uk/government/publications/regional-energy-data-guidance-note>.

consumption figures in NEED are not exactly the same as those in the sub-national consumption publication despite being based on the same source.

Gas consumption in the majority (99 per cent) of households is below 50,000 kWh and the relatively small number of properties with consumption greater than this have been excluded in order to avoid biasing the NEED analysis. This should reduce the likelihood of including non-domestic properties or domestic properties with invalid consumption in the analysis.

At the lower end of the distribution of consumption figures, there is a cluster of values around 1 kWh to 100 kWh. In 2014, 0.5 per cent of gas consumption records in the NEED sample fell into this category. These have also been excluded from all analysis, as they are likely to be households with gas supplies which are not used, or new build properties which are not yet occupied.

In addition, suspected estimated meter readings have been excluded from the data before analysis was undertaken. These take two forms. For any given year, an estimate is assumed if a household has a gas consumption value identical to the previous year. There are also a small number of cases which are suspected to be estimated readings used by suppliers (ie, estimates based on typical consumption for certain household types, sizes, etc). Here, the consumption figures appeared in the data more often than would be expected given the frequency of similar consumption values. Improvements to the data mean there were no assumed estimates on this basis for gas between 2011 and 2014.

The impact of removing these records on the data is small. It results in the mean for NEED being a little lower (3.0 per cent) than it would be if these filters were not applied due to elimination of a relatively small number of records with high consumptions. The median consumption remains almost the same.

Comparison with other sources

To check that the sample used for analysis is consistent with the other estimates of domestic consumption published by DECC—and therefore increase confidence in use of the data—the mean consumption for the NEED analysis sample¹⁰ has been compared with the data published by DECC in DUKES and as sub-national consumption statistics.

Figure A2 shows that when looking at gas consumption, the mean is very similar for the published sub-national statistics and the NEED sample at 14,100 kWh compared with 14,000 kWh respectively. Since both these sources are based on the same input data it is expected that these values should be similar. The mean for the NEED sample is slightly lower than that of the sub-national consumption statistics due to differences in address

¹⁰ The NEED sample covers England only for 2005 to 2010, and England and Wales from 2011 to 2014.

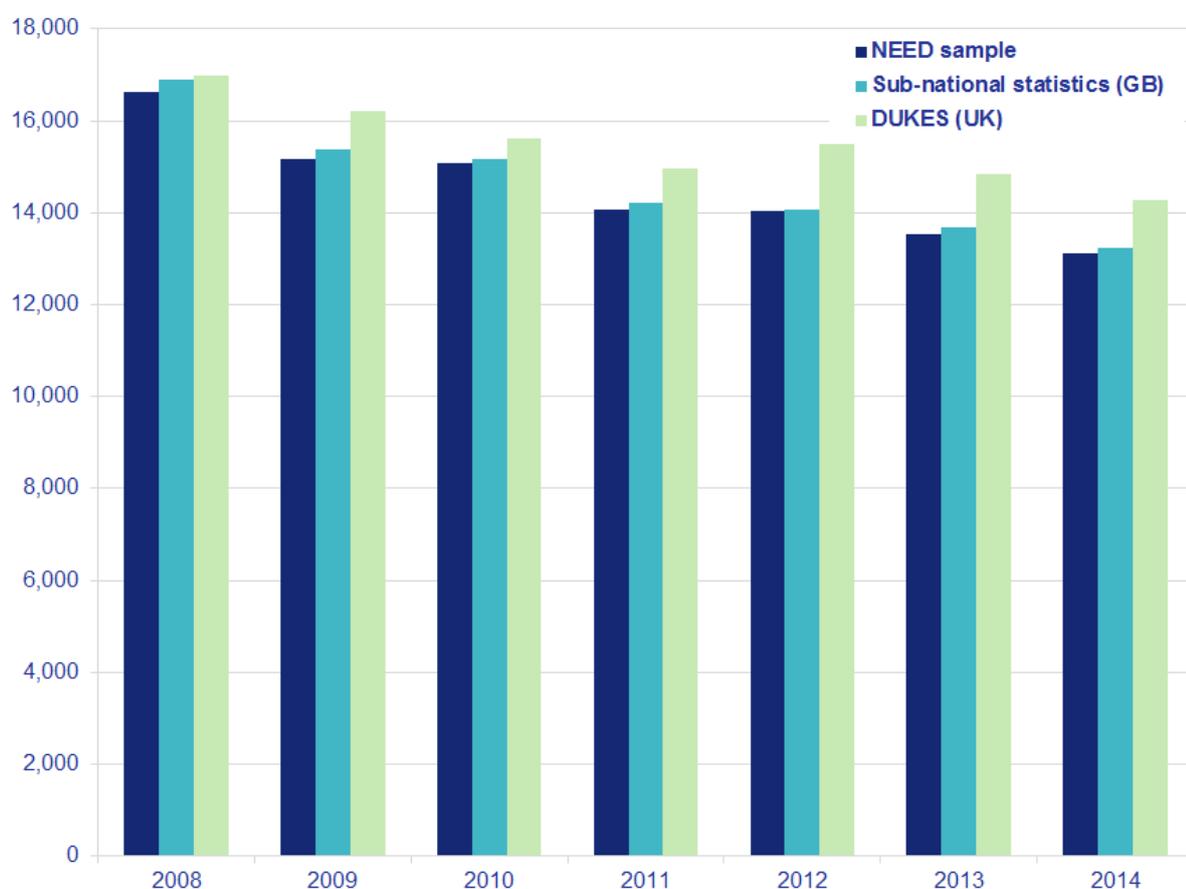
matching and the application of further cleaning and validation routines. Figure A.2 also illustrates that there is more variation when comparing the mean consumption in the NEED sample with that presented in DUKES¹¹; the mean consumption presented in DUKES being 1,200 kWh higher than that of the NEED sample. Gas consumption data in the NEED sample cannot be exactly reconciled with DUKES for a number of reasons:

- the consumption data in DUKES are based on a calendar year whereas the consumption data in the NEED sample cover 1st October to 30th September;
- there are differences in the weather correction method used for DUKES and that for the meter point consumption data;
- the consumption data in DUKES cover the United Kingdom, whereas the NEED sample covers England and Wales;
- the different sources of data used for these publications: DUKES estimates are based on aggregate estimates of energy supplied, while NEED is based on information from gas meters about energy consumed; and
- DUKES data are based on the number of customers. This differs from the number of meter points since it is possible for a property to have more than one meter installed.

¹¹ Source: DUKES estimates per household and weather corrected, as reproduced in Energy Consumption in the UK, table 3.07,

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/238797/domestic.xls.

Figure A2: Comparison of estimates of mean gas consumption (kWh) per household



Electricity consumption data

Data collection

Data are collected with the full co-operation of the electricity industry. Annualised consumption data are generated by the data aggregators, agents of the electricity suppliers, who collate/aggregate electricity consumption levels for each customer meter or Meter Point Administration Number (MPAN). In addition to this, address information for each meter is obtained from the Gemserv meter address file.

The electricity consumption data are generated for both Non-Half Hourly (NHH) meters (domestic and small/medium commercial/industrial customers) and for Half Hourly (HH) meters (larger commercial/industrial customers). There are just under 30 million NHH meters and 120,951 HH meters in Great Britain. For the NHH data, annualised estimates are based on either an Annualised Advance (AA) or Estimated Annual Consumption (EAC). The AA is an estimate of annualised consumption based on consumption recorded between two meter readings. In comparison, an EAC is used where two meter readings are not available and an estimate of annualised consumption is produced by the energy company using historical information and the profile information relating to the meter. These data provide a good approximation of annualised consumption, but do not exactly cover the calendar year. For example, 2014 annualised consumption estimates cover the period from 1st February 2014 up to 31st January 2015. For the half hourly meter consumption estimates, data aggregators are asked to produce a report for each MPAN for the relevant calendar year.

DECC publishes estimates of domestic and non-domestic consumption with aggregate and average consumption figures provided for each local authority. The domestic consumption is based on NHH meters with profiles 1 and 2 (these are the standard domestic and economy 7 meters respectively). Non-domestic consumption is based on NHH meters with profiles 3 to 8 and all HH meters¹². However, it should be noted that these assumptions differ from those used in NEED, where the use of the data mean it is more appropriate to use a slightly different approach to ensuring a property is domestic and has valid consumption. This is described in more detail in the data validation section below.

Coverage

These data cover all of Great Britain. Data for Northern Ireland are currently excluded from the dataset. Some very large industrial consumers with connection to high voltage lines of the transmission system are also excluded. These consumers are classified as Central Volume Allocation (CVA) users who have different arrangements with their electricity suppliers, compared with NHH and HH meter customers. CVA generally accounts for around 2% of electricity sales, but represents a much smaller number of customers.

Data validation

There are differences in the consumption records included in the sub-national consumption publications and those used in NEED.

¹² Non-domestic consumption also includes any nominally domestic meters with consumption of more than 100,000 kWh in a year or meters with consumption between 50,000 and 100,000 kWh with address information which suggests non-domestic use.

Electricity consumption in the majority of households is below 25,000 kWh. The relatively small number of properties with consumption greater than this have been excluded from NEED, in order to avoid biasing estimates. This should reduce the likelihood of including non-domestic properties or domestic properties with invalid consumption in the analysis.

At the lower end of the distribution, there is a cluster of values around 1 kWh to 100 kWh. These have also been excluded from all analysis, as they are likely to represent households with electricity supplies which are not used (or new build properties which are not yet occupied). Unlike in the sub-national consumption statistics, all negative meter readings are also excluded¹³.

In addition, suspected estimated values have been excluded from the data before analysis was undertaken. These take two forms. For any given year, if a household has a consumption value identical to the previous year it is assumed to be an estimate. There are also a small number of values which are suspected to be estimated readings used by suppliers. Similar to the gas estimations, these were assumed on the basis of values that appear in the data more often than would be expected given the frequency of similar consumption values. Two per cent were assumed to be estimates on this basis for electricity in 2013.

The impact of removing these records is small. It causes the mean for NEED to be slightly lower (1 per cent) than it would be if these filters were not applied, due to the elimination of a relatively small number of records with a high consumption. The median remains almost the same.

Comparison with other sources

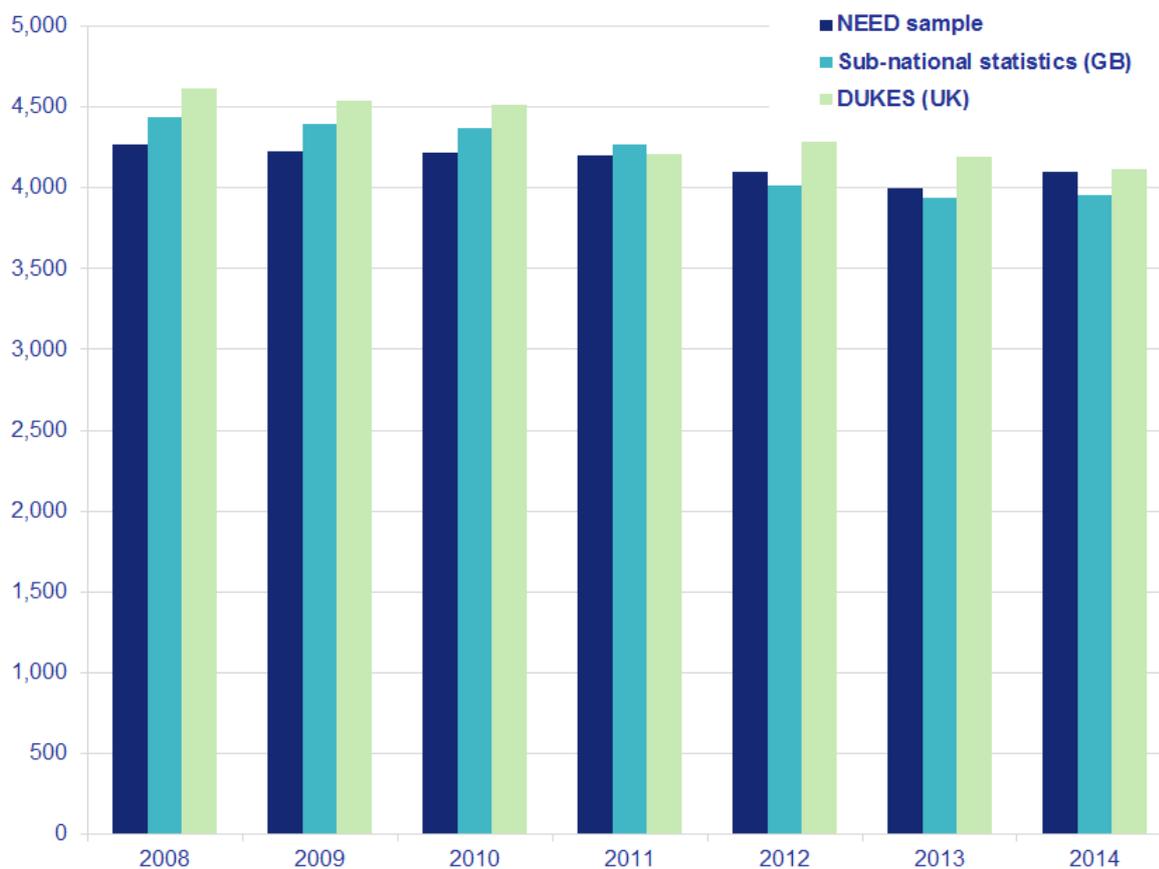
To assess the consistency of the analysis sample with the other estimates of domestic consumption published by DECC—and therefore increase confidence in use of the data—mean consumption for the NEED analysis sample¹⁴ was compared with the data published by DECC in DUKES and in sub-national consumption statistics.

Figure A3 below shows that the mean electricity consumption across all six years presented is similar for all three sources being compared. When looking at consumption in 2013, the difference between the mean electricity consumption in NEED and the other two sources is less than one per cent. The mean 2013 sub-national electricity consumption is lower than that of NEED as it is influenced by a small number of large negative electricity consumption readings which are included in the sub-national analysis, but excluded from the NEED sample.

¹³ As data are based on billed consumption, it is possible that a negative reading is valid if an estimated reading provided in a previous year was too high. However, these reading are not considered valid in NEED.

¹⁴ The NEED sample covers England only for 2005 to 2010, and England and Wales between 2011 and 2014.

Figure A3: Comparison of estimates of mean electricity consumption (kWh) per household



Conclusion

The consumption data are a rich source of data which form the core of NEED. Table A3 summarises the approaches taken towards using the meter point consumption data in domestic NEED and DECC's sub-national estimates.

The differences lead to small discrepancies in mean consumption, but are important to provide confidence in the detailed analysis carried out with NEED, particularly in relation to the impact of installing energy efficiency measures. The comparisons carried out against other data sources confirm that the consumption estimates based on NEED are consistent with other sources.

Table A3: Differences in domestic consumption data^{15,16}

NEED data	Sub-national consumption estimates
<ul style="list-style-type: none"> • The property must be included as a domestic property in the Valuation Office Agency property attribute dataset to be included in domestic NEED analysis. • Gas consumption between 100 kWh and 50,000 kWh. • Electricity consumption between 100 kWh and 25,000 kWh. • Data matched to other sources via unique property reference number (UPRN) at property level. • Suspected estimated readings removed. 	<ul style="list-style-type: none"> • Domestic properties classified based on consumption for gas (less than 73,200 kWh) and profile class for electricity (profiles 1 and 2 are domestic). • Gas consumption below 73,200 kWh. • Electricity consumption below 100,000 kWh and profile class 1 or 2 (including negative readings). • Data assigned to Lower Layer Super Output Area.

¹⁵ Electricity consumption of between 50,000 and 100,000 kWh is reviewed and if it has a likely non-domestic address then it is also excluded from the sub-national domestic estimates.

¹⁶ This means that for the sub-national consumption statistics some properties can be assigned accurately if the street is identified even if the exact property is not known.

Valuations Office Agency Data

Introduction

The Valuation Office Agency (VOA) is the central Government agency responsible for valuing homes for council tax purposes¹⁷. The VOA has had responsibility for valuing properties for council tax since it was first introduced in 1993 and, before then, for the earlier system of domestic rates. Property attribute data was originally introduced in the 1970s in order to provide a simple system for understanding the main features and attributes of a property.

In order to maintain accurate and fair lists of council tax bandings, the VOA needs to keep the information it holds about properties up to date. It does this in a number of ways, including:

- Getting information from the local authority when a home is extended or altered to the extent that planning permission is required.
- Using voluntary questionnaires to enable the occupier to confirm information about a property.
- Obtaining other sources of freely available and publicly published information. For example, a contract with Calnea Analytics to access the Residata website which contains details of properties marketed through mouseprice.com since 2007.

In addition, the VOA will sometimes ask to visit a property when the information it needs cannot be ascertained from other sources. This can often be at the occupier's request; for example when they have challenged the council tax banding of their property and wish the VOA to carry out a review.

There are 16 individual property attributes collected, four of which are used in NEED analysis:

- property type;
- property age;
- floor area (m²); and
- number of bedrooms.

¹⁷ It does not set the level of council tax nor collect the money, which is the task of local government.

Coverage

The VOA Council Tax Database covers properties in England and Wales. Table A4 shows the proportion of properties with missing data for each variable within the VOA dataset and the sample of data used in the latest NEED analysis.

Table A4: VOA property attribute dataset missing data

	Missing data in the dataset (%)			
	Property age	Property type	Number of bedrooms	Floor area
NEED dataset	1.0	0.8	1.5	1.7
NEED sample	0.0	0.0	0.0	0.3

Table A4 shows that the coverage on the VOA dataset is good for all variables. As three of the four variables were used to select the stratified random sample, all records in the sample have information for property age, property type and number of bedrooms. Less than half a per cent of records in the sample lacked information on floor area. These are included as "unknown" in published outputs.

Table A5 below shows the categories of data used in the analysis for each of the VOA variables (categories are those published in the English Housing Survey). In most cases, VOA has more detailed data; the VOA categories have been grouped into the categories used for the purposes of the NEED analysis and presentation of results. Full details of the breakdowns included in the VOA dataset are available on the VOA website¹⁸.

Table A5: VOA property attribute data

	Property age	Property type	Number of bedrooms	Floor area (m ²)
Categories	Pre 1919	Detached	1	1-50
	1919-44	Semi-detached	2	51-100
	1945-64	End terrace	3	101-150
	1965-82	Mid terrace	4	151-200
	1983-92	Bungalow	5 or more	Greater than 200
	1993-99	Purpose-built flat		
	Post 1999	Converted flat		

¹⁸ <http://www.voa.gov.uk/corporate/Publications/DwellingHouseCodingGuide/index.html>

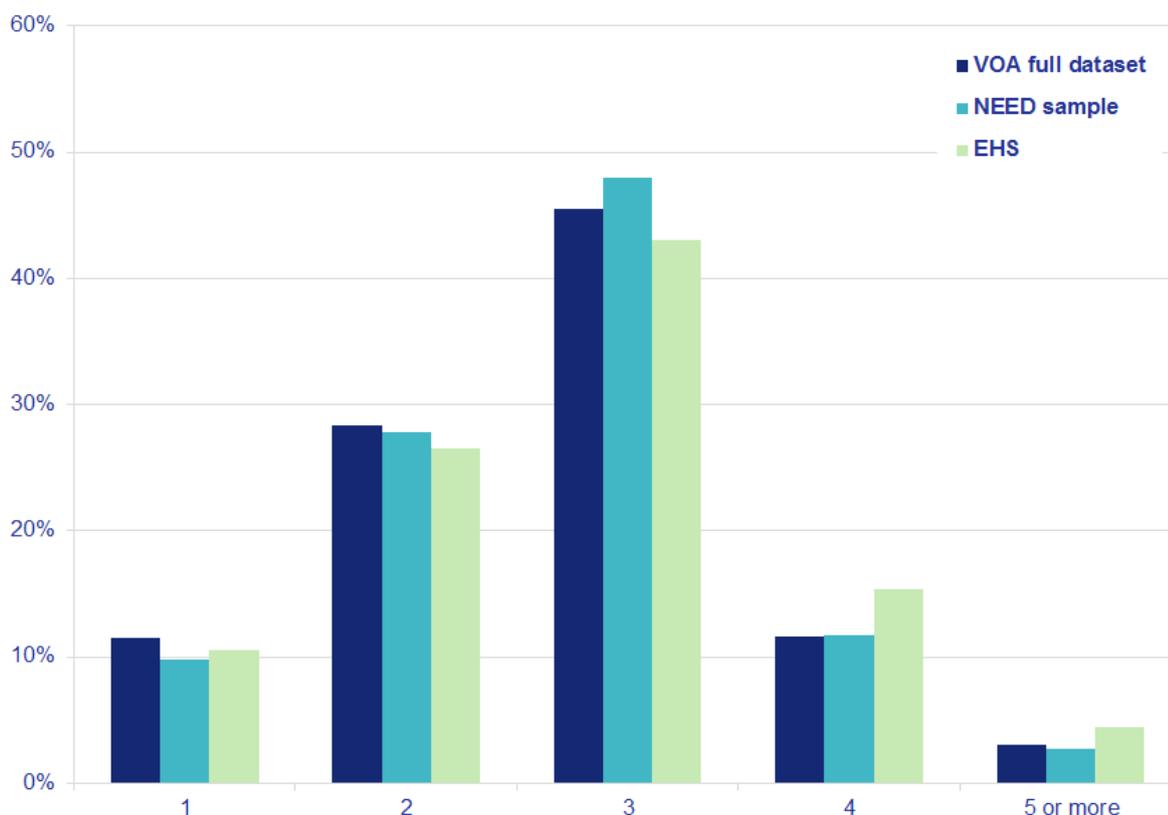
Summary of data and comparison with other sources

This section shows how the data in the NEED sample compare with the distribution of the data in the full VOA property attribute database and with the English Housing Survey (EHS)¹⁹. Differences between the NEED sample and VOA are a result of record loss as described in section 2. The selected sample was created to match the same distribution as the VOA dataset. However, with the removal of six per cent of records (which could not be matched to other sources) this has led to some differences in the distribution of the two datasets.

The EHS will vary compared with the VOA data as it is a sample survey and only covers England, whereas VOA data and the NEED sample cover England and Wales. However it still provides helpful context to validate the VOA data.

Figures A4 to A6 show the proportion of properties in each category for each of the three sources of data for the three variables used to stratify the NEED sample.

Figure A4: Comparison of distributions by number of bedrooms



¹⁹ EHS data are from the English Housing Survey Headline Report 2012-13:

<https://www.gov.uk/government/publications/english-housing-survey-2012-to-2013-headline-report>

Figure A5: Comparison of distributions by property type

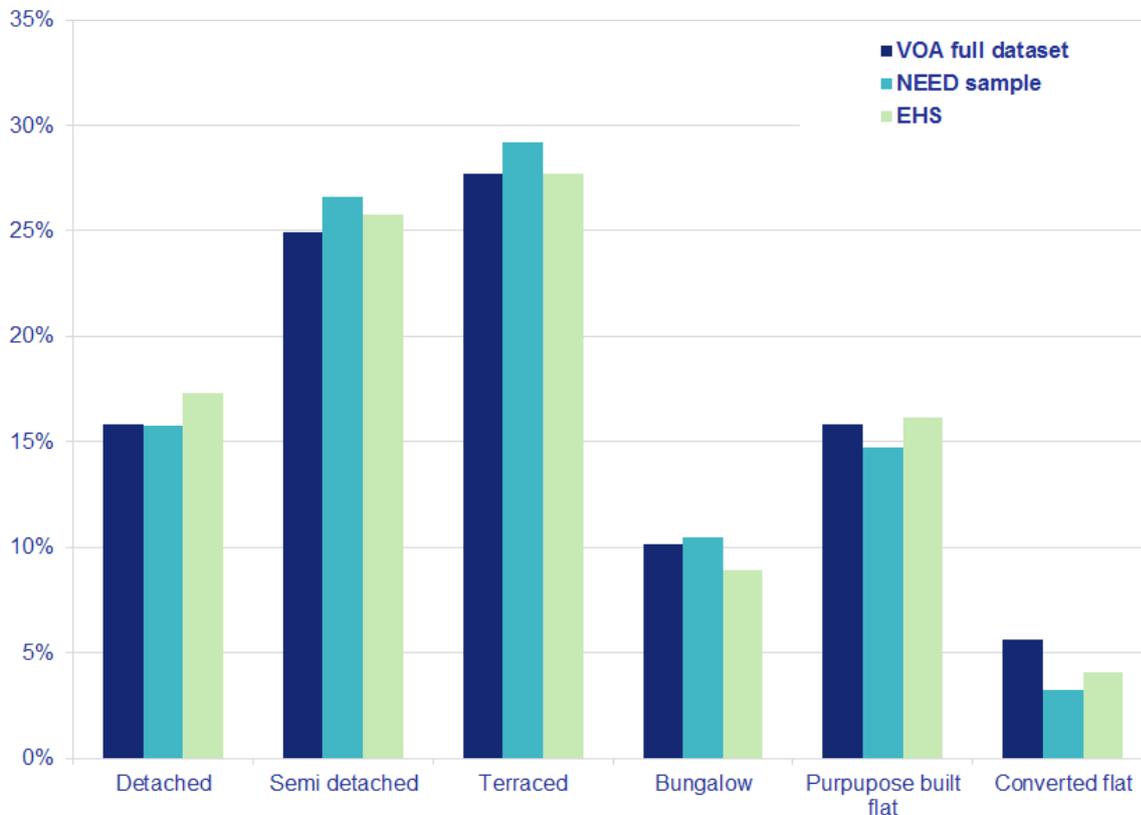
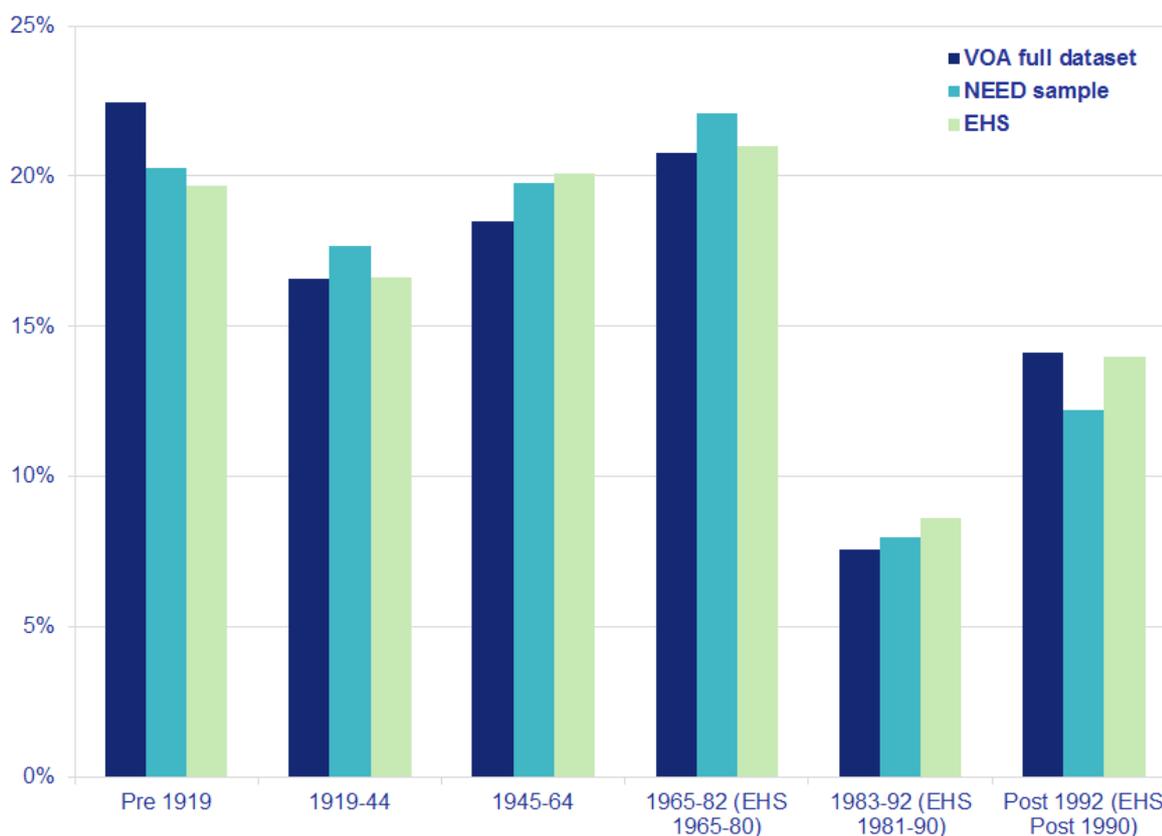


Figure A6: Comparisons of distributions by property age



Conclusion

The data in the VOA property attributes dataset have excellent coverage of properties in England and Wales as demonstrated by the comparisons with other sources in this section. Data from the EHS confirms that the distribution of data is consistent for all property attributes considered in the NEED analysis.

Experian data

Introduction

DECC purchased data from Experian for each property in the UK. Data are modelled by Experian based on other data sources including Experian surveys and aggregate published data (such as the Census). The data purchased by DECC are for 2013. A unique property reference number could be assigned to 95 per cent of records in the dataset provided by Experian, with 98 per cent of records in the NEED sample assigned an Experian record.

Coverage and comparison with other sources

The household characteristics data purchased include:

- household income;
- tenure; and
- the number of adult occupants.

Household income

The household income variable identifies the likely household income for each property. The data are based on results from responses to Experian's consumer survey, which is then used alongside other predictive data (including Experian's person and household level demographics and Mosaic) to build a model. Household income is available in ten income bands which are set out below in table A6.

Table A6: Distribution of households by income band using the full Experian dataset

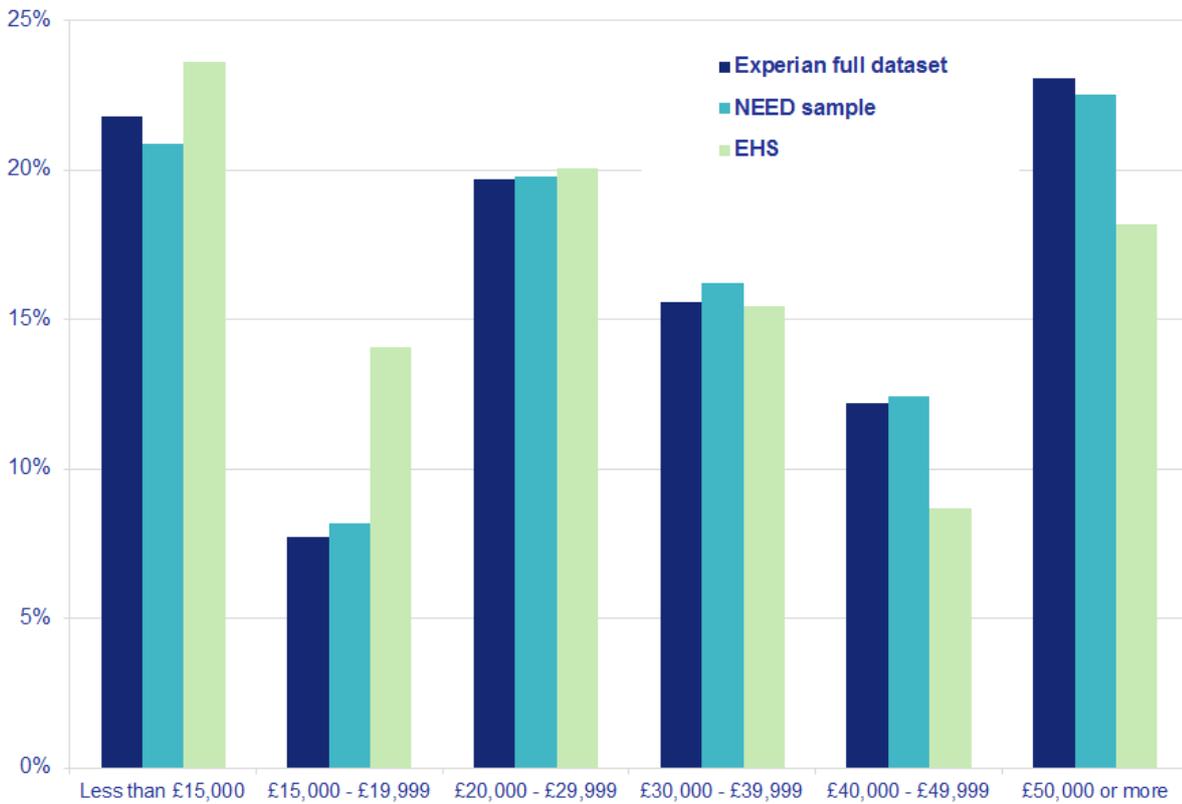
Band	Description	Households (%)
1	Less than £15,000	18.5
2	£15,000 - £19,999	8.2
3	£20,000 - £29,999	21.3
4	£30,000 - £39,999	16.9
5	£40,000 - £49,999	12.7
6	£50,000 - £59,999	7.4
7	£60,000 - £69,999	4.6
8	£70,000 - £99,999	6.4
9	£100,000 - £149,999	2.9
10	£150,000 or more	1.0

When interpreting any analysis of income in the NEED report it should be noted that data for each property are modelled and therefore are indicative of the income a household is likely to have rather than being an actual value for the current occupant of the property.

Experian have made an assessment of the quality of these data and conclude that on average household income is accurate to £16,500. Based on Experian's assessment of the data, 34 per cent of properties are in the correct category and 64 per cent of properties are assigned to within one band of the correct category. Figure A.7 shows how the distribution of income for the Experian dataset and the NEED sample compares with the income reported by the EHS. Note that some of the income categories from the Experian data have been grouped together to allow comparison with the categories used in the EHS.

Figure A7 shows that Experian appears to be under-assigning properties to the lowest income band and over-assigning them at the higher income bands. This is consistent with DECC's understanding that the Experian income data is less reliable at the extremes. However, it should also be noted that the EHS is a survey and therefore subject to variation. Income is a self-reported variable and therefore likely to be less reliable compared to the EHS variables considered in the previous section of this annex which are based on a physical survey on the property carried out by a trained surveyor.

Figure A7: Comparison of distributions by household income band



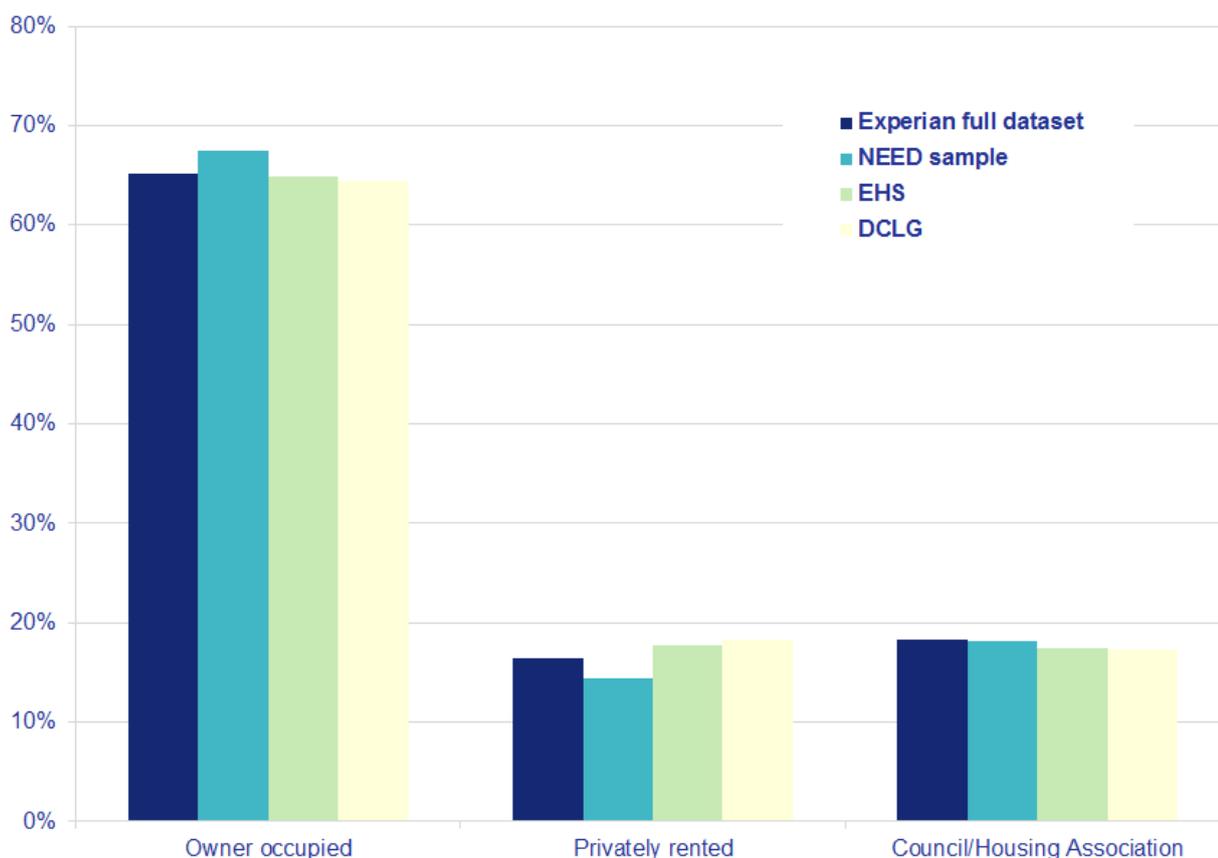
Tenure

Tenure data from Experian allocates each household in the UK to one of three categories; owner occupied, council/housing association or privately rented. The data are based on responses to Experian’s lifestyle survey which are then used to predict the status of all properties. As with the household income variable, a model is used to predict the tenure for each property.

Experian’s assessment of this variable is that 81 per cent of properties are allocated to the correct category. The accuracy of the assignment varies with the group. For example 90 per cent of properties described as owner occupied in Experian’s dataset are actually owner occupied, while only 42 per cent of properties allocated to privately rented are actually privately rented. The equivalent figure is 75 per cent for council/housing association housing.

Figure A8 shows how the Experian data compares with data from other sources at the national level²⁰. It shows that the proportion of properties assigned to each tenure category is similar for all sources. It appears that the Experian dataset as a whole and the NEED sample allocate too many properties to the owner occupied category and too few to privately rented. This is likely to be linked to the loss of flats and properties in London when the NEED sample selected at VOA was matched to other sources.

Figure A8: Comparison of distributions by tenure²¹



Number of adult occupants

The number of adults variable gives the number of adults over 18 living in a household. Experian takes the number of adults information from its ConsumerView database. Experian do not provide an assessment of the accuracy of these data, but note that any discrepancy between the value on the dataset provided and the true value will be due to

²⁰ Note that the Experian full dataset covers the UK, the NEED sample covers England and Wales, and the EHS covers England only.

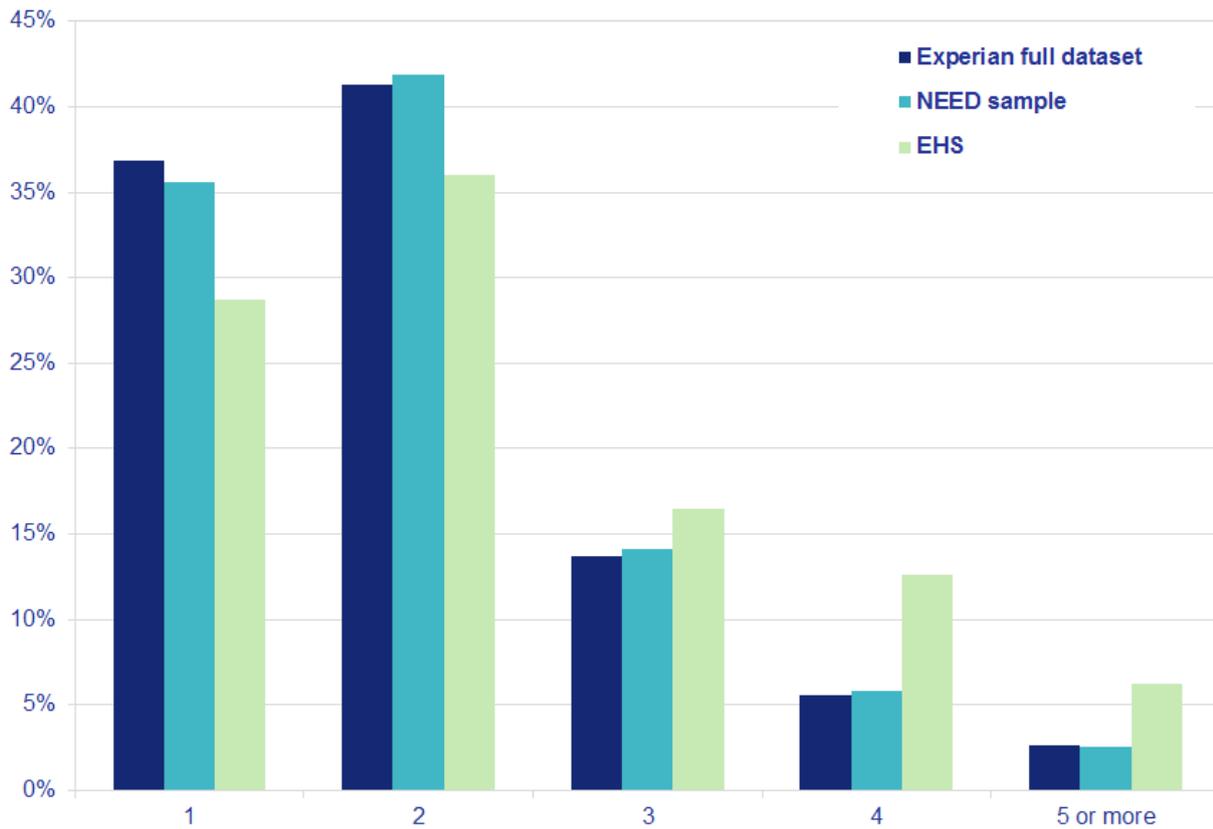
²¹ DCLG estimates from Tables 104 and 106: <https://www.gov.uk/government/statistical-data-sets/live-tables-on-dwelling-stock-including-vacants>

incomplete or erroneous data on the underlying source data. Figure A9 shows how the data in the NEED sample compares with other sources.

The variation in the distribution is likely to be because the EHS estimates are based on household size whilst the Experian data is based on the number of occupants aged 18 and over. This means a household with two adults and two children would be classified as two in the Experian data and four in the EHS. Therefore there are more properties with one or two occupants in the Experian data and more properties with three or more in the EHS.

While the Experian data are valuable since they provides an understanding of the properties in the NEED sample and how consumption and impact of energy efficiency measures vary for different types of properties, it is important that interpretation of results relating to income, tenure and number of adult occupants is in the context of the limitations of the data.

Figure A9: Comparison of distributions by number of adults in a household²²



²² EHS data are based on household size not number of adults. EHS 2011 to 2012, Household Report, Table T1.1: <https://www.gov.uk/government/publications/english-housing-survey-2011-to-2012-household-report>.

Homes Energy Efficiency Database

Introduction

The Homes Energy Efficiency Database (HEED) is a national database developed by the Energy Savings Trust (EST). It was set up to help monitor and target carbon reduction and fuel poverty work. It contains details of energy efficiency and micro-generation installations such as cavity wall insulation and solar hot water. It also includes information on the date each measure was installed. HEED also includes data about property attributes (such as property age and type) and heating systems. However due to coverage and quality these data are not used in NEED.

Coverage

Data have been recorded in HEED since 1995 including activity reported from Government programmes, such as the Energy Efficiency Commitment (EEC) and the Carbon Emissions Reduction Target (CERT), and activity reported by trade associations such as CORGI and FENSA.

Approximately 50 per cent of UK homes have a record in HEED. However there may not be complete information for each of these records. For example, if a measure has been installed through a Government scheme then there may be information on the measure installed but no information on what other energy efficiency measures the property has if they were not installed through a Government scheme.

Table A7 shows how many records in the NEED sample had a HEED record associated with it. It also sets out the number of measures recorded as being installed in properties in the NEED sample for each of the energy efficiency measures included in the analysis. These measures could have been installed in any year from 1995 to 2012.

Table A7: HEED data²³ coverage in NEED sample

Energy efficiency measures	Number of installations	Percentage (%)
HEED record	2,361,570	64
Cavity wall insulation	756,680	20
Loft insulation	759,880	20
Solid wall insulation	109,380	<1

However, there is no information on measures that households have installed themselves (DIY measures) or measures installed at the time the property was built.

Data in HEED

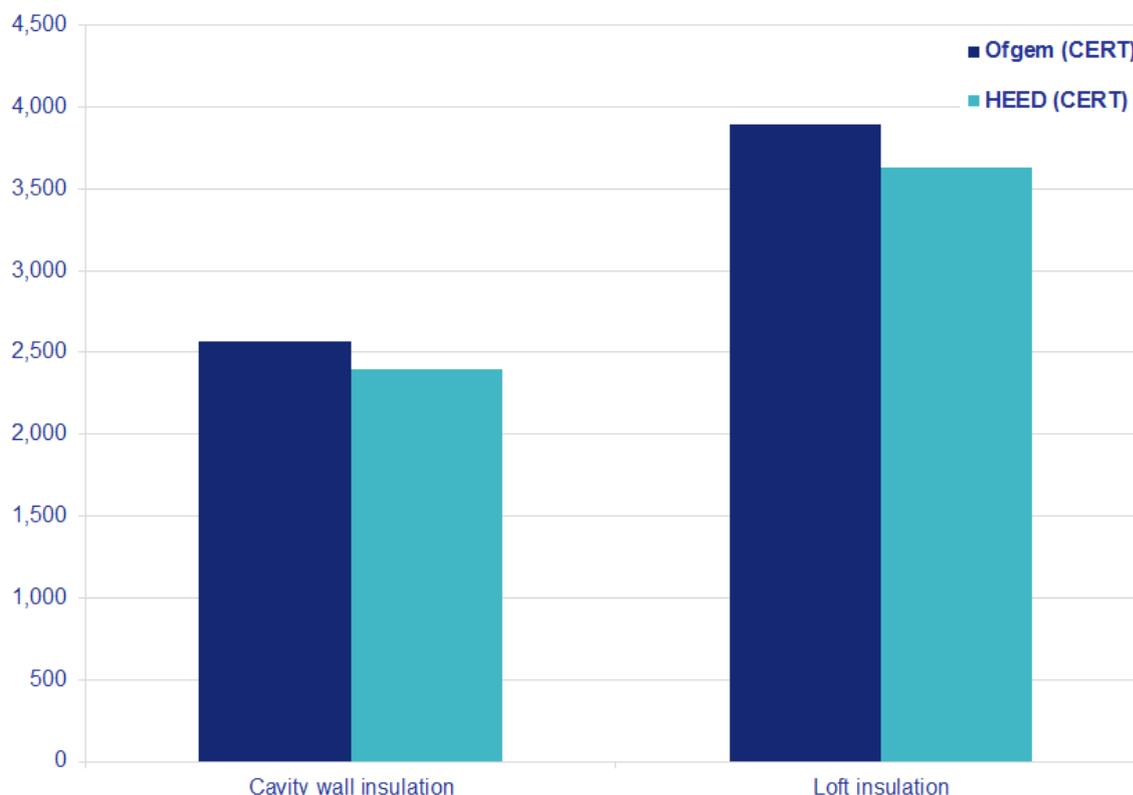
For the majority of data used in NEED analysis, information is based on data EST receive from energy suppliers which cover measures installed through Government schemes (e.g. EEC, CERT). These data undergo validation before they are included in the HEED database. For example, they are checked to ensure that the same measure has not already been installed in the specified household.

Because the majority of measures recorded in HEED are measures installed through Government schemes which are aimed at particular segments of the population, the types of households receiving measures are not representative of the populace or of the housing stock as a whole. However, HEED does have good coverage of properties which have received measures.

Figure A10 below shows the number of properties with cavity wall insulation and loft insulation installed under CERT. It shows that HEED includes a high proportion of the measures reported by suppliers to Ofgem. As no information is known about the specific properties receiving measures reported by Ofgem it is not possible to determine whether there is any bias in the HEED data, but the good coverage means that any bias should be small. The gap between data reported by Ofgem and data included in HEED has reduced now that CERT has ended and final measures have been reported.

²³ This figure is based on the complete NEED dataset rather the NEED sample due to the small number of properties which have had solid wall insulation installed.

Figure A10: HEED and Ofgem reported measures installed under CERT



Coverage of solid wall insulation is not as comprehensive as for cavity wall insulation and loft insulation, although following end of scheme data from CERT and Community Energy Saving Project (CESP) it is now more complete than it has been previously. EST does not publish figures showing how much solid wall insulation was installed under CERT so no comparison with Ofgem is shown. However, solid wall also differs from the other measures as a high proportion of the installations were installed through CESP. Ofgem reports show that approximately 22,000 solid walls were insulated through CERT or CESP in 2011: this compares with approximately 15,000 in the NEED full dataset.

Data relating to boilers come from a wider range of sources, however there are no data available for boiler installations prior to 2009. Typical savings following installation of a boiler are only available for 2009 to 2012 as a complete time series.

When considering the quality of HEED data included in NEED it should also be noted that the installation dates associated with records are of varying quality. This is particularly so for earlier installations of solid wall insulation where it is not possible to distinguish when between 2005 and 2008 measures were installed.

Central Feed-in Tariff Register

Introduction

The Central Feed-in Tariff Register (CFR) is an electronic, web-based system used to manage the Feed-in Tariff (FIT) scheme that Ofgem administers on behalf of DECC. Extracts from the CFR are provided to DECC on a monthly basis, and serve as the basis of a number of statistical publications on the FIT scheme, available at <https://www.gov.uk/government/collections/feed-in-tariff-statistics>. The CFR contains installation-level data on every microgeneration installation that is registered for and receives feed-in tariffs. The data used in this publication are from the March 2015 extract. A unique property reference number could be assigned to 98 per cent of records in the dataset provided by Ofgem.

Coverage and comparison with other sources

The FIT installation data recorded in the CFR include:

- geographical location;
- technology type;
- installed capacity; and
- type of installation (i.e. domestic, community, or commercial).

Conclusion

NEED is a valuable source of evidence on energy consumption and the impacts of energy efficiency measures but its value is dependent on the quality of data used to form it. This annex shows that the quality of these data is generally good, with excellent coverage of the target population. In all cases, the data are broadly consistent with the other sources they have been compared with. At a property level, data from the administrative sources are more reliable than data modelled by Experian. Table A8 summarises the strengths and weaknesses of the data used in NEED.

Table A8: Strengths and weaknesses of data used in NEED

Data source	Strengths	Weaknesses
Consumption data	<ul style="list-style-type: none"> • Covers Great Britain • Good coverage of almost all properties (post matching) • Data provided by energy suppliers • Gas data are weather corrected 	<ul style="list-style-type: none"> • Based on billing data (sometimes estimated) • Gas and electricity years do not cover the calendar year (or the same period as each other) • Domestic/non-domestic split
Valuation Office Agency (VOA)	<ul style="list-style-type: none"> • Covers every property in England and Wales • Excellent coverage—more than 99 per cent of properties in the NEED sample for all variables 	<ul style="list-style-type: none"> • No data for Scotland • Some data may not be up to date
Experian	<ul style="list-style-type: none"> • Data available for each household in the UK • Best source of data at property level on household characteristics 	<ul style="list-style-type: none"> • Modelled data with varying accuracy at property level
Energy efficiency measures (HEED, ECO, GD, RHI and FITs)	<ul style="list-style-type: none"> • Has data for measures installed in homes in the UK including the date of installation 	<ul style="list-style-type: none"> • Only covers measures installed through Government schemes; no information on measures installed by households themselves or installed when the property is built • Matching of (converted) flats not reliable