Department of Energy & Climate Change

Annex B: Quality Assurance

Contents

1. Summary and Introduction 2

2. Matched Sample 4

3. Consumption Data 8

3.1 Introduction 8

3.2 Gas consumption data 8

3.3 Electricity consumption data 11

3.4 Conclusion 13

4. Valuation Office Agency Data 14

4.1 Introduction 14

4.2 Coverage 14

4.3 Summary of data and comparison with other sources 15

4.4 Conclusion 17

5. Experian Data 18

5.1 Introduction 18

5.2 Coverage and comparison with other sources 18

6. Conclusion 22

1. Summary and Introduction

This annex provides information on the quality of data used in the production of analysis using the National Energy Efficiency Data-Framework (NEED) as well as further information on the revised sample used for analysis of 2011 consumption data. More information on NEED along with outputs from NEED are available here: https://www.gov.uk/government/organisations/department-of-energy-climate-change/series/national-energy-efficiency-data-need-framework.

The outputs from NEED are based on a sample of records. This sample was selected in order to be representative of the housing stock in England and Wales. Figure 1.1 shows how the distribution of properties in the NEED sample compares with the Department of Communities and Local Government (DCLG) estimates of dwelling stock in Wales and the English regions in 2011.

Figure 1.1: Distribution of NEED sample compared with DCLG dwelling stock estimates

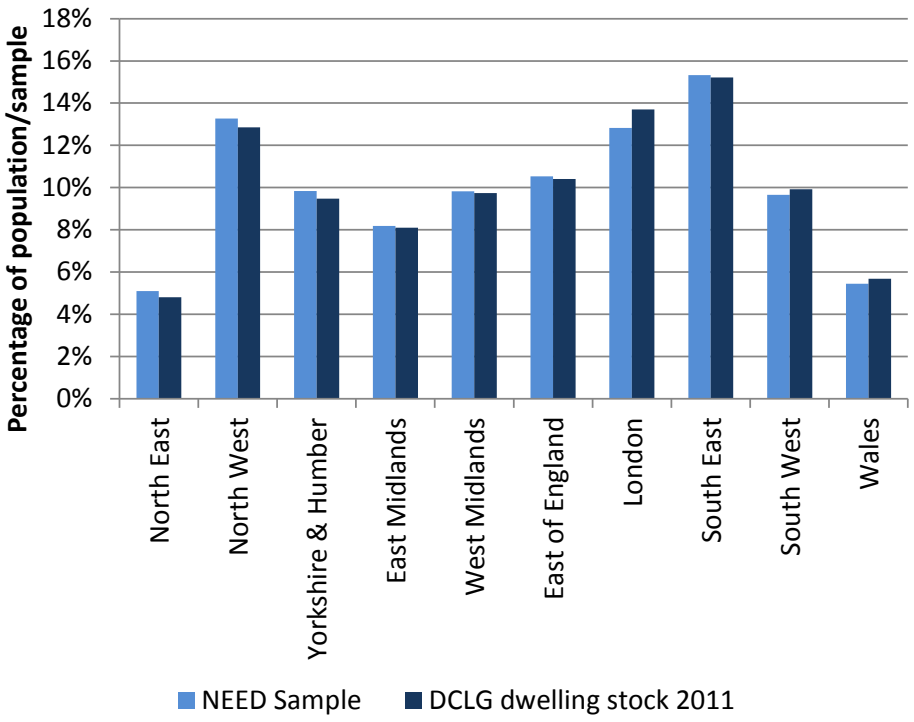


Table 1.1 summarises the strengths and weaknesses of each of the main data sources used for the June 2013 NEED publication. The quality and coverage of the data are good but any interpretation of results should be considered in the context of the strengths and weaknesses of each source.

Table 1.1: Strengths and weaknesses of data in NEED.

Data sources	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.

Data sources	Strengths	Weaknesses
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.

2. Matched Sample

In order to help increase processing speed, reduce cost and ensure that DECC is not processing more data than necessary, a sample is used for analysis.

To create the matched sample, address information in each dataset was matched to the address information on the National Land and Property Gazetteer (NLPG). The NLPG unique property reference number (UPRN) was then assigned to each record. Table 2.1 shows the proportion of records on each dataset which could be matched to the NLPG. 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 2.1: 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%

The sample which had been used for previous analysis was revised for the 2011 analysis. The previous sample was created in 2010 and has been used for all analysis outputs prior to this report. It was decided to create a new sample in order to:

- Include Wales;
- Ensure there was no significant bias in the original sample and therefore increase confidence in the results;
- Make the sample representative at local authority level; and
- Use a more up to date sample of households.

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 resulted 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.

---

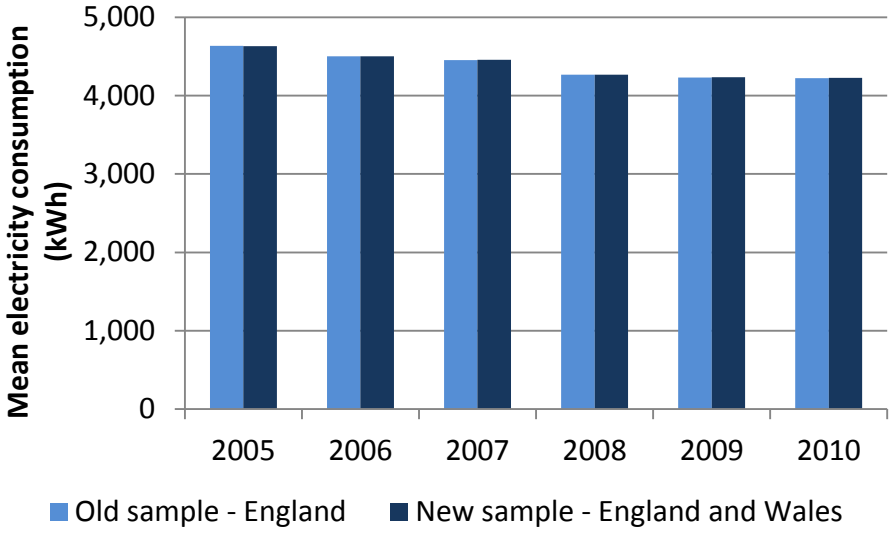
[1] A sub-building is a separate property within the same building, such as a flat within a converted property or an individual shop within a shopping centre.
[2] Property age consists of pre-1919, 1919-44, 1945-64, 1965-82, 1983-92, 1993-99 and Post 1999.
[3] Property type consists of detached, semi-detached, 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 of the other datasets.

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 lost in London than other areas of England and Wales.

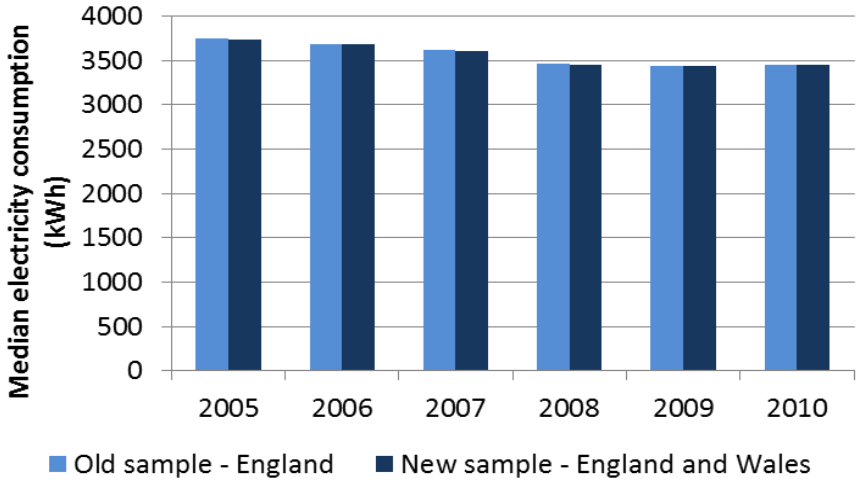
Once the data had been matched to other sources some further records were lost as a result of invalid or missing consumption values in the data (see Section 3 for details). For 2011, 95 per cent of the sample had a valid electricity consumption value and 78 per cent had a valid electricity consumption value. The lower rate observed for gas is expected as not all properties have a gas meter[6]. The impact of loss of records on the distribution of dwellings in the sample can be seen in Section 4.

Figures 2.1 and 2.2 shows how the mean and median consumption for valid records in each year on the new sample (covering England and Wales) compares to the equivalent values for the old sample (for England), for electricity and gas respectively.

Figure 2.1: Comparison of electricity consumption, new sample versus previous sample
(a) Mean



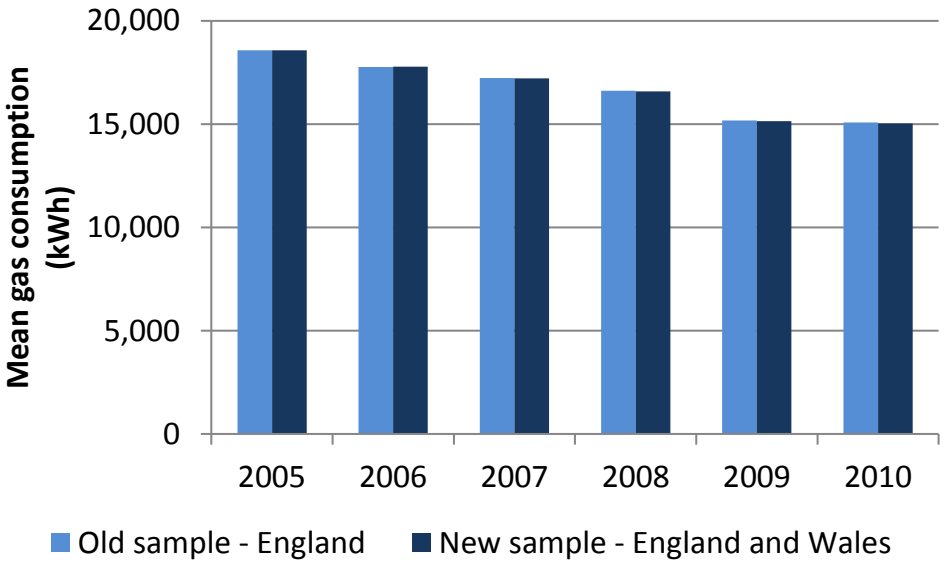
(b) Median



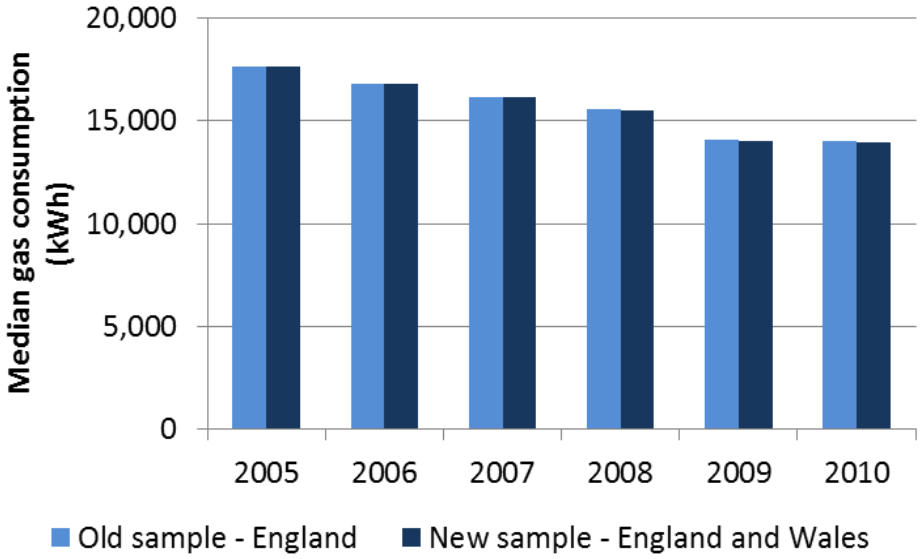
[6] It is estimated that 14 per cent of properties in England and 20 per cent of properties in Wales were not connected to the gas network in 2011. Source, DECC sub-national gas consumption fact sheet: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/175425/Gas_factsheet_2013.pdf.

For mean electricity consumption the largest difference in any year is less than 10kWh, for median consumption the largest difference is less than 20kWh.

Figure 2.2: Comparison of gas consumption, new sample versus previous sample
(a) Mean



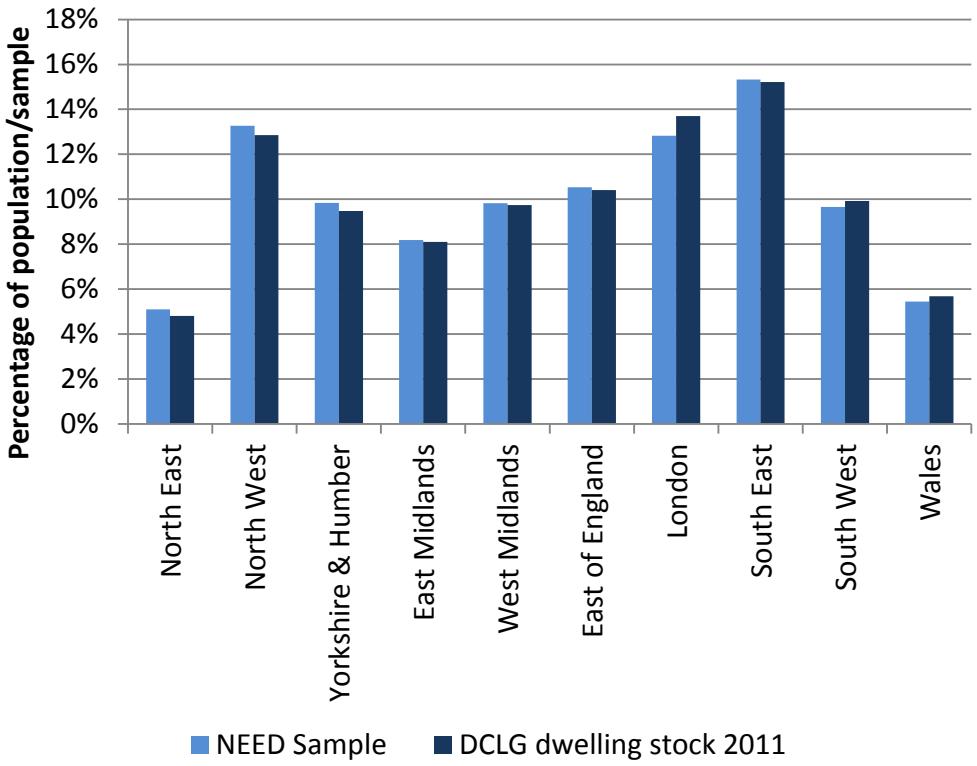
(b) Median



There is slightly more variation between the two samples for gas consumption than electricity (in part because of the higher values), but the biggest difference in mean or median consumption in any year is still less than 50kWh. This small differences reinforces the validity of past results and give confidence that trends can be considered as continuous, without revisions to historic data.

The distribution of the new sample was also compared with other sources to check that it is representative of the dwelling stock. Figure 2.3 shows that the dwellings in the NEED sample have similar distributions to those reported in the Department for Communities and Local Government (DCLG) Dwelling Stock estimates. Comparisons for other variables are shown in Sections 4 and 5.

Figure 2.3: Distribution of NEED sample compared with DCLG dwelling stock estimates[7]



[7] DCLG: http://www.communities.gov.uk/housing/housingresearch/housingstatistics/housingstatisticsby/stockincludingvacants/livetables/.

3. Consumption Data

3.1 Introduction

UK Government has collected and published energy consumption data within the Digest of UK energy Statistics (DUKES) since 1948,[8] and time series back to 1970 on how energy has been used is published in Energy Consumption in the UK[9]. However, data at individual meter point, as used in NEED, was first obtained in 2004 in order to produce local areas estimates of consumption – work that was awarded a Royal Statistical Society Award for innovation in 2010. These data cover consumption of gas and electricity for all homes and businesses within England, Scotland and Wales. There is no property level data available for other fuels which may be being used to heat homes, 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. The consumption data are published on the DECC website down to Lower Level Super Output Area (groups of approximately 400 homes)[10]. This section provides more detail on the electricity and gas consumption data used in NEED.

3.2 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 17 year 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[11].

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.

---

[8] https://www.gov.uk/government/organisations/department-of-energy-climate-change/series/digest-of-uk-energy-statistics-dukes.

[9] https://www.gov.uk/government/organisations/department-of-energy-climate-change/series/energy-consumption-in-the-uk.

[10] More detailed information about how these data are collected and compiled for DECC's sub-national publication available here: https://www.gov.uk/government/publications/regional-energy-data-guidance-note.

[11] The 2011 gas year runs from 1 October 2010 to 30 September 2011.

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 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 exclude 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. 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). 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.

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. 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. 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). Unlike the sub-national consumption statistics, all negative meter readings are also excluded[12].

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 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.

[12] 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.

The impact of removing these invalid records on the data is small. It means the mean for NEED is a little bit lower 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. Figure 3.1 shows the differences.

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 NEED analysis sample[13] has been compared with the data published in DUKES[14] and the data from the sub-national consumption statistics also published by DECC.

Figure 3.1: Comparison of estimates of mean gas consumption per household

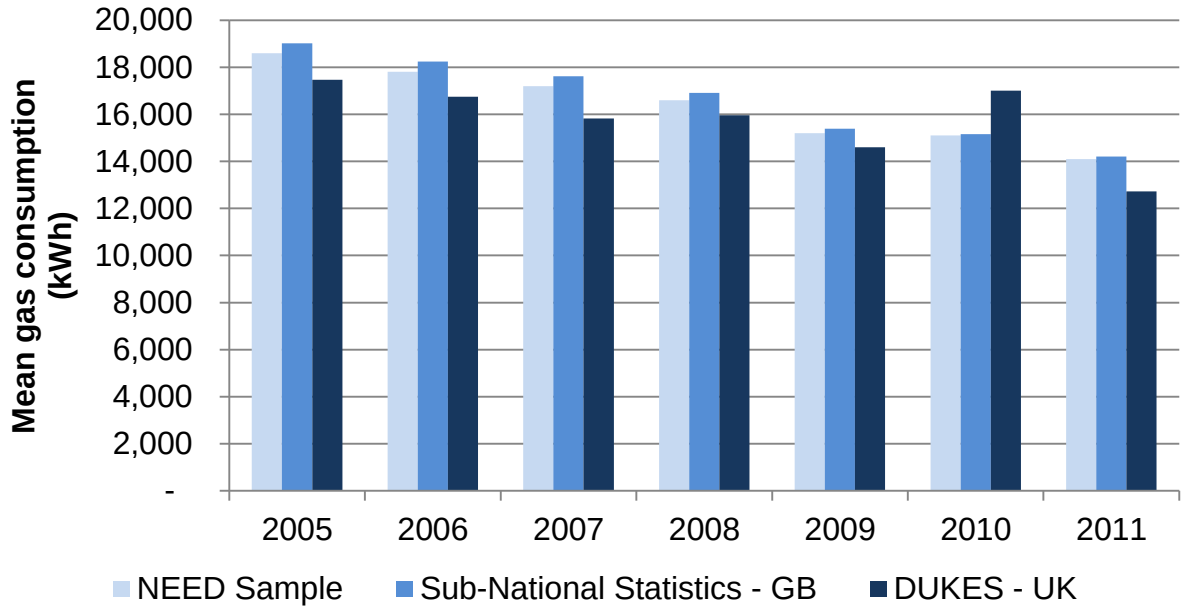


Figure 3.1 shows that the mean consumption is very similar for the published sub-national statistics and for the NEED sample. The mean for the NEED sample is slightly lower than the sub-national consumption estimates as a result of the way the NEED sample is selected and the further validation applied. For example NEED only includes properties which are included on the Valuation Office Agency property attribute database, and excludes any properties with consumption over 50,000 kWh. The sub-national statistics estimates include all properties with consumption up to 73,200 as a domestic property. This means a number of higher consuming properties are excluded from the NEED sample, bringing the mean consumption down.

There is more variation in the DUKES data between years. The primary reason for this variation is the difference between sources. DUKES estimates are not weather corrected, while the NEED and sub-national statistics data are. Therefore in a cold winter, the DUKES data shows higher mean consumption, while in a year that has a warmer winter – like 2011 – the DUKES mean consumption is lower than that given by NEED and the sub-national statistics. There will also be some differences between the DUKES and sub-national Statistics or NEED estimates because of the method of collection. While the later two are based on the same

[13] The NEED sample covers England only for 2005 to 2010 and England and Wales in 2011.
[14] The published DUKES data gives a total domestic consumption figure for the UK which has been converted into a mean consumption based on the number of dwellings in the UK and the proportion of households in GB with gas meters.

source data and built up from meter point data for each property, DUKES is based on aggregate data. There is also a small difference as a result of the different geographic coverage of the sources.

3.3 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 around 29 million NHH meters and 113,000 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, 2011 annualised consumption estimates cover the period from 28 January 2011 up to 27 January 2012. 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 type tariffs 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 data validation section below.

Coverage

These data cover all of Great Britain. Data for Northern Ireland are currently excluded from the dataset (though work is on going to produce data for Northern Ireland and experimental data has been published in Energy Trends[15]). 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.

[15] June 2013 Energy Trends: https://www.gov.uk/government/organisations/department-of-energy-climate-change/series/energy-trends.

Data validation

For the sub-national statistics estimates, it is generally assumed that any consumption of over 50,000 kWh is not domestic unless the address suggests otherwise. However, because of the nature of the analysis undertaken in NEED further cleansing and validation was undertaken to decide on what should be considered valid data 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[16].

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.

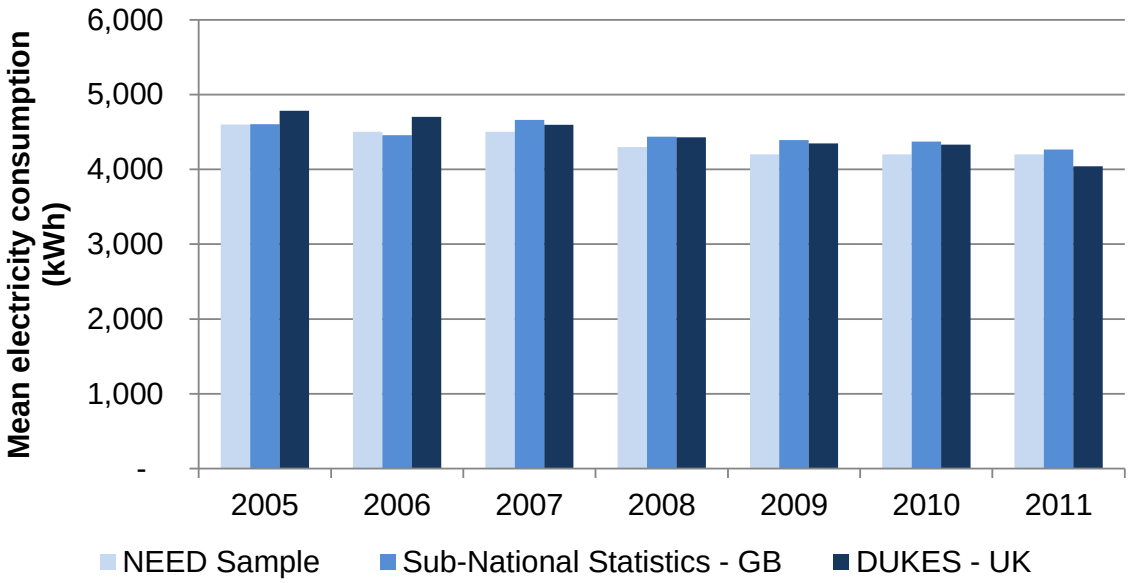
The impact of removing these invalid records on the data is small. It means the mean for NEED is a little bit lower than it would be if these filters were not applied, due to the elimination of a relatively small number of records with a high consumption. The median remains almost the same. Figure 3.2 shows the differences.

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 NEED analysis sample[17] has been compared with the data from DUKES[18] and the data from the sub-national consumption statistics published by DECC[19]. Figure 3.2 shows that the mean consumption is very similar for all three sources. There is some variation, which will result from the different geographic coverage and the difference in the way the data is collected from different sources. However, all three sources are consistent.

[16] 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.
[17] The NEED sample covers England only for 2005 to 2010 and England and Wales in 2011.
[18] The published DUKES data gives a total domestic consumption figure for the UK which has been converted into a mean consumption based on the number of dwellings in the UK. The sub-national consumption data is based on consumption per consumer in 2005 and 2006 and sales per household from 2007.
[19] The sub-national consumption data is based on consumption per meter in 2005 and 2006. From 2007 it is based on consumption per household.

Figure 3.2: Comparison of estimates of mean electricity consumption per household



3.4 Conclusion

The consumption data are a rich source of data which form the core of NEED. Table 3.1 summarises the differences in approach used in NEED and DECC's sub-national estimates.

Table 3.1: Differences in 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[20].
• Data matched to other sources via the NLPG UPRN at property level.	• Data assigned to Lower Level Super Output Area[21].
• 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. The comparisons with other sources confirm that the consumption estimates based on the NEED are consistent with other sources.

[20] 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.

[21] 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.

4. Valuation Office Agency Data

4.1 Introduction

The Valuation Office Agency (VOA) is the central government agency responsible for valuing homes for council tax purposes[22]. 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 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)
- Number of bedrooms

4.2 Coverage

The VOA Council Tax Database covers properties in England and Wales. The table below 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.

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

Table 4.1: 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.

The table below shows the categories of data used in the analysis for each of the VOA variables. 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 its website[23].

Table 4.2: 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+	Greater than 200
	1993-99	Purpose built flat		
	Post 1999	Converted flat		

4.3 Summary of data and comparison with other sources

This section shows how the data in the NEED sample compare with the distribution of data on the full VOA property attribute database and with the English Housing Survey (EHS)[24] . Differences between the NEED sample and the 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. However it still provides helpful context to validate the VOA data.

Figures 4.1 to 4.3 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.

[23] http://www.voa.gov.uk/corporate/Publications/DwellingHouseCodingGuide/index.html.
[24] EHS data are from the English Housing Survey: Homes Report 2011: https://www.gov.uk/government/publications/english-housing-survey-2011-to-2012-headline-report.

Figure 4.1: Comparison of distributions – number of bedrooms

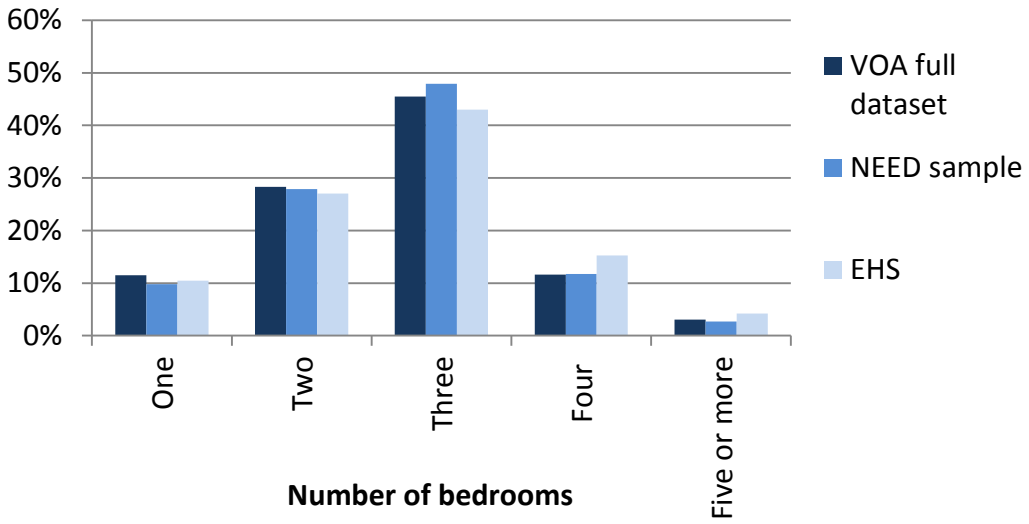


Figure 4.2: Comparison of distributions - property type

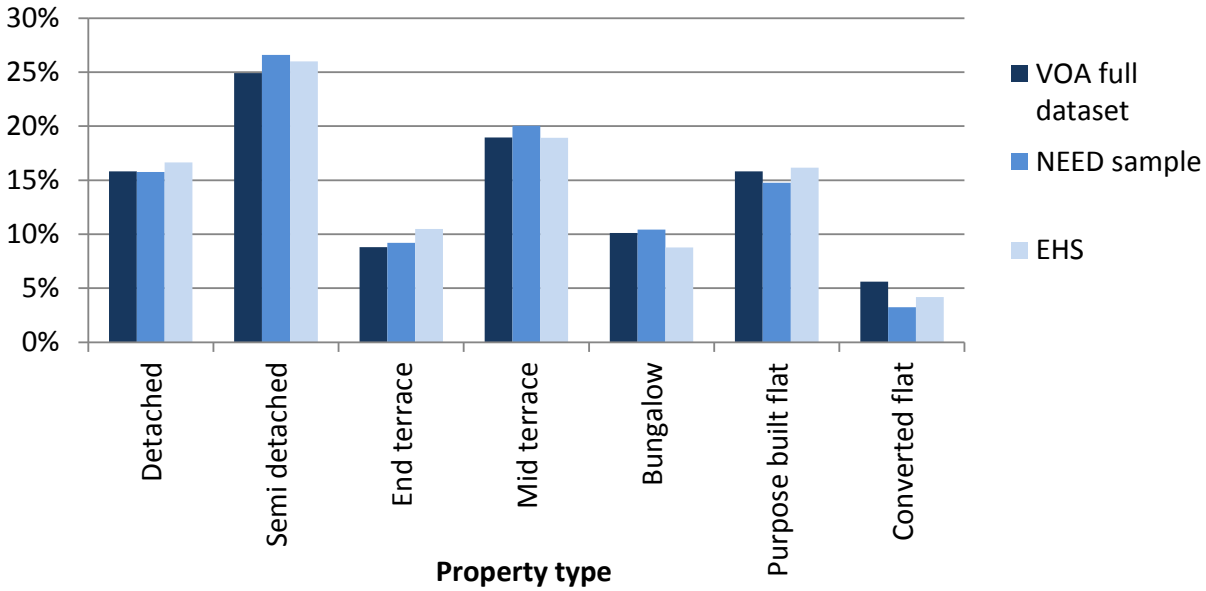
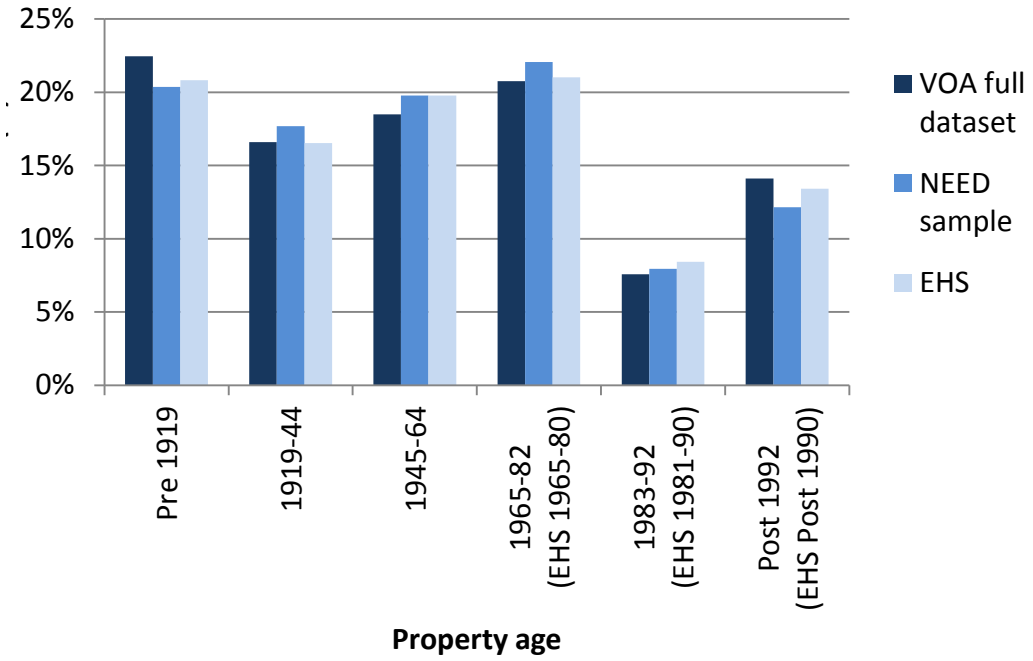


Figure 4.3: Comparison of distributions – property age



4.4 Conclusion

The data in the VOA property attribute dataset have excellent coverage of properties in England and Wales and the comparisons with the EHS confirm that the distribution of data is consistent for all property attributes considered in the NEED analysis.

5. Experian Data

5.1 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 provided in the Experian dataset, with 98 per cent of records in the NEED sample assigned an Experian record.

5.2 Coverage and comparison with other sources

The household characteristic data purchased include:

- Household income
- Tenure
- Number of adults

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. The income bands in 2011 have been revised, with the new bands set out in table 5.1.

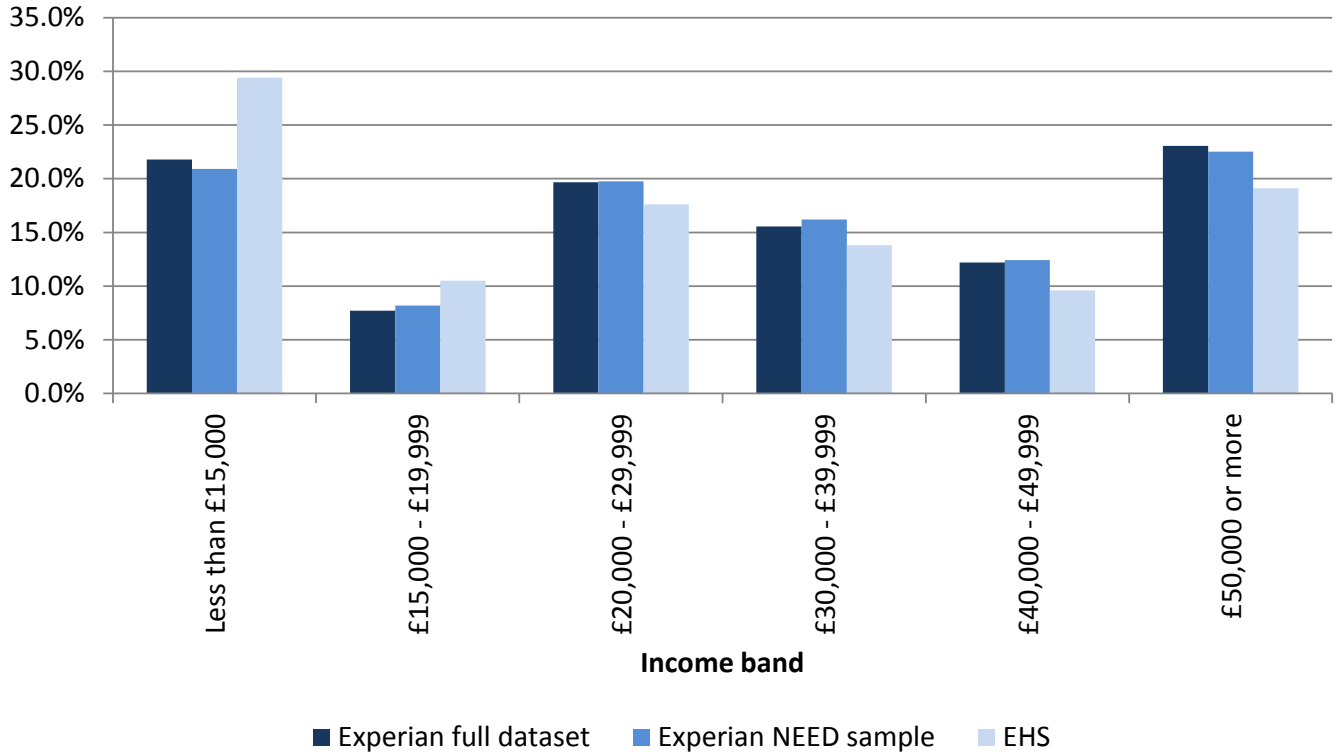
Table 5.1: Distribution of households by income band on the full Experian dataset

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 5.1 shows how the distribution of income for the Experian dataset and the NEED sample compares with the income reported by the English Housing Survey (EHS)[25]. Note that some of the income categories from the Experian data have been grouped together to allow comparison with the categories used in the English Housing Survey.

Figure 5.1: Comparison of distributions – household income band



The figure shows that Experian appears to be under assigning properties to the lowest income band. This is consistent with DECC's understanding that the Experian income data is least 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 of the property, carried out by a trained surveyor.

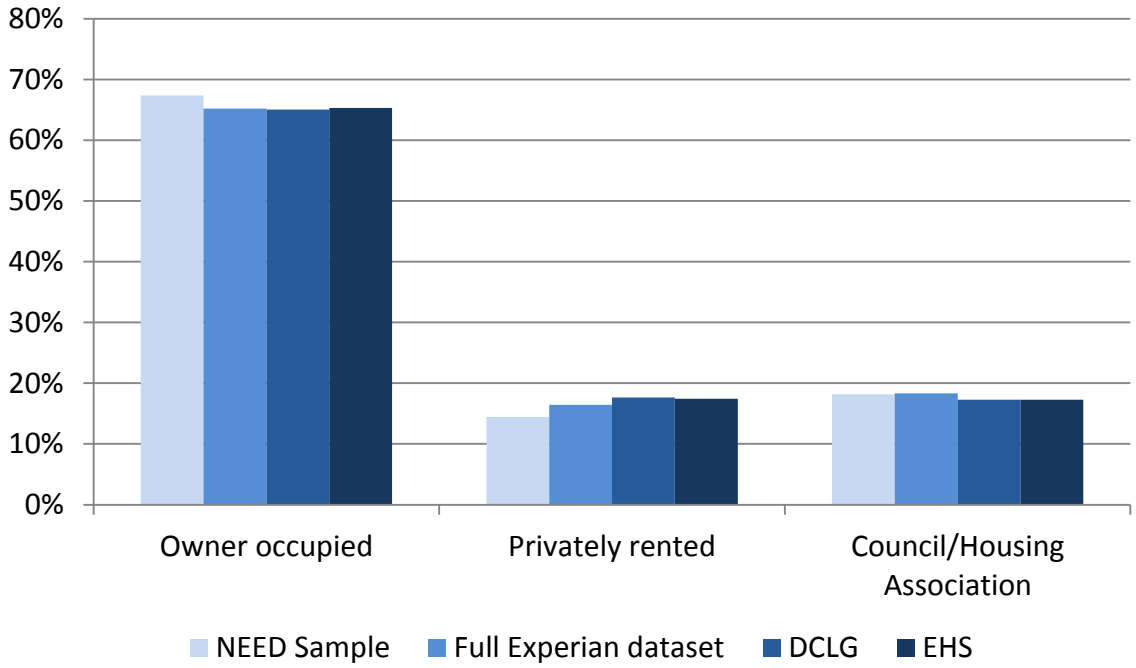
Tenure

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.

[25] The EHS 2010 homes report has been used for this variable as information is not available by this break down in the 2011 report.

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 5.2 shows how the Experian data compares with data from other sources at the national level[26].

Figure 5.2: Comparison of distributions – tenure[27]



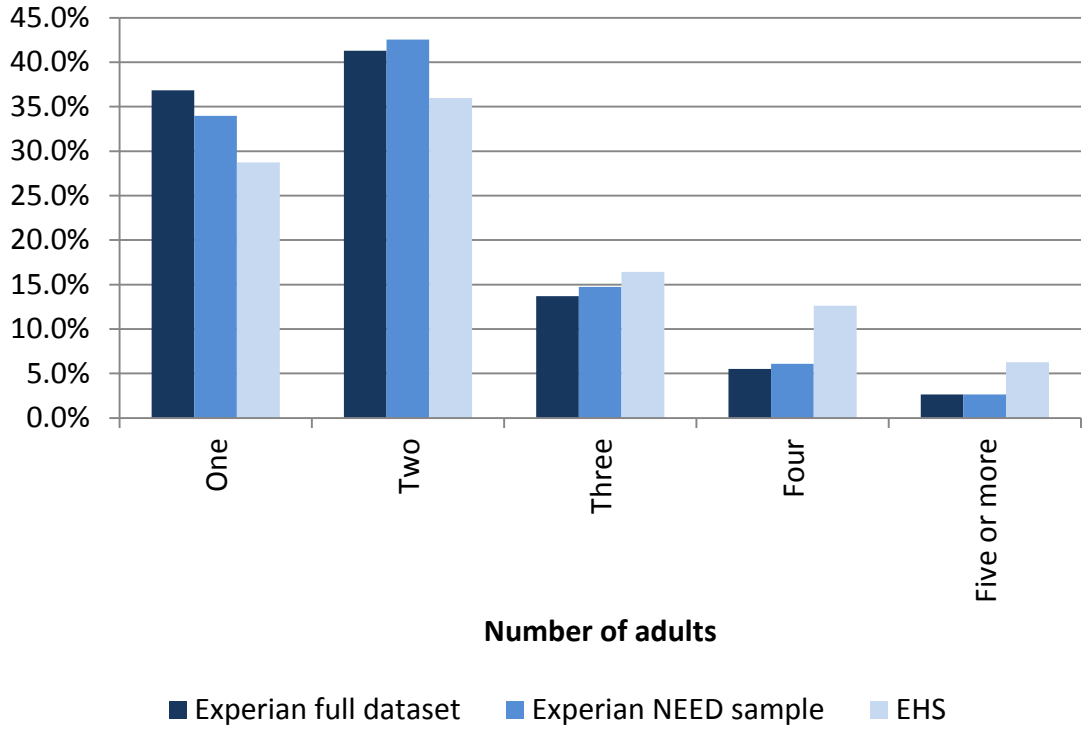
The figure shows that the proportion of properties assigned to each tenure category is similar for all sources. It appears that the Experian dataset as a whole and specifically the NEED sample allocates too many properties to the owner occupied category and too few to privately rented. This is likely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

Number of adults

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 5.3 shows how the data in the NEED sample compare with other sources.

[26] Note that the Experian full dataset covers the whole UK, while the NEED sample covers England and Wales and EHS covers England only.
[27] DCLG estimates from Tables 104 and 106: https://www.gov.uk/government/statistical-data-sets/live-tables-on-dwelling-stock-including-vacants.

Figure 5.3: Comparison of distributions – number of adults[28]



The variation in the distribution is likely to be because the EHS estimates are based on household size while the Experian data is based on occupants aged 18 or 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 in the Experian database and more properties with three or more in the EHS.

While the Experian data is valuable in order to provide an understanding of the properties in the NEED sample and how consumption and impacts of energy efficiency measures vary for different types of properties, it is important that interpretation of results relating to income and tenure is in the context of the limitations of the data.

[28] EHS data is based on household size (not number of adults) from Table AT1 or the 2011 headline report.

6. 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 the data used to form NEED.

This annex shows that in general 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 the property level, data from the administrative sources are more reliable than the data produced by Experian. Table 6.1 summarises the strengths and weaknesses of the data used in NEED.

Table 6.1: Strengths and weaknesses of data used in NEED

Data sources	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.

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 be undertaken to improve the quality of data in NEED. This will include using Energy Performance Certificate (EPC) data to further validate some of the data in NEED and looking at alternatives to the Experian data.