Department of Energy & Climate Change

Annex A: Quality Assurance

26th June 2014

URN 14D/192

Any enquiries regarding this publication should be sent to us at EnergyEfficiency.Stats@decc.gsi.gov.uk.

Contents

Summary and Introduction

The National Energy Efficiency Data-Framework (NEED) is based on data from a number of 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 A.1 shows how the distribution of properties in the NEED sample compares with the Department for Communities and Local Government (DGLG) estimates of the dwelling stock in English regions and Wales in 2011.

Figure A.1: Distribution of NEED sample compared with DGLG dwelling stock estimates

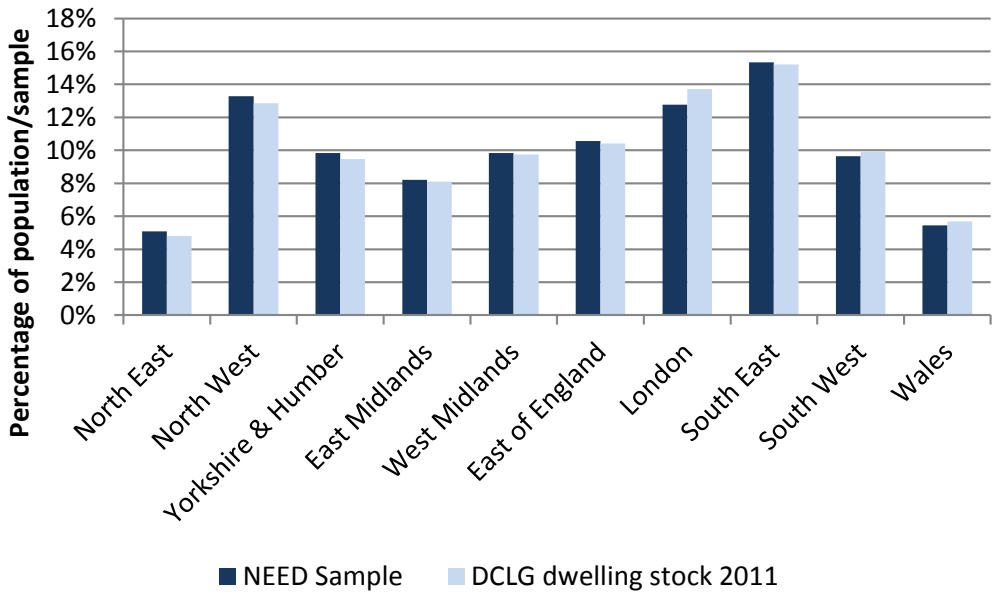


Table A.1 summarises the strengths and weaknesses of each of the main data sources used for the June 2014 NEED publication. The quality and coverage of the data are good but any interpretation of results should be considered in the context of the strengths and weaknesses of each source.

Table A.1: Strengths and weaknesses of data in NEED

Data source	Strengths	Weaknesses
Consumption data	• Covers Great Britain • Good coverage of almost all properties (post matching) • Data provided by energy suppliers • Gas data are weather corrected	• 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
Valuation Office Agency (VOA)	• Covers every property in England and Wales • Excellent coverage – more than 99 per cent of properties in the NEED sample for all variables	• No data for Scotland • Some data may not be up to date
Experian	• Data available for each household in the UK • Best source of data at property level on household characteristics	• Modelled data with varying accuracy at property level
Homes Energy Efficiency Database (HEED)	• Has data for measures installed in homes in the UK including the date of installation	• 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

Matched sample

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

To create the matched sample address information in each dataset was matched to the AddressBase Unique Property Reference Number (UPRN). Table A.2 shows the proportion of records on each dataset which could be matched to AddressBase. The electricity and gas consumption figures quoted cover domestic and non-domestic properties in Great Britain. The analysis sample was selected from records on the VOA dataset which had a valid UPRN; therefore the match rate for VOA was 100 per cent. All other match rates were high.

Table A.2: Match rates (sub-building[1] match rates in brackets)

Data source	Match rate
Electricity consumption	94% (87%)
Gas consumption	97% (93%)
Experian	95%
VOA property attribute data	100%
HEED	94%

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[2], property type[3] and number of bedrooms[4].

The sample selected was originally 17 per cent (one in six records) of the complete property attribute dataset held by VOA, this results in a sample containing approximately 4 million records. Matching this sample to consumption information held by DECC resulted in the loss of six per cent of records[5], so the final matched dataset is 16 per cent of the VOA property attribute dataset, or 3.7 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. Loss of records can also occur where there was no consumption record provided in the original dataset being matched.

Once the data had been matched to other sources some further records were lost as a result of invalid or missing consumption values in the data (see Section 3 of this annex for details). For 2012, 94 per cent of the sample had a valid electricity consumption value and 75 per cent had a valid gas consumption value. The lower rate observed for gas is expected as not all properties have a gas meter[6]. The impact of the loss of these records on the distribution of dwellings in the sample can be seen in Section 4 of this annex.

---

[1] 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.

[2] Property age consists of pre-1919, 1919-44, 1945-64, 1965-82, 1983-92, 1993-99 and Post 1999.

[3] Property type consists of detached, semi-detached, end terrace, mid terrace, bungalow, purpose built flat and converted flat.

[4] Number of bedrooms consists of 1, 2, 3, 4, and 5 or more.

[5] All VOA records had a UPRN assigned in order to be included in the sample selection. The loss of records resulted from the relevant UPRN not being present in any other datasets.

[6] It is estimated that 10 per cent of properties in England and 15 per cent of properties in Wales were not connected to the gas network in 2012. Source, DECC sub-national gas consumption fact sheet: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/267774/sub_national_gas_consumption_factsheet_2012.pdf

Consumption data

UK Government has collected and published energy consumption data within the Digest of UK energy statistics since 1948[7]. A time series including data back to 1970 on how energy has been used has also been published in Energy Consumption in the UK[8]. Data at individual meter point level (which makes up the consumption part of NEED) was 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. This meter point consumption data covers both gas and electricity consumption for all homes and businesses within England, Scotland and Wales. Property level data for other heating fuels, such as oil or coal, is not available. The electricity and gas data are from energy supplier's administrative systems and cover around 30 million electricity meters and 25 million gas meters. Consumption data based on these meter level readings is published by DECC down to Lower Layer Super Output Area (LSOA)[9], these are areas containing 400 to 1,200 homes in each group. This section provides more detail on the gas and electricity consumption data used in NEED.

Gas consumption data

Data collection

DECC obtain 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 the independent gas transporters. DECC are provided with annualised estimates of consumption for all the MPRN's (meter point reference numbers) in Great Britain based on an Annual Quantity (AQ). An AQ 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 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 September[10].

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 agreement by the gas suppliers.

The gas data has no reliable domestic and industrial/commercial flag to enable an accurate split between these sectors. The gas industry use 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. This means that in the sub-national estimates, there are a significant number of businesses (estimated to be around 2 million) misallocated. This is an issue which DECC are looking to resolve, but does not impact on data in NEED. NEED uses the allocation of property for council tax and non-domestic rates to define which customers are domestic and which are non-domestic. There are some limitations to this approach, particularly

---

[7] https://www.gov.uk/government/collections/digest-of-uk-energy-statistics-dukes

[8] https://www.gov.uk/government/collections/energy-consumption-in-the-uk

[9] 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.

[10] The 2012 gas year runs from 1 October 2011 to 30 September 2012.

for the non-domestic sector, however, it is believed to be considerably more accurate than the crude approach used by the gas industry.

Coverage

The gas data excludes properties in Northern Ireland, due to the market structure. In addition, a considerable amount of consumption relating to power stations and some very large industrial consumers is not included in the data.

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 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. It also excludes large loads fed directly from the National Transmission System (such as certain power stations and large industrial consumers). These two exclusions only affect 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 is assumed they are not domestic. However, because of the nature of the analysis undertaken in NEED further cleansing and validation was undertaken. This means that consumption figures in NEED are not exactly the same as those in the sub-national consumption publication; despite being based on the same source.

The gas consumption in the majority of households is below 50,000 kWh, when looking at 2012 gas consumption 1 per cent of records in the NEED sample had a consumption over 50,000 kWh. In order to avoid the relatively small number of properties with consumption over 50,000 kWh having a disproportionate impact on the analysis in NEED these have been excluded. 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 are a cluster of values around 1 kWh to 100 kWh when looking at 2012 gas consumption 0.5 per cent of records in the NEED sample had a consumption of less than 100 kWh. 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, if a household has a gas 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. 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; improvements to the data supplied mean there were no assumed estimates on this basis for gas in 2011 and 2012.

The impact of removing these invalid records on the data is small. It results in the mean for NEED being a little lower (3.1 per cent) than it would be if these filters were not applied, due to elimination of a relatively small number of records with a high consumption. The median 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 increased confidence in use of the data –

mean consumption for the NEED analysis sample[11] has been compared with the data published by DECC in DUKES and sub-national consumption statistics.

Figure A.2: Comparison of estimates of mean gas consumption (kWh) per household

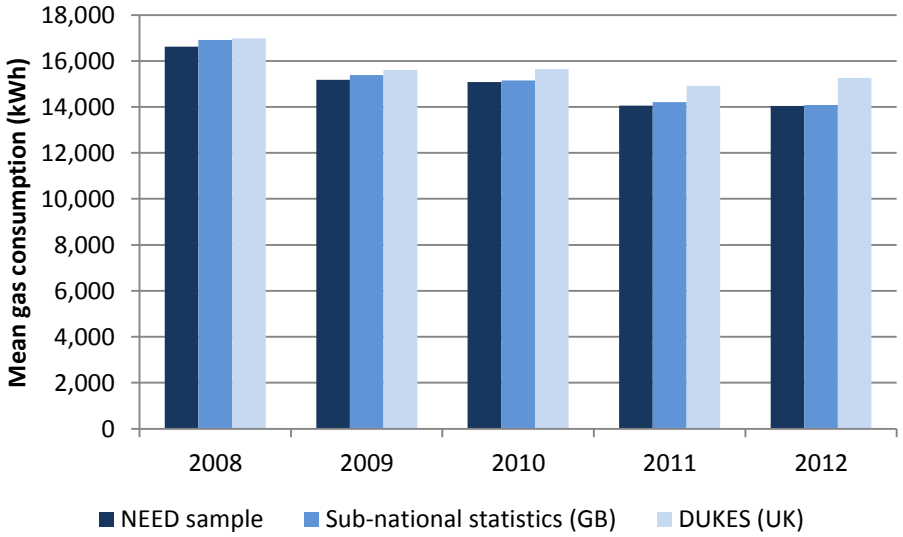


Figure A.2 shows that when looking at gas consumption, the mean is very similar for the published sub-national statistics and the NEED sample, 14,100 kWh compared to 14,000 kWh respectively. Since both these sources are based on the same input data it is expected that these values should be similar. The mean for the NEED sample is slightly lower than that of the sub-national consumption statistics due to differences with address matching and further cleaning and validation routines applied.

Figure A.2 also illustrates that there is more variation when comparing the mean consumption in the NEED sample to that presented in DUKES[12], with the mean consumption presented in DUKES being 1,200 kWh higher than that of the NEED sample. Gas consumption data utilised in the NEED sample cannot be exactly reconciled to DUKES for a number of reasons:

- the consumption data used in DUKES are based on a calendar year whereas the consumption data in the NEED sample covers 1 October to 30 September;
- differences in the weather correction method used for DUKES and the meter point consumption data;
- consumption data in DUKES covers the United Kingdom, whereas the NEED sample covers England and Wales;
- 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 on energy consumed; and
- DUKES data are based on number of customers which will differ from the number of meter points since it is possible for a property to have more than one meter installed.

[11] The NEED sample covers England only for 2005 to 2010, and England and Wales in 2011 and 2012.

[12] Source: DUKES estimates per household and weather corrected, as produced in Energy Consumption in the UK, table 3.07, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/238797/domestic.xls.

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 MPAN (meter point administration number). 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 123,000 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 cover exactly the calendar year. For example, 2012 annualised consumption estimates cover the period from 27 January 2012 up to 26 January 2013. For the half hourly meter consumption estimates, data aggregators are asked to produce a report for each MPAN for the relevant calendar year.

DECC publish estimates of consumption with domestic/non-domestic splits, 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 meters (and 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). However, it should be noted that these assumptions differ from those used in NEED, where the use of the data means 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 CVA or Central Volume Allocation users, who have different arrangements with their electricity suppliers, compared to NHH and HH meter customers. CVA generally accounts for around 2% of electricity sales, but a much smaller number of customers.

Data validation

Due to the nature of the analysis undertaken in NEED further cleansing and validation was undertaken to decide on what should be considered valid data specifically for this analysis. This means that consumption figures in NEED are not the same as those in the sub-national consumption publication, but are very similar.

Electricity consumption in the majority of households is below 25,000 kWh. In order to avoid the relatively small number of properties with consumption over 25,000 kWh having a disproportionate impact on the analysis in NEED these have been excluded. 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 are a cluster of values around 1 kWh to 100 kWh. These have also been excluded from all analysis, as they are likely to be households with electricity supplies which are not used (or new build properties which are not yet occupied). Unlike the sub-national consumption statistics, all negative meter readings are also excluded[13].

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. 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. There were no assumed estimates on this basis for electricity in 2012.

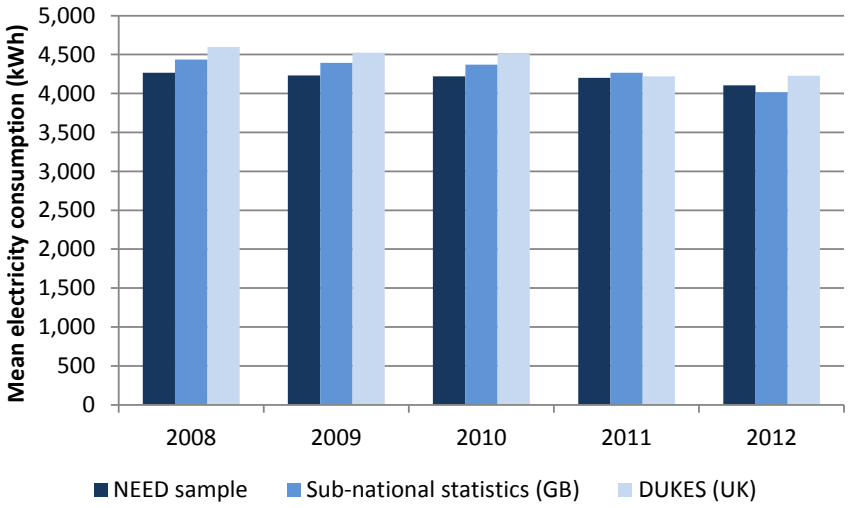
The impact of removing these invalid records on the data is small. It means the mean for NEED is slightly lower (5.0 per cent) than it would be if these filters were not applied, due to the elimination of a relatively small number of records with a high consumption. The median remains almost the same.

Comparison with other sources

To check that the sample used for analysis is consistent with the other estimates of domestic consumption published by DECC – and therefore increased confidence in use of the data – mean consumption for the NEED analysis sample[14] has been compared with the data published by DECC in DUKES and sub-national consumption statistics.

Figure A.3 below shows that the mean electricity consumption across all five years presented is similar for all three sources being compared. When looking at consumption in 2012 the difference between the mean electricity consumption in NEED to the other two sources was around 100 kWh. The mean 2012 electricity consumption for sub-national is lower than that of NEED as it is being 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.

Figure A.3: Comparison of estimates of mean electricity consumption (kWh) per household



[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 had been 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 in 2011 and 2012.

Conclusion

The consumption data are a rich source of data which form the core of NEED. Table A.3 summarises the approaches taken towards using the meter point consumption data in domestic NEED and DECC's sub-national estimates.

Table A.3: Differences in domestic consumption data

NEED data	Sub-national consumption estimates
The property must be included as a domestic property on the Valuation Office Agency property attribute dataset to be included in domestic NEED analysis.	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 between 100 kWh and 50,000 kWh.	Gas consumption below 73,200 kWh.
Electricity consumption between 100 kWh and 25,000 kWh.	Electricity consumption below 100,000 kWh and profile class 1 or 2[15] (including negative readings).
Data matched to other sources via unique property reference number (UPRN) at property level.	Data assigned to Lower Layer Super Output Area[16].
Suspected estimated readings removed.	

The differences lead to a small differences 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.

Valuations Office Agency Data

Introduction

The Valuation Office Agency (VOA) is the central government agency responsible for valuing homes for council tax purposes[17]. 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

[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.

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

domestic rates. Property attribute data was originally introduced in the 1970's 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.
- 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 (detached, semi-detached etc.);
- property age;
- floor area (m^2); and
- number of bedrooms.

Coverage

The VOA Council Tax Database covers properties in England and Wales. Table A.4 shows what proportion of properties are missing data for each of the variables used in this report. It shows the number of properties missing data for the VOA dataset as a whole (covering England and Wales) and for the sample of data used in the latest NEED analysis.

Table A.4: VOA property attribute dataset missing data

	Property Age	Property Type	No. of Bedrooms	Floor Area
Missing - Full Dataset	1.0%	0.8%	1.5%	1.7%
Missing - NEED Sample	0.0%	0.0%	0.0%	0.3%

It shows that for all variables the coverage on the VOA dataset is good. 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 did not have information on floor area. These are included as unknown in published outputs.

Table A.5 below shows the categories of data used in the analysis for each of the VOA variables (determined by categorise published in the English Housing Survey). In most cases VOA have more detailed data; the VOA categories have been grouped to the categories set out

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 website[18].

Table A.5: 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		

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 on 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 the lost records described in section 2, the selected sample had exactly the same distribution as the VOA dataset, however the six per cent of records which could not be matched to other sources were not evenly distributed and have 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 A.4 to A.6 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.

[18] http://www.voa.gov.uk/corporate/Publications/DwellingHouseCodingGuide/index.html

[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 A.4: Comparison of distributions – number of bedrooms

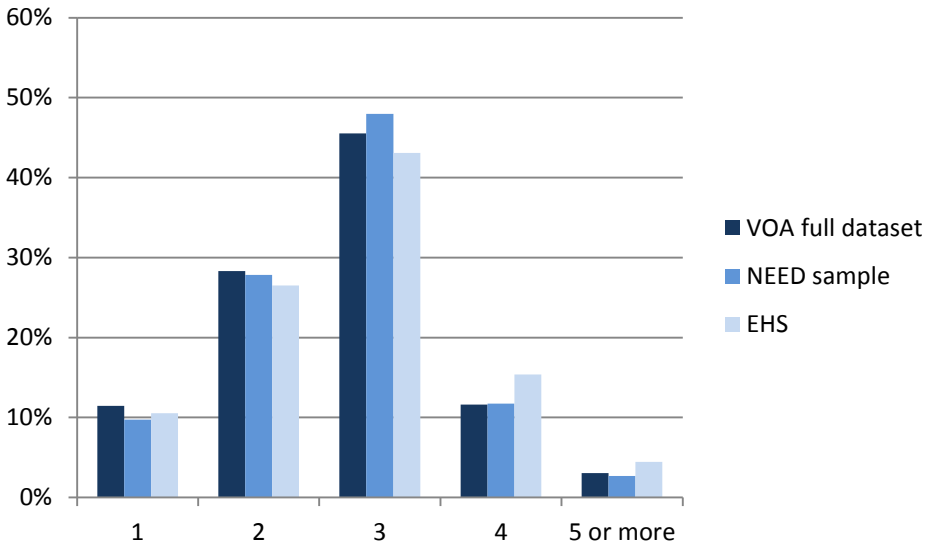


Figure A.5: Comparison of distributions – property type

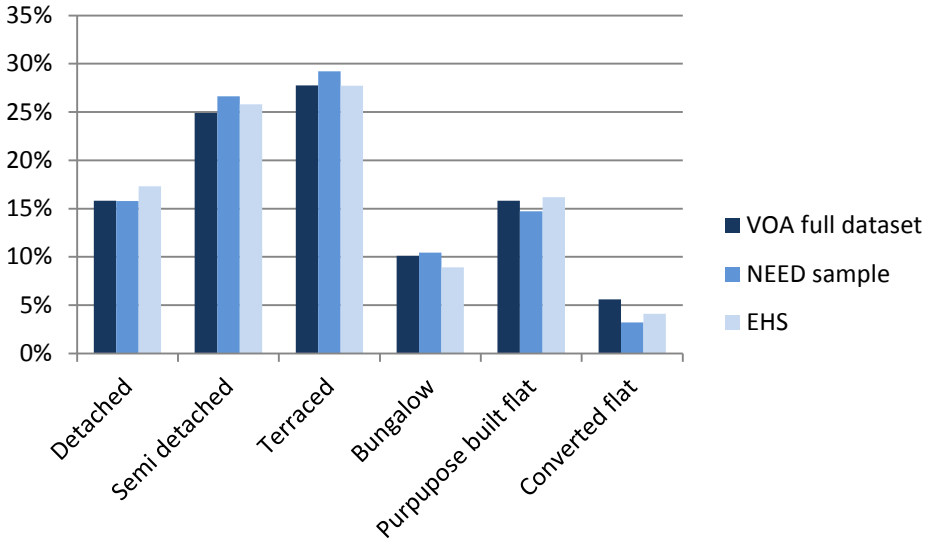
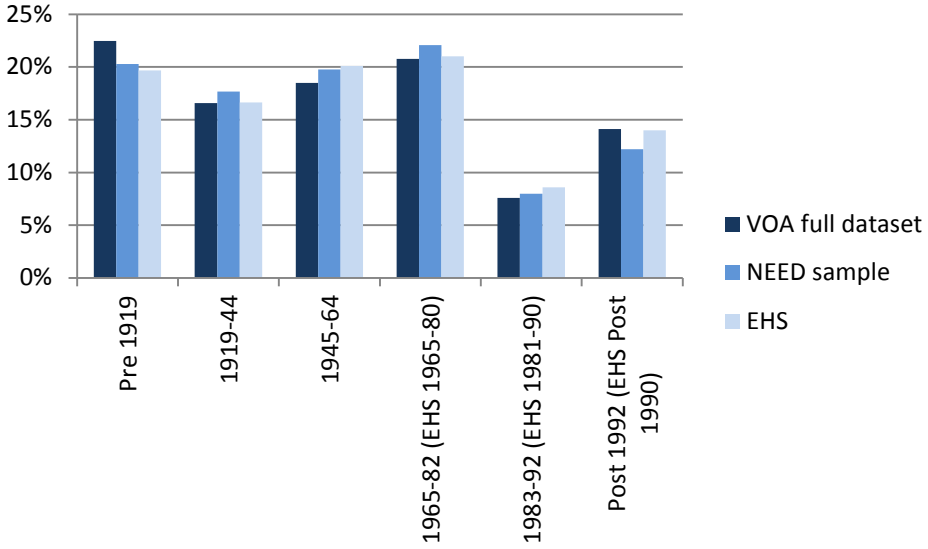


Figure A.6: Comparisons of distributions – property age



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 2011. 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
- 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 A.6.

Table A.6: Distribution of households by income band using the full Experian dataset

Band	Description	Households (%)
1	Less than £15,000	21.79%
2	£15,000 - £19,999	7.71%
3	£20,000 - £29,999	19.68%
4	£30,000 - £39,999	15.56%
5	£40,000 - £49,999	12.18%
6	£50,000 - £59,999	7.44%
7	£60,000 - £69,999	4.71%
8	£70,000 - £99,999	6.60%
9	£100,000 - £149,999	3.19%
10	£150,000 or more	1.13%

It should be noted when interpreting any analysis of income in the NEED report that data for each property are modelled and therefore are indicative of the income a household is likely to have rather than 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 A.7: Comparison of distributions – household income band

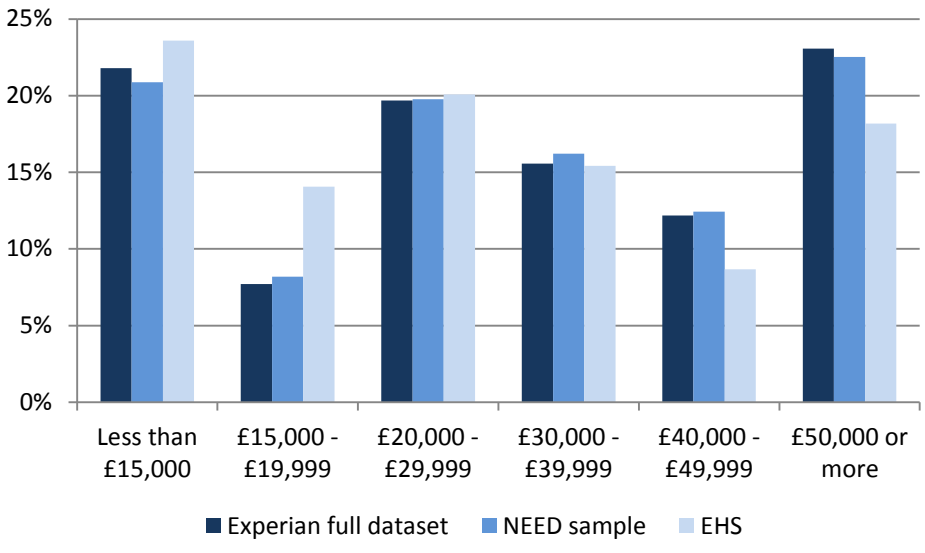


Figure A.7 shows that Experian appears to be under assigning properties to the lowest income band and over assigning them at the higher income bands. This is consistent with DECC's understanding that the Experian income data is less reliable at the extremes. However, it should also be noted that the 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.

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 suggests that 81.1 per cent of properties are allocated to the correct category. The accuracy of the variable varies within groups. 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. For council/housing association housing the equivalent figure is 75 per cent. Figure A.8 shows how the Experian data compares with data from other sources at the national level[20].

Figure A.8: Comparison of distributions – tenure[21]

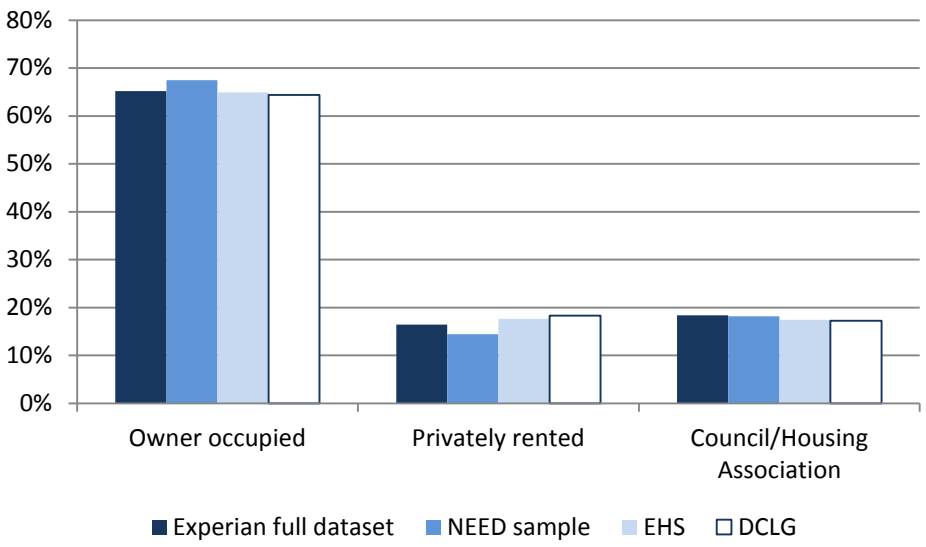


Figure A.8 shows that the proportion of properties assigned to each tenure category is similar for all sources. It appears that the Experian dataset as a whole and the NEED sample allocate too many properties to the owner occupied category and too few to privately rented. This is likely to be linked to the loss of flats and properties in London when the NEED sample selected at VOA was matched to other sources.

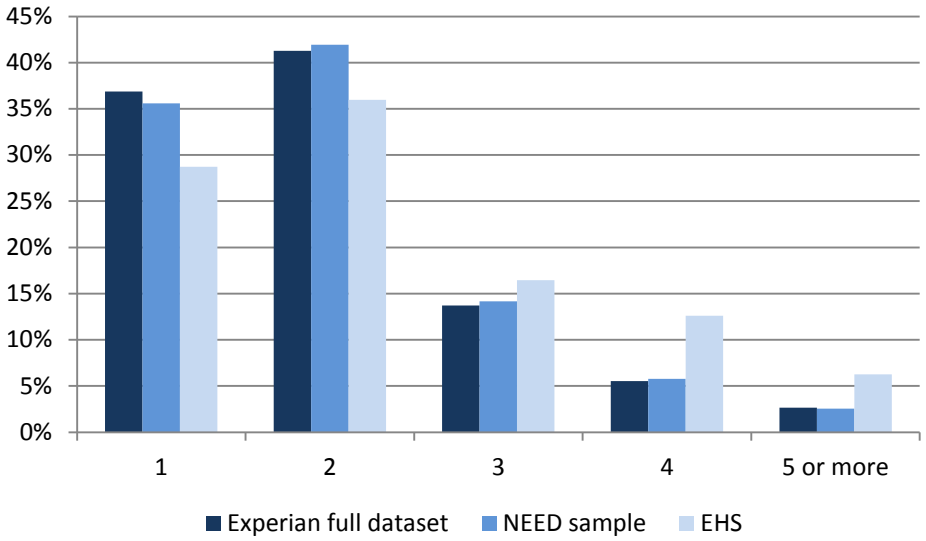
[20] Note that the Experian full dataset covers the UK, while the NEED sample covers England and Wales, and the EHS covers England only.

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

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 incomplete or erroneous data on the underlying source data. Figure A.9 shows how the data in the NEED sample compares with other sources.

Figure A.9: Comparison of distributions – number of adults[22]



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.

The Experian data is valuable since it 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.

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

[22] EHS data is 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.

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 A.7 shows how many records in the NEED sample had some kind of 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 A.7: HEED data coverage in NEED sample

	Count	%
HEED record	2,361,570	64%
Cavity wall insulation	756,680	20%
Loft insulation	759,880	20%
Solid wall insulation[23]	109,380	0.4%

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 and covers measures installed through Government schemes (e.g. EEC, CERT). These data undergo validation before they are included in the HEED database. For example, checking that the same measure has not already been installed in the specified household.

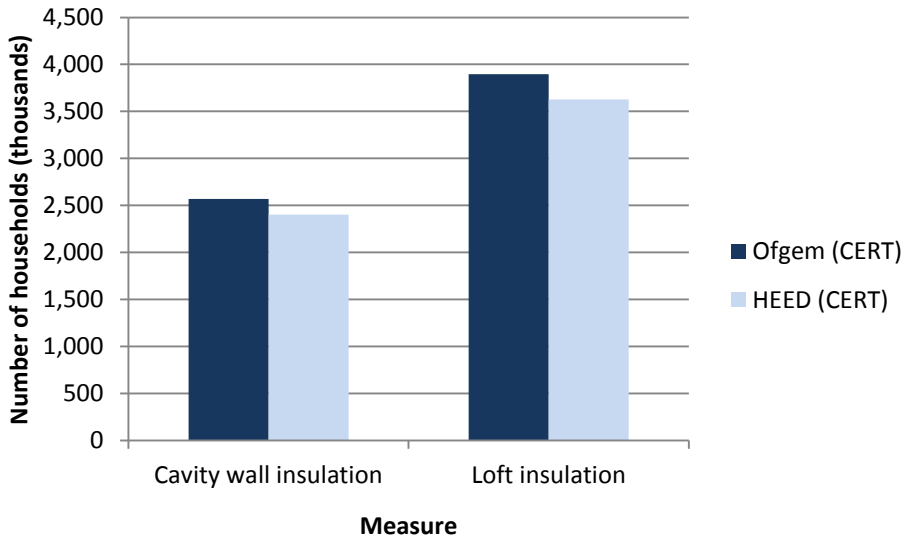
Because the majority of measures recorded in HEED are measures installed through Government schemes, the types of households receiving measures are not representative of

[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.

the population or housing stock as a whole. However, HEED does have good coverage of properties which have received measures.

Figure A.10 below shows the number of properties with cavity wall insulation and loft insulation installed under CERT.

Figure A.10: HEED and Ofgem reported measures 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.

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 comes from a wider range of sources and until recently there was no boiler data available for 2008 and 2009. Data relating to boilers has not been included in the latest published results due to data quality concerns with new historic data on boiler installations received late in the publication process. Further quality assurance is required before these data can be used with confidence. It is planned that further quality assurance of these new boiler data will be carried out which will then enable publication of typical savings following installation of a boiler for 2005 to 2011 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, particularly for earlier installations of solid wall insulation where it is not possible to distinguish when between 2005 and 2008 measures were installed.

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 NEED. This annex shows that generally the quality of data used in NEED are good, with excellent coverage of the population. In all cases, the distribution of data is broadly consistent with the other sources it has been compared with. At a property level data from the administrative sources are more reliable than data modelled by Experian. Table A.8 summarises the strengths and weaknesses of the data used in NEED.

Table A.8: Strengths and weaknesses of data used in NEED

Data source	Strengths	Weaknesses
Consumption data	• Covers Great Britain • Good coverage of almost all properties (post matching) • Data provided by energy suppliers • Gas data are weather corrected	• 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
Valuation Office Agency (VOA)	• Covers every property in England and Wales • Excellent coverage – more than 99 per cent of properties in the NEED sample for all variables	• No data for Scotland • Some data may not be up to date
Experian	• Data available for each household in the UK • Best source of data at property level on household characteristics	• Modelled data with varying accuracy at property level
Homes Energy Efficiency Database (HEED)	• Has data for measures installed in homes in the UK including the date of installation	• 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

Overall the data in NEED are of good quality. However, there are some weaknesses, and given the importance of the quality of the input data on the reliability of analysis, work will continue to monitor and improve the quality of data in NEED.

Department of Energy & Climate Change
3 Whitehall Place
London SW1A 2AW
www.gov.uk/decc
URN 14D/192