Vulnerable Consumers Market Segmentation Report

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 754051

Table of contents

1. Presentation of ASSIST Project

1.1 ASSIST overview and introduction

ASSIST is a 36-months European 'market activation and policy orientation' project to tackle fuel poverty and support vulnerable consumers. It intends to actively engage consumers in the energy market and positively change behaviour in relation to energy consumption and to influence design of policy at all levels to tackle fuel poverty issues.
Based on the conclusion of the Energy Citizens' Forum and of the European Vulnerable Consumers Working Group, the project intends to combine activities addressing both energy and social dimensions as fuel poverty is not only an energy issue nor can it be tackled in isolation of the bigger issue of poverty. More specifically, ASSIST strategic objectives are to contribute to:

- tackle energy poverty;
- reduce the main barriers of the energy market faced by vulnerable consumers;
- support vulnerable consumers to be more efficient with their domestic energy consumption (electricity and gas).

To fulfil its goals, the project foresees diversified and correlated research, networking and in-field actions, consistent with the relevant national and European scenarios. Among them, ASSIST intends to create a network of innovative professional figures supporting vulnerable consumers in their domestic energy consumption: **"Home Energy Advisor (HEA)"**.

1.2 "WP5 – ASSIST 2gether Action"

The fifth work package aims to design, implement and evaluate innovative support services for vulnerable consumers/energy poor on a country based level with a market-oriented and flexible approach (as recommended "projects based on behaviour change and efficiency should align with existing local initiatives)", through the implementation of pilot actions. The actions will be defined at National level with the support of the National Steering Committee and on the basis of the results of the national context analysis carried out in WP2 and of the market segmentation carried out at the beginning of the work package. The ASSIST actions will address 4.500 vulnerable consumers (750 per country). The tasks foreseen in the work package include:

- Market segmentation, that aims at assigning vulnerable consumers in market groups with common characteristics. It will act as a driver for the design of the actions.
- Designing, initiating and delivering the actions in each country.
- Evaluation of results: in order to assure reaching the set objective and performance indicators, the partners will closely monitor the development of and results achieved within the action.

1.3 Document overview and structure

The market behavioural segmentation represents a key preparatory phase for the development of the national and local actions. It will lead to a complete understanding of the actual situation under different aspects: socio-demographic, economic and dwelling characteristics of consumers, but also consumers' degree of knowledge, common needs and priorities regarding energy. The final objective is to assign vulnerable consumers into market groups who share common characteristics. It is strictly important to obtain a reference baseline of the national situation in order to better understand the effectiveness of the actions locally implemented, and also to have the possibility of comparisons, in time and space. It is important to underline that the absence of a common and recognised definition of energy poverty makes it difficult to characterize different consumers, in particular those who are facing energy related problems. The market segmentation intends to divide population into homogeneous clusters, without having the necessity of a definition, but relying only on household's characteristics. The importance of having an analysis which include the whole national population, is given by the fact that, by starting with a complete picture, it is then possible to identify different sub-categories of vulnerable consumers, who could have different peculiarities and necessities, so different actions could be required in such a way that the actions bring to effective results.
In order to pursue the planned objectives, all the available information shall be considered. The characteristics of the population at a national level are available thanks to the national statistical agency, while more specific data in terms of energy poverty are given by the market survey developed in the Task 2.5 and with data collected by the HEA network specifically in households who are part of the pilot ASSIST program.
The document is structured country-by-country, and then a final conclusion is drawn.

2. Belgium

2.1 Methodology

2.1.1 General

The Belgian consumer market segmentation is based on the 2017 edition of the annual energy poverty barometer published by the King Baudouin Foundation[1]. In this annual barometer, the width and depth of the energy poverty problem in Belgium is measured using three different indicators:

- The 'measured energy poverty' indicator: is based on the Boardman approach (used in the UK to define fuel poverty). Each year, the median value of the ratio between energy expenditures and equivalent household income (corrected for the household size)[2] is calculated. A 'boundary value' is defined as twice the value of the median value. If a household spends more on energy than indicated by the boundary value (in %), this household is considered as an 'energy poor' household in an objective sense.
- The 'hidden energy poverty' indicator: concerns the fraction of the Belgian households that is reducing their energy use to the extent that it might have a negative impact on living conditions and quality of life in general. The hidden energy poverty indicator is calculated based on the comparison of a household's energy expenditures with the average energy expenditure of a comparable household (with the same number of inhabitants) living in a comparable dwelling (with the same number of rooms). If a household spends less than half of the average of a comparable household living in a comparable dwelling, and if this household belongs to the 50% of households with the lowest equivalent incomes in Belgium, this household is considered to be in a situation of hidden energy poverty.
- The 'subjective energy poverty' indicator: is based on the percentage of households that report having difficulties to adequately heat their dwelling.

Data on energy expenditures are derived from the EU-SILC inquiry on living conditions in the EU. This inquiry is performed on a yearly basis and includes a statistically representative sample of about 6.000 Belgian households.

2.1.2 Results

[1] Available at https://www.kbs-frb.be/nl/Activities/Publications/2017/20170313NT1.
[2] The highest equivalent incomes are excluded from the calculation of the median, and housing expenditures are subtracted from the household income.

For the year 2015 (latest results available), the following percentages are reported for the different indicators:

- Measured energy poverty (14,5% on average in Belgium):

Flanders region: 10,8%

Brussels region: 12,8%

Walloon region: 22%

- Hidden energy poverty (3,9% on average in Belgium):

Flanders region: 3%

Brussels region: 9,8%

Walloon region: 2,2%

- Subjective energy poverty (5,1% on average in Belgium):

Flanders region: 2,4%

Brussels region: 8,1%

Walloon region: 7,9%

The higher measured energy statistics for Wallonia (compared to Flanders) can be explained by an on average lower disposable income, the size (larger) and quality (lower) of the dwellings, and a lower average temperature. The average income in the Brussels region is even lower than in Wallonia, but the share of flats is very high in this region so that heating expenditures are generally lower.

Figure 1 shows the relation between measured energy poverty (gEA), hidden energy poverty (vEA) and subjective energy poverty (sEA). From the figure, it becomes clear that there is only a limited overlap between the different categories of energy poverty in Belgium. All in all, taking into account the overlaps, 21% of Belgian households are potentially at risk of falling into energy poverty according to one of the three definitions.

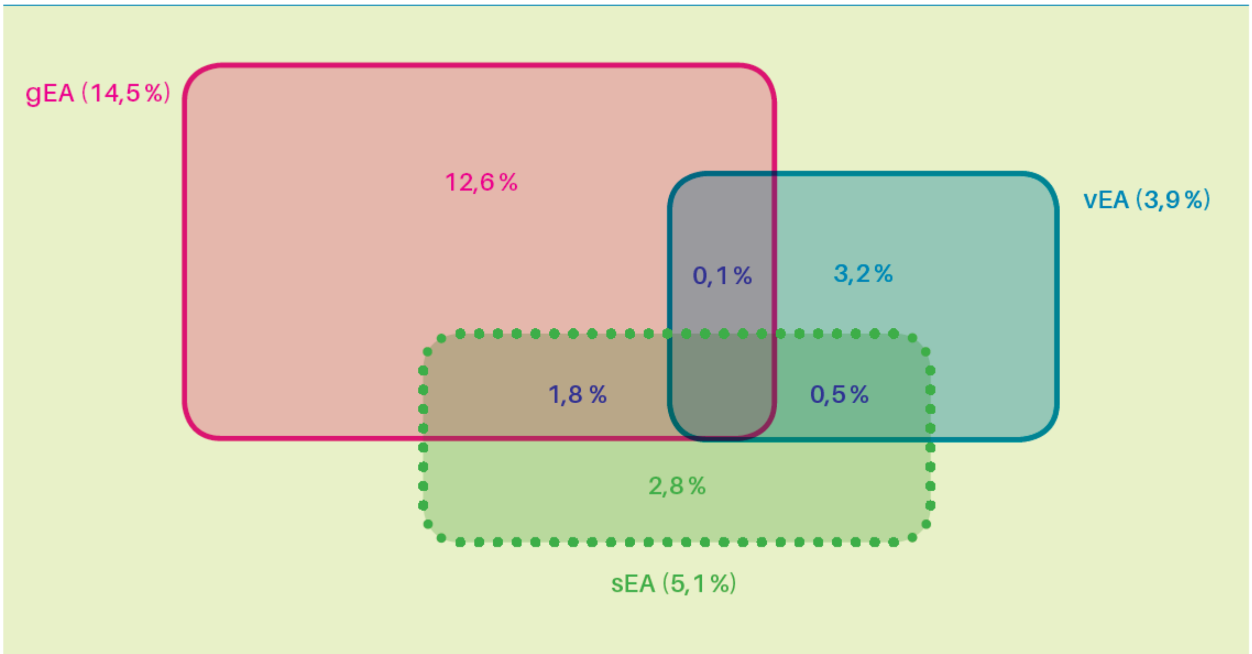


Figure 1 - Relation between measured energy poverty (gEA), hidden energy poverty (vEA) and subjective energy poverty (sEA) in Belgium.

2.2 Cluster analysis of the population

The energy poverty barometer shows the following interesting correlations.

About 70% of households that are at risk of poverty in general[3] also fall into one or more of the above-mentioned categories of energy poverty. The three lowest disposable income deciles run the highest risk of falling into energy poverty. On the other hand, the overlap between energy poverty and poverty in general is not complete, as 40% of energy poor households (according to at least one of the above-mentioned definitions) is not at risk of poverty in a general sense.

Tenants are clearly at a higher risk of energy poverty than owners. 21,9% of tenants are energy poor (according to the 'measured energy poverty' indicator), while only 11% of owners fall into this category.

Owners without a mortgage loan are at a higher risk of energy poverty than owners. 15,4% of owners without a mortgage loan are energy poor (according to the 'measured energy poverty' indicator), while only 6,2% of owners with a mortgage loan fall into this category. This correlation appears to be counter-intuitive, but is explained by the fact that single person households and senior owners are over-

[3] According to the EU definition, a household is at risk of falling into poverty when disposable equivalent income <60% of the median value of Belgian households.

represented in the category of owners without a mortgage loan, and these types of households are at a higher risk of falling into energy poverty (cf. infra).

Single person households and single parent households are clearly at a higher risk of energy poverty than others. 26,9%/20,3% of single person/single parent households are energy poor (according to the 'measured energy poverty' indicator), compared to the average of 14,5% for all households. Single parent households are predominantly (84%) composed of a single mother + child(ren).

Single person households account for 57,7% of the 'measured energy poor' households. Out of the energy poor single person households, 48,4% are seniors (older than 65 years). Old age in combination with living alone clearly represents a risk factor regarding energy poverty. This is explained by the fact that senior people often experience a significant drop in disposable income, while they often live in old energy-inefficient dwellings that are too big for their needs.

Not surprisingly, households without a labour income are over-represented in the category of 'measured energy poor'. 26,3% of households without a labour income are energy poor (according to the 'measured energy poverty' indicator), compared to the average of 14,5% for all households.

Finally, 16,2% households living in an energy-inefficient[4] dwelling are energy poor (according to the 'measured energy poverty' indicator), compared to the average of 14,5% for all households.

2.3 Vulnerable consumers characterization

The ASSIST action in Belgium (Flanders region) will be carried out by the already existing network of 'energy cutters' (energiesnoeiers), who will receive an additional ASSIST training. In Flanders, certain well-defined target groups of energy consumers are eligible to receive a free energy scan. An energy scan is a quick audit of the energy consumption of a building. The 'energy cutters' perform a quick inspection of the dwelling (not comparable to a full-scale energy audit) and give simple tips on how to save energy, for example by isolating. They also give tips about heating, lighting or household appliances. The 'energy cutter' can also immediately carry out small energy-saving measures, such as installing energy-efficient light bulbs.

For the ASSIST action, we will work together with the umbrella organization KOMOSIE. KOMOSIE coordinates a.o. the working of 29 social economy

[4] An indicator was constructed based on the EU-SILC data. A household was considered to live in an energy-efficient home if the roof, walls, windows and floors of the dwelling were at least partly insulated.

organisations of ‘energy cutters’. 20 of these organisations perform energy scans, 9 organisations are mainly concerned with implementing social home improvement actions. The fact that the employees of these organisations are recruited from people in vulnerable situations adds to their effectiveness: they can easily gain the trust of vulnerable customers. In general, the barriers during the initial contact stages (e.g. shame etc.) are lower.

About 20.000 free energy scans are performed on a yearly basis. There are 6 target groups that qualify for receiving a free energy scan

- ‘protected’ customers that are entitled to a social maximum energy tariff (customers that enjoy certain social welfare benefits or payments are eligible, e.g. seniors with a minimum pension allowance, people with a labour incapacity, etc.);
- customers with a budget meter (a budget meter is installed if the customer has problems in paying the bill of the energy provider of last resort – i.e. the DSO);
- customers that are eligible for a social energy loan;
- customers who are at risk of being disconnected (such a request always has to be filed with an judged by a ‘local advisory committee’);
- customers living in a dwelling belonging to a social housing corporation;
- tenants paying a maximum rental price of 462,72 euro, or 512,74 euro in certain cities.

Except for the category of social housing tenants, which are clearly at a higher energy poverty risk (23,7% fall into the category of ‘measured energy poor’, compared to the average of 14,5%), there is no statistical analysis available of the energy poverty prevalence (according to the three measurements) among recipients of free energy scans. However, many of the selection criteria (e.g. being a ‘protected’ customer, customers eligible for a social energy loan, tenants paying a maximum rental price) relate to the overall income of the customer, which has to be low enough to enjoy the benefit of a free energy scan. Even though the overlap between poverty in general and energy poverty is not perfect (cf. supra), 70% of poor people are also considered to be vulnerable energy customers.

On the other hand, customers with a budget meter or customers at risk of being disconnected might not be suitable candidates for the ASSIST action. These customers are highly likely to be very aware of energy saving opportunities as the budget meter functions on prepaid cards that have to be recharged regularly. Opportunities for further energy saving through the ASSIST action seem to be limited in this case.

Furthermore, at a later stage of the research insights on the recipients of free energy scans will be enlarged by the customer survey developed in WP2 (Task 2.5). These data are collected through the network of the 'energy cutters', therefore they should represent in detail which are the characteristics of the vulnerable consumers involved in the Flemish ASSIST Action, both in terms of socio-economic characteristics and behavioural attitudes towards energy management.

2.4 Conclusions

From the Belgian energy poverty barometer (2017 edition, based on data for 2015), it appears that at the national level in Belgium, the households who can be considered most at risk of energy poverty include the following:

- Households that are at risk of poverty in general (i.e. equivalent disposable income < 60% of the Belgian median value).
- Tenants (and especially social housing tenants) are clearly at a higher risk of energy poverty than owners.
- Owners without a mortgage loan are at a higher risk of energy poverty than owners.
- Single person households and single parent households are clearly at a higher risk of energy poverty.
- Old age in combination with living alone clearly represents a risk factor regarding energy poverty.
- Not surprisingly, households without a labour income are over-represented in the category of 'measured energy poor'.
- Finally, households living in an energy-inefficient dwelling have a higher risk of being energy poor.

The ASSIST action in Belgium (Flanders region) will be carried out by the already existing network of 'energy cutters' (energiesnoeiers), who will receive an additional ASSIST training. Certain well-defined target groups of energy consumers (i.e. those who are eligible to receive a free energy scan) will be targeted for the ASSIST action. No systematic statistical data are available on the classification of our target group into the categories listed above. Nevertheless, when defining the ASSIST action in Flanders some categories from the national market segmentation could be targeted specifically in the action, e.g. the elderly who live alone, tenants, or single parent households. On the other hand, certain segments, such as customers with a budget meter or customers at risk of being disconnected might not be suitable candidates for the ASSIST action. Further knowledge on the target group will be developed on the basis of the results of the customer survey developed in WP2.

The detailed planning of the ASSIST actions in Belgium (Flanders) and the further obtained results will be reported in deliverable D5.2.

3. Finland

3.1 Methodology

3.1.1 General

In the Nordic countries, the issue of energy poverty is relatively new. There is still very few studies, statistical data and political measures to address energy poverty. In Finland two previous studies by the Ministry of Environment have explored the risk of energy poverty: Ympäristöministeriön raportteja 21/2013: Selvitys energiaköyhyydestä and as follow up Ympäristöministeriön raportteja 6/2015: Pienituloisen omistusasujan energiaköyhyys. The first study by the Ministry of Environment estimated the risk of fuel poverty and energy costs in different type of households. It identified that the people most at risk of energy poverty are low-income households, such as low-income families and pensioners, living outside the urban area in large-non-energy efficient dwellings (p. 38, YM 21/2013). The second study as a follow up investigated the residential heating renovations and energy costs in relation to energy poverty.

Currently there is no official definition, statistics or specific information sources for measuring energy poverty in Finland. In order to estimate the risk for energy poverty and the characteristics affecting the energy expenditure of a household, in the current research, we have analysed the results of the Market survey on vulnerable consumers' needs, expectations and interests, (further described in the Vulnerable Consumers Fuel Poverty Report) that was conducted as part of Assist project. The survey was distributed both electronically with an email link to home owners via Home Owners Association and with paper version to the elderly that does not have access to an electronic survey via The Association for welfare of the elderly. In total the survey was distributed to 24 484 people and 4660 answers were collected.

It should be noted that the sample for the survey in Finland was selected based on previous research in to energy poverty, and the survey was targeted to homeowners and elderly. Thus, the socio-demographic aspects in the data acquired through the survey, such as age employment status, dwelling characteristics cannot be considered statistically representative of the whole Finnish population.

The purpose of the analysis was to identify categories/clusters of customers that might be vulnerable, taking under consideration their Annual Electricity and Heating costs as a share of their income (Electricity/Heating costs). For this reason, a feature selection and clustering methodology has been designed and implemented, in order to first select the most important variables and then use them as the criteria for separating customers into clusters. The step-to-step methodology and analysis results are described in detail in the following subsections.

3.1.2 Description of the data

The consumer survey contained in total 38 questions, that were either multiple choice or open text questions: 8 questions were multiple choice questions where multiple answers were possible.

In our analysis, we have taken under consideration 36 variables obtained from 22 questions in the survey. The selection has been made based on the completeness of the data and the relevance with the basic targets of the analysis. The 36 attributes used in the analysis can be separated in 4 categories, according to their content:

- Socio-Demographic (Table 1)
- Building (Table 2)
- Heating Appliances (Table 3)
- Consumer's Energy Sensitivity (Table 4)
- Summary statistics of each variable have been added in the tables, including number of samples per category for categorical and ordinal variables (in brackets) and mean and median values for numerical data.

Table 1 - Description and basic statistics of Socio-Demographic Variables.

SOCIO – DEMOGRAPHIC VARIABLES		
Variable name	**Variable type**	**Summary statistics**
1. Nationality	Categorical	• Finnish [4574] • Other[41]
2. Marital status	Categorical	• Married or cohabitant [3702] • Divorced or separated [377] • Widow [332] • Unmarried [218]
3. Education level	Ordinal	1. No Education [66] 2. Basic Level [408] 3. Upper secondary level [1942] 4. Lower-Degree Level Tertiary [1279] 5. Higher-Degree Level Tertiary [858] 6. Post-Graduate [79]
4. Employment situation	Categorical	• Retired [2573] • Employed [1830] • Unemployed [140] • Student [24] • Unable to work [23] • Home maker [14] • Inactive (not seeking employment) [18]
5. # people in household	Numerical	mean value: 2.37, median:2
6. # kids in household	Numerical	mean value: 0.26, median:0
7. # elderly people in household	Numerical	mean value: 0.84, median:1
8. Diseases that are worsened by damp or cold	Yes/No	Yes [1325], No [3303]
9. Lifepreserving equipment	Yes/No	Yes [145], No [4468]
10. Perception of financial situation	Ordinal	1. Not enough money for primary needs (food and energy bills) [91] 2. Enough money for primary needs, but not for non-basic expenses [576] 3. Enough money for primary needs and sometimes for non-basic expenses [2073] 4. Enough money for primary needs and often for non- basic expenses [1888]
11. Ability to maintain adequate temparature	Yes/No	Yes [4417], No [219]
12. Social benefits	Yes/No	Yes [536], No [4125]

Table 2 - Description and basic statistics of Building Variables

BUILDING VARIABLES		
Variable name	**Variable type**	**Summary statistics**
1. Region (Regional State)	Categorical	• Southern Finland [2352] • Southwestern Finland [872] • Eastern Finland [797] • Western and Inland Finland [357] • Northern Finland [127] • Lapland [67]
2. Region (Eurostat NUTS)	Categorical	• Helsinki-Uusimaa [1887] • Etela-Suomi [1078] • Pohjois- ja Ita-Suomi [974] • Lansi-Suomi [469]
3. Location type	Categorical	• Inside a city/town (urban) [2261] • On the outskirts of a city/town [1321] • In the countryside (rural) [1028] • Other [26]
4. Housing type	Categorical	• Detached/Semi-detached house [4280] • Terraced house [202] • Mid floor flat/Apartment [79] • Top floor flat/Apartment [47] • Ground floor flat/Apartment [21]
5. Year of building	Ordinal	1. Before 1945 [342] 2. 1946-1960 [659] 3. 1961-1970 [476] 4. 1971-1980 [856] 5. 1981-1990 [1066] 6. 1991-2000 [534] 7. 2001-2005 [262] 8. After 2005 [440]
6. Housesize (m2)	Ordinal	1. Less than 40 m^2 [15] 2. 41-60 m^2 [97] 3. 61-90 m^2 [500] 4. 91-120 m^2 [1398] 5. 121-150 m^2 [1279] 6. Over 150 m^2 [1347]
7. Renovated household	Yes/No	Yes [3605], No [1001]
8. Renovation to improve energy efficiency	Yes/No	Yes [2943], No [1684]
9. Non-structural energy efficiency improvements	Yes/No	Yes [2943], No [1684]

Table 3 - Description and basic statistics of Heating Appliance Variables.

APPLIANCE VARIABLES		
Variable name	**Variable type**	**Summary statistics**
1. Central heating (no cooling)	Yes/No	Yes [3305], No [1356]
2. Central heating and cooling	Yes/No	Yes [396], No [4265]
3. Independent heating and cooling unit	Yes/No	Yes [1751], No [2910]
4. Independent heater unit	Yes/No	Yes [2280], No [2381]

Table 4 – Description and basic statistics of Energy-Sensitivity Variables

ENERGY – SENSITIVITY VARIABLES		
Variable name	**Variable type**	**Summary statistics**
1. Aim to reduce energy bill by changing electricity/heating use	Yes/No	Yes [3789], No [822]
2. Energy efficiency measure: Switch off the heating in part of the house	Yes/No	Yes [1025], No [3636]
3. Energy efficiency measure: Reduce temperature in living/ dining room or bedrooms	Yes/No	Yes [2555], No [2106]
4. Energy efficiency measure: Reduce temperature when house is empty	Yes/No	Yes [1210], No [3451]
5. Energy efficiency measure: Decreased ventilation	Yes/No	Yes [843], No [3818]
6. Energy efficiency measure: Reduce use of appliances	Yes/No	Yes [570], No [4091]
7. Energy efficiency measure: Turn off lights when not needed	Yes/No	Yes [4036], No [625]
8. Energy efficiency measure: Turn off TV/stereos instead of putting them on standby	Yes/No	Yes [2778], No [1883]
9. Believe they would benefit from energy advisory services	Yes/No	Yes [1988], No [2595]

As we can see, 2 alternative ways of specifying region have been used (Building variables 1 and 2), in order to examine the importance of both of them during feature selection. They were both created by grouping the collected regional data from the survey according to Eurostat NUTS (based on population) and according to the division of Regional State Administrative Agencies in Finland (Figure 2 and Figure 3 show the two alternative divisions). The original data collected with the questionnaire were the following:

- Uusimaa [1783]
- Varsinais-Suomi [720]

- Pohjois-Karjala [436]
- Pohjois-Savo [230]
- Kanta-Häme [174]
- Kymenlaakso [133]
- Keski-Suomi [113]
- Satakunta [111]
- Pohjois-Pohjanmaa [111]
- Etelä-Pohjanmaa [100]
- Etelä-Savo [98]
- Päijät-Häme [88]
- Pirkanmaa [85]
- Lappi [65]
- Etelä-Karjala [53]
- Pohjanmaa [37]
- Kainuu [13]
- Keski-Pohjanmaa [6]

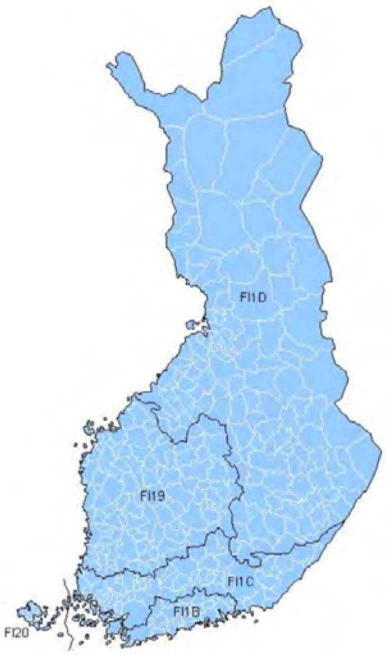


Figure 2 - Division of Finland according to Eurostat NUTS (based on population).

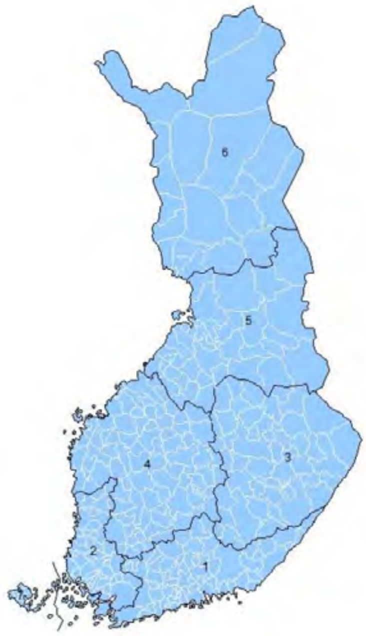


Figure 3 - division of Regional State Administrative Agencies in Finland.

Annual Electricity and Heating costs as share of income (from now on we will refer to it as Electricity/Heating costs) was considered as the output variable for both feature selection and clustering. This variable (summarised in

Table 5) was calculated from the collected data, by dividing the sum of annual electricity and heating costs by the average income of the annual income category that corresponded to each sample:

Annual Electricity and Heating costs as share of income (Electricity/Heating costs):

$$\frac{Cost\ of\ Electricity + Cost\ of\ oil + Cost\ of\ gas + Cost\ of\ solid\ fuels\ (wood/pellets) + Other\ fuel\ costs}{Average\ annual\ income}$$

Table 5 - Description and basic statistics of Output Variable.

OUTPUT VARIABLE		
Variable name	**Variable type**	**Summary statistics**
Annual Electricity and Heating costs as share of income (Electricity/Heating costs)	Numerical	mean value: 6.44%, median: 4.19%, std: 0.10

3.1.3 Approach

The data described in the previous subsection were analysed to create categories of consumers, based on their similarity on Electricity/Heating costs and other demographic characteristics. For this reason, a feature selection and clustering methodology has been designed, implemented and applied to the data. Before getting into technical details, it was considered important to get an overview of the analysis approach, by focusing on the overall structure of the procedure and the interaction of the analysis components. A step-to-step representation of the methodology can be found in Figure 4.

As shown in the diagram, the 12 socio-demographic, 9 building and 4 heating appliance variables served as input of the algorithm, while the output variable was Electrictity/Heating costs. Ordinary Least Squares (OLS), Variable Inflection Factor (VIF) and Least Absolute Shrinkage and Selection Operator (LASSO) techniques were combined during feature selection, in order to identify the variables that can be best used to explain the variations of the output variable. The features identified as important were then used by a Decision Tree Regressor, for the creation of the clusters. The Decision Tree algorithm was selected as the most appropriate clustering approach for the specific problem, because it performs directed clustering. The target variable used by the Regressor was again Electricity/Heating costs. The clusters were the end-leaves of the Decision Tree that was created based on the automatically generated rules, after omitting outliers. The technical aspects of the methodology components are described in greater detail in the following subsections.

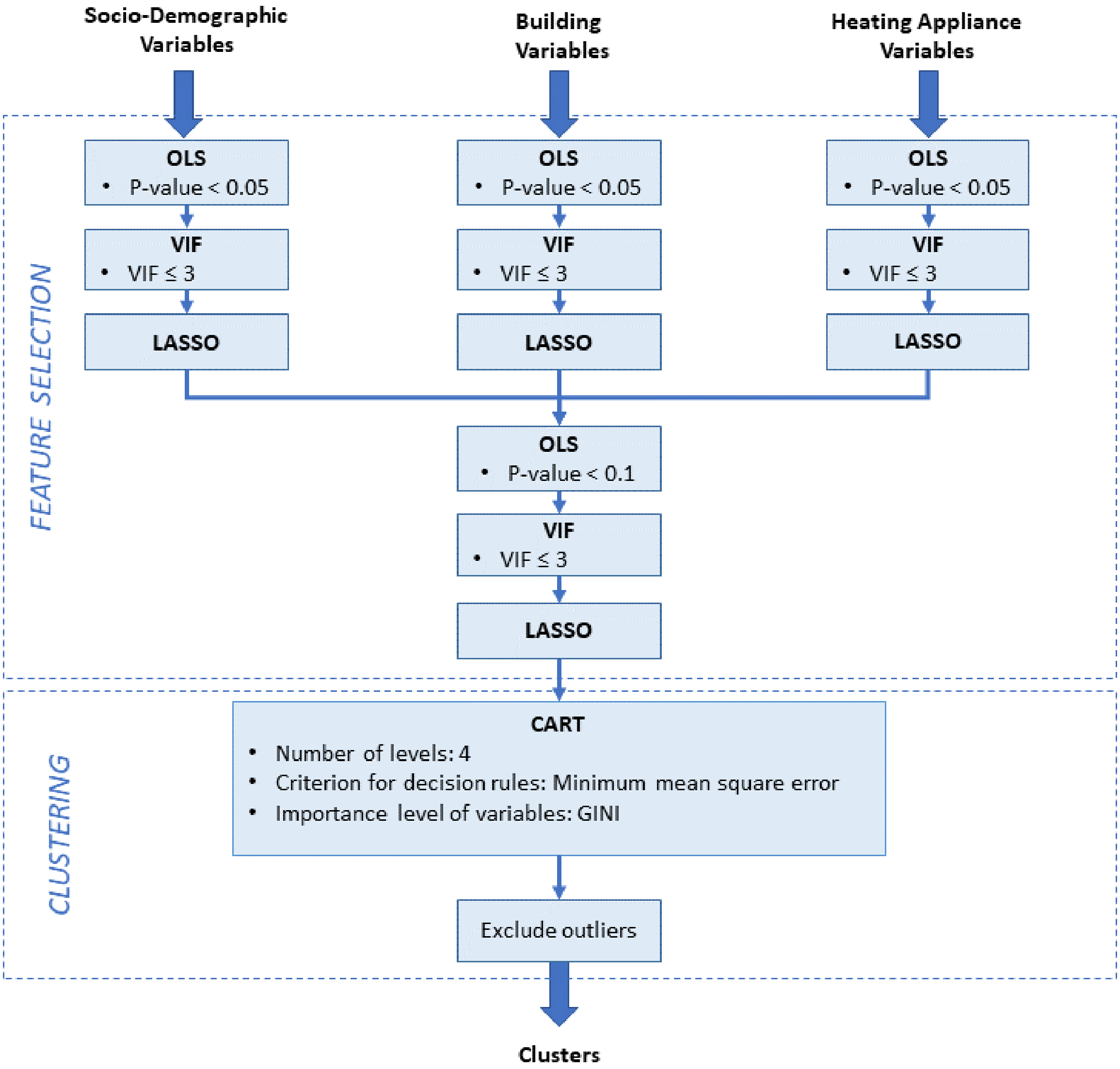


Figure 4 – Step-to-step methodology diagram.

3.1.4 Data cleaning/preprocessing

Data cleaning is a necessary first step before the application of any statistical procedure, in order to ensure reliability and limit bias of the analysis results. Data that originate from questionnaires often include erroneous values that can be related with participants' misunderstanding of the question (e.g. insert monthly instead of annual costs), typos in open-text questions (e.g. 100 instead of 1000), while sometimes non-obligatory fields are left blank. In the current analysis, data cleaning and pre-processing were performed in steps, as shown below.

1. **Dealing with outliers of output variable:** Outliers in Electricity/Heating costs, which served as the output variable during all stages of the analysis, could be critical for the analysis results. Thus, cases where the output variable took values that were higher than 100% were treated as outliers and

omitted from the dataset. Two box-plots of the output variable are shown below, prior to (left side - Figure 5) and after (right side - Figure 6) omitting the outliers.

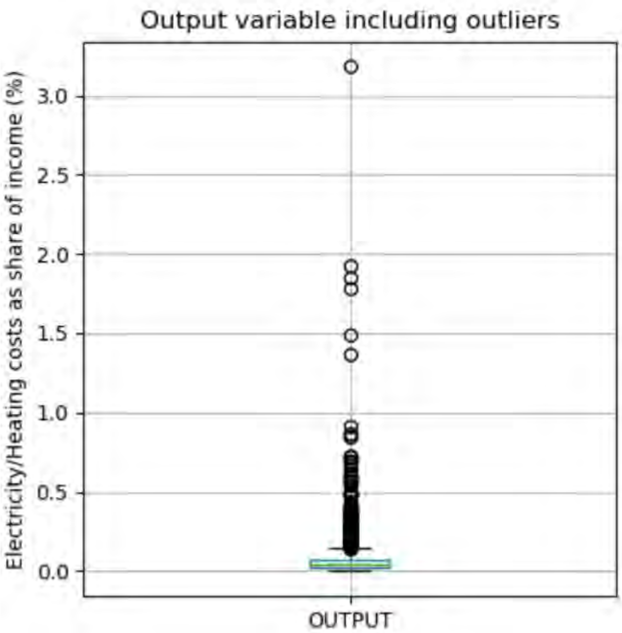


Figure 5 - Box-plot of Electricity/Heating costs (%) including outliers.

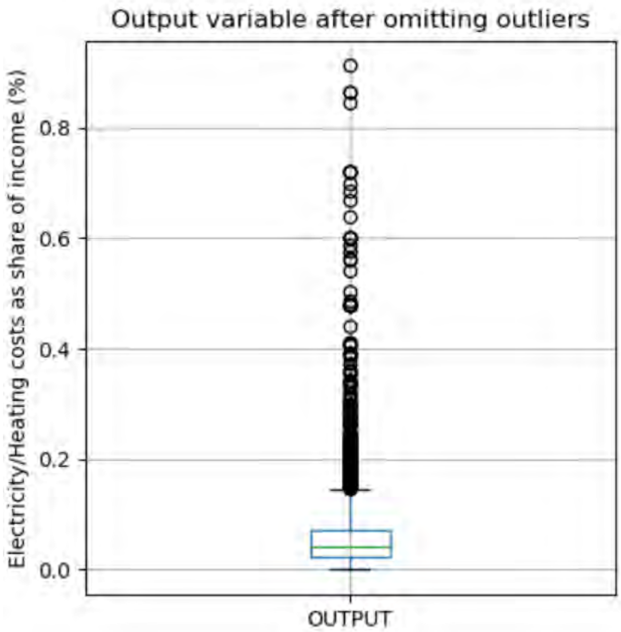


Figure 6 - Box-plot of Electricity/Heating costs (%) after omitting outliers.

During the analysis, more outliers were detected and removed based on the results of the Decision Tree Regressor, before the creation of the final clusters. Those outliers were related with inconsistences on other data variables that made certain samples dissimilar with the rest of the dataset and could be related with either erroneous input or very uncommon characteristics.

2. **Managing missing values:** Even in cases that the number of available data is big, we cannot afford throwing away samples just because of some missing values. Instead of this, we followed the most common practice, by filling in missing values with some basic statistic (i.e. mean values, median, most common value). Specifically, missing values in numerical variables were filled in using the mean value of all samples, ordinal data were filled in using the median and categorical data were filled in using the most common category.
3. **Transform categorical variables into dummy variables:** Regression based techniques cannot recognise categorical variables, as they usually use distance-based algorithms as criteria. For this reason, categorical variables were transformed in binary dummy variables.
4. **Regularisation of the data:** Since the regression techniques used all have distance-based criteria, it was important to ensure that conclusions won't be biased from differences in the scaling of variables. Thus, all data were

regularised by centring them (set mean value to zero) and then scaling them to unit variance.

3.1.5 Feature Selection

The main tool used during feature selection was a sequence of three techniques: Ordinary Least Squares (OLS), Variable Inflection Factor (VIF) and Least Absolute Shrinkage and Selection Operator (LASSO).

Ordinary Least Squares (OLS) was performed using annual electricity and heating cost as the dependent variable. The p-values of the individual variables were used as a criterion of whether they influence the dependent variable in a statistically significant manner. Statistical significance of 95% (p-value<0.05) was used as a threshold during OLS application on categories of variables (i.e. socio-demographic, building, heating appliances), while a less strict 90% (p-value<0.1) for the OLS on the combined variables.

Variable Inflection Factor (VIF) is a metric that quantifies the degree of multicollinearity in OLS and was used to detect correlations between the variables identified as influential by OLS. There is no formal threshold for critical values of VIF, but a range of 5 – 10 is most commonly used in literature. A stricter threshold was imposed in the current analysis, setting the critical value to 3. Thus, for variables with VIF > 3 after OLS, the LASSO regression was employed.

Least Absolute Shrinkage and Selection Operator (LASSO) is a penalised regression analysis technique that can be used for feature selection, as it sets redundant predictors to zero. In our analysis, it was used to determine the variables with multicollinearity that can be sacrificed, being the least-relevant ones and at the same time allowed us to find out the ones that contain the most information.

As shown in Figure 4, this 3-step sequence took place 4 times. One time for each category of variables (i.e. socio-demographic, building, heating appliances) and one final time using as input the merged outcomes of the other three. The set of variables that was determined by the forth application, were the input variables that were used for the clustering analysis. Though, as it will be discussed below, the set of influential variables was further reduced during clustering analysis, after the application of the Decision Tree Regressor.

3.1.6 Clustering Analysis

A Decision Tree Regressor was selected as the optimal approach for clustering survey participants, based on their Electricity/Heating cost and taking under consideration the variables identified during feature selection. The algorithm works by automatically creating a set of rules that are used to divide the sample based on an optimisation criterion, until a termination criterion is satisfied. If no stopping criterion is provided, the generated tree can be complicated and provide a level of detail that is unnecessary for the current analysis. Thus, we have limited the depth of the tree by setting a maximum number of levels (4) as a termination criterion. The optimisation criterion used was minimal Mean Square Error (mse), which is the most popular in literature. The clusters were then the end-leaves of the tree. Clusters with significantly small number of samples (number of sample<10) were considered as outliers, since they contained cases with large dissimilarity from the rest of the samples.

A great advantage of the particular algorithm is the fact that it also ranks feature based on their importance, by setting weights to input variables according to their contribution to the clustering (and outlier detection), using on GINI importance metric. So, after the completion of the clustering, the variables identified as important via feature selection were further reduced and weighted based on their importance.

After the completion of clustering, the created participant groups were studied by examining the decision rules and doing some basic profiling. They were also investigated in relation with a set of variables that can be used to identify the ones that have greater possibility to contain vulnerable consumers:

- Electricity/Heating costs
- Ability to maintain adequate temperature in the households
- Social benefits
- Perception of Financial situation

3.2 Analysis results

3.2.1 Feature Selection

The list of variables that were identified as important, as output of feature selection can be found in

Table 6. After the end of feature selection, an extra OLS was implemented to further study the influence of the selected features and to find out their statistical significance. The results are summarised in Table 7, where the statistical significance is represented in the last column using stars, as shown below:

- * statistically significant in 90% confidence level
- ** statistically significant in 95% confidence level
- *** statistically significant in 99% confidence level

Table 6 - List of variables determined as important, as output of feature selection algorithm.

Category	Variables
Socio-demographic	Marital status
	Education level
	Employment
	Financial situation
	Ability to maintain adequate temperature
	Number of kids
Building	Location
	Housing type
	Year of building
	Renovation to improve energy efficiency
Appliances	Central heating and cooling

Table 7 - Basic OLS statistics for selected variables.

Variables	coef	std err	t	P>|t|	Stat. Significance
Marital status: Married or cohabitant	-0.0108	0.001	-7.333	0.000	***
Education level: Higher-Degree Level Tertiary	-0.0041	0.002	-2.662	0.008	***
Education level: Lower-Degree Level Tertiary	-0.0027	0.002	-1.773	0.076	*
Employment situation: Retired	0.0133	0.002	8.556	0.000	***
Employment situation: Unable to work	0.0042	0.001	2.939	0.003	***
Employment situation: Unemployed	0.0084	0.001	5.759	0.000	***
Ability to maintain adequate temperature	-0.0061	0.001	-4.117	0.000	***
Perception of Financial situation	-0.0139	0.002	-9.082	0.000	***
# Kids in the household	0.0046	0.002	3.037	0.002	***
Location type: Inside a city or town (urban)	-0.0043	0.002	-2.349	0.019	**
Location type: On the outskirts of a city or town	-0.0030	0.002	-1.655	0.098	*
Housing type: Ground floor flat/Apartment	-0.0035	0.001	-2.473	0.013	**
Housing type: Mid floor flat/Apartment	-0.0066	0.001	-4.593	0.000	***
Housing type: Terraced house	-0.0059	0.001	-4.107	0.000	***
Housing type: Top floor flat/Apartment	-0.0041	0.001	-2.859	0.004	***
Year of building	-0.0082	0.002	-5.082	0.000	***
Renovation to improve energy efficiency	-0.0034	0.002	-2.149	0.032	**
Central heating and cooling	-0.0024	0.001	-1.647	0.100	*

We need to clarify that variables significant in 90% significance level were acceptable only during the last stage of feature selection, where OLS was performed with merged input variables. For the three OLS applications for each one of the variable categories a stricter threshold of $p<=0.05$ was used. We can also see that different statistics have been calculated for specific option of the categorical variables. This is due to the fact that they were transformed to dummy variables during pre-processing.

3.2.2 Clustering

Clustering was implemented with the application of a Decision Tree Regressor taking under consideration the selected variables (

Table 6) and Electricity/Heating costs as the target-variable. The rules of the decision tree that was created are summarised in Figure 7. The (8) end-leaves of the decision tree with number of samples greater than 10 were the actual clusters. End-leaves with smaller than 10 samples were considered as outliers, as they contained samples with such great dissimilarities with the rest of the dataset that they could not be grouped into any of the clusters. A closer look on those outliers confirms the fact that they include uncommon data points that in some cases could even be related with erroneous recordings. More specifically, the samples omitted as outliers included:

- 1 sample with number of kids ≥ 13
- 1 sample with Perception of Financial Situation 1 or 2 with Central heating and cooling
- 7 unemployed and unmarried people with Perception of Financial Situation 1 or 2, that live in households built prior to 1959
- 8 unemployed and unmarried people with Perception of Financial Situation 1 or 2, that live in households built after 1959

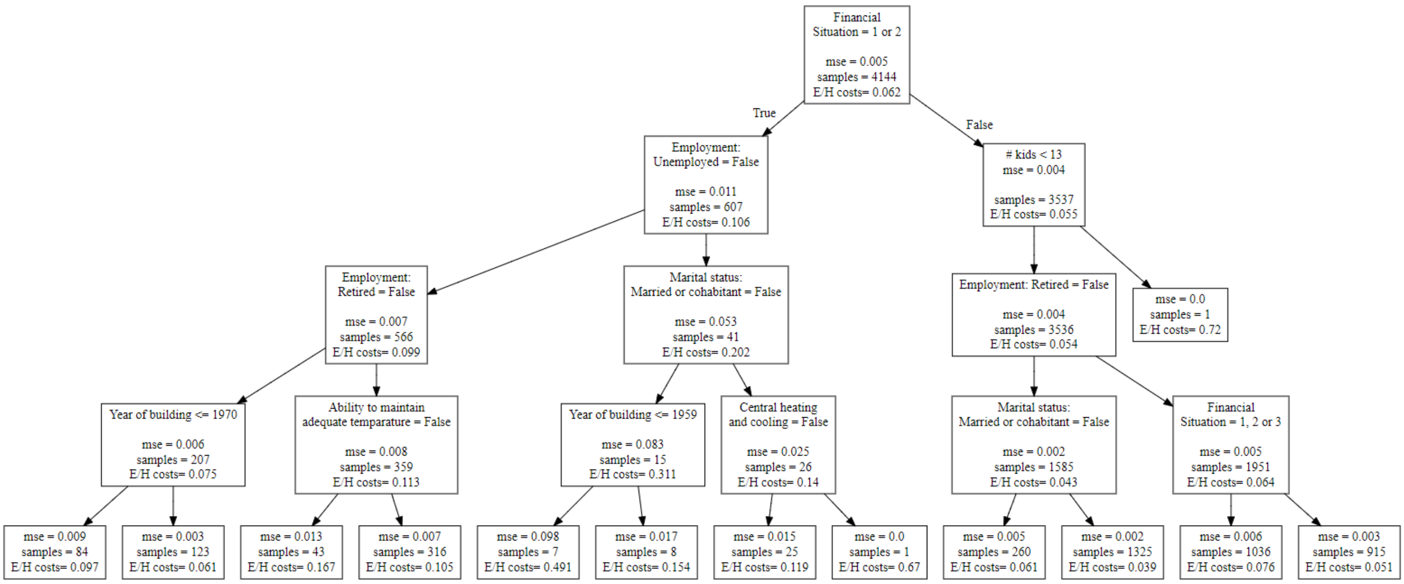


Figure 7 - Decision Tree Regressor Rules.

During clustering, the list of important characteristics got further reduced because the termination criterion we set did not allow a full development of the tree. Decision Tree Regressor has assigned weights to the variables according to their importance, using GINI importance metric. The final set of variables identified as important can be found in Table 8.

It is important to clarify that certain of those variables were used for outlier detection and not for clustering (i.e. # kids in household, central heating and cooling). This fact does not make those variables less important, as they are critical for ensuring the success of the clustering. This is also depicted from their importance weights.

The fact that Perception of Financial situation has been identified as by far the most important variable (GINI importance 0.38 versus 0.22 of the second highest) confirms that - despite the fact that people's perspective can often be subjective - this variable is highly related with Electricity/Heating costs. The variable with the second highest importance is Employment situation, which is a common factor for identifying people at risk. Year of building, the third on the raw, is directly related with heating costs since older houses are usually more difficult to warm up due to older isolation technology or humidity on the walls and at the same time leaving in old houses is often an indication of worse financial situation since old houses tend to be cheaper.

Table 8 - Ranged variables used for clustering by Decision Tree Regressor, based on their importance (GINI Importance).

Variable	Category	Importance
Perception of Financial situation	Socio – demographic	0.38
Employment situation	Socio – demographic	0.22
Year of building	Building	0.11
# kids in household	Socio – demographic	0.10
Marital status	Socio – demographic	0.09
Central heating and cooling	Appliances	0.07
Ability to maintain adequate temparature	Socio – demographic	0.03

After excluding the outliers from the dataset, the decision tree took the form in Figure 8, where we can see the structure of the decision rules for the creation of 8 clusters. The characteristics of people that belong to each cluster based on the decision rules, can be found in Table 9. Since Perception of Financial situation was considered as the most important variable by the Decision Tree Regressor, we can see that there are no people with Financial Situation 1 or 2 in the same cluster with people with Financial Situation 3 or 4.

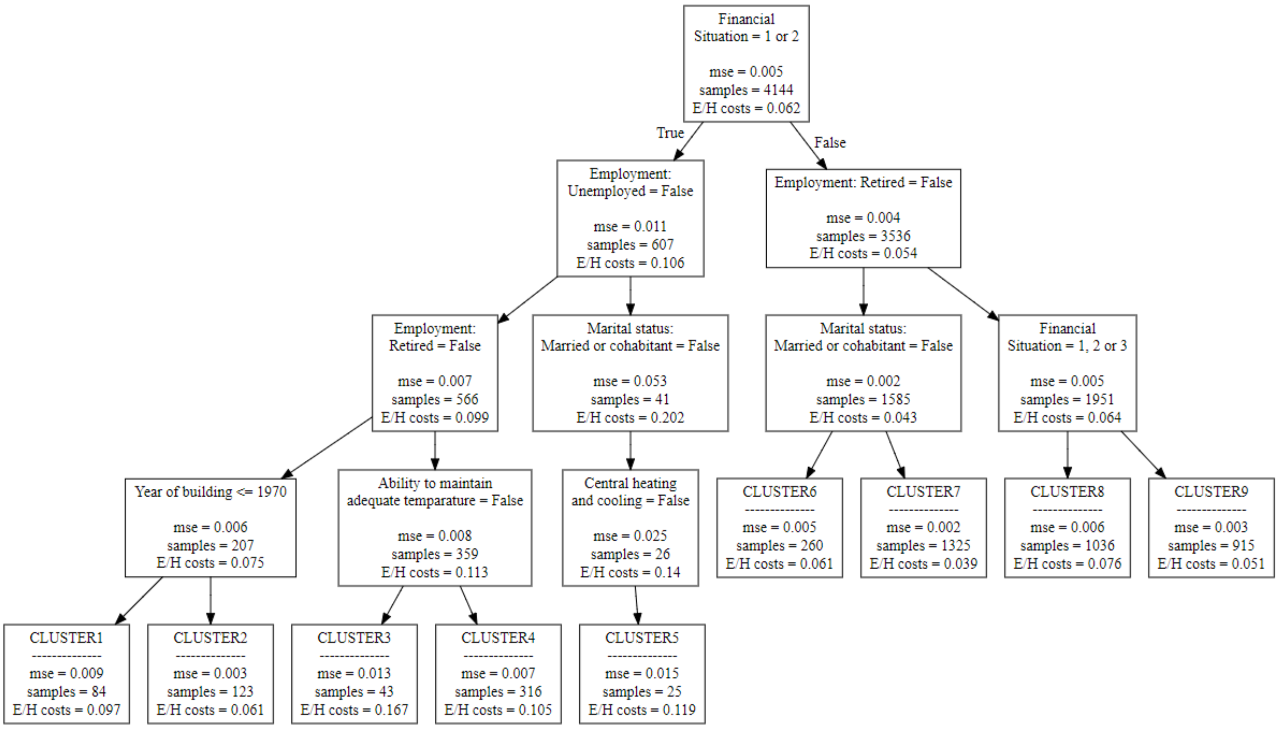


Figure 8 - Clusters as outcome of Decision Tree.

As a result, the following 9 clusters were identified:

Table 9 - Characteristics of cluster participants based on decision rules.

CLUSTERS	# SAMPLES	% OF SAMPLE	CHARACTERISTICS
CLUSTER1	84	2%	• Financial situation: *1 or 2* • Employment status: *Employed, Student, Unable to work, Home-maker or Inactive* • Year of building: *<=1970*
CLUSTER2	123	3%	• Financial situation: *1 or 2* • Employment status: *Employed, Student, Unable to work, Home-maker, Inactive* • Year of building: *>1970*
CLUSTER3	43	1%	• Financial situation: *1 or 2* • Employment status: *Retired* • Ability to maintain adequate temperature: *No*
CLUSTER4	316	8%	• Financial situation: *1 or 2* • Employment status: *Retired* • Ability to maintain adequate temperature: *Yes*
CLUSTER5	25	1%	• Financial situation: *1 or 2* • Employment: *Unemployed* • Marital status: *Married or cohabitant* • Central heating and cooling: *No*
CLUSTER6	260	6%	• Financial situation: *3 or 4* • Employment status: *Employed, Student, Unable to work, Home-maker, Inactive or Unemployed (not Retired)* • Marital status: *Divorced/separated, Unmarried or widow (not*

			married or cohabitant)
CLUSTER7	1325	32%	• Financial situation: *3 or 4* • Employment status: *Employed, Student, Unable to work, Home-maker, Inactive or Unemployed (not Retired)* • Marital status: *Married or cohabitant*
CLUSTER8	1036	25%	• Financial situation: *3* • Employment status: *Retired*
CLUSTER9	915	22%	• Financial situation: 4 • Employment status: *Retired*

In Table 10, some vulnerability criteria of the clusters are examined, to identify clusters that have a high possibility to include vulnerable customers. These criteria are:

- The proportion of people in the dataset with Electricity/Heating costs greater than 10%. We have used 10% as threshold, following the definition of energy poverty in several EU countries such as UK, even though this is not an official criteria or threshold in Finland.
- The proportion of people in the dataset with Electricity/Heating costs greater than 20%.
- Average proportion of Electricity/Heating costs per cluster.
- Share of survey participants that declared inability to maintain adequate temperature in their household.
- Share of survey participants that receive social benefits.
- Proportion of people that declared they do not have enough money for primary needs (Perception of Financial situation = 1)
- Proportion of people that declared they have enough money for primary needs, but not for non-basic expenses (Perception of Financial situation = 2)

The classification of the clusters into vulnerable and non-vulnerable was done based on the combination of the examined characteristics and is shown in Table 10 using colours (light red for vulnerable, light green for non-vulnerable).
A first observation is the fact that all clusters, including the ones classified as non-vulnerable, contain a certain proportion of people that receive social benefits, which is not linearly correlated with any of the examined variables. Though, the 2 clusters (5 and 1) with the highest proportion of people that receive social benefits are the ones with the lowest proportion of people with perception of financial situation = 1 among the vulnerable clusters. This can be interpreted as a confirmation that social benefits are indeed helpful for improving Finnish people’s financial situation. This can also mean that some vulnerable might not be seeking social security measures even when in need. This has especially been the concern with elderly.

In overall, the clusters that were classified as not at risk include 85% of the survey participants. In greater detail, clusters 3, 4, 5 and 1 all include a significant proportion of households with Electricity/Heating costs > 10% (over 30% in all cases, while about 70% in cluster 3 that has the highest) and they have average Electricity/Heating costs higher than 10%. All those 4 clusters contain people with Perception of Financial situation 1 or 2. Specifically, clusters 3 and 4, that have the highest proportion of people with Electricity/Heating costs > 10% have higher proportion of people with Perception of Financial situation 1 than the other two clusters, while cluster 3 contains by far the highest proportion of people with Electricity/Heating costs > 20%. A possible reason for the large percentage of Electricity/Heating costs in those two clusters is the fact that they only contain retired people, that tend to spend more time at home and are usually less resistant in low temperatures in comparison with younger people. Cluster 5 contains exclusively unemployed people and is the cluster with highest proportion of people that receive social benefits, 1.8 times higher than cluster 1 that follows. On the other hand, clusters 8, 6, 9 and 7 contain participants with Perception of Financial situation 3 or 4, while the proportion of people with Electricity/Heating costs > 10% is lower than 25%. They also contain only a non-negligible but significantly smaller proportion of people with inability to maintain an adequate temperature in the household (lower than 4%).

Cluster 2 is an interesting case, since it contains people with relatively low proportion of people with Electricity/Heating costs > 10%, but on the same time a high proportion of people with Perception of Financial situation 1. Specifically, this proportion is on the same level with clusters 3 and 4. At the same time, they contain a significant proportion of people with inability to maintain an adequate temperature in the household (15%) and proportion of people that receive social benefits (23%). All those reasons contribute into classifying the specific cluster as vulnerable and it can serve as a great example showing that Electricity/Heating costs alone do not give a complete picture of financial vulnerability but need to be combined with other factors. Clusters 1 and 2 include people with similar demographic characteristics, that mainly differentiate to the age of their houses. The fact that people of cluster 1 live in old houses (year of building prior to 1970) is probably the reason for their increased Electricity/Heating costs and higher inability to maintain an adequate temperature on the household.

Another interesting observation comes by comparing clusters 3 and 4. Although they both contain a high proportion of people with Perception of Financial situation =1 (about 14% and 15%) and Electricity/Heating costs > 10%, they have a surprisingly high difference in ability to maintain adequate temperature in the household (100% versus 0%). It seems that the reason for this is the fact that they prioritise differently, so cluster 3 gives higher priority to the household's temperature and use of electricity appliances, while cluster 4 is more likely to sacrifice some of their heating/electricity needs to satisfy better other prior needs.

Table 10 – Identification of vulnerable clusters (Red: vulnerable, Green: not at risk).

CLUSTERS	E/H costs > 10%	E/H costs > 20%	Average E/H costs	No ability to maintain adequate temperature	Social Benefits	Financial situation=1	Financial situation=2
CLUSTER3	70%	28%	17%	100%	26%	14%	86%
CLUSTER4	44%	9%	11%	0%	10%	15%	85%
CLUSTER5	36%	12%	12%	20%	64%	4%	96%
CLUSTER1	35%	8%	10%	33%	35%	6%	94%
CLUSTER8	24%	4%	8%	3%	6%	0%	0%
CLUSTER2	17%	3%	6%	15%	23%	15%	85%
CLUSTER6	13%	3%	6%	4%	10%	0%	0%
CLUSTER9	10%	2%	5%	1%	4%	0%	0%
CLUSTER7	5%	1%	4%	4%	16%	0%	0%

3.3 Vulnerable consumer characterisation

The Clustering analysis described in the previous sections was used to identify the characteristics affecting the energy expenditure of a household and vulnerability. The identified clusters are further examined to estimate the differences in energy behaviour and attitude in the next section. Further characterisation of vulnerable consumers identified from the survey can be found in the Vulnerable Consumers Fuel Poverty Report.

In order to obtain more specific information to be used in designing the ASSIST ACTION a short questionnaire was created for the vulnerable consumers. The results of the questionnaire is presented in this chapter.

3.3.1 Vulnerable consumer characterisation (per cluster)

The following figures are used to the further examine the clusteres based on their energy habits, using the energy-sensitivity variables. The same colour-coding as in Table 10 has been used for separating clusters that have been classified as vulnerable from the ones that are not at risk. Table 10 and Figure 10 study each group's attitude towards saving and study whether participants are willing to accept external help from some specialist to decrease their electricity/heating costs.

More specifically, according to Figure 9, significant proportions of participants from vulnerable and non-vulnerable clusters are willing to adjust their energy consumption habits in order to reduce their energy bills. The lowest proportion (76%) is in cluster 5 (classified as vulnerable) while the highest (87%) in cluster 6 (identified as non-vulnerable). Clusters 1 and 3 have equally high proportions

(86%), showing that despite the fact that they include the highest share of cases that cannot maintain an adequate temperature at home (see Table 10), they are willing to further reduce their heating/electricity usage. The proportions related with the vulnerable clusters are slightly lower, probably due to the fact that sometimes in households most at risk people have already limited heating/electricity consumption (in some cases too low to fully cover their needs). This difference is very small though. In overall, the small range between the proportions and the high similarity between the vulnerable and non-vulnerable clusters can serve as a proof that the willingness to save energy is independent from people's financial situation.

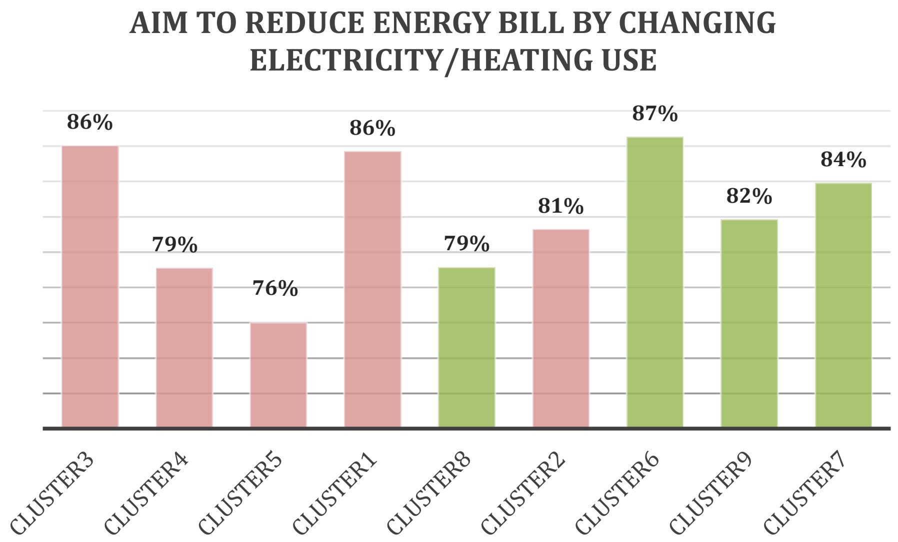


Figure 9 - % of survey participants aiming to reduce energy bill by change electricity/heating use, per cluster.

The proportion of people that think they would benefit from energy advisory services has bigger variation among clusters. Cluster 3 has the highest proportion (63%), followed by cluster 1 (58%), while cluster 9 has the lowest (37%). Here, the proportion is on average higher in the vulnerable clusters. The only cluster from the non-vulnerable clusters with more than half participants believing an energy advisor could help them is cluster 6. This might be related with the demographic characteristics of participants of cluster 6, as it does not contain any elder (no retired people) or married people, so it is more likely to contain young adults and students. Those groups tend to be open in attaining new knowledge and have a better understanding of technology.

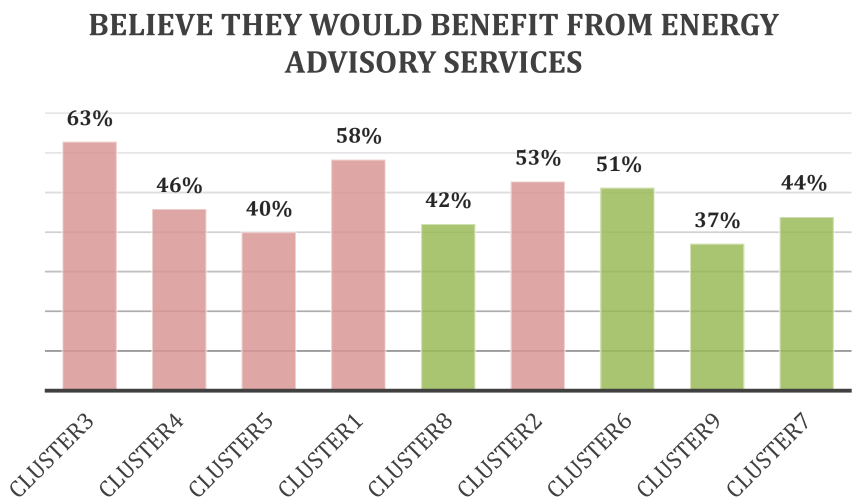


Figure 10 - % of survey participants that believe they would benefit from energy advisory services, per cluster.

3.3.2 Short questionnaire to Vulnerable Consumers

In order to obtain more specific information to be used in designing the ASSIST ACTION a short questionnaire was created for the vulnerable consumers. In Finland, due to advanced social security system, energy poverty (as well as poverty in general) is relatively low. This highlights also the significance of individuals perception of their situation. Therefore perception will also be used in identifying the vulnerable. In order to limit the bias due to the fact that people's perception can be subjective, we have ensured that the identification of vulnerable and non-vulnerable groups was in accordance with the clustering results, that took under consideration multiple factors and used the share of Electricity/Heating costs as target-variable.

Out of the original survey respondents 591 where in the clusters identified as vulnerable, out of these 517 had given their contact information for further questionnaires. Out of these we received 184 responses. In the following paragraph, the main findings on Vulnerable Consumers preferences about energy advisory services are reported.

According to the received responses, the portion of vulnerable consumers that have never heard of programs or services offered to help consumers optimize their energy use, even though energy advisory services are offered by both the National Energy Agency Motiva and NGO's. Majority of the vulnerable consumers that had heard of these services still did not know where to find them.

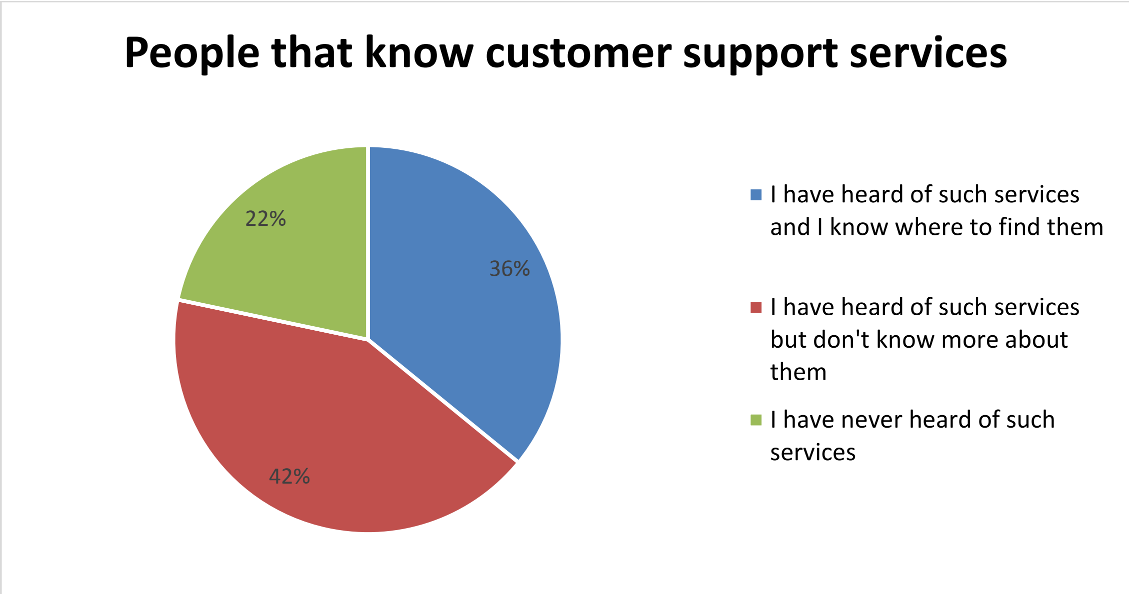


Figure 11 - % of participants by knowledge level of programs by professionals, who provide external help to optimize energy consumption.

Majority of the vulnerable consumers felt that their current knowledge level is sufficient to understand enough about actions suggested in such programs to improve and optimize their energy consumption.

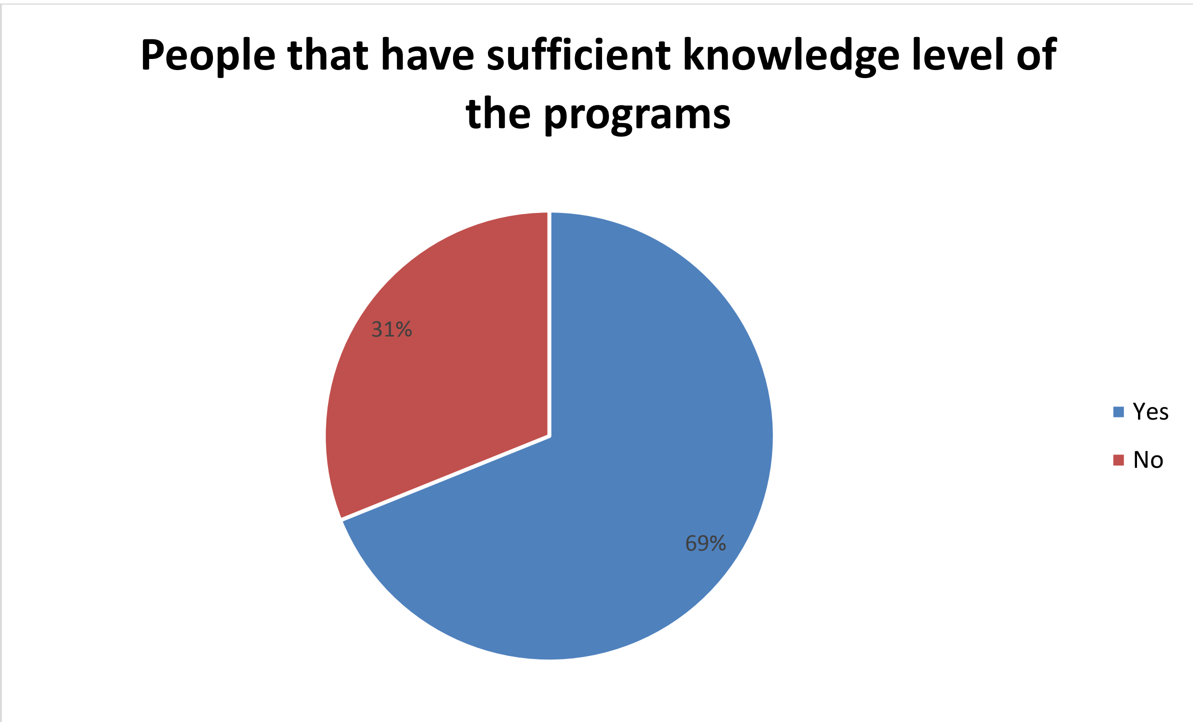


Figure 12 - % of people who felt their current knowledge level was sufficient in order to benefit from such programs.

When inquiring their willingness to participate in such programs that would help them optimize energy consumption, majority were interested in participating.

Interestingly the willingness was not dependent of prior knowledge of these programs. On the other hand, people it was not dependent of the feeling of their current knowledge level either, meaning that people who felt that their current knowledge level was not sufficient to benefit of such programs were as interested or even slightly more interested than people who felt their knowledge level were sufficient.

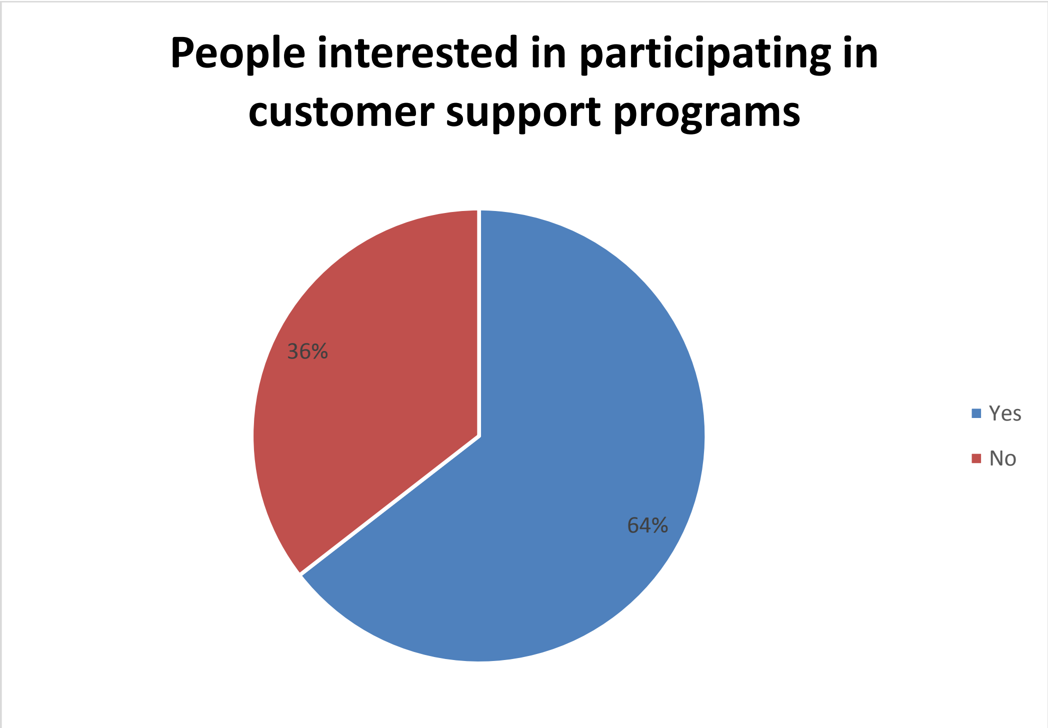


Figure 13 - % of people interested in participating in such programs.

More than one third of people were not interested in participating in such programs for variety of reasons described in Figure 13. Biggest reason that they felt they know enough already. 16% felt that they don't have time and as many that they cannot decrease their usage any more.

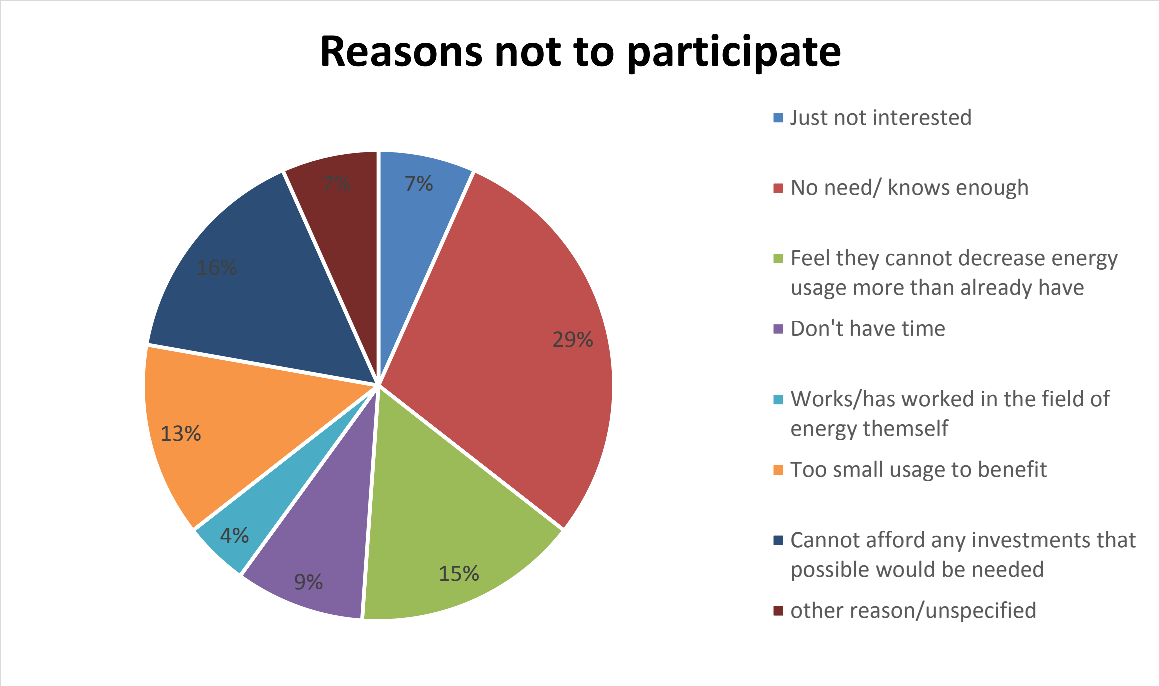


Figure 14 - % of different reasons why people did not want to participate in such programs.

When asked who the vulnerable wanted most to organise such programs, majority thought that consumer organisation would be most suited. Also, University or research entity or current energy company were thought to be suited in organising such programs.

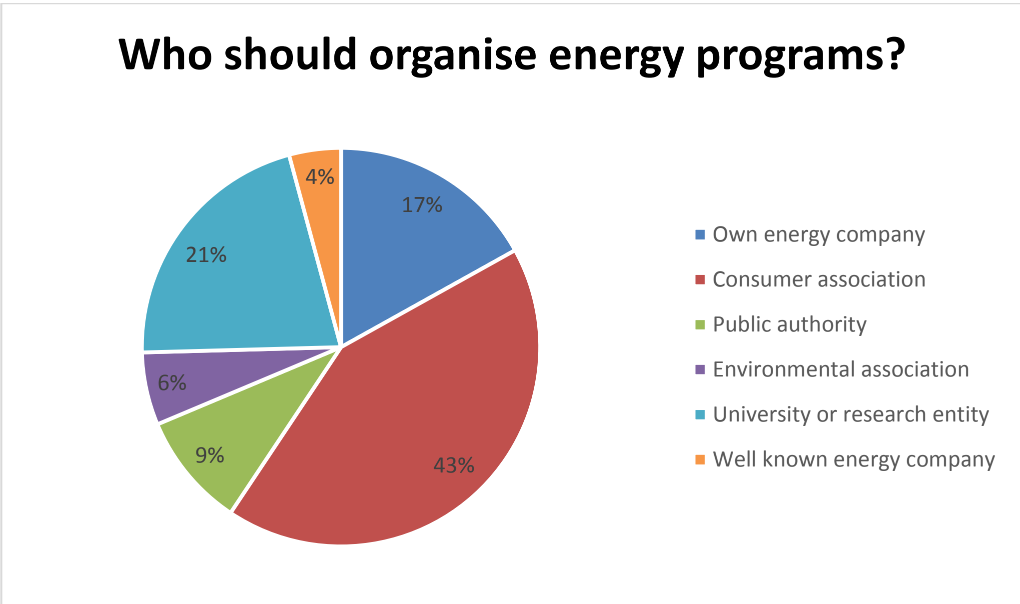


Figure 15 - Who the vulnerable would want to organise such energy programs.

Even though majority of people were interested in participating in such programs, they were less keen on advisors visiting their home. Slight majority of all respondents said that they would not want an advisor to visit their home.

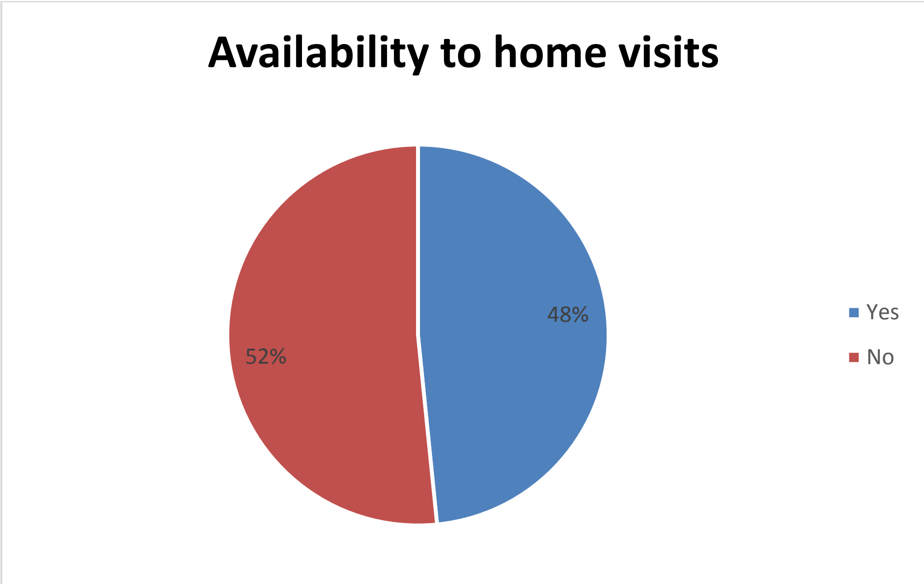


Figure 16 - would participants want receive house visit during such program.

The channels through which people felt they would want to participate and interact in such programs, varied a lot. Via internet page, in person and via email were the most common options. Very few, only 0.5% wanted to participate through social media.

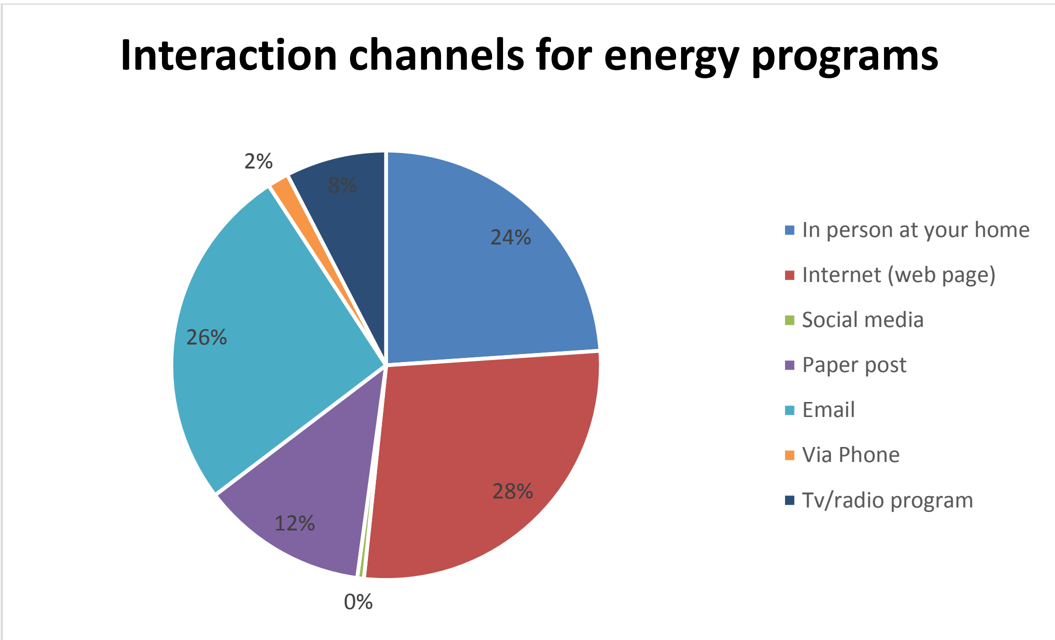


Figure 17 - favourite channels through which people felt they would want to participate and interact in such programs.

When inquiring about the most convenient time to receive advise on how to optimize their energy use, majority felt that the best time would be when receiving their energy bill or signing up for a new contract.

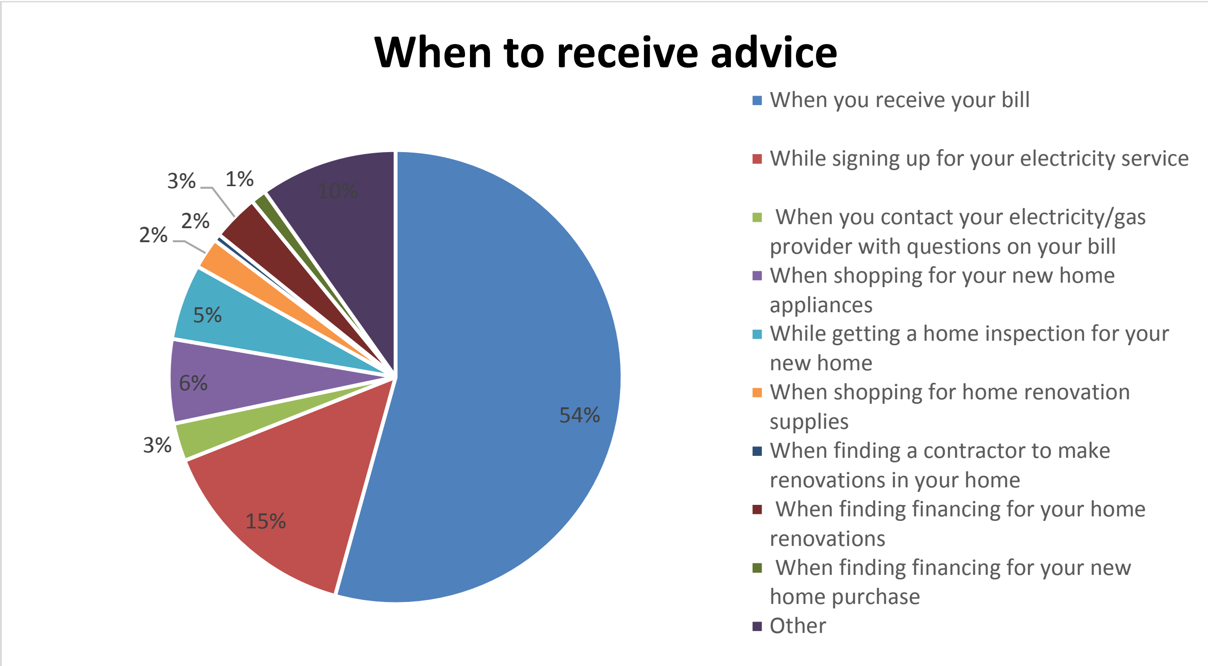


Figure 18 - most convenient time to learn more about the actions to optimize energy consumption.

The things people felt most discouraging to joining such programs varied, most common reason was that it would increase the energy consumption or that it would allow third parties greater access to personal consumption data. The fact that people feel uncomfortable that third parties have access to their consumption data is challenging as personalised advising on energy use requires knowledge of individual circumstances.

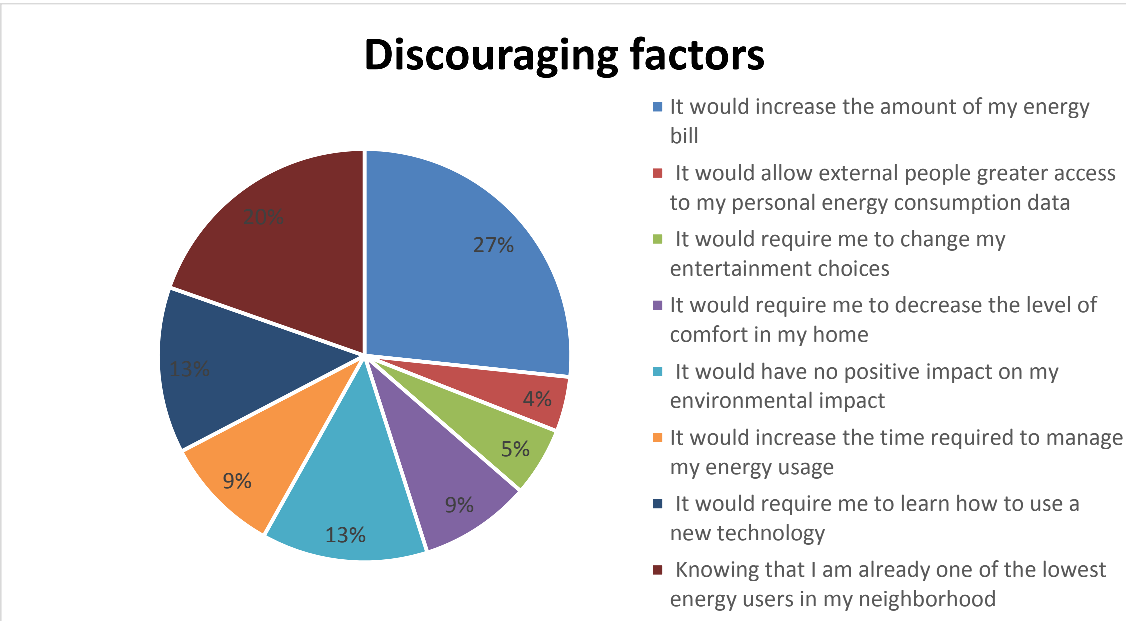


Figure 19 - Most discouraging factor to joining such programs varied.

On the other hand, when asked what people felt was the most encouraging factor, majority felt that it was the possibility to decrease in their energy bill.

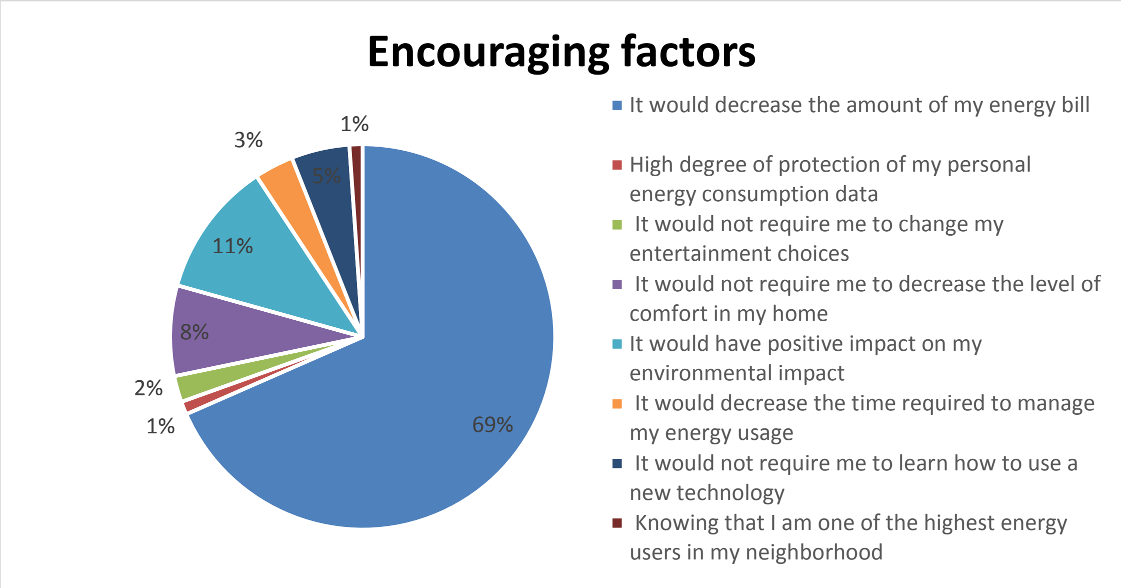


Figure 20 - Factors that would most encourage to join such programs.

The most valuable output the people were expecting to receive varied, but the most common options were customized energy advise to improve energy consumption and audit of energy behaviour. Support service available was considered least important.

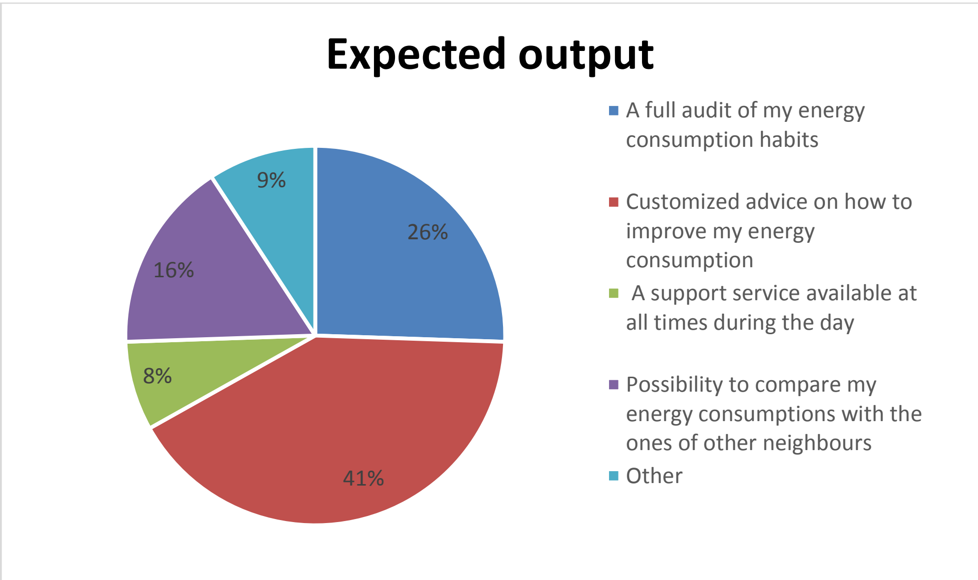


Figure 21 - Most valuable output expected to receive from such programs.

3.4 Conclusions

It appears, from the statistic clustering analysis, that the households that can be considered most at risk of being vulnerable or energy poor are households that have retired, unemployed, student, unable to work, home-maker or inactive and feel that their financial situation is so weak that they cannot afford the basic needs such as food, heating and electricity or that they can just manage but are unable to afford anything more. The result puts emphasis on perception of one’s situation.
In Finland the advanced social security system mitigates poverty as a whole, but there is a structural problem with existing aid schemes: they do not adequately encourage energy efficiency measures. To improve Energy efficiency, energy advisory services are offered by both the National Energy Agency Motiva and NGO’s. Although according to the responses received from the vulnerable, large portion of vulnerable consumers have never heard of programs or services offered to help consumers optimize their energy use, which signifies the need for advisory services to be targeted at these groups. Challenges might occur from people being reluctant for home visits or giving their personal consumption data and these issues need to be carefully addressed when designing the action.

4. Italy – Methodology

4.1 General approach

The general approach that has been developed in this phase of the project arises from the necessity to establish pilot actions to fight energy poverty, that could bring to significant impacts. In order to do that, the objective imposed requires the implementation of a series of different activities in different realities of the Country; this in terms of different areas of the territory, but also different living situations and different urbanization levels (from little municipalities to metropolitan areas). The objective of creating an Action distributed in such a way that the all Country is characterized has also come out from the VCSC - Vulnerable Consumers Steering Committee and the ASSIST Action think tank event (that will be further investigated in D5.2 and D7.1 and D7.2). Given the fact that the different locations in which the actions will be implemented are not known in advance, it is necessary to obtain a complete characterization of the different consumers in the whole territory. On the basis of the results obtained, and taking into account the different resources and availabilities of consumers, associations and municipalities, the locations for the ASSIST Action will be identified.

In accordance with the objective of defining the characteristics of the entire country, the analysed sample (collected by the national statistical agency – ISTAT) can be considered statistically representative of the whole Italian population and describes socio-demographic aspects, characteristics of the dwellings, appliance owned and spending capacity (distributed for energy, food, house, etc.). The methodology developed in order to obtain the required market segmentation is a cluster analysis of 15.000 families and their characteristics (more than 1.000 variables), extracted through a two-step methodology.

Before presenting the methodology,it is necessary to give a description of the available dataset, followed by a detailed description of the procedures to get the National segmentation and the vulnerable consumers analysis; respective results are presented in Chapter 3 and Chapter 4.

4.1.1 Dataset

As previously mentioned, the dataset used for the national market segmentation comes from the consumption expenditure survey on the Italian population prepared by ISTAT (reference year: 2015). In particular, micro data have been collected from approximately 15.000 families, in 502 different municipalities, to obtain the monthly equivalent expenditures. For each household, more than 1.000 variables are available; they concern (a) socio-demographic aspects, (b) characteristics of the dwellings, (c) owned appliances and (d) spending capacity of the inhabitants. Based on the dataset structure as well as on the previous literature, it is observed that the characterisation concerning the vulnerable consumers will be based on the expenditure capacity; this aspect may be a limitation, as the condition of vulnerability could be originated from other different reasons. Another limitation concerns the dataset itself: it refers to 2015 data only; however, since in the last years the national situation has been quite stable, these data can be still considered as representative of the current situation. This statement is verified by comparing the aggregated data of 2015 and 2016, as displayed in Figure 22, that shows the expenditure percentage of the Italian families for house, water, electricity, gas and other fuels, divided into 5 macro-geographical areas, in 2015 and 2016.

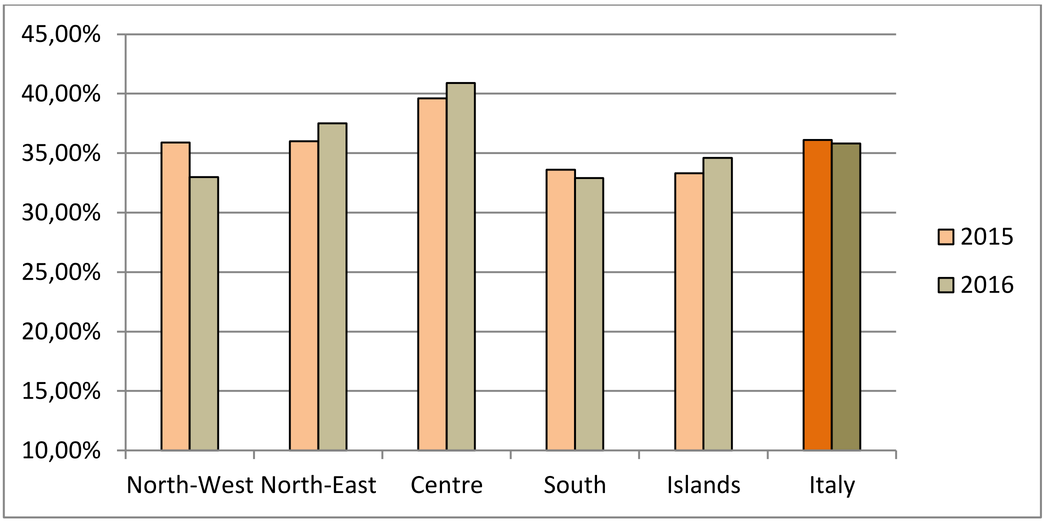


Figure 22 - Expenditure percentage for house, water, electricity, gas and other fuels.

As observed in Figure 22, there are small variations from 2015 to 2016 for the expenditure percentage; it is also worth noting that the above-mentioned statement is supported by the data concerning the poverty level of the Country. These data are displayed in Figure 23, which shows that the incidence of the absolute poverty index and the relative poverty index are very similar in the two years, with a small increase. These indicators are both defined and presented in the report about Italian poverty of 2017.

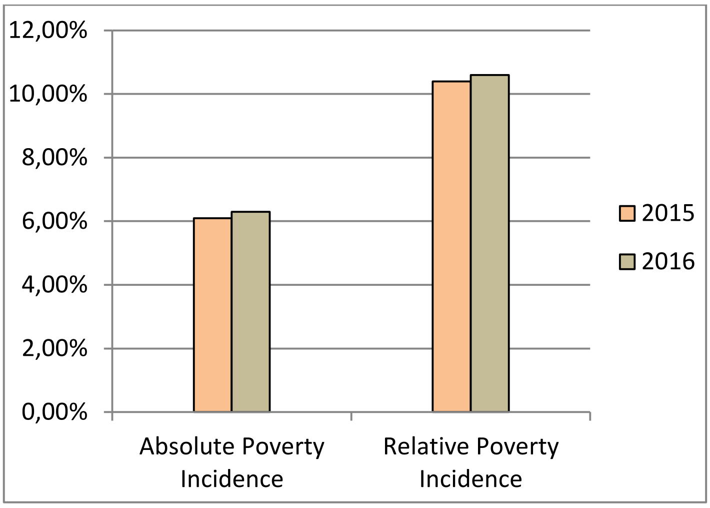


Figure 23 - Incidence of absolute and relative poverty indexes.

Before introducing the detailed analysis, some preliminary investigation on the micro-data can help to frame the situation. For example, some relevant information concerning energy poverty could be related to the absence of an heating system, as declared by approximately 5% of the population (which indicates a class of the population in non-negligible risk of health problems and bad living conditions). Another interesting aspect is that approximately 20% of the entire sample is not connected to the gas network, so, in Italy, it doesn't have the possibility to access the Natural Gas Bonus. The presence of the above-mentioned issues may create more difficulties in detecting vulnerable consumers who could need economical support, i.e., the electricity and gas bonus (refer to D2.1 and D2.2). It is known from several previous studies that energy poverty is not only related to the limited income capacity of an household, but it is also related to other factors (i.e., the characteristics of the building, in particular its energy performances, the characteristics of a family, the personal perceptions of single subjects, the energy prices and contracts, etc.). This multidimensionality of the problem is one of the main reasons why it is so difficult to define and measure energy poverty with a numeric index. However, these different aspects can be interesting and valuable. As an example, the histogram in Figure 24 shows, in percentage of families, the personal perception of the economic resources available, subdivided into four classes: optimal (less than 2% of the population), adequate (approximately half of the population), scarce (more than one third) and insufficient (almost 8%).

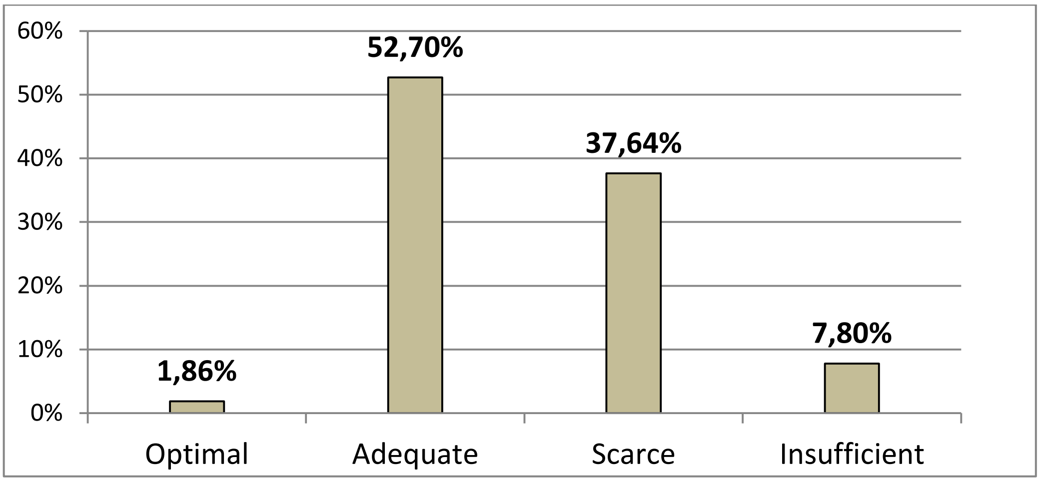


Figure 24 - Personal evaluation of the economic resources available.

Finally, it is useful to present how the energy expenditure is distributed in the population; in particular, Figure 25 and Figure 26 display the quartiles of heating and electricity expenditure, on a monthly basis: one fourth of the population spends on average (a) around 15 € a month for heating (with a national average of 78 €) and (b) around 18 € a month for electricity (with a national average of 47 €). It is far more probable to find vulnerable consumers in these sub-classes of population than in others.

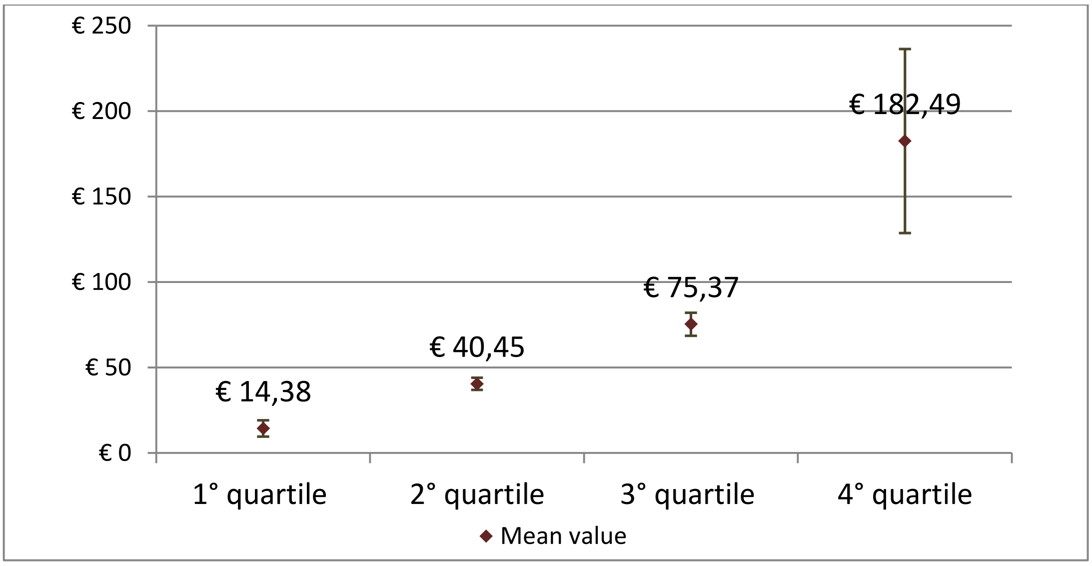


Figure 25 - Quartiles of heating expenditure on a monthly basis.

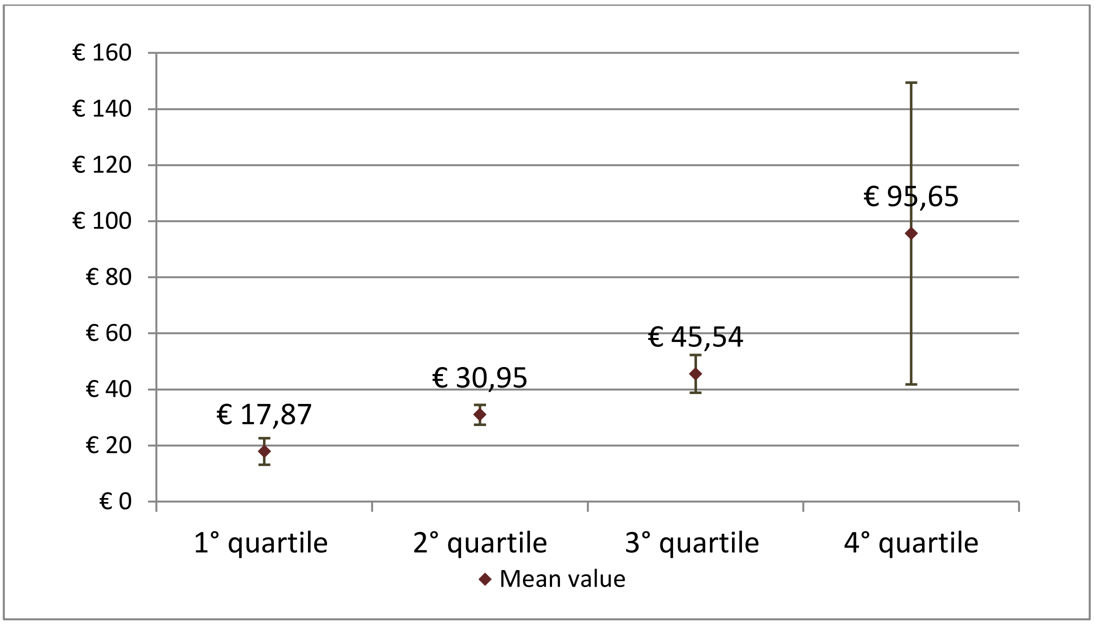


Figure 26 - Quartiles of electricity expenditure on a monthly basis.

4.1.2 National segmentation

A national segmentation is obtainable with a two-step analysis: (a) a regression analysis and (b) a cluster analysis (Figure 27).

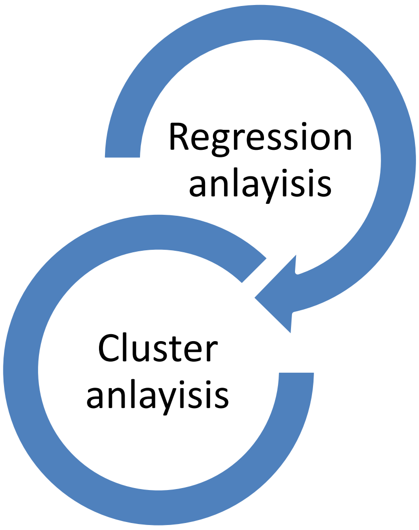


Figure 27 - Approach used.

Regression Analysis

The first analysis concerns the evaluation of the variables, which mostly contribute to determine the energy expenditure. The results of this analysis represent a guideline for the selection of the most appropriate data to describe energy consumption and identify vulnerable consumers. This analysis has been performed by using a statistical approach, applied in similar case studies, concerning (a) OLS (Ordinary Least Square) analysis (considering the energy expenditure as dependent variable), (b) VIF (Variance Inflaction Factor) analysis and (c) LASSO

(Least Absolute Shrinkage and Selection Operator) Analysis. The steps are shown in Figure 28.

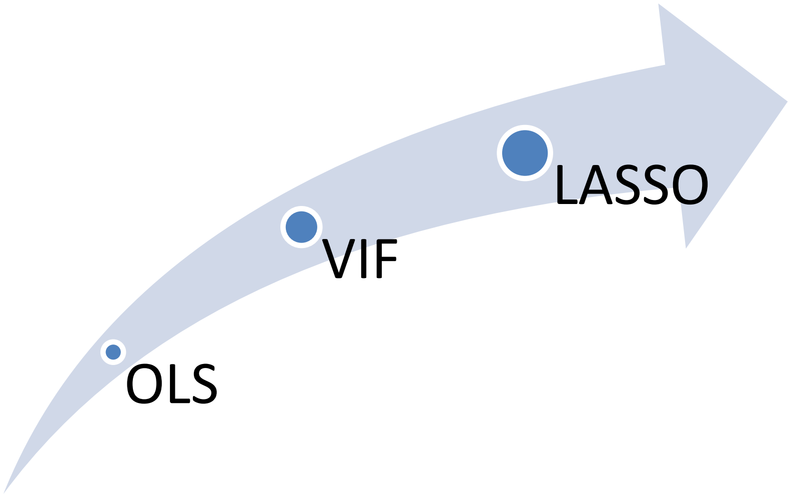


Figure 28 - Regression approach.

Cluster Analysis
The second phase starts from the results of the regression and applies the Classification And Regression Trees technique (CART), to classify into sub-groups the Italian population (Figure 29). The results of the *CART* algorithm have a tree-shape structure, where the ends ("leaves") are the terminal nodes. This operation highlights which are the specific characteristics of a household, that lead to a higher vulnerability, when defined by low energy consumption.

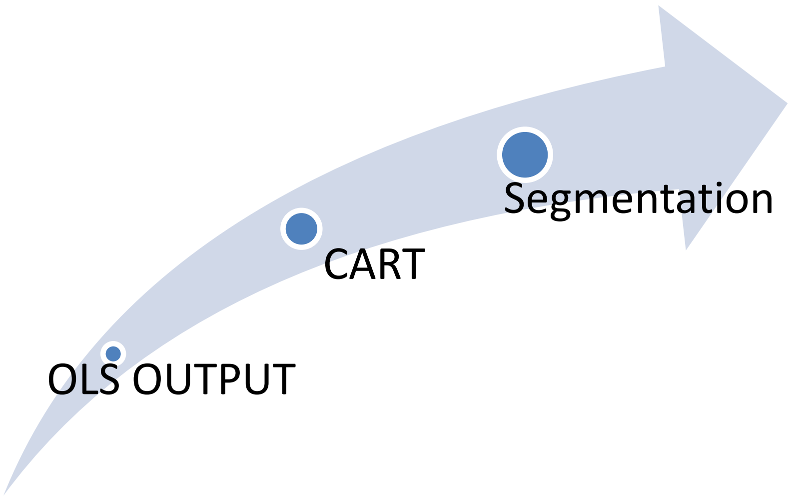


Figure 29 - Clusterization approach.

4.2 Regression Analysis

The aim of this analysis is to determine which are the variables affecting the energy consumption (through the energy expenditure) in households. In general, we try to answer to the following question: does changes in household socio-economic circumstances translate in changes in energy expenditures? In particular, three main predictors have been applied: (a) socio-demographic characteristics, (b) building characteristics and (c) owned appliances. In other terms, the study analyses the impact that these three classes of variables have on the energy expenditure of the household.

In literature, it is possible to find different studies concerning regression analysis on the energy expenditure/consumption, coming from different Countries and different dataset dimensions; however, at present there is no agreement on variable selection and an agreement is far from being reached. The strength of the procedure implemented in the present study is that it doesn't only concern a single OLS analysis, but it also checks for multicollinearity issues. Multicollinearity occurs when two or more predictors in an OLS are highly correlated. A simple example could be the presence of variables such as "*surface of building*" and "*number of* rooms", which could explain the same part of the energy expenditure. Thanks to this analysis it is possible to identify which variable is the more significant, while the other one can be neglected. The presence of multicollinearity means that regression coefficients cannot be reliably interpreted; hence for each regression analysis first Variance Inflaction Factors (VIF) are inspected to see whether there is multicollinearity and then, if it exists, LASSO regression is carried out to sets redundant predictors to zero (and removing multicollinearity).

The above-mentioned procedure is repeated four times: first the regression models of the three variables categories has been performed individually, and then, with the relative results, the final step combines all the variables together in one regression analysis.

More specifically, for each of the classes of variables, and finally for the whole set, the procedure starts with a linear ordinary least squares (OLS) regression. Then, given the suspected issue of multicollinearity, the VIF analysis is implemented and it returns, for each variable, a factor that indicates how much the variance of the estimated coefficient increases if the explanatory variables are correlated. This means that, if VIF is equal to one, the variable is uncorrelated with the others, while if it is higher than one multicollinearity starts to be present. There is no formal cut-off point for critical values of VIF, however in this study the threshold imposed is VIF > 3, with some exception in the case of variables whose elimination implies an excessive reduction of the adjusted R^2 of the regression. The chosen threshold is anyway lower than the one applied in most of the literature, where the critical value is considered in the range of 5 - 10. When VIF is greater than 3 after the first OLS, the LASSO regression is employed. LASSO regression is built on a linear model, but uses an alternative procedure to calculate regression coefficients. It is a penalised analysis promoting a sparser model: by using a fitting procedure, it sets some coefficients to zero, suggesting to the user which are the variables that can be sacrificed being the non-relevant ones. After eliminating the variables affected by multicollinearity, another OLS regression is carried out. At this point, to further refine the result, a trial and error procedure is applied to exclude additional variables; the trial and error procedure is based on a progressive exclusion of variables by checking the adjusted R^2 values.

The following sections describe in detail the variables included in the regression, starting from the dependent variable, namely the one to be predicted, and then all the predictors used to describe this dependent variable.

- **Dependant variable**

The dependent variable is the household energy expenditure on a monthly basis. This quantity is computed by the sum of the electricity expenditure and the heating expenditure (considering both households who are connected to the gas network and households that use other fuels). This variable can be considered as a proxy of energy consumption and is the best available indicator to define the condition of vulnerability of a family, regarding energy poverty. The energy expenditure has been log-transformed (natural log) to achieve greater symmetry of the distribution and of the residuals in the regression analysis. It is worth noting that families having energy expenditure equal to 0 have been excluded from this analysis.

- **Predictors**

As previously stated when describing the dataset, there are three main categories of variables that have been exploited for the regression analysis: (a) socio-demographics, (b) dwelling characteristics and (c) owned appliances . Since the whole dataset consists of more than 1.000 variables, a first skimming to identify the ones to involve in the procedure has been applied, on the basis of the ones considered in literature. Thus, the following tables show the three categories, each one with their selected variables. It is possible to observe that there are two types of variables: (a) categorical and (b) continuous. The majority of the predictors are categorical, with a number of categories that is variable, starting from two options (binaries variables) to more than 10; for these ones, in the tables are reported the number of families that fall into every categories. For the continuous variables, the tables contain the mean value in the whole sample and the relative variance. The last aspect to highlight is that, for every categorical variable, one of the categories is printed in bold: these are the reference categories used for the subsequent regressions. The choice of the reference category doesn’t affect the final results; the only difference lies in the parameters of the regressions, which are referred to the reference category, but every parameter significance, and obviously the global one, is unchanged. Anyway, in most of the cases the reference has been set on the more frequent category.

Table 11 shows socio-demographic aspects (14 categorical), Table 12 concerns the building variables (10 categorical and 1 continuous) and Table 13 shows the appliances (7 categorical and 1 continuous). In bold the established baseline.

Table 11 - Socio-demographic variables.

Variable	*Summary statistics*
2. Birth place of the components	**(a) Only born in Italy [13274]**, (b) At least one born abroad [965], (c) Only born abroad [547]
3. Citizenship of the components	**(a) Only Italian citizens [13985]**, (b) At least one foreign citizens [256], (c) Only foreign citizens [545]
4. Marital status of the principal component	(a) Unmarried [2450], **(b) Married or cohabitant with the spouse [8220]**, (c) Married but not cohabitant with the spouse [336], (d) Legally separated [603], (e) Divorced [689], (f) Widow or widower [2488]
5. Qualification of the components	(a) No member has a qualification [371], (b) At least one member with elementary school [1936], (c) At least one member with junior high school [3053], **(d) At least one member with high school [6399]**, (e) At least one member with a degree [3027]
6. Working time of the components	**(a) One or more full time [6640]**, (b) One or more part time [749], (c) No one working and no pension [2449], (d) No one working and one or more pension [4948]
8. Source of income of the components	(a) There is no income [75], (b) At least one maintained [370], (c) At least one pension [4858], **(d) At least one income [9483]**
Enrolment in study courses	**(a) No members enrolled in a course [10735]**, (b) At least one in no title school [418], (c) At least one in elementary school [740], (d) At least one in junior high school [583], (e) At least one in high school [1236], (f) At least one in a degree or post-degree course [1074]
10. Presence of elderly or disabled people who need assistance	(a) Yes [98], **(b) No [14688]**
11. Current economic resources	(a) Optimal [277], **(b) Adequate [7822]**, (c) Scarce [5566], (d) Insufficient [1121]
12. Changing in economic resources compared to the previous year	(a) Much improved [29], (b) A little bit improved [496], **(c) More or less the same [8387]**, (d) A little worsened [4563], (e) Much worsened [1311]
13. Absolute poverty	(a) Yes [779], (b) **No [14007]**
14. Households structure	(a) Single person 18-34 years [346], (b) Single person 35-64 years [1744], (c) Single person 65 years and more [2184], (d) Couple without children with r.p 18-34 years [174], (e) Couple without children with r.p. 35-64 years [1341], (f) Couple without children with r.p. 65 years and more [2155], **(g) Couple with 1 child [2268]**, (h) Couple with 2 children [2176], (i) Couple with 3 children and more [494], (l) Mono parent family [1021], (m) Others [883]

* r.p. reference person, for whom relationship are defined

Table 12 - Building variables.

Variable	*Summary statistics*
1. Period of construction	**(a) After 2009 [187]**, (b) Between 2000 and 2009 [1206], (c) 1990s [1392], (d) 1980s [2280], (e) 1970s [3382], (f) 1960s [2693], (g) 1950s [1450], (h) Between 1900 and 1949 [1385], (i) Before 1900

	[811], (l) Don't know/don't remember [0]
2. Surface	**Continuous variable** [Mean = 98 / Variance = 1343]
3. Number of rooms	(a) 1 [420], (b) 2 [2795], **(c) 3 [5839]**, (d) 4 [4007], (e) 5 [1200], (f) 6 [332], (g) 7 and more [193]
5. Geographic area	(a) North-west [3211], (b) North-east [3352], **(c) Centre [2750]**, (d) South [4315], (e) Islands [1158]
6. Type of housing	(a) Single family villa [2708], (b) Multifamily villa [4499], **(c) Apartments in building with less than 10 apartments [3671]**, (d) Apartments in building with 10 or more apartments [3898], (e) Other [10]
7. Type of municipality	**(a) Centre of metropolitan area [1865]**, (b) Periphery of metropolitan area and municipalities with 50.001 inhabitants and more [3982], (c) Other municipalities until 50.000 inhabitants. [8939]
8. Gas from network	(a) **Yes [12178]**, (b) No [2608]
9. Type of heating system and fuel	(a) **Central heating - Gas from network [1351],** (b) Central heating – Gas oil/kerosene/other liquid fuels [261], (c) Central heating – Other fuels [53], (d) Autonomous heating - Gas from network [9313], (e) Autonomous heating – Gas oil/kerosene/other liquid fuels [263], (f) Autonomous - Gas cylinder [477], (g) Autonomous - Wood/pellet/ Other solid fuels [958], (h) Autonomous - Other [40], (i) District heating - Gas from network [70], (l) District heating - Other [39], (m) Individual devices - Gas from network [100], (n) Individual devices – Gas cylinder [186], (o) Individual devices - Wood/pellet/ Other solid fuels [831], (p) Individual devices - Other [101], (q) Other [3], (r) No heating system [740]
10. Domestic hot water system	(a) Electric boiler [1968], (b) Gas boiler [3550], (c) **Heating system [9007]**, (d) Other [24], (e) Solar panel [197], (f) No hot water system [40]
11. Gas/Electricity cooking	(a) **Yes [14717]**, (b) No [69]

Table 13 - Appliances variables.

Variable	*Summary statistics*
1. Self-consumption	(a) Yes [2016], **(b) No [12770]**
2. Expenses for appliances and electric tools	***Continuous variable*** [Mean = 22 / Variance = 1664]
3. Owner of television	**(a) Yes [14255]**, (b) No [531]
4. Owner of computer	**(a) Yes [9319]**, (b) No [5467]
5. Owner of fridge	**(a) Yes [14757]**, (b) No [29]
6. Owner of washing machine	**(a) Yes [14545]**, (b) No [241]
7. Owner of dishwasher	(a) Yes [7150], **(b) No [7636]**
8. Owner of air conditioner	(a) Yes [4997], **(b) No [9789]**

4.3 Cluster Analysis

Once the whole data set is reduced, by applying the above-mentioned procedure, in terms of both number of useful records and number of significant fields, the market segmentation procedure can be applied. In order to obtain an efficient decision support system for the design of the ASSIST actions, a "*decision tree method*" is applied. The "*decision tree method*" approach helps understanding the energy expenditure patterns. Within this modelling approach, an empirical tree represents a segmentation of the whole dataset; such segmentation is performed by applying a series of different rules (aimed to identify different sub-categories of the sample, through a repetitive process of splitting). One of the most common "*decision tree method*" is the Classification And Regression Trees (*CART*[5]), which is implemented in this work. it is worth noting that it is a non-parametric procedure, so it is not necessary to test for the normality of the data (as it happened in the previous analysis).
The "*decision tree method*" is one of the most commonly used data mining methods, due to its many advantages (viz. ease of use and the possibility to generate accurate models with understandable and interpretable structures). On the practical point of view, a "*decision tree method*" uses a flowchart-like tree structure to segregate a dataset into different classes, with different characteristics. Being a logical model, the decision tree shows how the value of the target variable (i.e., the energy expenditure), can be predicted by using values of a set of predictors (i.e., the ones obtained in the above-mentioned OLS-VIF-LASSO analysis). Moreover, the "*decision tree method*" allows to process both numerical and dummy variables (i.e. the variables in the ISTAT dataset) and, thus, to perform a classification without large computation efforts; finally, the provided result gives clear information on the relevance and role of the significant predictor variables in the classification.
The *CART* technique displays the results under the form of an inverted tree, with three kinds of nodes: a root node, internal nodes and leaf nodes (see Figure 30);

[5] The CART method includes both Classification and Regression procedures; the choice of using either the former or the latter depends on the characteristics of the target variable: in the case of a categorical variable the procedure takes the name of classification; conversely, if numerical, like the energy expenditure (viz. the present case), the results are regression trees.

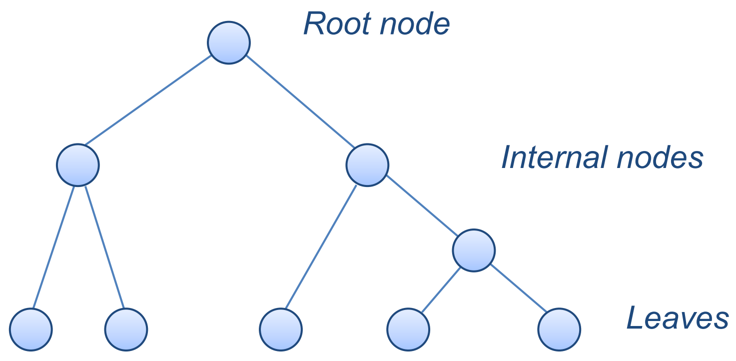


Figure 30 - Decision tree structure example.

The root node denotes a binary split based on a certain rule; conversely, the leaves represent the outcomes of the classification. The parameters used in the splitting of the dataset at each step are automatically selected. In this respect, it is worth noting that the more a split is close to the root, the larger is the influence of the variable on the dataset splitting procedure. This method not only provides the combination of relevant predictors to determine a certain energy expenditure, but it also details the threshold values of these factors. Moreover, the algorithm may use the same variables more than once, in different parts of the tree, with different values to split the data; this capability can uncover complex interdependencies between sets of variables.

On the practical point of view, when building a regression tree it is important to define some rules to control the significance of the result. In the present case, the approach is implemented in two phases: (a) first, the tree is constructed with a non-restrictive stopping rule (leading to an "overgrown" decision tree); (b) second, the option of "pruning" the tree is applied to avoid over-fitting and to find the best compromise between error and number of leafs. Please note that the retrospective approach of the second step allows minimizing the probability that stopping too soon the subdivision doesn't overlook important structures in the data set.

In particular, the two phases are implemented in four steps, as follows:

a) Generation of a tree with a non-restrictive stopping rule

As previously mentioned, it is widely recommended to start the procedure with an "*overgrown*" tree. To this end, the *complexity parameter* (*cp*) of the algorithm is modified. The *complexity parameter* controls the growth of the tree as follows: the overall R-squared must increase by *cp* at each subdivision. Hence, by starting with a low *cp* value, the split is permitted also when it doesn't imply a very significant increase of the result, leading to a very elaborated tree with a lot of variables and a lot of sub-groups of households at the end of the tree-structure.

b) Generation of a tree with minimum cross-validation error

Once the complete development of the tree is built, it is possible to search for the best *cp* value. This value is the one that minimizes the cross-validation error (*xerror*). The cross-validation error is related to the *PRESS* statistic (*Predicted Residual Error Sum of Squares*) and it is computed as the sum of the squares of all the resulting prediction errors; hence it is used exactly to understand at which point the tree starts to be over-grown (viz. when the regression model is over-fitted). Thus, identifying the minimum *xerror* leads to the optimal number of splits.

c) Generation of a trade-off tree

At this point, the previously-obtained tree is further processed to find the one that represents the best compromise between error and complexity. To this end, the cutting point is changed and the modifications to the tree structure are observed (i.e., R-squared, cross-validation error, number of variables involved, number of leafs, graph of the residual of the model).

d) Analysis of the sub-groups of households identified

The final tree obtained at the previous step is, thus, used as market segmentation. At this point, the characteristics of the sub-groups are studied and post-processed.

4.4 Vulnerable Consumers Analysis

It is nowadays well known that one of the first barriers to fight the energy poverty phenomenon is not only how to solve or alleviate the problem, but first of all, just reaching or classify vulnerable consumers results very difficult (also because of the lack of a recognised definition).

In this framework, to obtain the best characterization possible, beyond the national segmentation, vulnerable consumers are analysed also incorporating other sources of information, specifically collected for ASSIST:

- Specific questionnaire of Task 5.1
- Conclusions from the market survey of Task 2.5

From the national level segmentation, some peculiarities and characteristics of the families, that are more at risk of poverty, are extracted; at this point of the analysis, these are integrated with the information coming from the questionnaire of Task 5.1. These data are collected through the network of the HEAs – Home Energy Advisors, therefore they should represent in detail which are the characteristics of the vulnerable consumers involved in the Action, both in terms of socio-economic characteristics and behavioural attitudes towards energy management. Finally, also the information collected in the market survey of task 2.5 are included to enrich the analysis.

At the end, an important advantage of this procedure is that, in a second moment, it could be possible to make comparisons with the established baseline, and also to create relations with the analysis conducted at different scales.

5. Italy – First proposal for a national market segmentation

In this Chapter, the results obtained for the national market segmentation are presented and discussed. In sub-section 1, the outcomes of the regression analysis on the energy expenditure are presented; in sub-section 2, the outcomes of the clusterization process in sub-groups of consumers are presented. Finally, the results of the questionnaire for Vulnerable Consumers are analysed.

5.1 Regression analysis

Following the above-mentioned cluster procedure, the results are presented first for the regressions based on the different predictor groups and, afterwards, for the global regression model. It is observed that buildings characteristics appear to have the strongest impact in determining changes in energy expenditures, followed by socio-demographic characteristics of the households, which still have a significance influence and are able to describe energy expenditures with a lower number of variables. Finally, appliance ownerships have a lower influence in determining the energy expenditure (as expected, given the appliance variable available in the ISTAT dataset).

5.1.1 Regression based on the different predictor groups

In the following sub-sections, the results of the individual regression models are presented and discussed: (a) Socio-demographic characteristics, (b) Building characteristics, (c) Appliances owned. R^2, adjusted R^2, the number of predictors excluded are presented and commented. For all models, the residuals were inspected, by the Q-Q plot inspection.

a) Socio-demographic characteristics

The socio-demographic model explains R^2 = 11.39 % (adjusted R^2 = 11.29 %) for the variability in energy consumption expenditure. After confirming the presence of multicollinearity with the VIF analysis in 7 variables, implementing LASSO regression and checking for the significance of the single variables, the final regression with socio-demographic aspects leads to 4 of the 12 initial variables available; these final variables are reported in the following Table 14:

Table 14 - Regression variables of the Socio-demographic group of predictors.

	Variable	N° categories
1	Households structure	11
2	Absolute poverty	2
3	Source of income of the components	4
4	Current economic resources	4

b) Building characteristics

The building model explains R^2 = 20.11 % (adjusted R^2 = 19.93) of the variability for the energy expenditure. In this second case, with only two variables presenting multicollinearity, VIF and LASSO techniques, plus significance control, lead to 7 of the 10 initial variables, listed in Table 15:

Table 15 - Regression variables of the Building group of predictors.

	Variable	N° categories
1	Surface	cont.
2	Number of rooms	7
3	Geographic area	5
4	Type of housing	5
5	Type of municipalities	3
6	Gas from network	2
7	Type of heating system and fuel	16

c) Appliances

The appliances model explains R^2 = 8.55 % (adjusted R2 = 8.52 %). It is worth nothing that this model is the one with the lower R^2, mainly because the selected predictors. The LASSO procedure is not applied in this case, because the VIFs are all lower than the threshold of 3; some variables are neglected because of the low significance, passing from 8 to the following 6 variables, reported in Table 16):

Table 16 - Regression variables of the Appliances group of predictors.

	Variable	N° categories
1	Self-consumption	2
2	Expenses for appliances and electric tools	cont.
3	Owner of television	2
4	Owner of computer	2
5	Owner of washing machine	2
6	Owner of dishwasher	2

5.1.2 Regression based on the aggregated groups

The previous models are combined together to test for increments in explanatory power when joining the three categories. Obviously, the variables now involved are only the ones resulted as significant in the single regressions. After merging together all the predictors obtained in the previous regressions (Tables Table 14-Table 15-Table 16), multicollinearity issues has been detected and it has been necessary to exploit the LASSO regression and some other variables are removed, due to their low significance. In the final model of the energy expenditure, the variability explained is R^2 = 25.59% (adjusted R^2 = 25.38 %), with 12 variables involved, of which 2 are socio-demographic, 6 are related to the building and 4 concern appliances (Table 17):

Table 17 - Regression variables of the global model.

	Variable	*N° categories*	*Class*
1	Households structure	11	*Socio-demographic*
2	Absolute poverty	2	*Socio-demographic*
3	Surface	cont.	*Building*
4	Geographic area	5	*Building*
5	Type of housing	5	*Building*
6	Type of municipalities	3	*Building*
7	Gas from network	2	*Building*
8	Type of heating system and fuel	16	*Building*
9	Expenses for appliances and electric tools	cont.	*Appliances*
10	Owner of television	2	*Appliances*
11	Owner of washing machine	2	*Appliances*
12	Owner of dishwasher	2	*Appliances*

In addition, for the final model (Table 17), we propose a more detailed description of the statistics obtained; in particular, Table 18 displays the aggregate statistics of the model. Finally, Table 19 presents all the single variables and their categories are described.

Table 18 - Summary statistics of the energy expenditure regression model.

Residuals				
Min	1st Quartile	Median	3rd Quartile	Max
-3.4223	-0.3603	-0.0041	0.367	2.9004
Residual standard error:		0.5815 on 14743 DF		
Regression				
Multiple R^2:	25.59 %		Adjusted R^2:	25.38 %
F-statistic:	F (42,14743) = 120.7		*p*-value:	< 2.2e-16

Table 19 - Predictors of the final energy expenditure regression model.

Final Regression Model on Energy Expenditure	Estimate	Std. Error	t value	Pr(>|t|)	Significance
(Intercept)	4.32101	0.030681	140.835	<	2.00E-16
Households structure **Single person 18-34 years**	-0.41509	0.034277	-12.11	<	2.00E-16
Households structure **Single person 35-64 years**	-0.35785	0.019053	-18.782	<	2.00E-16
Households structure **Single person 65 years and more**	-0.28319	0.018002	-15.731	<	2.00E-16
Households structure **Couple without children with r.p 18-34 years**	-0.3134	0.045914	-6.826	9.08E-12	***
Households structure **Couple without children with r.p. 35-64 years**	-0.12129	0.020067	-6.045	1.53E-09	***
Households structure **Couple without children with r.p. 65 years and more**	-0.05725	0.017646	-3.244	0.00118	**
Households structure **Couple with 2 children**	0.018862	0.017536	1.076	0.282098	
Households structure **Couple with 3 children and more**	0.083577	0.028997	2.882	0.003954	**
Households structure **Mono parent family**	-0.11389	0.022026	-5.171	2.36E-07	***
Households structure **Others**	-0.03003	0.023266	-1.291	0.196834	
Absolute poverty **Yes**	-0.25842	0.022188	-11.647	<	2.00E-16
Surface	0.00295	0.000153	19.348	<	2.00E-16
Geographic area **North-west**	0.138155	0.015543	8.889	<	2.00E-16
Geographic area **North-east**	-0.02274	0.015281	-1.488	0.136719	
Geographic area **South**	-0.17804	0.014756	-12.065	<	2.00E-16
Geographic area **Islands**	-0.14276	0.023011	-6.204	5.66E-10	***
Type of housing **Single family villa**	0.121416	0.016413	7.398	1.46E-13	***
Type of housing **Multifamily villa**	0.112972	0.013518	8.357	<	2.00E-16
Type of housing **Apartments in building with 10 or more apartments**	-0.10036	0.014759	-6.8	1.08E-11	***
Type of housing **Other**	0.065093	0.184342	0.353	0.724011	
Type of municipality **Periphery of metropolitan area and municipalities with 50.001 inhabitants and more**	0.068235	0.017455	3.909	9.30E-05	***
Type of municipality **Other municipalities until 50.000 inhabitants**	0.06096	0.017753	3.434	0.000597	***
Gas from network **No**	-0.21237	0.020551	-10.334	<	2.00E-16
Type of heating system and fuel **Central heating - Gas oil/kerosene/other liquid fuels**	0.343667	0.0423	8.124	4.85E-16	***
Type of heating system and fuel **Central heating - Other fuels**	0.104144	0.083076	1.254	0.210007	
Type of heating system and fuel **Autonomous heating - Gas from network**	0.091348	0.018717	4.88	1.07E-06	***
Type of heating system and fuel **Autonomous heating - Gas oil/kerosene/other liquid fuels**	0.460427	0.043656	10.547	<	2.00E-16
Type of heating system and fuel **Autonomous - Gas cylinder**	0.338394	0.036961	9.155	<	2.00E-16
Type of heating system and fuel **Autonomous - Wood/pellet/other solid fuels**	0.186687	0.028863	6.468	1.02E-10	***
Type of heating system and fuel **Autonomous - Other**	0.074235	0.094475	0.786	0.432023	
Type of heating system and fuel **District heating - Gas from network**	0.140181	0.071433	1.962	0.049735	*

Type of heating system and fuel **District heating - Other**	0.048173	0.095591	0.504	0.614305	
Type of heating system and fuel **Individual devices - Gas from network**	-0.14623	0.061349	-2.384	0.017156	*
Type of heating system and fuel **Individual devices - Gas cylinder**	0.034852	0.049471	0.704	0.481141	
Type of heating system and fuel **Individual devices - Wood/pellet/ Other solid fuels**	0.16231	0.031048	5.228	1.74E-07	***
Type of heating system and fuel **Individual devices - Other**	0.019971	0.06305	0.317	0.751436	
Type of heating system and fuel **Other**	-0.05355	0.336481	-0.159	0.873558	
Type of heating system and fuel **No heating system**	-0.17487	0.032823	-5.328	1.01E-07	***
Expenses for appliances and electric tools	0.00086	0.000122	7.057	1.77E-12	***
Owner of television **No**	-0.10211	0.026333	-3.878	0.000106	***
Owner of washing machine **No**	-0.13966	0.038787	-3.601	0.000318	***
Owner of dishwasher **Yes**	0.051338	0.010847	4.733	2.23E-06	***

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*'*

5.1.3 Conclusions based on regression analysis

Using a large, nationally representative sample of around 15.000 households, this study presents a statistical analysis and the outcomes can be considered representative of the whole Italian population.
Beside the methodology implication of this work (i.e., OLS-VIF-LASSO approach), the main conclusions are as follows. First, it has been found that, inside a specific category of variables (i.e., Socio-demographic, Building, Appliance), there are some aspects that are more significant than others in determining variations in household's energy expenditure. Secondly, the study highlights that building characteristics are the most influential variables, followed by socio-demographic characteristics and ownedappliances. Finally, the combination of the three categories leads to a description of the energy expenditure variation equal to 25%, meaning that acting on these variables could bring a significant variation in the energy expenditure; anyway it is important to draw attention on the fact that there might be many other variables that could enhance the description and that these analysis is strictly dependent on the available dataset.
In conclusion, the main outcomes of the carried out regression analysis is the definition of the more significant variable from which the energy expenditure depend, but also the opposite aspect of which are the non-significant characteristics in this kind of analysis, that instead could leave space to others which could improve the description.

5.2 Cluster Analysis

Once the selected variables are defined, it is possible to proceed with the market segmentation of the Italian families with respect to the total energy expenditure. The cluster analysis and the main results are described in the following paragraphs.

5.2.1 Generation of a tree with a non-restrictive stopping rule

The regression tree with a non-restrictive stopping rule is obtained with *cp* = 0.0001. This phase starts considering the 12 variables resulting from the linear regression model (Table 17). The resulting tree is composed by > 500 leafs, which is not an overgrown tree. It is worth noting that, in this phase, the predictor "*Owner of washing machine*", belonging to the Appliance macro-category, is not used.

5.2.2 Generation of a tree with minimum cross-validation error

The *CART* algorithm displays, for different *cp*s > 0.0001, the changes in the tree characteristics; in particular, the one to be selected at this step is the one that corresponds to the minimum cross-validation error, (viz. *xerror* = 0.8714 corresponding, *cp* = 0.001197, number of leafs = 38). The variables now involved in the regression tree procedure are 9, out of the 12 used as the predictors: (a) "*Owner of washing machine*", (b) "*Owner of television"* (Appliance) and (c) "*Absolute poverty*" (Socio-demographic) are not used.

5.2.3 Generation of a trade-off tree:

To perform a trade-off analysis, it is possible to make some considerations on the graphs in Figure 31.

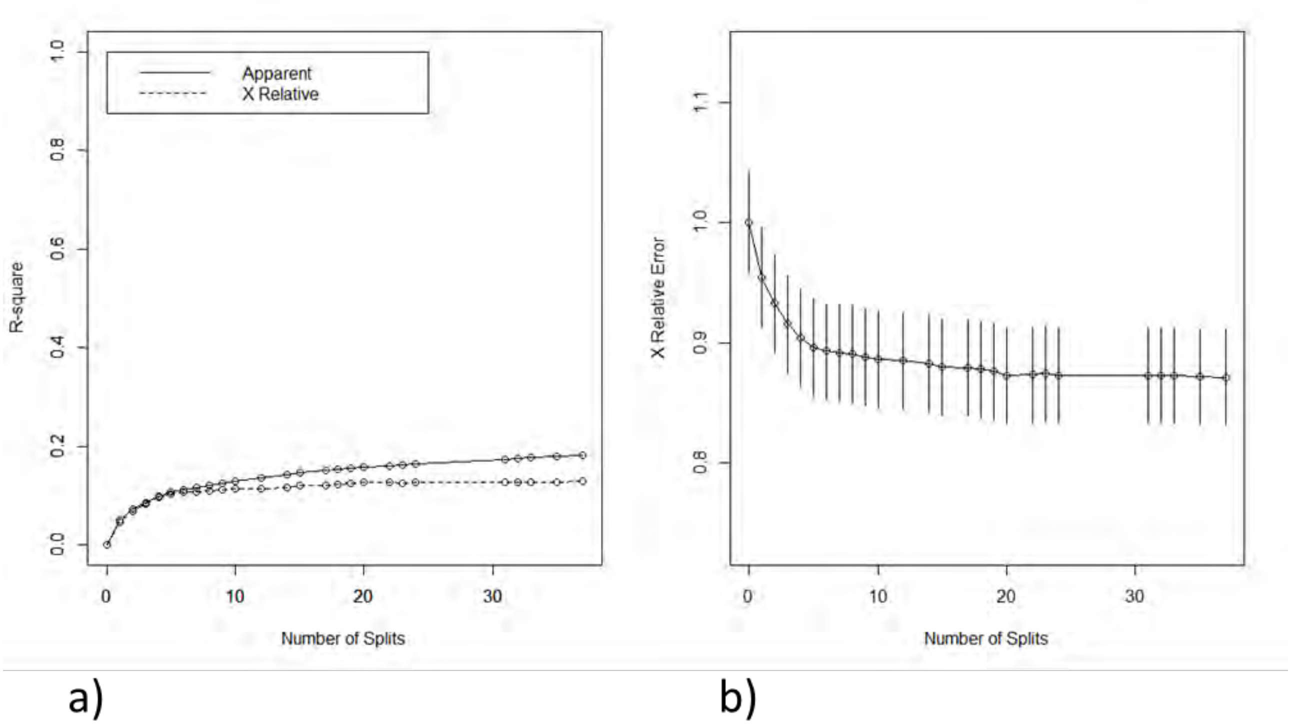


Figure 31 - R^2 (a) and Relative error (b) of the regression with different n. of split.

Figure 31a displays the change in the value of R^2 when increasing the number of split (and, thus, the number of leaves); the dashed line is the same calculation on the validation sample. Conversely, Figure 31b presents an opposite behaviour, since it represents the variation in the *Relative Error* (1-R^2) in cross-validation, with respect to different number of splits. Based on Figure 31, in the present case, 10 split has been selected as, for a larger number of splits, the increment in R^2 and, consequently, the reduction of the *Relative Error* are negligible (thus, meaning adding new splits and new variables doesn't bring to significant improvement in the model).

As a consequence of the selected number of split, the resulting number of sub-groups of consumers is equal to 12. In Table 20 there is a summary of the statistical characteristics of this final model, with a comparison with the previous optimal one.

Table 20 - Comparison between final regression tree and optimal regression tree.

Parameter	Final tree	Optimal tree
Cp	0.0030	0.0012
n° of leaves	12	38
relative error	0.8673	0.8175

In addition, Figure 32 and
Table 21 display all the sub-groups. In the next section the characteristics of the 12 subgroups are discussed.

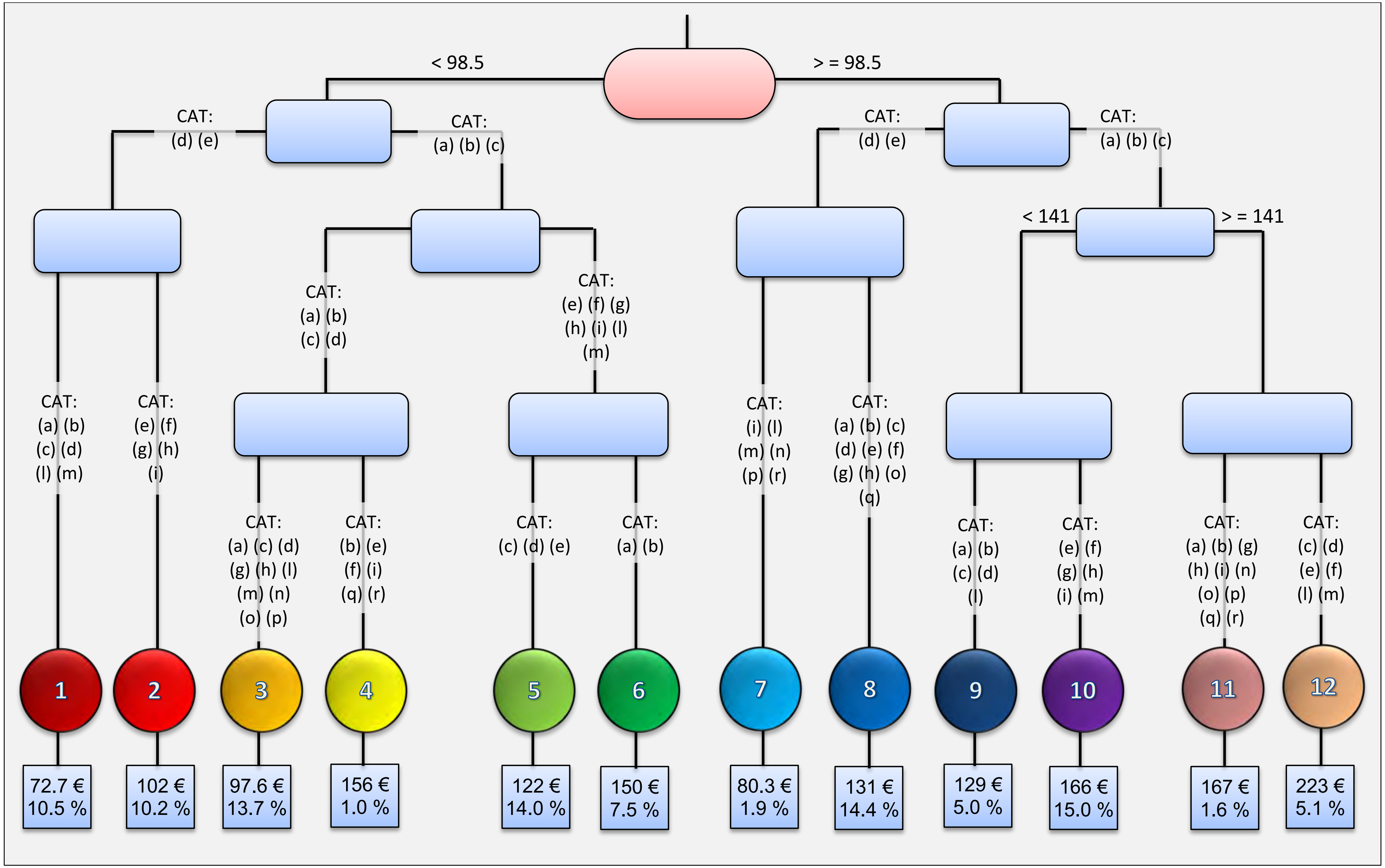


Figure 32 - Regression tree process with CART algorithm - Code references for labels in Table 22.

Table 21. Details on the groups of cluster analysis

CLUSTER	Surface	Geographic Area	Households structure	Type of heating system and fuel	Type of housing
1	< 98.5 m^2	▪ South ▪ Islands	▪ Single person 18-34 years ▪ Single person 35-64 years ▪ Single person 65 years and more ▪ Couple without children with r.p. 18-34 years ▪ Mono parent family ▪ Others	/	/
2	< 98.5 m^2	▪ South ▪ Islands	▪ Couple with 1 child ▪ Couple without children with r.p. 35-64 years ▪ Couple without children with r.p. 65 years and more ▪ Couple with 2 children ▪ Couple with 3 children and more	/	/
3	< 98.5 m^2	▪ Centre ▪ North-West ▪ North-East	▪ Single person 18-34 years ▪ Single person 35-64 years ▪ Single person 65 years and more ▪ Couple without children with r.p. 18-34 years	▪ Central heating (Gas from network, Other) ▪ Autonomous heating (Gas from network, Wood, Pellet, Other solid fuels, Other) ▪ District heating (Other) ▪ Individual devices (Gas from network, Gas cylinder, Wood, Pellet, Other solid fuels, Other)	/
4	< 98.5 m^2	▪ Centre ▪ North-West ▪ North-East	▪ Single person 18-34 years ▪ Single person 35-64 years ▪ Single person 65 years and more ▪ Couple without children with r.p. 18-34 years	▪ Central heating (Gas oil, Kerosene, Other liquid fuels) ▪ Autonomous heating (Gas oil, Kerosene, Other liquid fuels, Gas cylinder) ▪ District heating (Gas from network) ▪ No heating system	/
5	< 98.5 m^2	▪ Centre ▪ North-West ▪ North-East	▪ Couple with 1 child ▪ Couple without children with r.p. 35-64 years ▪ Couple without children with r.p. 65 years and more ▪ Couple with 2 children ▪ Couple with 3 children and more ▪ Mono parent family ▪ Others	/	▪ Apartments in building with less than 10 apartments ▪ Apartments in building with 10 or more apartments ▪ Other

6	< 98.5 m^2	▪ Centre ▪ North-West ▪ North-East	▪ Couple with 1 child ▪ Couple without children with r.p. 35-64 years ▪ Couple without children with r.p. 65 years and more ▪ Couple with 2 children ▪ Couple with 3 children and more ▪ Mono parent family ▪ Others	/	▪ Single family villa ▪ Multifamily villa
7	≥ 98.5 m^2	▪ South ▪ Islands	/	▪ District heating (Gas from network, Other) ▪ Individual devices (Gas from network, Gas cylinder, Other) ▪ No heating system	/
8	≥ 98.5 m^2	▪ South ▪ Islands	/	▪ Central heating (Gas from network, Gas oil, Kerosene, Other liquid fuels) ▪ Autonomous heating (Gas from network, Gas oil, Kerosene, Other liquid fuels, Gas cylinder, Wood, Pellet, Other solid fuels, Other) ▪ Individual devices (Wood, Pellet, Other solid fuels)	/
9	≥ 98.5 m^2 & < 141 m^2	▪ Centre ▪ North-West ▪ North-East	▪ Single person 18-34 years ▪ Single person 35-64 years ▪ Single person 65 years and more ▪ Couple without children with r.p. 18-34 years ▪ Mono parent family	/	/
10	≥ 98.5 m^2 & < 141 m^2	▪ Centre ▪ North-West ▪ North-East	▪ Couple with 1 child ▪ Couple without children with r.p. 35-64 years ▪ Couple without children with r.p. 65 years and more ▪ Couple with 2 children ▪ Couple with 3 children and more ▪ Others	/	/
11	≥ 141 m^2	▪ Centre ▪ North-West ▪ North-East	/	▪ Central heating (Gas from network, Gas oil, Kerosene, Other liquid fuels) ▪ Autonomous heating (Wood, Pellet, Other solid fuels, Other) ▪ District heating (Gas from network) ▪ Individual devices (Gas cylinder, Wood, Pellet, Other solid fuels, Other) ▪ Other ▪ No heating system	/
12	≥ 141 m^2	▪ Centre ▪ North-West ▪ North-East	/	▪ Central heating (Other) ▪ Autonomous heating (Gas from network, Gas oil, Kerosene, Other liquid fuels, Gas cylinder) ▪ District heating (Other)	/

	▪ Individual devices (Gas from network)

5.2.4 Analysis of the identified sub-groups of households

The 12 groups of consumers are defined by 5 variables, as shown in Figure 32 and

Table 21 and Table 22 (please note that the code for the labels in Figure 32 is listed in Table 21). It is worth noting that all the variables belong to the Building group, except for one from Socio-Demographic group, while there is no variable from the Appliances group. The last column of Table 22 shows a score for each variable that represents the level of importance in the classification, with the *Surface* variable on top. Since one variable may appear more than once in the tree, the importance level is given by the sum of the goodness of split measures for each split.

Table 22 - Final variables of the regression tree.

Variable	Category	Type	Importance
Surface	Building	Continuous	38
Geographic area	Building	5 categories	20
Type of heating system and fuel	Building	16 categories	13
Household structure	Socio-demographic	5 categories	12
Type of housing	Building	11 categories	8

In detail, the procedure starts with the root node in which the whole sample of 15.000 families has been splitted based on the *Surface* of the dwelling; since this one is a continuous variable, the algorithm returns the value in m^2 for which the subdivision into two subgroups brings to the more significant improvement of the regression. Hence, when the surface of the building is lower than 98.5 m^2, the relative family falls on the left side of the tree, otherwise it goes on the right. At the second level of the tree, the variable involved in the splitting is the *Geographic area* for both of the branches. Depending on the categories of the *Geographic area,* the two splits bring to four subgroups of consumers. At the third level of the tree, with the four groups just obtained, the variables now involved for the splitting are different: on the left side the *Households structure* leads to Cluster 1 and 2; *Households structure* plus *Type of heating system and fuel* lead to Cluster 3 and 4; *Households structure* plus *Type of housing* define Cluster 5 and 6. Moving to the right side of the tree, a splitting defined by *Type of heating system and fuel* brings to Cluster 7 and 8; Surface plus Households structure leads to Cluster 9 and 10; *Surface* plus *Type of heating system and fuel leads to Cluster 11 and 12.* The entire procedure here described is represented in Figure 32, where, inside the nodes, is contained the variable name that define a splitting, and on the branches there are the categories (or the value if it continuous) of that variable that move a household on the left or on the right. It is now clear why the *Surface* variable had the highest score of importance: it is the first variable called to create a subdivision and there is also a second call of it at the third level of the tree, where the splitting value is 141 m^2. The *Geographic area* was also important and in fact it is called twice on the second level of the tree. It is also very clear the reasons why the *Type of housing* was the less important: it is involved at the fourth level and only for one splitting. The last important result shown in Figure 32 is the label attached to the 12 leaves of the tree; these labels show two values: the first one is the mean energy expenditure of a specific cluster, the other one is the percentage of families that fall into that cluster.

As an example of reading the tree-structure, in Cluster 1 there are all the households in which the *Surface* of the building is lower than 98.5 m^2, the *Geographic area* is of categories (*d*) or (*e*), which correspond to *South* or *Island,* and the *Households structure* is of categories (*a*) or (*b*) or (*c*) or (*d*) or (*l*) or (*m*), which correspond to *Single person 18-34 years, Single person 35-64 years*, *Single person 65 years and more,* Couple without children with reference person (r.p.) 18-34 years, Mono parent family, Others. To appreciate in a more descriptive way the characteristics of each cluster, in Table 22 are listed all the variables and categories defining each group.

5.2.5 Post-processing of the market segmentation with respect to energy poverty

The 12 groups of consumers have been defined in such a way that, given a set of characteristics, the households which satisfy them have a certain regression model which explains their energy expenditure. However. through the implemented approach, it is possible to enrich the classification ex-post. This means that, for each group of consumers, in which there is a specific relation between energy expenditure and class variables, one might evaluate other aspects and indicators; in particular, for the purposes of the ASSIST project these evaluations concerns the conditions of vulnerability with respect to energy poverty. Given the lack of a unique definition of energy poverty and, at the same time, the characteristics of the database, the approach followed here is intended to capture all the possible different aspects that could be related to energy poverty, and then define a level of vulnerability for each cluster.

Table 23 summarizes the 5 aspects evaluated for each cluster in order to identify vulnerable clusters. To help the comprehension of the results, to each variable is associated a scale-colour that goes from green, when the number of households that meet the condition are low, to red when the condition of vulnerability affects a higher percentage of families. A more detailed analysis of the results is proposed in the following paragraphs.

Table 23 - Clusters characteristics related to energy poverty.

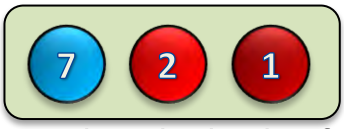
CLUSTER	families	MEAN ENERGY EXPENDITURE	percentage of families that inside a cluster satisfy the specific condition				
			ENERGY EXPENDITURE > FOOD EXPENDITURE	ABSOLUTE POVERTY	SCARCE ECONOMIC RESOURCES	INSUFFICIENT ECONOMIC RESOURCES	ENERGY EXPENDITURE > 10 % TOTAL EXPENDITURE
1	10.5%	72.70 €	6.0%	9.3%	48.4%	14.9%	16.0%
2	10.2%	102.06 €	3.3%	9.4%	45.9%	11.1%	11.2%
3	13.7%	97.59 €	10.5%	4.6%	38.7%	6.6%	16.3%
4	1.0%	156.48 €	24.7%	1.9%	37.0%	13.6%	31.8%
5	14.0%	121.68 €	4.2%	6.5%	39.1%	7.1%	8.9%
6	7.5%	150.09 €	5.6%	5.4%	39.0%	6.8%	16.7%
7	1.9%	80.33 €	3.8%	11.2%	46.2%	14.7%	9.8%
8	14.4%	130.60 €	6.5%	4.3%	36.0%	6.6%	15.1%
9	5.0%	129.42 €	11.9%	1.1%	33.8%	6.4%	15.6%
10	15.0%	166.25 €	4.8%	2.6%	29.1%	3.8%	10.9%
11	1.6%	167.46 €	8.0%	3.8%	32.4%	6.7%	9.2%
12	5.1%	222.82 €	9.5%	0.7%	22.0%	2.0%	13.3%

MEAN ENERGY EXPENDITURE

The mean energy expenditure, which characterizes each group, is a first indication of a possible condition of vulnerability (i.e., a lack of resources in reaching adequate temperatures or other comforts related to energy). Hence, in this single case, the red colour is associated to low values. These values are the same as the ones reported in the leaves labels of the regression tree and indicate that the more vulnerable groups are cluster 1 and 7, but also in group 2 and 3 there is a high risk. Furthermore, considering as a threshold the global mean energy expenditure, evaluated on the whole sample and equal to 127.2 €, clusters much below the limit are 1 and 7, slightly above them there are 2 and 3, and slightly below the limit there is number 5.

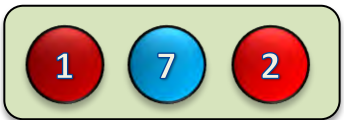
ENERGY EXPENDITURE > FOOD EXPENDITURE

This indicator may help capturing another facet of energy poverty: it highlights the families that, in order to maintain an adequate level of comfort, have to spend more for energy than for food, developing another form of vulnerability. In this case the more vulnerable group is the 4^{th}, but also number 3 and 9 contain a high percentage, with more than 10%, of households falling into this condition.



ABSOLUTE POVERTY

The absolute poverty indicator is defined by ISTAT and is calculated on the basis of a threshold that represents the minimum monthly expenditure necessary to buy a basket of essential goods and services, specifically defined for the Italian context and for different family structures, geographic areas and type of municipalities. In the entire database, around 6% of the population falls in this condition. It is widely recognised that energy poverty and poverty in general are different phenomena, but still they could be also strongly related, therefore the application of this indicator can help the characterization. Looking at the table we can see that, in terms of absolute poverty, risky situation are located in clusters 1 and 2, but above allin the 7^{th}.



SCARCE ECONOMIC RESOURCES

Energy poverty is certainly related to the economic condition of a household: the indicator here applied counts the number of families that consider their economic situation as scarce. This variable has been proposed in the initial description of the dataset, on aggregate level, and it is a subjective indication of the single perception of the income available, on a scale that includes 4 options: optimal, adequate, scarce and insufficient. Almost 40% of the population feels to have scarce economic resources, and this is the reasons why also on a cluster level the percentages are higher than the other cases, but anyway, it appears that there is more concentration of this condition in clusters 1, 2, and 7.

INSUFFICIENT ECONOMIC RESOURCES

The condition of insufficient economic resources is related to the same aspects of the previous case, but now indicates a perception of the income that is insufficient for households needs. In this case the global percentage is about 8% of the population, but we can find them in higher concentration inside clusters 1, 7, 4 and 2.

ENERGY EXPENDITURE > 10% TOTAL EXPENDITURE

The most famous indicator for energy poverty, applied in the vast majority of the studies around Europe, is the one defined by Boardman in 1991, that defines that there is a condition of "fuel poverty" when a household has to spend more than 10% of its income in energy to maintain an adequate level of temperature. This indicator is here adapted exploiting the total expenditure of a household, instead of the income. Hence, the ration between energy expenditure and total expenditure is evaluated, and it shows how much of the spending capacity is dedicated to energy. If the limit is set on the 10% it appears that the most vulnerable class is the 4th, but also in cluster 1, 3, 6, 8 and 9 there is a concentration around 15% of households.

GLOBAL APPROACH

The results in Table 23 can be read with two approaches: (a) by looking at each column searching for the most vulnerable groups, for fixed indexes; (b) by looking at each row and identifying when a single group is characterized by worse conditions. In this respect, the more vulnerable clusters can be considered 1, 2, 4 and 7. In other words, this means that, when a household has the specific characteristics of that cluster, it also has a higher probability to find itself in a particular condition of vulnerability in relation to energy.

DESCRIPTIVE APPROACH

Figure 33 proposes another approach to read the results; in particular, from this graphs it's easily visible which is the contribution of the different indicators in determining the vulnerabilities. The same indicators as in Table 23 are used: the higher is the bar the higher is the vulnerability. Considering in aggregated way the whole indicators, the more at risk is cluster 4, followed by 1 and 7.

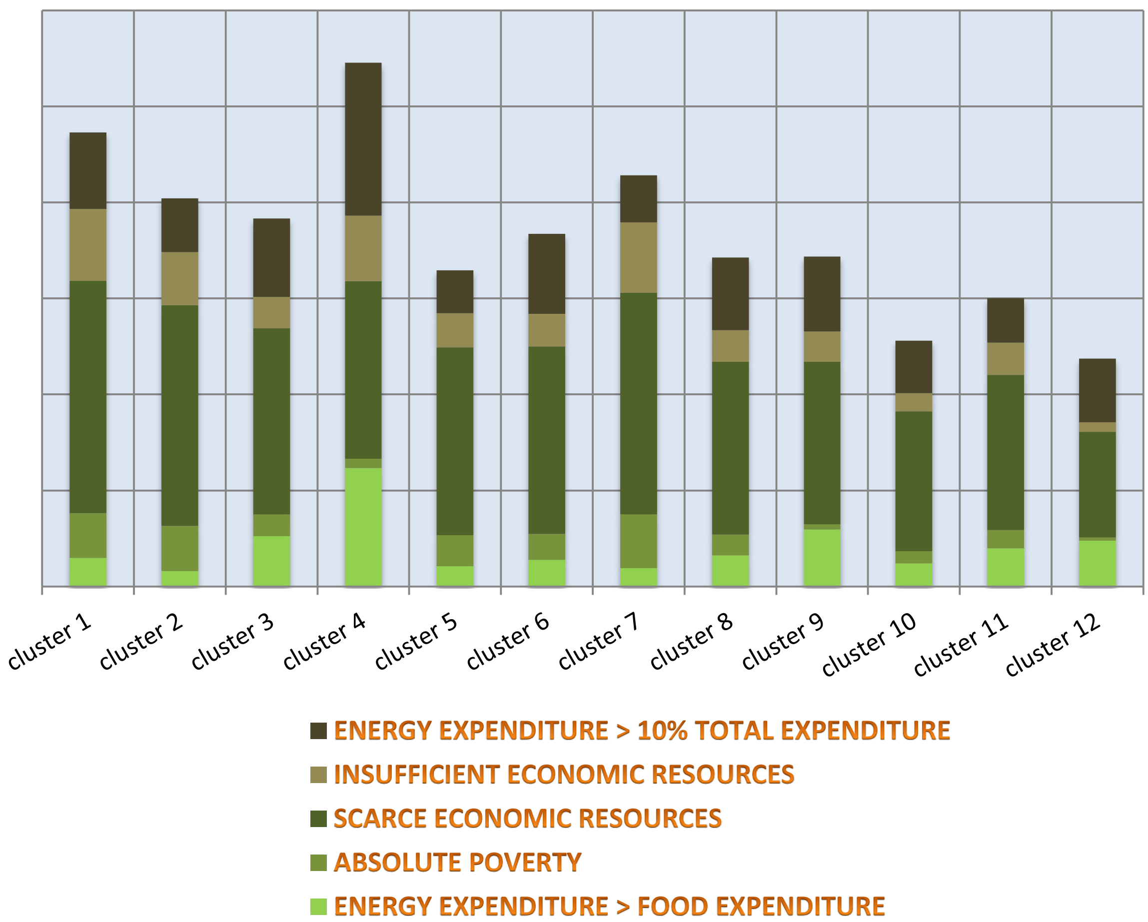


Figure 33 - Clusters aggregate characteristics related to energy poverty.

Figure 34 - Figure 38 display the distribution inside clusters of the different indicators. This approach is more sensitive to the total number of households contained in a group, but it still can give useful indications. As an example, inside cluster 4 the percentage of households with energy expenditure higher than food expenditures is equal to 24.7%, almost one fourth of the group. However, since this is a particular cluster composed by a reduced number of families, in Figure 34 is shown that the bar relative to cluster 4 is the lowest. By combining the results reported in the two figure, it is possible to conclude that the majority of the population satisfying the above condition is in cluster 1, 2, 3 and 8, already identified as vulnerable; moreover, also in cluster 5 and 10, where there are less vulnerable people in %, the number of vulnerable families is very high, due to the fact that around 30% of the total population falls into these two clusters. The same tendency can be found reproduced also when analysing the other indicators, with one interesting exception for what reported in Figure 35, related to absolute poverty indicator, in which, despite cluster 4 is one with the lowest number of households, the most of the population satisfying the condition is concentrated in that group, in particular one fourth of the sample.

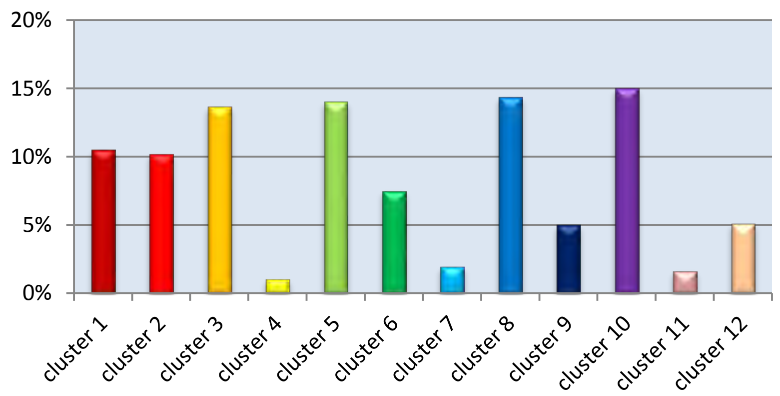


Figure 34 - Distribution inside cluster of the households with energy expenditure higher than food expenditure.

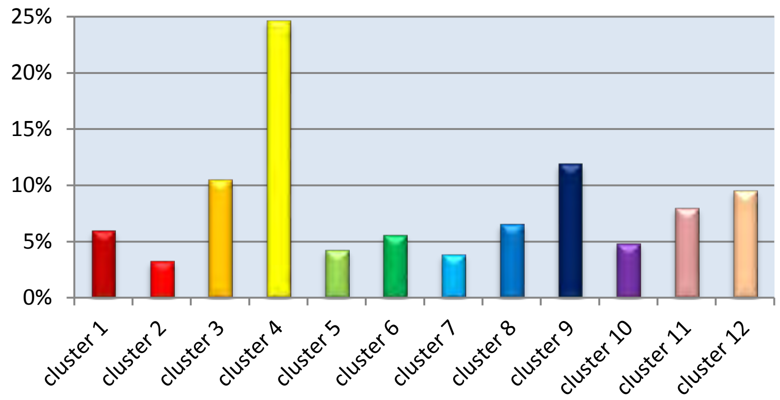


Figure 35 - Distribution inside the clusters of the households in absolute poverty.

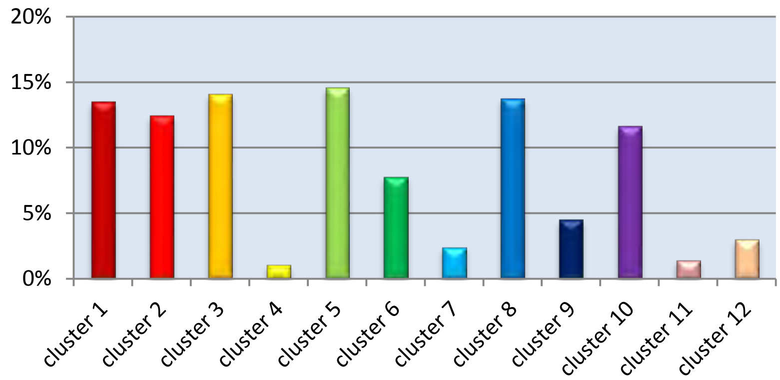


Figure 36 - Distribution inside the clusters of the households with scarce economic resources.

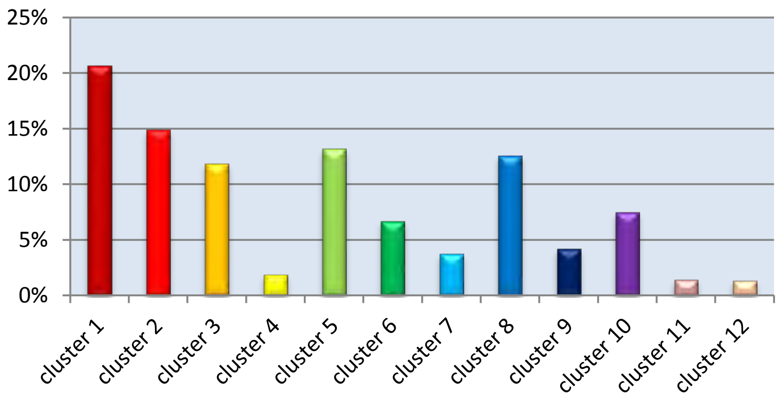


Figure 37 - Distribution inside the clusters of the households with insufficient economic resources.

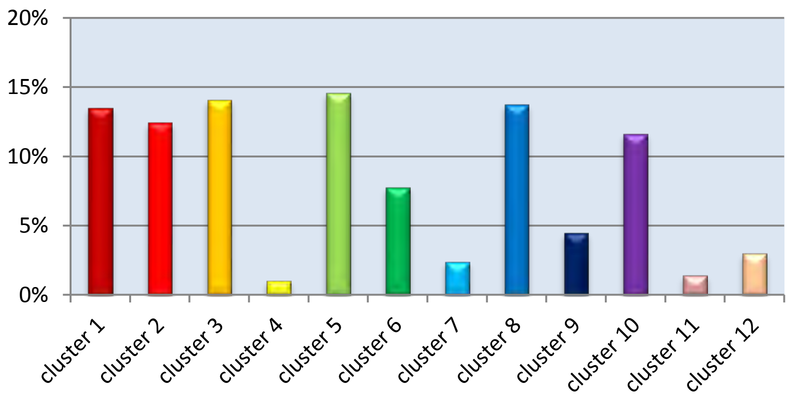


Figure 38 - Distribution inside the clusters of the households with energy expenditure higher than 10 % of total expenditure.

6. Italy – Refined analysis for the Italian household segmentation

The goal of this section is to propose an alternative approach for the segmentation of the Italian households, based on the per-capita household energy expenditure on an annual basis. The present analysis is based on independent variables which are modified compared with the ones available in the ISTAT dataset; conversely, the analysis presented in the previous sections is based on variables which are included in the ISTAT dataset, without any further modification. This analysis has the advantage of being more refined: it uses a larger set of independent variables, e.g. including also transport costs, and the dependent variable is the annual energy consumption, so giving a more detailed picture of Italian household expenses.

6.1 Regression Analysis methodology

The following sections describe in detail the variables included in the regression, starting from the dependent variable, namely the one to be predicted, and then all the predictors used to describe this dependent variable. Please note that the regression approach (viz. the coupling between OLS-VIF-LASSO approach) is the same as the one previously described.

- **Dependent variable**

The dependent variable is the per-capita household energy expenditure on an annual basis. This quantity is computed by the ratio of (a) the sum of the annual electricity expenditure and the annual heating expenditure (considering both households who are connected to the gas network and households that use other fuels) to (b) the number of the components in the household. The annual electricity expenditure is computed by multiplying the monthly expenditure for 12 (it was verified that the annual electricity expenditure has a limited dependence upon the time variable). The annual heating expenditure is computed by estimating (based on the ISTAT dataset), a typical annual energy consumption profile; based on this information, the monthly energy heating is corrected into the annual one. A similar procedure was also applied by Faiella et al. (2017). The per-capita household energy expenditure on an annual basis can be considered as a proxy of energy consumption and is the best available criterion to define the condition of vulnerability of a family, regarding energy poverty. The dependent variable has been log-transformed (natural log) to achieve greater symmetry of the distribution and of the residuals in the regression analysis. It is worth noting that families having energy expenditure equal to 0 have been excluded from this analysis.

Predictors

Table 24 displays the socio-demographic variables considered in the analysis (22 variables),

Table 25 displays the building variables considered in the analysis (11 variables) and

Table 26 displays the appliances and related variables considered in the analysis (10 variables). The variables marked in bold in Table 24,

Table 25 and

Table 26 are the baseline for the categorical variables.

Table 24 - Socio-demographics variables.

Variable	*Summary statistics*
1. Gender of the reference person	**(a) Male [10193]**, (b) Female [4820]
2. Current economic resources	(a) Optimal [279], **(b) Adequate [7912]**, (c) Scarce [5651], (d) Insufficient [1171]
3. Changing in economic resources compared to the previous year	(a) Much improved [30], (b) A little bit improved [512], **(c) More or less the same [8488]**, (d) A little worsened [4626], (e) Much worsened [1357]
4. Absolute poverty	(a) Yes [834], **(b) No [14179]**
5. Birth place of the components	**(a) Only born in Italy [13456]**, (b) At least one born abroad [973], (c) Only born abroad [584]
6. Citizenship of the components	**(a) Only Italian citizens [14176]**, (b) At least one foreign citizens [257], (c) Only foreign citizens [580]
7. Marital status of the reference person	(a) Unmarried [2551], **(b) Married or cohabitant with the spouse [8252]**, (c) Married but not cohabitant with the spouse [355], (d) Legally separated [625], (e)) Divorced [698], (f) Widow or widower [2532]
8. Qualification of the components	(a) No member has a qualification [377], (b) At least one member with elementary school [1978], (c) At least one member with junior high school [3108], **(d) At least one member with high school [6483]**, (e) At least one member with a degree [3067]
9. Work contract of the components	**(a) There is neither temporary job nor permanent job [7536]**, (b) At least one temporary job [1125], (c) At least one permanent job [6352]
10. Source of income of the components	(a) There is no income [83], (b) At least one maintained [413], (c) At least one pension [4911], **(d) At least one income [9606]**
11. Enrolment in study courses	**(a) No members enrolled in a course [[10930]**, (b) At least one in no title school [419], (c At least one in elementary school [747], (d) At least one in junior high school [584], (e) At least one in high school [1244], (f) At least one in a degree or post-degree course [1089]
12. Expenditure for elderly/disabled people	(a) Yes [100], **(b) No [14913]**
13. Households	(a) Single person 18-34 years [391], (b) Single person 35-64 years [1817],

structure	(c) Single person 65 years and more [2240], (d) Couple without children with r.p 18-34 years [178], (e) Couple without children with r.p. 35-64 years [1350], (f) Couple without children with r.p. 65 years and more [2164], **(g) Couple with 1 child [2276]**, (h) Couple with 2 children [2184], (i) Couple with 3 children and more [495], (l) Mono parent family [1033], (m) Others [885]
14. Workers in the primary sector	**(a) No one [13622]**, (b) One [1100], (c) More than one [291]
15. Workers in the secondary sector	**(a) No one [9766]**, (b) One [4098], (c) More than one [1149]
16. Workers in the tertiary sector	(a) No one [4577], **(b) One [6195]**, (c) More than one [4241]
17. Managers and employees	**(a) No one [8227]**, (b) One [4739], (c) More than one [2047]
18. Workers and similar	**(a) No one [8166]**, (b) One [4741], (c) More than one [2106]
19. Entrepreneurs and freelancers	**(a) No one [13696]**, (b) One [1172], (c) More than one [145]
20. Self-employed workers	**(a) No one [11876]**, (b) One [2583], (c) More than one [554]
21. Age of the reference person	(a) Up to 34 years [995], (b) From 25 to 44 years [2343], **(c) From 45 to 54 years [3059]**, (d) From 55 to 64 years [2934], (e) From 65 to 74 years [2841], (f) From 75 years [2841]
22. Transport expenditures	**Continuous variable** [Mean = 130 / Variance = 18725]

** r.p. = reference person of the household*

*** Summary statistics evaluated on the whole data-set*

**** Reference categories highlighted in bold*

Table 25 - Building variables.

Variable	Summary statistics
1. Period of construction	(a) After 2009 [190], (b) Between 2000 and 2009 [1222], (c) 1990s [1411], (d) 1980s [2326], **(e) 1970s [3426]**, (f) 1960s [2724], (g) 1950s [1481], (h) Between 1900 and 1949 [1404], (i) Before 1900 [829]
2. Number of rooms	(a) 1 [443], (b) 2 [2876], **(c) 3 [5912]**, (d) 4 [4044], (e) 5 [1212], (f) 6 [333], (g) 7 and more [193]
3. Gas from network	**(a) Yes [12359],** (b) No [2645]
4. Gas/Electricity cooking	**(a) Yes [14940]**, (b) No [73]
5. Occupation title	(a) Rent or sublease [2358], **(b) Property [11169]**, (c) Usufruct [352], (d) Free use [1134]
6. Type of housing	(a) Single family villa [2738], (b) Multifamily villa [4587], **(c) Apartments in building with less than 10 apartments [3733]**, (d) Apartments in building with 10 or more apartments [3939], (e) Other [16]
7. Type of municipalities	(a) Centre of metropolitan area [1889], (b) Periphery of metropolitan area and municipalities with 50.001 inhabitants and more [4032], **(c) Other municipalities until 50.000 inhabitants [9092]**
8. Domestic hot water system	(a) Electric boiler [2009], (b) Gas boiler [3603], **(c Heating system [9135]**, (d) Other [24], (e) Solar panel [199], (f) No hot water system [43]
9. Geographic area	(a) North-west [3284], (b) North-east [3382], **(c) Centre [2791]**, (d) South [4385], (e) Sicilia [753], (f) Sardegna [418]

10. Type of heating system and fuel	(a) Central heating - Gas from network [1381], (b) Central heating – Gas oil/kerosene/other liquid fuels [264], **(c) Autonomous heating - Gas from network [9443]**, (d) Autonomous heating – Gas oil/kerosene/other liquid fuels [270], (e) Autonomous - Gas cylinder [483], (f) Autonomous - Wood/pellet/ Other solid [964], (g) District heating [109], (h)) Individual devices - Wood/pellet/ Other solid [841], (i) Individual devices - Other fuels not solids [396], (l) Other [101], (m) No heating system [761]
11. Surface	**Continuous variable** [Mean = 98 / Variance = 1342]

** r.p. = reference person of the household*
*** Summary statistics evaluated on the whole data-set*
**** Reference categories highlighted in bold*

Table 26 - Appliances and related variables.

Variabile	Descrizione
1. Owner of fridge	**(a) Yes [14982],** (b) No [31]
2. Owner of washing machine	**(a) Yes [14750],** (b) No [263]
3. Owner of dishwasher	(a) Yes [7220], **(b) No [7793]**
4. Owner of air conditioner	(a) Yes [5043], **(b) No [9970]**
5. Self-consumption	(a) Yes [2029], **(b) No [12984]**
6. Number of cars	(a) No car [2761], **(b) One car [7324],** (c) Two cars [4226], (d) Three or more cars [702]
7. Number of mobile phones	(a) No mobile phones [1565], **(b) One mobile phones [4373]**, (c) Two mobile phones [5225], (d) Three or more mobile phones [3850]
8. Number of televisions	(a) No TV [648], **(b) One TV [6600],** (c) Two TVs [5785], (d) Three or more TVs [1980]
9. Number of computers	(a) No pc [5607], **(b) One pc [6771],** (c) Two pc [2087], (d) Three or more pc [548]
10. Expenses for appliances and electric tools	***Continuous variable*** [Mean = 22 / Variance = 1650]

** r.p. = reference person of the household*
*** Summary statistics evaluated on the whole data-set*
**** Reference categories highlighted in bold*

6.2 Cluster Analysis methodology

The cluster analysis is the same as the one previously described in chapter 5, without any further modification. In particular, the regression tree is built accordingly with the CART approach, by using the following phases:

e) Generation of a tree with a non-restrictive stopping rule

The procedure is started with an "*overgrown*" tree. To this end, the *complexity parameter* (*cp*) of the algorithm is modified. The *complexity parameter* controls the growth of the tree as follows: the overall R-squared must increase by *cp* at each subdivision. Hence, by starting with a low *cp* value, the split is permitted also when it doesn't imply a very significant increase of the result, leading to a very elaborated tree with a lot of variables and a lot of sub-groups of households at the end of the tree-structure.

f) Generation of a tree with minimum cross-validation error

Once the complete development of the tree is built, it is possible to search for the best *cp* value. This value is the one that minimizes the cross-validation error (*xerror*). The cross-validation error is related to the *PRESS* statistic (*Predicted Residual Error Sum of Squares*) and it is computed as the sum of the squares of all the resulting prediction errors; hence it is used exactly to understand at which point the tree starts to be over-grown (viz. when the regression model is over-fitted). Thus, identifying the minimum *xerror* leads to the optimal number of splits.

g) Generation of a trade-off tree

At this point, the previously-obtained tree is further processed to find the one that represents the best compromise between error and complexity. To this end, the cutting point is changed and the modifications to the tree structure are observed (i.e., R-squared, cross-validation error, number of variables involved, number of leafs, graph of the residual of the model).

h) Analysis of the sub-groups of households identified

The final tree obtained at the previous step is, thus, used as market segmentation. At this point, the characteristics of the sub-groups can be studied and post-processed with respect to vulnerability criteria.

6.3 Regression Analysis

In this section, the results of the regression models (both the partial and the aggregated models) are presented and discussed: (a) Socio-demographic characteristics, (b) Building characteristics, (c) Appliances owned. Adjusted R^2 (R_{ady}^2), the number of predictors excluded are presented and commented. For all models, the residuals were inspected, by the Q-Q plot inspection. In particular, the main results are presented in Table 27, Table 28, Table 29, Table 30 and Table 31. Table 27 summarizes the R_{ady}^2 values for the predictor groups and their aggregation. Table 28 displays the socio-demographics variables used in the partial model and in the aggregate model. Table 29 displays the building variables used in the partial model and in the aggregate model. Table 30 displays the appliances and related variables used in the partial model and in the aggregate model. Table 31 provide all the details concerning the final regression model on energy expenditure.
The socio-demographic model explains R_{ady}^2 = 27.7 % for the variability in energy consumption expenditure (Table 27). After confirming the presence of multicollinearity with the VIF analysis, implementing LASSO regression and checking for the significance of the single variables, the final regression with socio-demographic aspects leads to (a) 9 of the 22 initial variables available in the partial model and (b) 6 of the 22 initial variables available in the aggregate model (Table 28). The building model explains R_{ady}^2 = 11.7 % for the variability in energy

consumption expenditure (Table 27). After confirming the presence of multicollinearity with the VIF analysis, implementing LASSO regression and checking for the significance of the single variables, the final regression with socio-demographic aspects leads to (a) 8 of the 11 initial variables available in the partial model and (b) 5 of the 11 initial variables available in the aggregate model (Table 29). The appliances and related variables model explains R_{ady}^2 = 15.8 % for the variability in energy consumption expenditure (Table 30). After confirming the presence of multicollinearity with the VIF analysis, implementing LASSO regression and checking for the significance of the single variables, the final regression with socio-demographic aspects leads to (a) 6 of the 10 initial variables available in the partial model and (b) 1 of the 10 initial variables available in the aggregate model (Table 28). The previous models are combined together to test for increments in explanatory power when joining the three categories. Obviously, the variables now involved are only the ones resulted as significant in the single regressions and are listed in Table 31. After merging together all the predictors obtained in the previous regressions, multicollinearity issues has been detected and it has been necessary to exploit the LASSO regression and some other variables are removed, due to their low significance. In the final model of the energy expenditure, the variability explained is R_{ady}^2 = 38.0 %, with 12 variables involved, of which 6 are socio-demographic, 5 are related to the building and 1 concern appliances.
In conclusion, the study highlights that socio-demographic characteristics are the most influential variables, followed by owned appliances characteristics and, finally, building variables. Finally, the combination of the three categories leads to a description of the energy expenditure variation equal to 38%, meaning that acting on these variables could bring a significant variation in the energy expenditure.

6.4 Cluster Analysis

Once the selected variables have been defined (as listed in Table 31), it is possible to proceed with the market segmentation of the Italian families with respect to the total per-capita annual energy expenditure. Applying the CART procedure, the resulting number of sub-groups of consumers is equal to 12 (please refer to the structure of the tree as presented in Figure 39). In detail, the procedure starts with the root node in which the whole sample of 15.000 families has been spitted based on the Household structure: if the household structure belongs to the categories d, e, f, g, h, i, l (Table 24), the relative family falls on the left side of the tree, otherwise it goes on the right side. On the resulting left side, at the second level of the splitting, the variable involved in the splitting is, again, the Household structure: if the household structure belongs to the categories d, g, h, i, l (Table 24), the relative family falls on the left side of the tree, otherwise it goes on the right side. The former families are, then, classified by the geographic area (if the belong to categories d, e, f the lead to cluster#1; otherwise, they belong to cluster#2); conversely, the latter are classified by the surface area: the families with an household surface area above 141 m2 belong to a single cluster (cluster#6). On the other hand, families with a with an household surface area below 141 m^2 are again classifies based on the Geographic area and, finally, on the type of housing: (a) households in the geographical area d,e and f belong to cluster #3, (b) households in the geographical area a, b and c and with a type of housing c, d and g belong to cluster 4; (c) households in the geographical area a, b and c and with a tpe of housing a, b belong to cluster 5. On the left side, at the second level of the splitting, the variable involved the Geographic area (if the household belongs to the categories d, e, f the relative family falls on the left side of the tree, otherwise it goes on the right side); subsequently, families are classified by the *type of heating system and fuel: (a)* households in the geographical area d, e and f and the *type of heating system and fuel* l, m belong to cluster#7; (b) households in the geographical area d, e and f and the *type of heating system and fuel* a, b, c, d, e, f, g, h, I belong to cluster#8; (c) households in the geographical area a, b and c and the *type of heating system and fuel* b, d, e and g belong to cluster#12. Finally, families belonging to the geographical area a, b, c with *type of heating systems and fuel* a, c, f, h, I, I , m are spitted based on the type of housing and for the surface area: *(a)* households with *type of heating system and fuel* c, d belong to cluster#9; (b) households with *type of heating system and fuel* a, b and e and an household surface area below 98 m^2 belong to cluster#10; (c) households with *type of heating system and fuel* a, b and e and an household surface area above 98 m^2 belong to cluster#11. The entire procedure here described is represented in Figure 39, where, inside the nodes, is contained the variable name that define a splitting, and on the branches there are the categories (or the value if it continuous) of that variable that move a household on the left or on the right.

Table 27 - Adjusted R^2 for the predictor groups and their aggregation.

Predictors	Adjusted R^2
Socio-demographics variables	27.7%
Building variables	11.7%
Appliances variables	15.8%
Aggregated model	**38.0%**

Table 28 - Socio-demographics variables used in the partial model and in the aggregate model.

Variable	Partial model	Aggregate model
Gender of the reference person	✓	✓
Current economic resources	✓	
Changing in economic resources		
Absolute poverty	✓	✓
Birth place of the components		
Citizenship of the components		
Marital status of the reference person		
Qualification of the components		
Work contract of the components		
Source of income of the components	✓	✓
Enrolment in study courses	✓	✓
Expenditure for elderly/disabled people		
Households structure	✓	✓
Transport expenditures	✓	✓
Workers in the primary sector		
Workers in the secondary sector	✓	
Workers in the tertiary sector		
Managers and employees		
Workers and similar		
Entrepreneurs and freelancers	✓	
Self-employed workers		
Age of the reference person		

Table 29 - Building variables used in the partial model and in the aggregate model.

Variable	Partial model	Aggregate model
Period of construction	✓	
Number of rooms	✓	
Gas from network	✓	✓
Gas/Electricity cooking		
Occupation title	✓	
Type of housing	✓	✓
Type of municipalities		
Domestic hot water system		
Surface	✓	✓
Type of heating system and fuel	✓	✓
Geographic area	✓	✓

Table 30 - Appliances and related variables used in the partial model and in the aggregate model.

Variable	Partial	Aggregate model
Owner of fridge		
Owner of washing machine		
Owner of dishwasher	✓	✓
Owner of air conditioner		
Self-consumption	✓	
Number of cars	✓	
Number of mobile phones	✓	
Number of televisions	✓	
Number of computers		
Expenses for appliances and electric tools	✓	

Table 31 - Final regression model on energy expenditure.

Variable	Estimate	Std.Error	t value	Pr(>|t|)	Significance
(Intercept)	5.8660	2.460e-	238.49	< 2e-16	***
Owner of dishwasher **Yes**	0.0642	1.080e-	5.94	2.95e-	***
Gas from network **No**	-0.2090	1.981e-	-10.55	< 2e-16	***
Type of housing **Single family villa**	0.1089	1.615e-	6.74	1.60e-	***
Type of housing **Multifamily villa**	0.1112	1.328e-	8.37	< 2e-16	***
Type of housing **Apartments in**	-0.1134	1.384e-	-8.19	2.77e-	***
Type of housing **Other**	0.1115	1.826e-	0.61	0.5416	
Surface	0.0028	1.515e-	18.26	< 2e-16	***
Type of heating system and fuel	-0.1072	1.835e-	-5.85	5.15e-	***
Type of heating system and fuel	0.2358	3.950e-	5.97	2.44e-	***
Type of heating system and fuel	0.3473	3.947e-	8.80	< 2e-16	***
Type of heating system and fuel	0.2231	3.199e-	6.98	3.16e-	***
Type of heating system and fuel	0.0620	2.221e-	2.79	0.0053	**
Type of heating system and fuel	0.0028	5.620e-	0.05	0.9606	
Type of heating system and fuel	0.0392	2.503e-	1.57	0.11755	
Type of heating system and fuel	-0.1293	3.258e-	-3.97	7.28e-	***
Type of heating system and fuel **Other**	-0.0281	6.055e-	-0.46	0.6425	
Type of heating system and fuel **No**	-0.2694	2.789e-	-9.66	< 2e-16	***
Geographic area **North-west**	0.1724	1.527e-	11.29	< 2e-16	***
Geographic area **North-east**	0.0240	1.498e-	1.60	0.1099	
Geographic area **South**	-0.1539	1.448e-	-10.63	< 2e-16	***
Geographic area **Sicilia**	-0.1536	2.646e-	-5.80	6.65e-	***
Geographic area **Sardegna**	-0.0203	3.320e-	-0.61	0.5410	
Gender of the reference person **Female**	0.0572	1.252e-	4.56	5.06e-	***
Absolute poverty **Yes**	-0.2708	2.229e-	-12.15	< 2e-16	***
Source of income of the components	-0.2141	6.714e-	-3.19	0.0014	**
Source of income of the components **At**	-0.0353	3.144e-	-1.12	0.2622	
Source of income of the components **At**	0.0967	1.599e-	6.05	1.48e-	***
Enrolment in study courses **At least**	-0.2036	3.077e-	-6.62	3.83e-	***
Enrolment in study courses **At least**	-0.1997	2.486e-	-8.03	1.03e-	***
Enrolment in study courses **At least**	-0.1544	2.728e-	-5.66	1.53e-	***
Enrolment in study courses **At least**	-0.0921	2.100e-	-4.39	1.16e-	***
Enrolment in study courses **At least**	-0.0600	2.108e-	-2.85	0.0044	**
Households structure **Single person**	0.5914	3.452e-	17.13	< 2e-16	***
Households structure **Single person**	0.6538	2.050e-	31.89	< 2e-16	***
Households structure **Single person 65**	0.6681	2.381e-	28.06	< 2e-16	***
Households structure **Couple without**	0.0328	4.595e-	0.71	0.4758	
Households structure **Couple without**	0.2189	2.101e-	10.42	< 2e-16	***
Households structure **Couple without**	0.2234	2.226e-	10.04	< 2e-16	***
Households structure **Couple with 2**	-0.2267	1.839e-	-12.33	< 2e-16	***
Households structure **Couple with 3**	-0.3975	2.983e-	-13.33	< 2e-16	***
Households structure **Mono parent**	0.1114	2.362e-	4.72	2.40e-	***
Households structure **Others**	-0.1561	2.340e-	-6.67	2.64e-	***
Transport expenditures	0.0003	4.015e-	7.88	3.57e-	***

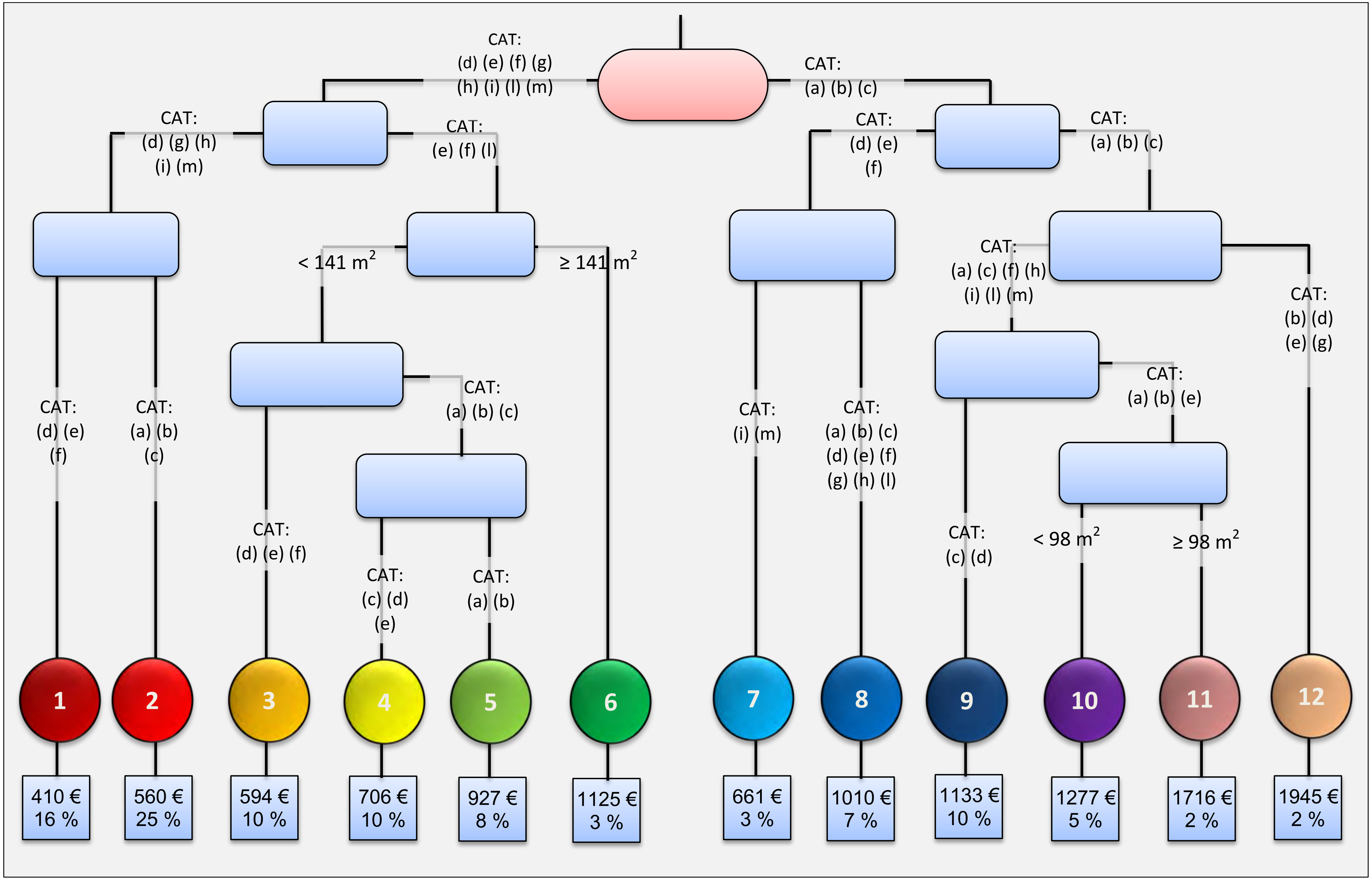


Figure 39 - Household segmentations based on the pro-capite anjual total energy expenditure.

Table 32 - Clusters charateristics.

CLUSTER	*Household structure*	*Geographic area*	*Type of heating system and fuel*	*Surface [m²]*	*Type of housing*
1	▪ Couple without children with r.p. 18-34 years ▪ Couple with 1 child ▪ Couple with 2 children ▪ Couple with 3 children and more ▪ Others	▪ South ▪ Sicily ▪ Sardinia	/	/	/
2	▪ Couple without children with r.p. 18-34 years ▪ Couple with 1 child ▪ Couple with 2 children ▪ Couple with 3 children and more ▪ Others	▪ North West ▪ North East ▪ Centre	/	/	/
3	▪ Couple without children with r.p. 35-64 years ▪ Couple without children with r.p. 65 years or more ▪ Single parent	▪ South ▪ Sicily ▪ Sardinia	/	< 141	/
4	▪ Couple without children with r.p. 35-64 years ▪ Couple without children with r.p. 65 years or more ▪ Single parent	▪ North West ▪ North East ▪ Centre	/	< 141	▪ Apartment in building with less than 10 apartments ▪ Apartment in building with 10 or more apartments ▪ Other
5	▪ Couple without children with r.p. 35-64 years ▪ Couple without children with r.p. 65 years or more ▪ Single parent	▪ North West ▪ North East ▪ Centre	/	< 141	▪ Single family villa ▪ Multifamily villa
6	▪ Couple without children with r.p. 35-64 years ▪ Couple without children with r.p. 65 years or more ▪ Single parent	/	/	≥ 141	/
7	▪ Single person 18-34 years ▪ Single person 35-64 years	▪ South ▪ Sicily	▪ Individual devices, other ▪ No heating system	/	/

	▪ Single person 65 years or more	▪ Sardinia			
8	▪ Single person 18-34 years ▪ Single person 35-64 years ▪ Single person 65 years or more	▪ South ▪ Sicily ▪ Sardinia	▪ Central heating, gas from network ▪ Central heating, gasoil, kerosene, other liquid fuels ▪ Autonomous, gas from network ▪ Autonomous, gasoil, kerosene, other liquid fuels ▪ Autonomous, gas cylinders ▪ Autonomous, wood and other solid fuels ▪ District heating ▪ Individual devices, wood and other solid fuels ▪ Other	/	/
9	▪ Single person 18-34 years ▪ Single person 35-64 years ▪ Single person 65 years or more	▪ North West ▪ North East ▪ Centre	▪ Central heating, gas from network ▪ Autonomous, gas from network ▪ Autonomous, wood and other solid fuels ▪ Individual devices, wood and other solid fuels ▪ Individual devices, other ▪ Other ▪ No heating system	/	▪ Apartment in building with less than 10 apartments ▪ Apartment in building with 10 or more apartments
10	▪ Single person 18-34 years ▪ Single person 35-64 years ▪ Single person 65 years or more	▪ North West ▪ North East ▪ Centre	▪ Central heating, gas from network ▪ Autonomous, gas from network ▪ Autonomous, wood and other solid fuels ▪ Individual devices, wood and other solid fuels ▪ Individual devices, other ▪ Other ▪ No heating system	< 98	▪ Single family villa ▪ Multifamily villa ▪ Other
11	▪ Single person 18-34 years ▪ Single person 35-64 years ▪ Single person 65 years or more	▪ North West ▪ North East ▪ Centre	▪ Central heating, gas from network ▪ Autonomous, gas from network ▪ Autonomous, wood and other solid fuels	≥ 98	▪ Single family villa ▪ Multifamily villa ▪ Other

			▪ Individual devices, wood and other solid fuels ▪ Individual devices, other ▪ Other ▪ No heating system		
12	▪ Single person 18-34 years ▪ Single person 35-64 years ▪ Single person 65 years or more	▪ North West ▪ North East ▪ Centre	▪ Central heating, gasoil, kerosene and other liquid fuels ▪ Autonomous, gasoil, kerosene and other liquid fuels ▪ Autonomous, gas cylinder ▪ District heating	/	/

6.5 Post-processing of the market segmentation with respect to energy poverty

The 12 groups of consumers (different from the groups identified with the first analysis, due to the changed variables) have been defined in such a way that, given a set of characteristics, the households which satisfy them have a certain regression model which explains their energy expenditure. Considering energy poverty criteria, the following indicators have been used (already described in par. 5.2):

- Criterion n#1: ENERGY EXPENDITURE LOWER THAN HALF OF THE MEDIAN
- Criterion #2: Incidence of energy expenditure higher than double the median
- Criterion #3: ABSOLUTE POVERTY
- Criterion #4: Energy expenditure higher than double the average
- Criterion #5: Energy expenditure higher than food expenditure
- Criterion #6: SCARCE ECONOMIC RESOURCES
- Criterion #7: INSUFFICIENT ECONOMIC RESOURCES
- Criterion #8: Not achieved 75% of the minimum thermal comfort[6]

Table 33 shows the results: in the first column there is the identification number of the clusters; second column summarize the characteristics of the regression tree of Figure 39, in particular the mean per-capita energy expenditure and the percentage of families included; the remaining columns contain the 8 aspects evaluated for each cluster, in order to identify the vulnerable ones. The way of reading the table is the same of the previous section, with red colour indicating that the condition of vulnerability affects a higher percentage of families.

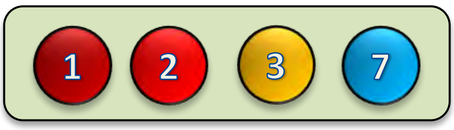
Table 33 - Market segmentation with respect to energy poverty.

CLUSTER	*Mean energy expenditure and percentage of families inside the cluster*	**Criterion #1**	**Criterion #2**	**Criterion #3**	**Criterion #4**	**Criterion #5**	**Criterion #6**	**Criterion #7**	**Criterion #8**

[6] Evaluated following the approach described by Faiella et al., (ref. [30]).

1	410 € 15.8%	39.7%	10.0%	8.2%	1.2%	2.9%	40.9%	10.6%	19.0%
2	560 € 24.8%	21.2%	9.8%	5.8%	2.3%	4.2%	34.7%	5.8%	23.0%
3	594 € 9.8%	16.2%	17.2%	8.0%	3.1%	4.6%	42.0%	10.2%	27.2%
4	706 € 9.7%	13.2%	10.1%	2.3%	5.7%	4.7%	35.2%	5.3%	27.3%
5	927 € 7.9%	4.5%	19.4%	2.0%	11.0%	6.6%	32.6%	4.5%	26.2%
6	1125 € 3.2%	3.4%	15.2%	1.3%	20.4%	7.6%	21.5%	4.2%	41.5%
7	661 € 3.0%	16.2%	13.0%	10.1%	4.3%	6.5%	53.0%	18.2%	44.7%
8	1010 € 7.4%	4.2%	24.2%	5.2%	13.4%	10.7%	46.5%	8.8%	34.2%
9	1113 € 10.0%	5.4%	13.9%	3.9%	17.9%	9.9%	36.5%	7.3%	33.2%
10	1277 € 4.6%	2.1%	25.2%	5.0%	27.0%	10.6%	39.7%	5.9%	33.8%
11	1716 € 2.4%	2.0%	22.1%	1.1%	45.1%	17.0%	33.0%	4.3%	43.0%
12	1945 € 1.5%	5.8%	31.9%	0.9%	50.9%	22.6%	34.5%	11.1%	20.8%

For each of the indicators, following Table 33, the more sensible clusters to energy poverty measures have been evaluated.



Criterion#1

When the vulnerability is defined by energy expenditure lower than half of the median, the more at risk is cluster 1 (39.7%), but also cluster 2 (21.2%), cluster 3 (16.2%) and cluster 7 (16.2%) contain a high percentage of families satisfying the condition.

Criterion#2

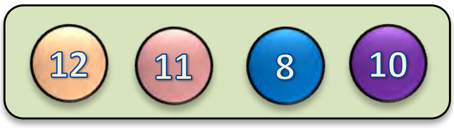
The incidence of energy expenditure, with respect to the total expenditures of a family, is above twice the median particularly in cluster 12 (31.9%), followed by cluster 10 (25.2%), cluster 8 (24.2%) and cluster 11 (22.1%).

- Criterion#3

Clusters with the highest percentage of families in absolute poverty, as defined by ISTAT, are cluster 7 (10.1%), cluster 1 (8.2%) and cluster 3 (8.0%).

- Criterion#4

Families with a very high energy expenditure, in particular, above the double of the mean, are situated with higher percentage in cluster 12 (50.9%) and in cluster 11 (45.1%); also in cluster 10 (27.0%) there is a high value compared to the other groups.

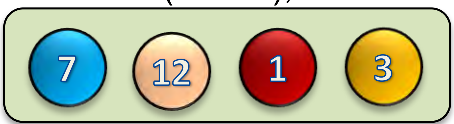


- Criterion#5

The indicator that compare the amount of money used to energy needs with the amount of money for food, shows that there is a higher percentage of households which spend more for energy in cluster 12 (22.6%), followed by cluster 11 (17.0%), cluster 8 (10.7%) and cluster 10 (10.6%).

- Criterion#6

Families who perceive to have scarce economic resources define a higher vulnerability in cluster 7 (53.0%), cluster 8 (46.5%), cluster 3 (42.0%), cluster 1 (40.9%).



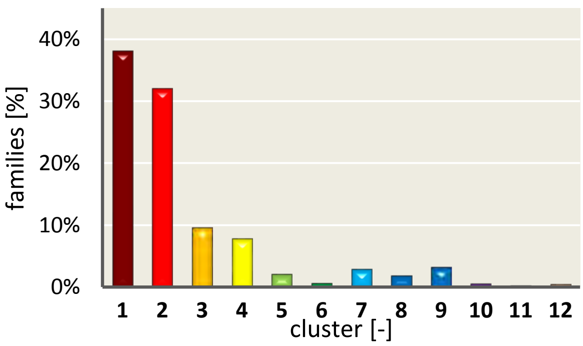
- Criterion#7

The condition of insufficient economic resources indicates that households more at risk from the economical point of view are distributed in higher concentrations inside cluster 7 (18.2%), cluster 12 (11.1%), cluster 1 (10.6%), cluster 3 (10.2%).

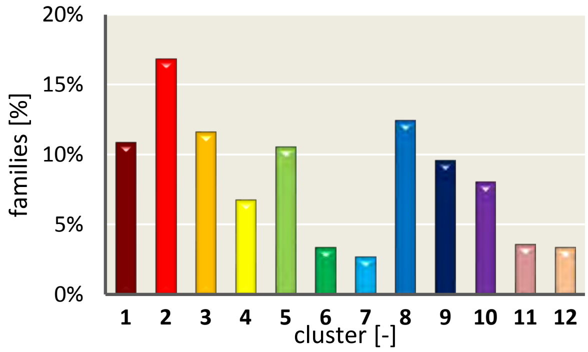
- Criterion#8

The risk of not reaching the 75% of the minimum level of thermal comfort is higher inside cluster 7 (44.7%), cluster 11 (43.0%) and cluster 6 (41.5%).

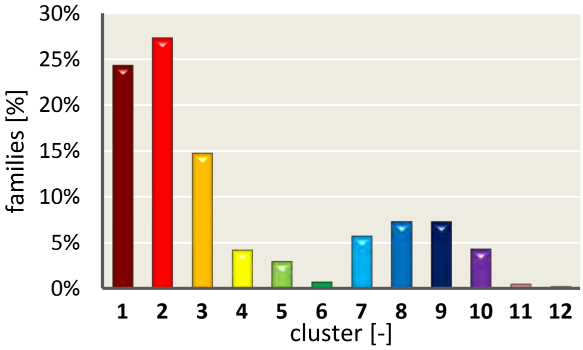
As in the previous analysis, Table 33 can be also observed looking at each row to identify in aggregate terms which are the more vulnerable groups. In this case, the households more at risk of energy poverty can be identified in clusters 1, 3, 7, 11 and 12; this means that, when a household has the specific characteristics of that cluster, it also has a higher probability to find itself in a particular condition of vulnerability in relation to energy. In the same way of the previous section, Figure 40 shows, for each one of the 8 indicators, how families satisfying the specific conditions are distributed. The interpretation of these graphs remains the same and the principal outcomes are now described. The households with annual per-capita energy expenditure lower than half of the median are mostly situated in clusters 1 and 2; those with incidence of the energy expenditure on total expenditure, over twice the median, are in the highest percentage in cluster 2, but a large part is also in clusters 1, 3, 5 and 8; with respect to the absolute poverty indicator, most vulnerable families are located in clusters 1, 2 and 3; by evaluating if the energy expenditure exceeds twice the median, it is noted that the highest fraction is in cluster 9; households which spend more for energy than food are principally located in cluster 2, 9 and 8; the distributions of households with scarce and insufficient economic resources are very similar, for both of them the highest values are in cluster 1 and 2; finally, the thermal comfort indicator identifies that families who not satisfy 75% of the minimal comfort are mostly situated in cluster 2 and 9. Finally, Figure 41 presents an alternative way to present vulnerable clusters; in particular, it provides a graphical rapresentation of the percentage of families, within each cluster that satisfy at least 2, 3 or 4 criteria. It is worth noting that this approach has the main drawback of not considering the possible correlations between the different criteria.



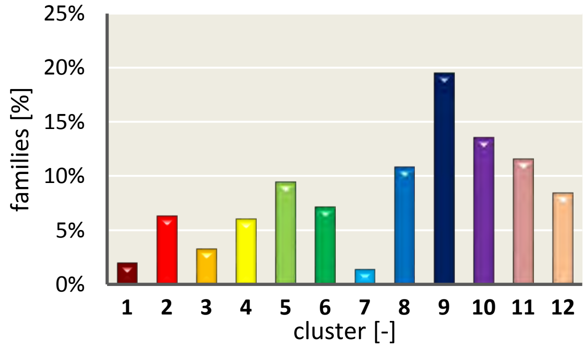
(*a*) Energy expenditure lower than half of the median



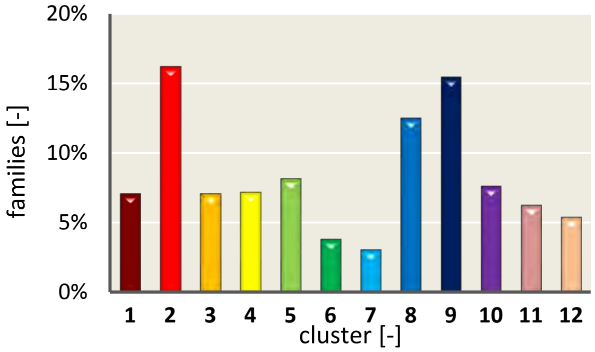
(*b*) Incidence of energy expenditure higher than double the median



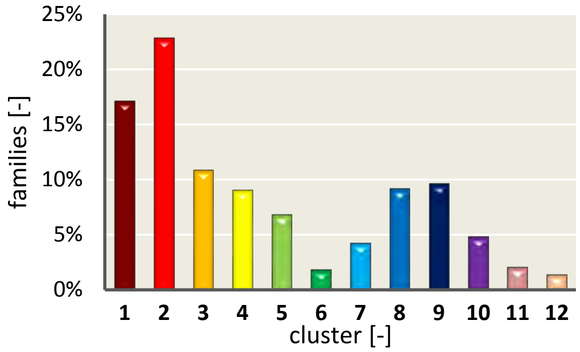
(*c*) Absolute poverty



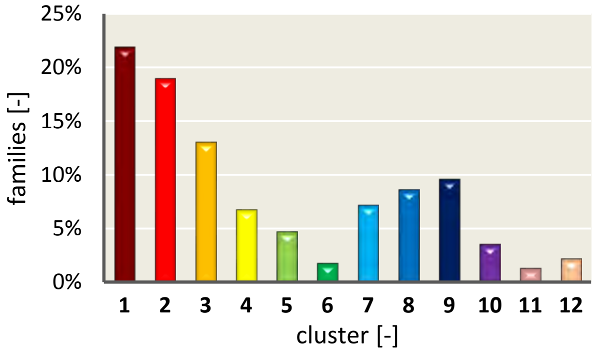
(*d*) Energy expenditure higher than double the average



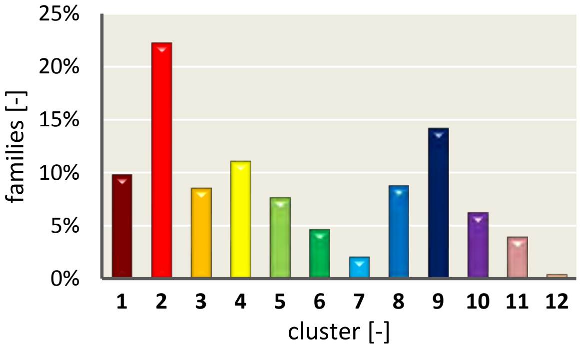
(*e*) Energy expenditure higher than food expenditure



(*f*) Scarce economic resources



(*g*) Insufficient economic resources



(*h*) Not achieved 75% of the minimum thermal comfort

Figure 40 - Distribution within the clusters of families that meet a certain condition.

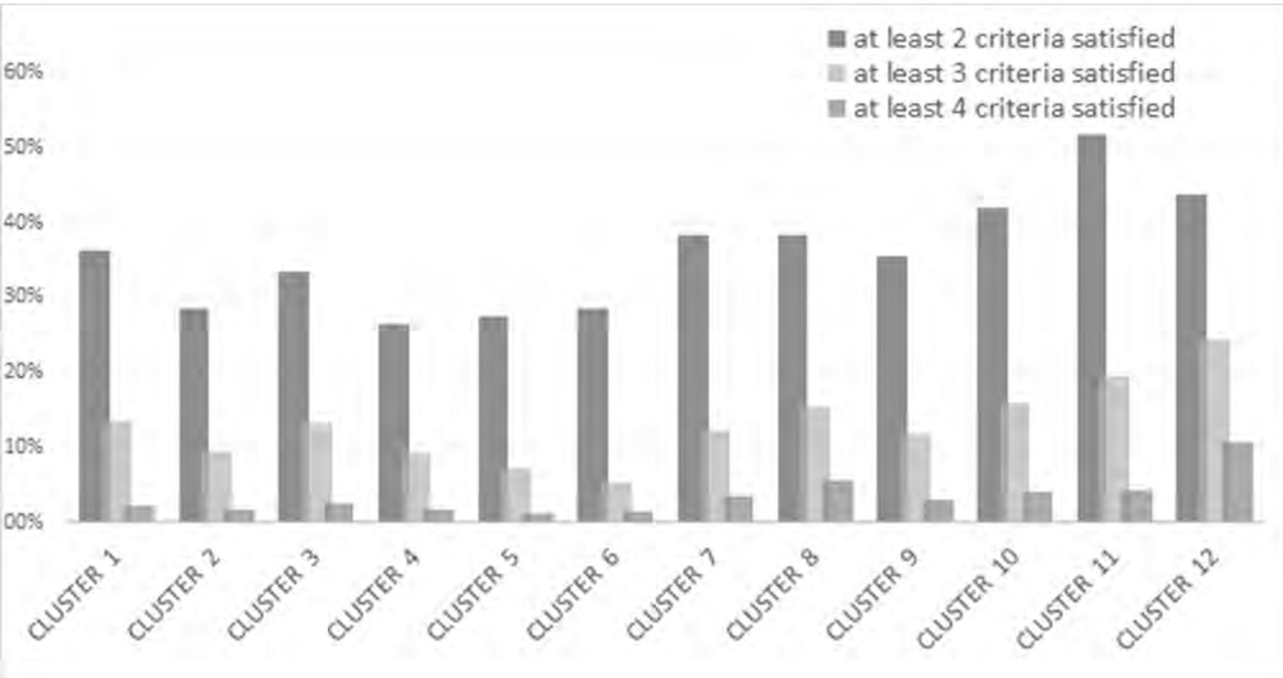


Figure 41 - Details of the clusters of families that meet a certain condition.

6.6 Conclusions

The presented analyses, first with the monthly expenses and then with the yearly expenses, are very useful to understand which types of families are more at risk of energy poverty. However, the main issue is the lack of a shared criterion to define an energy poor family, like, for example, a LIHC. This is mostly due to the unavailability of statistics that can correlate energy expenses to income. We have tried to solve the issue, by using different criterion to evaluate energy poverty, e.g. the expense to reach a minimum level of thermal confort or the fact that a household spends more in energy than in food. However, future studies should work on the collection of more “energy poverty focused” data and on the possible correlations between the different energy vulnerability criteria.
The results of the segmentation will be used to identify the targets for ASSIST Action, whose detail swill be presented in D5.2.

7. Italy – Vulnerable consumers questionnaire

The main advantage of the approach proposed in the previous sections is given by the fact that the use of different indicators may highlight different conditions of vulnerability and this is a very desirable procedure since energy poverty, with all its facets, is generated by different drivers (i.e., income, energy efficiency, energy prices, etc.) in different measures depending on households characteristics and needs. The results of this analysis are of fundamental importance for the knowledge of the Italian situation on the entire territory and constitute a reference and also a guide for the detection of possible households that could be part of the ASSIST action phase, in particular to be sure of including all the possible situation that the Country contains. Anyway, it is important to underline that this is a statistical approach and it is not possible to consider particular situation or exclude the possibility to find vulnerable consumers in other groups.
For this reason, to pursue the objective of the study, and in particular to obtain other important characteristics related to energy poverty and vulnerable consumers, it is required to consider also the results of the questionnaire specifically developed in the ASSIST activities and presented in the following paragraphs.
In this way, primarily, from the national level segmentation, some peculiarities and characteristics of the more at risk of poverty families are extracted, then the information from the questionnaire of Task 5.1 are considered. These data are collected through the network of the HEA – Home Energy Advisors; therefore they should represent in detail which are the characteristics of the vulnerable consumers involved in the ASSIST Action, both in terms of socio-economic characteristics and behavioural attitudes toward the energy management. Finally, also the information collected in the market survey of task 2.5 will be included to enrich the analysis.

7.1 Short questionnaire to Vulnerable Consumers

A short questionnaire has been developed in order to understand the needs of vulnerable consumers. It is composed by ten questions, mostly related to how they see the figures of the Energy Advisors and how they would prefer to receive information about optimizing their energy consumption. Moreover, some specific questions about their living conditions have been added, in order to be able to classify them according to the above reported statistics.
The questions are reported in the next paragraphs together with the results.
Since the HEA network is not developed at the time of writing this report, it has been decided to give the questionnaires to potential vulnerable consumers through the support of different actors, especially two types: the consumers associations and the social housing cooperatives.

These two types of actors have been chosen due to the fact that:

- In social houses, it is more likely to find people with low income, often with health issues, and with old and non-renovated houses with a low energy efficiency: this puts them in a potential vulnerable situation;
- At the consumers associations offices, where people go to understand their energy bills and to claim that the energy suppliers are charging them with high bills, it is possible to find potential vulnerable consumers, unable to afford to pay for their electricity and natural gas consumptions.

Moreover, a positive cooperation had been already established with several cooperatives and associations during previous projects, so it was easy to obtain their support.

The questionnaire has been provided in paper format, and has been given to the people while waiting for the electricity/natural gas bills consultancy at the consumers associations offices and at the inhabitants of some social housing blocks.

The main results are reported in the next paragraphs.

7.2 Socio-demographic and living conditions results

Six preliminary questions were related to the characteristics of the household. There were 113 responses to the questionnaire, that was mostly spread in the areas around Milano and Roma (north-west and central Italy, as shown in question 6).

a) Which is your nationality?

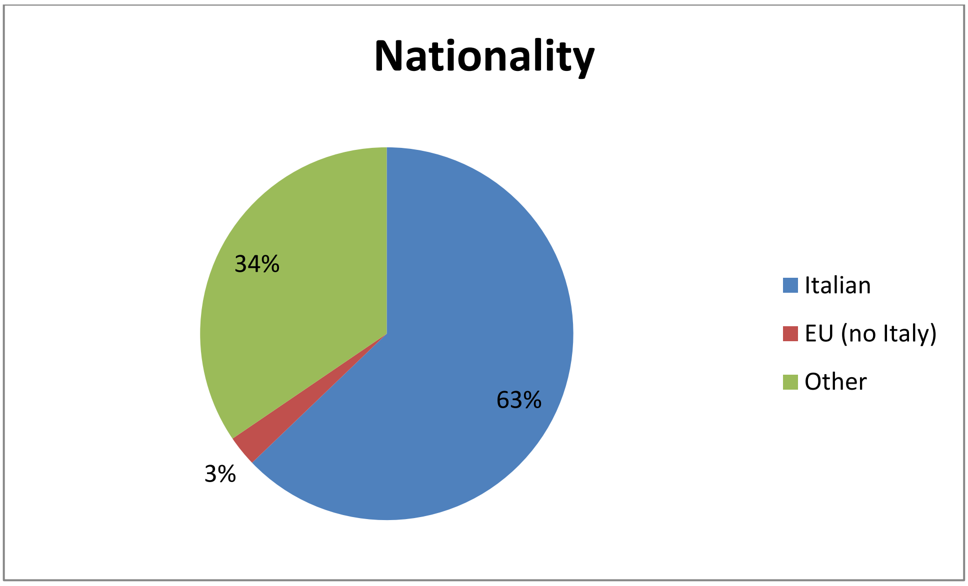


Figure 42 - Nationality of the responders.

Most of the responders (around two thirds) were Italian, while around one third was coming from outside the EU. Only 3% was coming from other EU countries.

b) How many people live in your house (including yourself)?

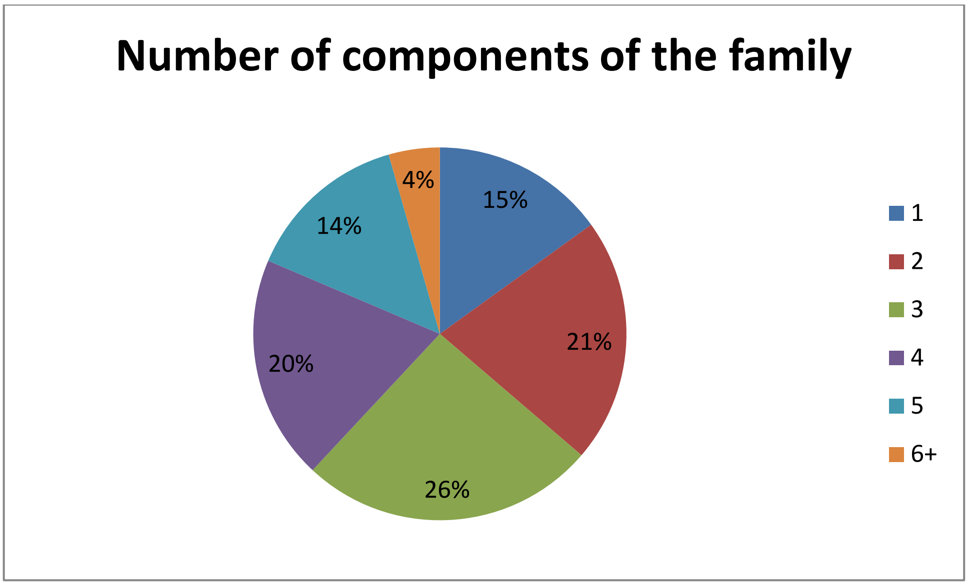


Figure 43 - Family composition of the responders.

The number of components is quite variable in the responders, with most families composed by 2-4 members.

c) What is the highest school qualification of the components of your family?

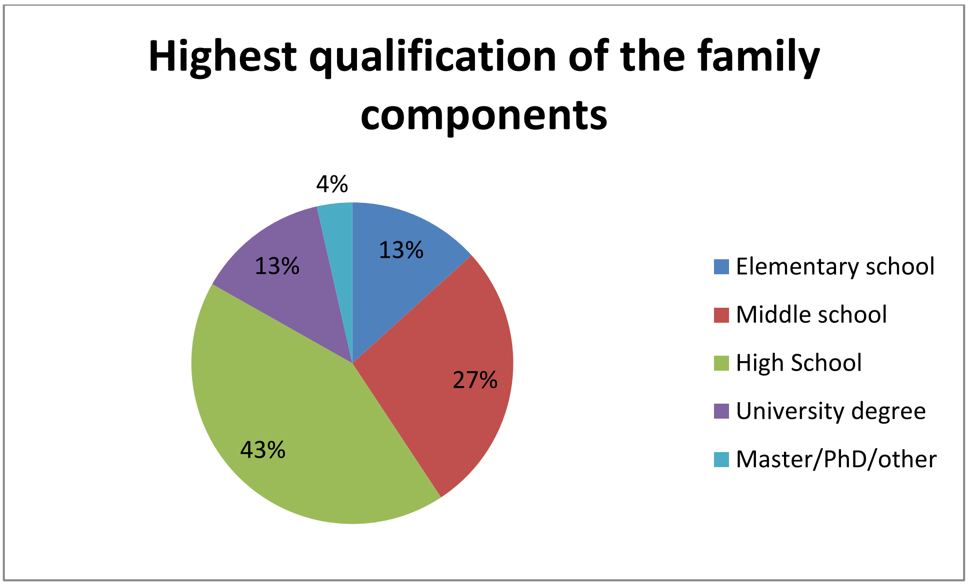


Figure 44 - Education level of the family.

Most families have a low-middle education level, with no university degree (in more than one fourth of the cases, not even high school degree). Some of them have only elementary school degree (most of them in the "older" group of the interviewed population).

d) How old are you? (considering the interviewed person the reference person of the family)

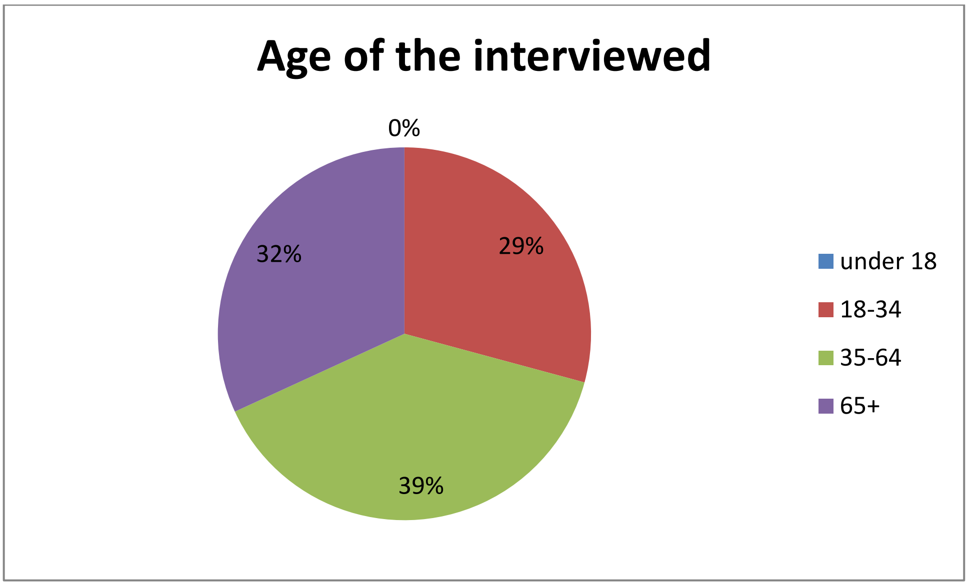


Figure 45 - Age of the interviewed reference person.

Most answers came from middle-aged people, while younger ones represented less than one third and old people around one third. It has to be noticed that most of the people in the social housing fell into these last categories.

e) What is the surface of your house?

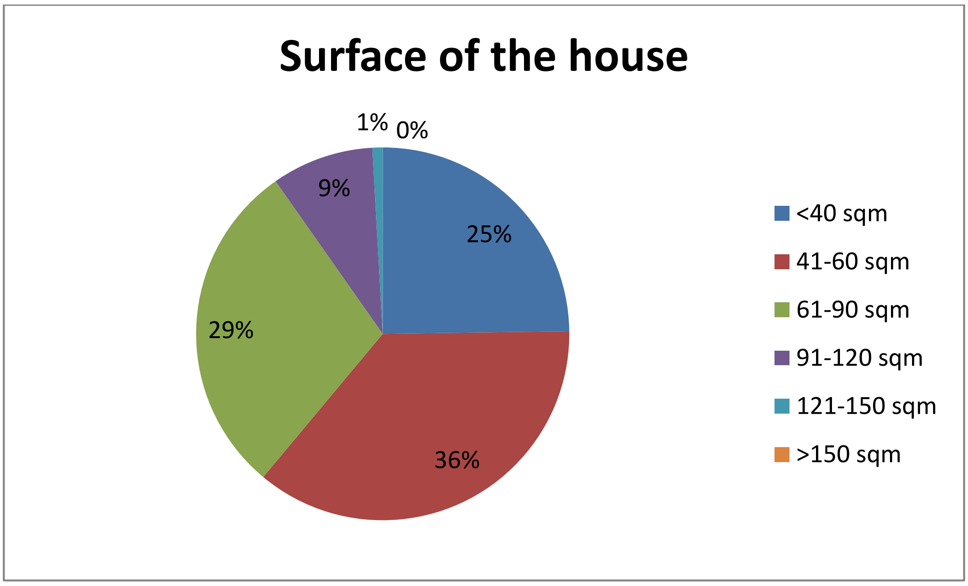


Figure 46 - Surface of the house.

Most houses are small or medium-small, with only one exception above 120 square meters. The smallest houses are the ones in the social housing context.

f) Where do you live?

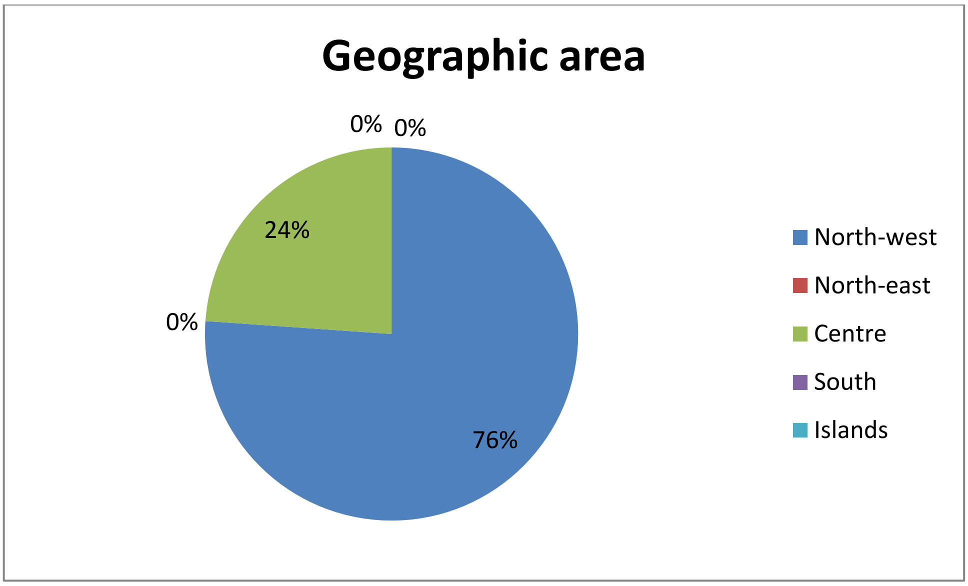


Figure 47 - Geographic area of the house.

As anticipated, most people live in the North-West of Italy (Milano area) and Central Italy (Roma area). This is due to the possibility of spreading the questionnaire, that were given by the consumers associations and social housing cooperatives contacted by the project partners.

7.3 Energy Consultancy preferences

In this paragraph, the main questions and results about energy consultancy preferences are reported, together with some considerations about how the HEAs should interpret these results.

1. “Have you ever heard of programs by professionals in the energy field (such as dedicated web sites, newsletters, social accounts,…) who provide you with external help to optimize your energy consumption?”

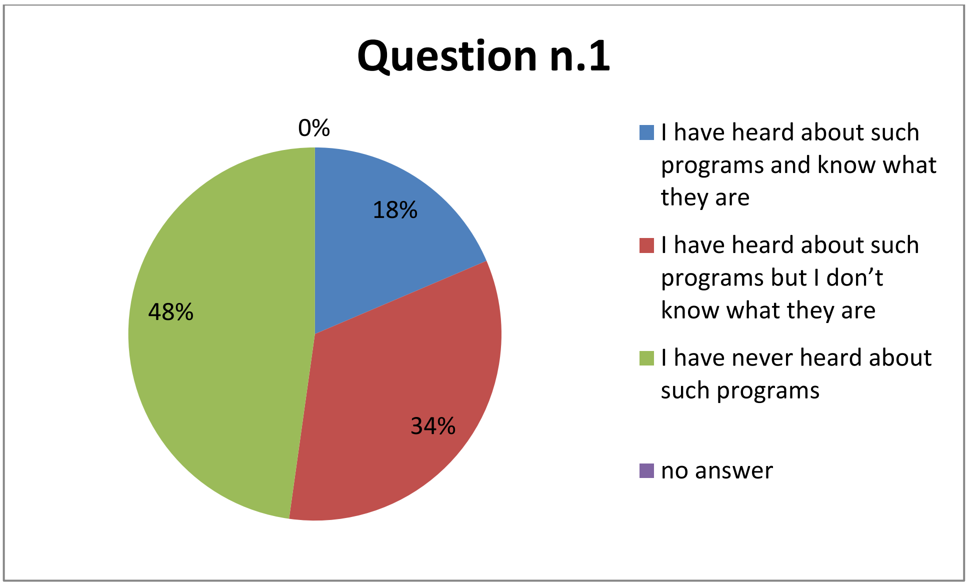


Figure 48 - Answers to question n.1.

Most of the interviewed consumers are not aware about the existence of energy consultancy services, or are not able to understand or know what they are offering. This means that the first step of the action performed by the HEA shall be to properly communicate with the target groups and to increase their awareness about energy saving programmes.

2. "Based on our knowledge about energy, do you think you would understand enough about the suggested actions to improve and optimize you energy consumption by the above mentioned programs?"

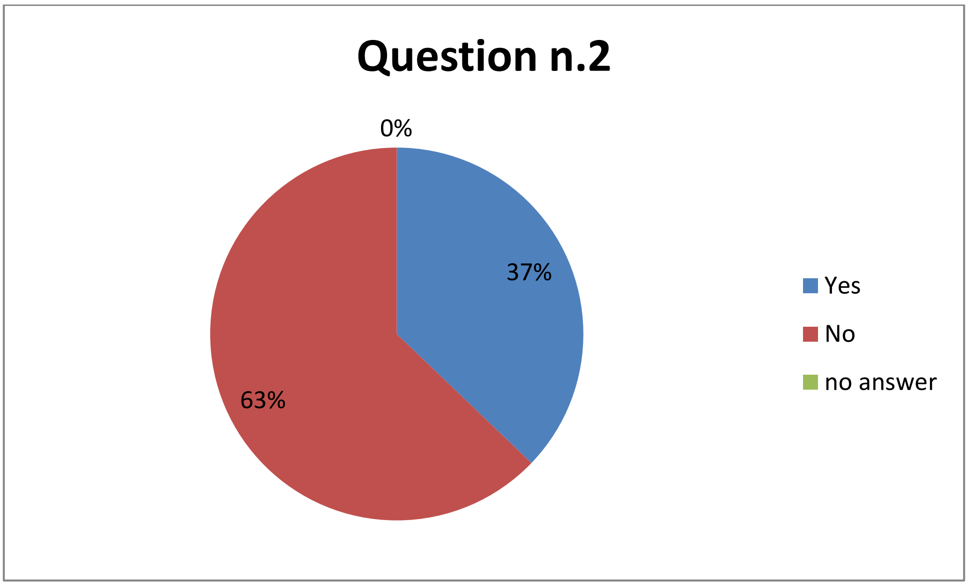


Figure 49 - Answers to question n.2.

Most people feel uncomfortable with their knowledge about energy sector and feel to be unable to understand the information provided by energy consultancy programs. It has to be remarked that the HEA will have to focus on their communication skills and to make the information they are passing as simple and complete as possible.

3. "Would you be willing to participate for free in programs which help you to optimize your energy consumption?"

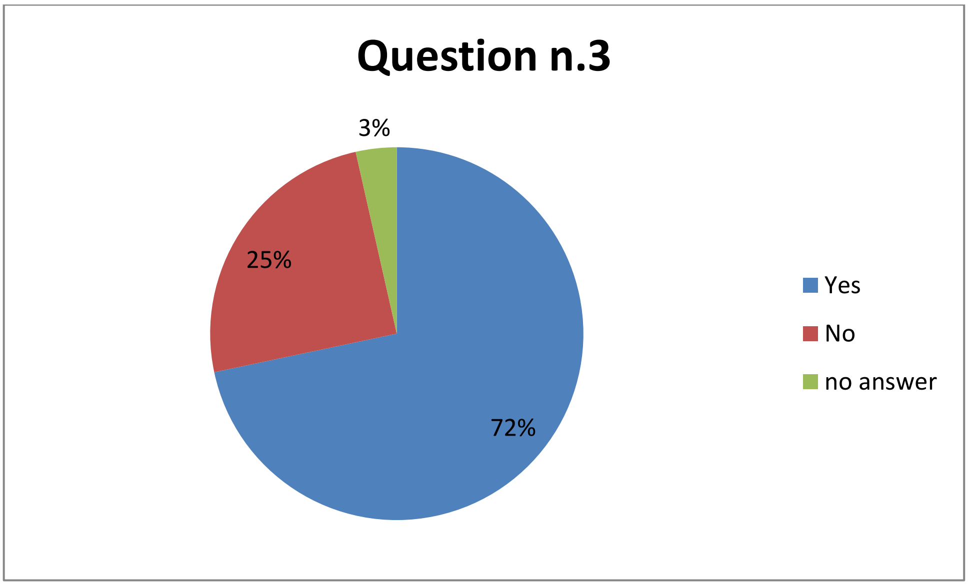


Figure 50 - Answers to question n.3.

Around one fourth of the interviewed persons seems not interested in participating to energy saving programs. In order to understand why, a further question has been proposed to those who answered "no": 3.1"If no, why?"
The answers to this question have been grouped as per the following diagram.

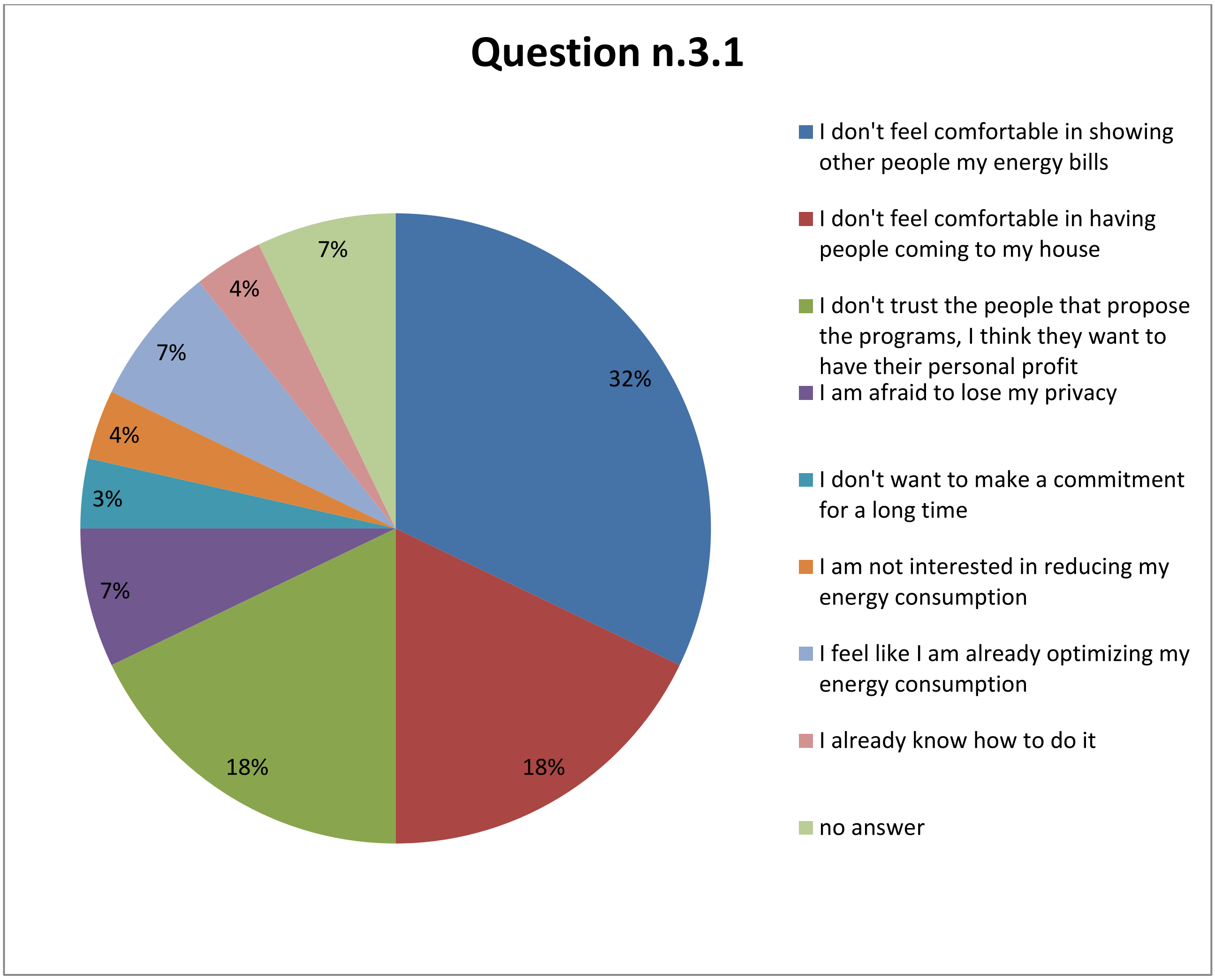


Figure 51 - Answers to question n.3.1.

Most people are afraid of losing their privacy rights (around 57% of the interviewed) or have trust issues (around 18% of the interviewed).
This means that the HEA, with the most reluctant people, will have to be persuasive and earn their trust before proposing any privacy-related action (e.g. audit in the house or help in reading the energy bills).

4. If yes, who would you like to manage them?

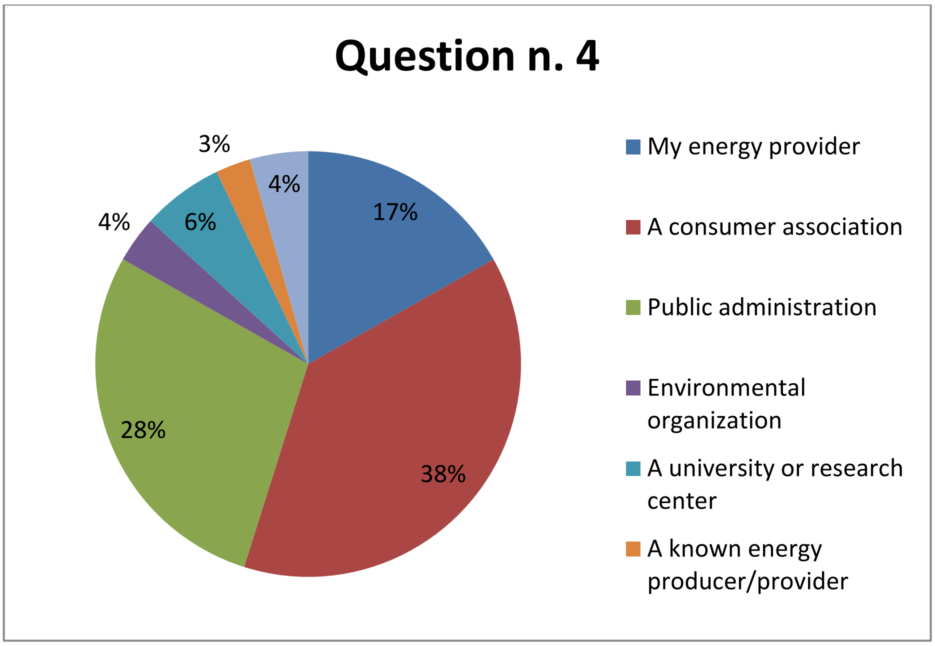


Figure 52 - Answers to question n.4.

Most people would feel comfortable if a “third party” is managing the energy saving programs, while most of them are not confident in having a research center/university managing it or an environmental association. Around one sixth (17%) feels comfortable in having their own energy provider managing such a program (probably related to a trust relationship with them) while very few people would not like another energy producer to handle it. Based on these results, it appears that the HEAs shall be viewed as a third party representatives, with no links with energy providers.

5. Would you be willing to receive visits at your home by professionals during such programs to discuss your energy consumption and try to find a way to improve it?

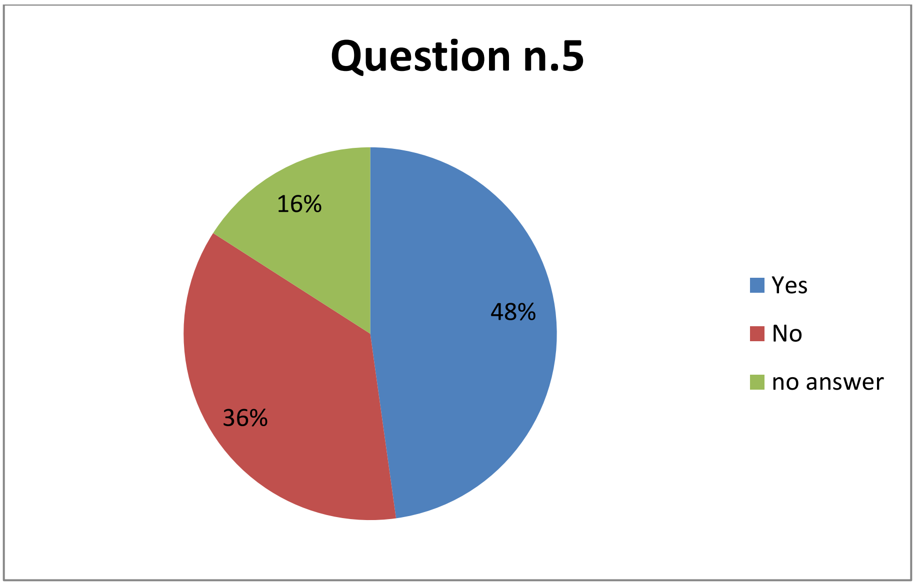


Figure 53 - Results from question n.5.

From this question, the already highlighted trust issue appears: 36% of the interviewed are not willing to let an energy consultant inside their house, while 16% did not answer this question, probably indicating an unwillingness to give access to their house as well.

6. What are your favourite channels through which participate and interact in such programs to help you to optimize your energy consumption?

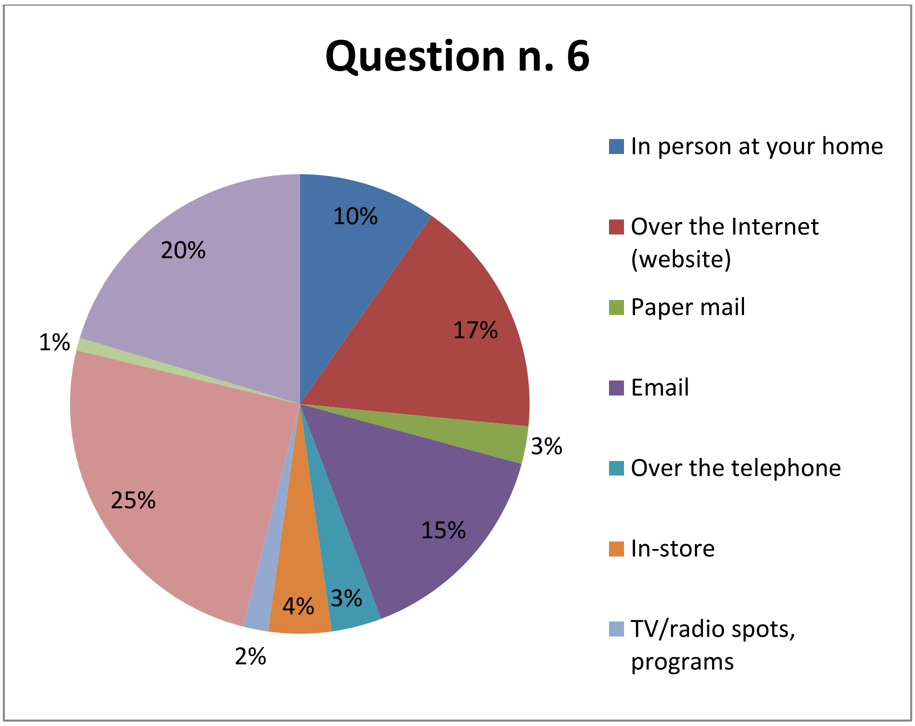


Figure 54 - Answers to question n.6.

From this question, the HEAs should take advice on how to communicate with their target groups. In particular, it has to be noted that the preferred method would be an in-home energy consumption display, that can be directly read and used by the family. Around 32% of the interviewed consumers are available to use telematics methods to receive the information (internet or e-mails), while paper mail are mostly not used. Only 10% of the interviewed is available to receive information in person at their home as a preferred communication channel. Again, it has to be noted that around one fifth of the interviewed did not answer this question.

7. When is it most convenient time for you to learn more about the actions to optimize your energy consumption?

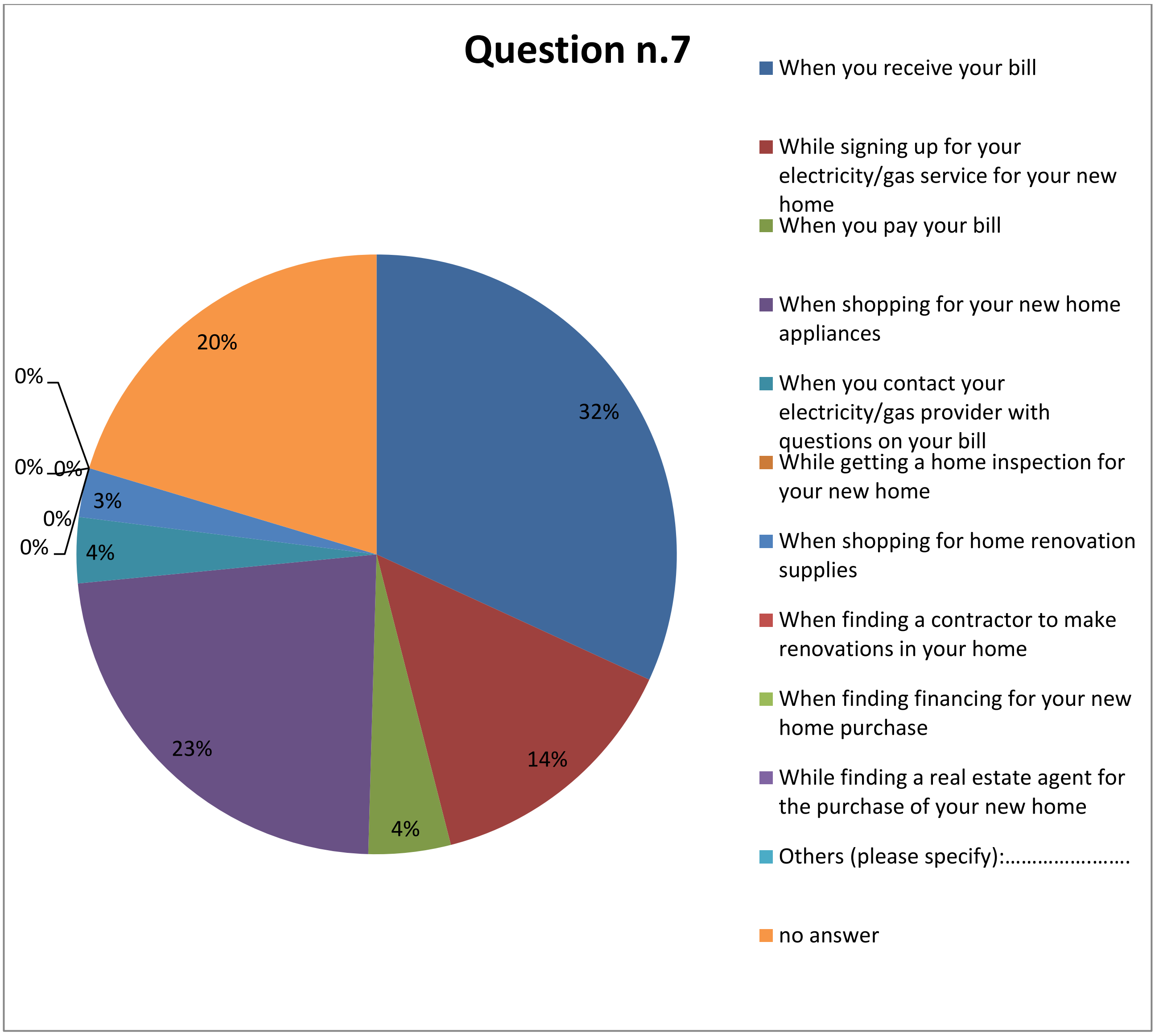


Figure 55 - Answers to question n.7.

Most of the interviewed persons prefers to receive the information through their energy bills or when shopping for home appliances (e.g. appliances energy performance tags). Around one fifth of the interviewed did not answer this question. The HEAs should take into account this information, showing the importance to interact with energy providers to work in the direction of displaying the energy consumption information inside the energy bills. Moreover, they should be able to provide advice when the vulnerable consumers ask for support in reading their bills and understanding how they consume.

8. Which factors would most discourage you from joining such programs to help you to optimize your energy consumption (3 answers max)?

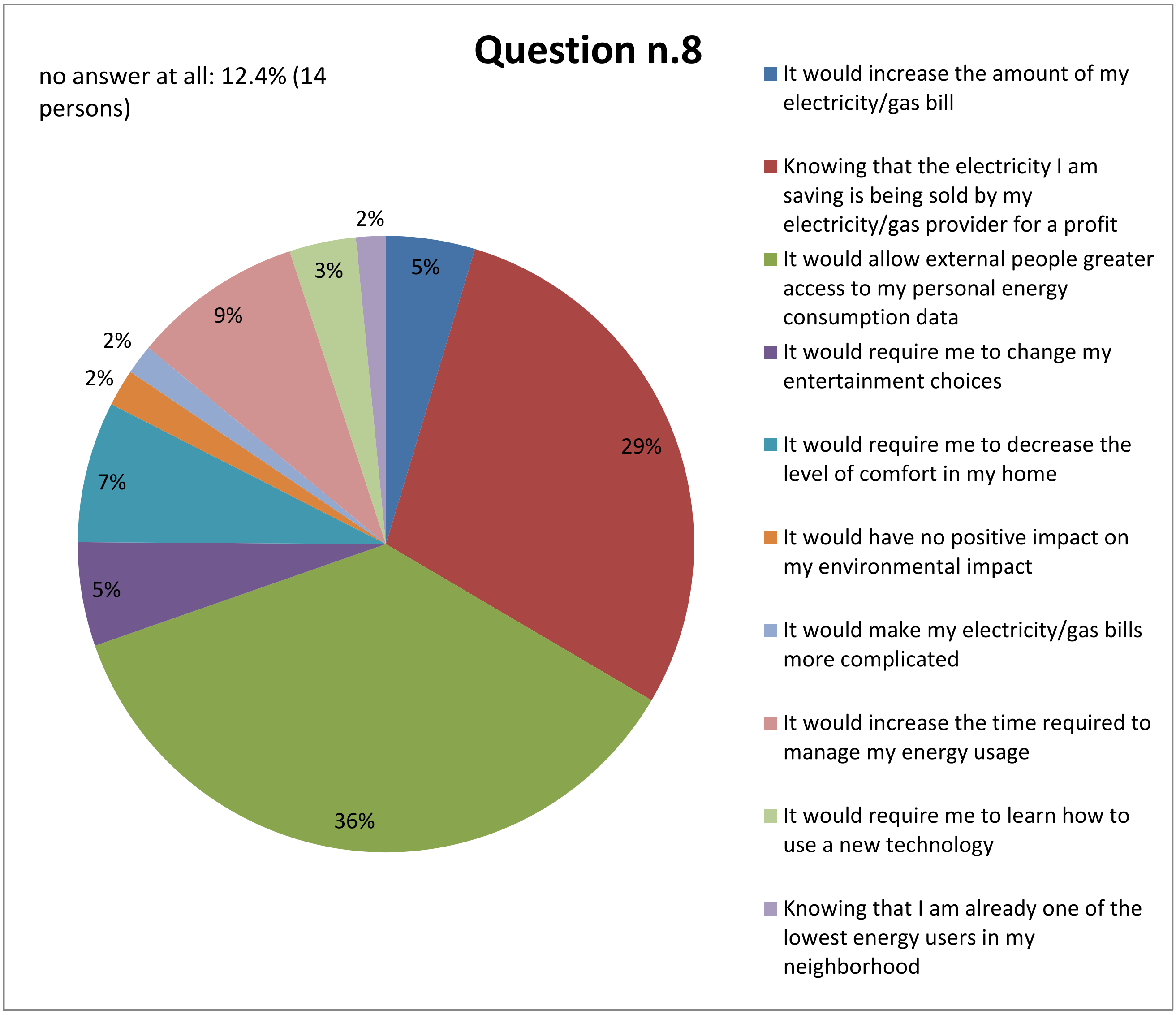


Figure 56 - Answers to question n.8.

From these answers, it is clear that the first barrier the HEAs will encounter are, again, trust issues and fear of sharing personal data with external people. Moreover, there are some that are not confident in their energy provider and a relevant percentage that are afraid to lose time/change habits/learn how to use energy in a different way. In this case, the amount of “no answer” is reduced when compared to the previous case.

These results suggest that HEAs, after earning the trust of the vulnerable consumers, will have to suggest solutions that each consumer can easily implement, without feeling to lose time or being afraid of changing their habits or their comfort in a too stressful way.

9. Which factors would most encourage you to join such programs to help you to optimize your energy consumption (3 answers max)?

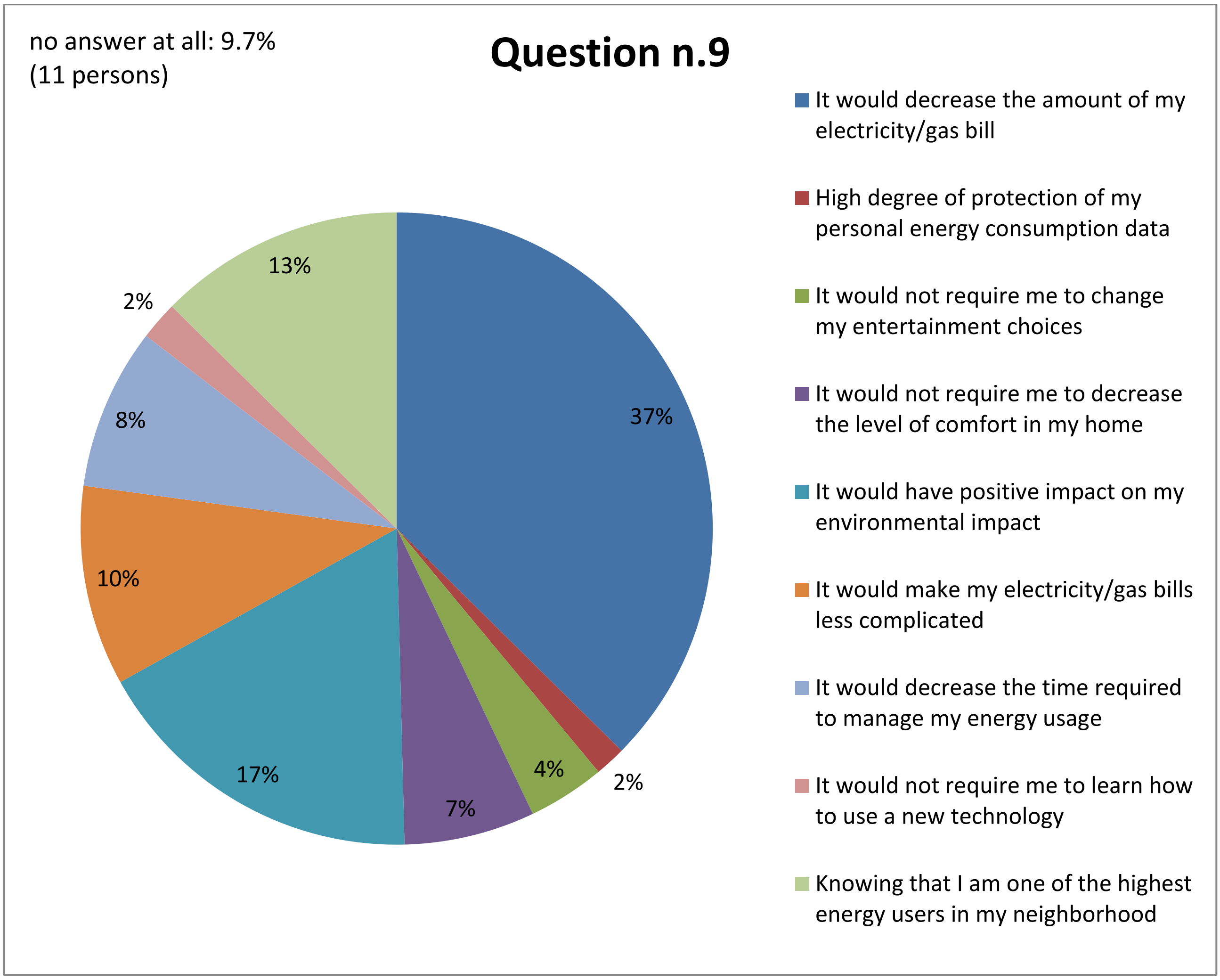


Figure 57 - Answers to question n.9.

From these results, it appears that the greatest benefit the users see in energy optimization programs would be related to economic savings and to environmental consciousness. Moreover, there is a high influence of the comparison with the neighbours, that could be a driver for the success of the proposed actions. Also in this case, fortunately, a lower number of people did not to answer (with respect to the previous questions).

10. Which would be the most valuable output you expect to receive from such programs to help you to optimize your energy consumption?

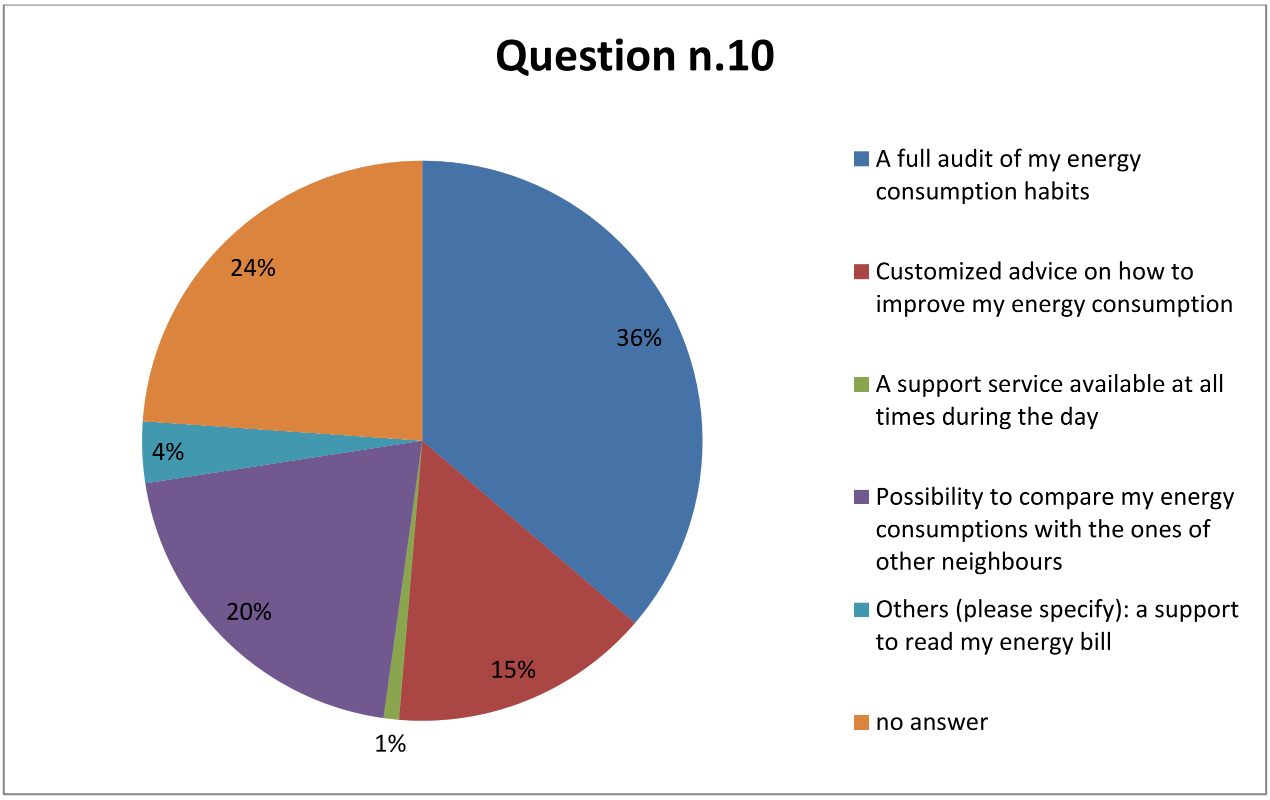


Figure 58 - Answers to question n.10.

In this case, people that are interested in the action (around 75% of the interviewed consumers decided to answer) would like to understand how they use energy. A significant percentage is interested by the comparison with neighbours, while around one sixth would like a customized advice on how to reduce their consumption. The few answers related to the “Others” option could be all grouped in the issue of reading energy bills.

7.4 Conclusions for the questionnaire

The interviewed groups are quite respondant to the national average, with some exception (e.g. the location of the household, that is influenced by the availability of the cooperatives/associations).

From the questionnaire, it appears that several people could be interested in receiving energy advice. However, several trust issues appear to be scaring the consumers, which range from giving access to their households to external people to privacy issues related to spreading their energy bills. The most interesting result is that, however, many of them would be interested in getting energy audits or customized advice.

The detailed planning of the ASSIST actions and the further obtained results will be reported in deliverable D5.2.

8. Poland

8.1 Methodology

In the Polish public debate, there is a lot of talk about the energy poverty. However, this problem remains quite marginal due to the lack of complex public intervention in that field. There is no clear definition of the energy poverty in Poland, however many institutions showed the need for the creation of such definition (ref.[1] and [2]). In 2013 the Energy Law Act was amended and included the definitions of vurnelable electicity and gas conusmers. Moreover, there are studies analysing the scale of the problem that through the statystical analysis show the groups that can be afected by the problem.

8.1.1 Direct impact

To improve the impact of the ASSIST Action we provide detailed information on various actions, aimed at mitigating the phenomena of the energy poverty, that can be identified at:

- national level,
- local level.

A significant number of actions, particularly at the local level, is dedicated to the improvement of air quality and indirectly contribute to the elimination of the energy poverty. At first look, no one expects promising results from chaotic actions, but somehow they seem to work out well.

As there is little hard data on the results of individual action finally, we show that even though most of the activities are not well coordinated in the end, thanks to the synergy of the uptaken actions, the number of people touched by the energy phenomena is decreasing. Also, the socio-economic problems have to be taken into account as over the years the unemployment rate has been falling significantly, and as a result, more people are capable of affording the energy/fuel costs. The presented means for fighting with the energy poverty highlight diversity of this problem which in lions share touch families whose households are in the poor technical conditions which lead to the increased energy demand for heating.

8.1.2 Statistical evaluation

The studies aiming to evaluate the energy povetry in Poland use many different methodologies. The different approaches for the estimation were shown in the study supported by KAPE (ref. [3]) in which the amount of energy poor consumers reached between 18% and 10% (LIHC) or even between 45% and 32% (10% or

13% of household expenses) of the whole population in 2013 depending on the methodology used. The main differences in approach result from the need to adapt the methodologies on the international level or used in other European countries with more expirience in targeting energy poverty to Polish conditions and needs.
The first presented methodology based on the expiriences from Great Britain uses the 10% expenses threshold to estimate the number of energy poor consumers. Each consumer whose expenses for the energy excceds the 10% of total household expenses is considered as energy poor, however the expenses for this calculation are hypothetical to include also those consumer that limit their use of energy.
The second approach is a modyfication of the first methodology that increases the threshold to 13% to the better fit the methodology to the way the statystical data is gathered in Poland.
Other methotology was the use of "Low Income High Costs" LIHC indicator to estimate the energy poverty. This factor estimates the household as energy poor when its model equivalent energy expenditure exceed the median in Poland and it reaches poverty limit which is set below 60% of median of equivalent Income After Housing Costs (IAHC) after subtracting the energy expenditure. The report presents also calculations for this factor when the poverty limit is set in different ways:

- Below 50% of average income;
- As set in the Polish Social Assistance Act (ref. [4])

The LIHC factor was also presented in the raport with a slighly modyfied calculation approach when the High Costs cryterium does not use the equivalent energy expenditure, but the expenditure per 1 m^2 of the dwelling.
The report was based on the Household Budget Survey (ref. [5]) from 2013 and the estimation of model energy expenditure prepared by experts from KAPE.
A statistical evaluation of the energy poverty in Poland is also described in one study presented by IBS in January 2018 (ref. [6]). The report aims to propose a statystical method of estimation of energy poverty in Poland, along with proposing some definitions of energy poverty and related concepts.
In the report the general definition of energy poverty is proposed. The definition states that: The households face energy poverty when they have difficutlies in satisfying their energy needs due to low income or bad characteristics of their dwelling. The authors of the report propose the adapted indicator LIHC (Wysokie Koszty - Niskie Dochody WK-ND) to estimate the energy poverty. This indicator is tailorefd to polish needs and constitutes of 3 main parts:

- hypothetical, equivalent energy expenditure of the dwelling (this factor aims to establish the potential energy needs rather than the current expenditure as many energy poor consumers tend to limit their expenditure even if this affects their comfort of living)

- equivalent income of the dwelling
- equivalent income AHC of the dwelling (the income per person lowered by taking the hosing costs into account as the owners usually cannot influence significantly those costs)

The indicator includes both the electricity and heating use.
The figure showing the use of LIHC indicator is presented below.

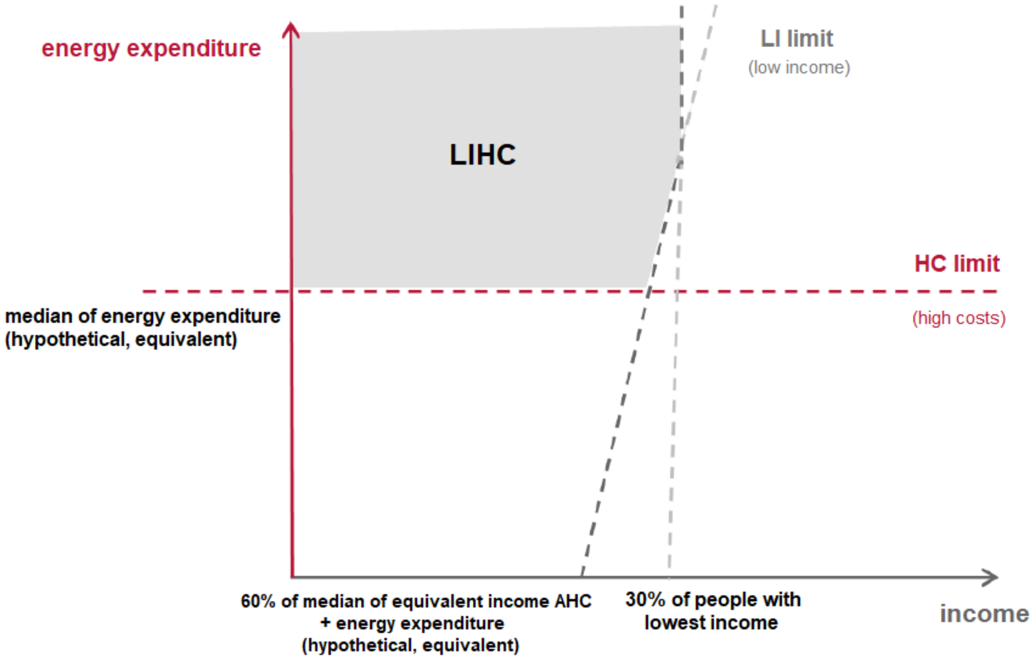


Figure 59 - LIHC indicator.

The data that was used for the study constitued of the analysis of approximetly 37 thousand households and was presented in the Household Budget Surveys (ref. [5]). Additionally the results of the survey about energy carriers used for household heating purposes supported the analysis of the data. However the sample of the results of the survey was relatively small, thus the results could not be used to establish the LIHC indicator.
The differences in methodologies for statistical evaluation of energy poverty result in relatively big spread of quntitative results of people affected by energy poverty, however many similarities can be found in the groups that can be indentyfied as the most affected by energy poverty in each study.

Nevetheless, the statystical analyses present only the general overview and direct field actions need to be taken for the precise indication of the energy poor consumers.

8.2 Cluster analysis of the population

Each of the described means for fighting with the energy poverty, even indirectly, was described in a way that gives a brief outlook of how does it work. However, due to the lack of precise data, they have somewhat descriptive character.

8.2.1 National level

At the national level three forms of financial support can be identified:

- housing allowances and energy supplements,
- thermos-modernization bonus,
- special purpose allowance.

These three forms of support are dedicated to various beneficiaries which has been presented below.

Housing allowances and energy supplements and special purpose allowance

The destitute residents can benefit from the housing allowance – a subsidy that partly covers accommodation-related expenses (i.e., energy expenses).

The potential beneficiaries must meet specific criteria to apply for the subsidy. Among them, the most important one is associated with the household income. Anyone whose income per person is lower than 175% (for the one-person household) or 125% (for the multiple person households) of the lowest level of pension can benefit from this form of the financial support.

Energy-related expenses can be subsidized by the special purpose allowance which is a form of support for those in the miserable financial condition. This allowance can be used to cover any most basic existence needs.

This form of support is available to those who meet the following criteria:

- the applicant must proof his entitelment to the dwelling,
- the applicant must be the one who is responsible for the household,
- the dwelling of the applicant must meet the surface criteria mentioned in the Act of Housing Allowances,
- the applicant must have a low income according to the Act of Housing Allowances

There are two additional criteria, over these which come out of the housing allowances, the potential beneficiaries must meet to apply for the energy supplement:

- having a contract for the energy supply with an energy company,
- occupying the area where the energy is delivered as a part of the contract.

Both forms of the financial support, i.e., housing allowances, and energy supplement are of the moderate use for the households. Unluckily the provided amount of money is quite low, and the administrative procedures might be discouraging for the potential beneficiaries who are often not familiar with the paperwork.

Thermo-modernization bonus

One of the possible ways to fight with the energy poverty is by increasing the building energy efficiency. That will lead to the reduced energy demand of the building and therefore fewer expenses on the energy bills.

The challenge is that most of the population touched by the phenomena of energy poverty doesn't manage to spend a significant amount of money to increase the building energy efficiency and as a result limit the thermal energy loss. Partial financing for energy efficiency investments can be received as a thermo-modernization bonus which is awarded by the Bank Gospodarstwa Krajowego (BGK) as a part of the Thermo-modernization and Renovation Fund. The establishment of this found was related to the successive agreement of governments and political parties on the necessity for improving the energy efficiency standards for industrial installations, public sector, and households.

A financial subsidy granted as a bonus is equal to 20% of the loan dedicated to the thermo-modernization investments, but can't exceed:

- 16% of the thermo-modernization cost.
- double amount of money foreseen to be saved thanks to the energy savings according to the energy audit.

The thermo-modernization bonus is particularly eligible for:

- legal entities (e.g., housing co-operatives or companies),
- local government units,
- housing associations,
- individuals.

Such financial subsidy is accessible to the wide range of investment type. On the other hand, the necessity to take a loan to benefit from the bonus might be a limiting factor for the energy poor who may be more reluctant to make investments.

8.2.2 Local level

The initiatives elaborated at the local level are complementary to these at the national one. Moreover, at the lower level, the support mechanism can be adapted to the local characteristic and potential beneficiares.

Low Emmission Reduction Plan

Nowadays many municipalities face problems related to the air pollution. To counteract this negative phenomenon is a complex task and Low Emission Reduction Plan (PONE) is a structured form that may help to improve the air quality.

PONE is a personalized plan for each municipality as the scale and characteristic of the pollution problem varies over the country. Cracow which is one the most polluted cities in Poland as a part of PONE offers a subsidy that can be spent on:

- upgrade of the boiler,
- connection to the district heating,
- installation of renewable energy sources.

To potential beneficiary of the PONE must be entity not related to the public finance sector, in particular:

- natural person,
- housing communities,
- legal person,
- entrepreneurs.

Also, the public finance sector entities being a municipal or county legal entities.

The PONE gives an opportunity to establish legal actions along with the Local Fund for Environmental Protection and Water Management. This has been done in Opole and there several ways of financing for natural persons and entrepreneurs who belong to the SMEs (small and medium-sized enterprises). The loans can be obtained for thermo-modernization of the buildings, replacement of the heating source or implementation of RES (renewable energy sources). Companies and legal entities are not entitled to be subsidized.

Local Shelter Programme

Another means to fight with the air pollution, and therefore energy poverty is a subsidy for those who decide to switch from fossil fuel-based heat source to less polluting energy source like gas, electricity, oil or district heating. That kind of programmes is established locally, i.e., in the city of Cracow. The beneficiaries receive a partial refund of their heating bills.

The potential recipients are required to participate in the Low Emission Reduction Plan (PONE) and must meet the same criteria as for the PONE.

Boiler Modernisation Programme

Another example of the financial mechanism aimed at reducing the local low emission is the Boiler Modernisation Programme established the city of Warsaw. The Warsaw residents can receive a subsidy for performing a switch from coal-fueled boilers to gas or oil boilers or district heating. As the programme is local, the applicant must be a Warsaw resident or a private investor or company.

Boiler Exchange Programme

The Local Fund for Environmental Protection and Water Management (WFOŚiGW) in Wrocław offers a programme which aims at reducing the air pollution by exchanging old solid-fueled heat sources for more efficient ones.

To benefit from the programme the applicant must be:

- local government unit or is association from the Lower Silesian Voivodship,
- final beneficiaries of the programme: natural person, including natural person conducting economic activity in the dwelling, housing communities whose members benefit from the heat generated in a shared boiler room, local government units related to the municipal housing construction.

8.2.3 General statistics

The lack of standardised definition and methods for calculating energy poverty results in different estimations of the phenomenon in Poland in different studies presented. The differences in estimation approaches are significant as the amount of people that could be described as energy poor ranges between 17.2 million and 3.8 million for the same reference year (2013).

The calculations using the LIHC methods presented in two studies show smaller differences, however still significant, as the amount of energy poor consumers ranges between 3.8 to 6.7 million.

Table 34 - Energy poverty estimation in Poland in 2013.

Low Income criterion	High Costs criterion - equivalent energy expenditure		High Costs criterion - energy expenditure per m^2	
	% of the population	number of people	% of the population	number of people
60% median	17.1	6.4	17.9	6.7
50% average	16.4	6.2	17.1	6.4
Poverty line	10	3.8	10.8	4.1

Similar numbers are presented in the study by IBS (ref. [6]), which uses a different approach for the calculation of the LIHC indicator.

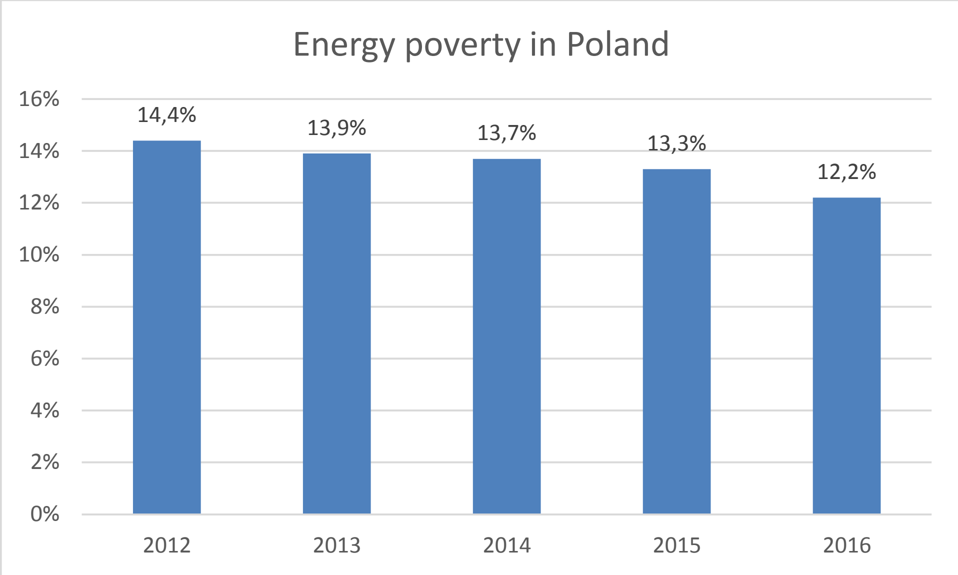


Figure 60 - Energy poverty estimation in Poland.

The difference in calculation procedure changes also the mostly affected type of consumers by the energy poverty, however some general clusters can be specified.

Table 35 - Type of households affected by energy poverty in Poland.

	Energy poverty LIHC			Alternative LIHC		
	60% median	50% average	Poverty Line	60% median	50% average	Poverty Line
Poland	17.1	16.4	10	17.9	17.1	10.8
Type of dwellers						
Couple with no children	10.5	10	5.4	10.2	9.8	5.2
Couple with 1 child	12.4	11.9	7.2	13.4	12.9	7.9
Couple with 2 or more children	18.5	17.8	12.6	21.1	20	14.9
Single parent with children	14.4	13.8	10.9	22.9	22	18.3
Parents, children and other dwellers	22.4	21.5	11.6	22	20.9	11.8
Single person	13.7	12.9	12.5	15.7	14.8	13.3
Other	18.4	17.5	8.3	17.7	16.6	7.8

Number of people in households						
One person	13.7	12.9	12.5	15.7	14.8	13.3
Two people	12	11.4	6.5	12.6	11.9	6.6
Three people	14.9	14.2	7.8	15.4	14.7	8.3
Four people	18.7	18	10.7	18.7	17.8	11.1
Five or more people	21.9	21	12.7	23.4	22.1	14.5
Number of children in households						
No children	17.3	16.5	9	16.9	16.1	8.7
One child	15	14.3	8.7	16.7	15.9	10.1
Two children	17.9	17.2	12.4	19.9	19	13.9
Three children	19.9	18.9	13.4	25.3	23.3	19.3
Four children	22.4	21	19	23.6	23.2	22.4
Five or more children	26	24.4	26	26.4	24.2	30.5
Socio-economic groups						
Farmers and self-employed	26.7	25.9	17.7	20.1	19.5	13.1
Blue-collar workers	19	18.2	9.9	22	20.8	12.4
White-collar workers	7.6	7.2	3.7	7.8	7.2	4
Retirement pensioners	15.5	14.7	8.5	15.4	14.5	8.4
Disability pensioners	29.1	27.5	20.3	33.4	31.7	23.5
Living on social benefits	23.8	22.6	19.2	37.2	36.7	30.8
Living on other non-wage benefits	14.6	13.7	11.6	25.2	24.3	18.9

General trend can be noticed that the more dwellers live in the apartment the more chance that they suffer from energy poverty. Similar situation can be noticed in terms of number of children living in the apartment. The smallest share of energy poor households is amongst couples with no children. As per socio-economic groups the highest share of disability pensioners, farmers and self-employed and living on social benefits is suspected to suffer from energy poverty. This situation is similar to what is presented in other reports as the most commonly named groups that could suffer from energy poverty are: farmers, people living on social benefits and disability pensioners (ref. [6]).

Table 36 - Share of people affected by energy poverty in Poland by city/town size and voivodship.

	Energy poverty LIHC			Alternative LIHC		
	60% median	50% average	Poverty Line	60% median	50% average	Poverty Line
Poland	17.1	16.4	10.0	17.9	17.1	10.8
City/town size						
500 thous. or more inhabitants	3.3	3.1	1.3	7.1	6.4	3.8
200-499 thous. inhabitants	5.7	5.5	2.7	11.1	10.4	5.9
100-199 thous. inhabitants	4.7	4.6	2.6	10.4	10.0	6.4
20-99 thous. inhabitants	7.2	6.7	4.1	12.2	11.3	7.2
under 20 thous. inhabitants	14.2	13.6	8.1	17.1	16.1	10.1
village	32.0	30.7	19.1	27.0	25.9	16.6
Voivodship						
Dolnośląskie	8.7	8.1	4.6	11.9	11	6.4
Kujawsko-pomorskie	19.4	18.6	11.4	22.4	21.6	13.8
Lubelskie	29.1	28.5	19.5	26.3	25.4	17
Lubuskie	14.1	13.5	6.8	14.6	13.7	7.4
Łódzkie	17,2	16,1	10	19,1	17,8	11,3

Małopolskie	23,1	21,3	13,5	20,4	19,2	12,3
Mazowieckie	15,3	14,7	9,5	16,5	15,6	10,6
Opolskie	18,4	17,4	11,9	17,0	16,1	11,1
Podkarpackie	28	26,9	15,8	25,8	24,4	15,0
Podlaskie	19,2	17,5	10,1	20,3	18,9	10,6
Pomorskie	12,5	12,3	7,8	14,5	14,1	9,7
Śląskie	10,9	10,6	5,7	14,2	13,8	8,1
Świętokrzyskie	24,4	23,2	13,0	22,1	20,9	12,3
Warmińsko-mazurskie	13,5	13,1	8,2	18,9	18,3	12,4
Wielkopolskie	18,5	17,7	10,4	17,5	16,7	10,3
Zachodniopomorskie	10,0	9,8	5,5	12,1	11,7	7,4

Almost 1/3 of people living in villages is suspected to suffer from energy poverty. This share is over twice as big as the next for the towns under 20 thous. inhabitants. The region that is likely to suffer from energy poverty the most is the south-east of Poland as the voivodships with over 1/5 of the inhabitants that can suffer from energy poverty are Lubelskie, Podkarpackie, Świętokrzyskie I Małopolskie.

Table 37 - Share of people affected by energy poverty in Poland by building type.

	Energy poverty LIHC			Alternative LIHC		
	60% median	50% average	Poverty Line	60% median	50% average	Poverty Line
Poland	17.1	16.4	10.0	17.9	17.1	10.8
Type of building						
a multi-family house	2,4	2,2	1,1	9,5	8,8	5,4
terraced house	16,6	15,8	9,6	15,4	14,9	9,3
detached single-family house	34,5	33,1	20,5	28,3	27,1	17,4

Construction date						
before 1946	18,3	17,3	10,9	21,3	20,2	13,6
in 1946-1960	31,9	31,2	20,0	36,6	35,5	23,2
in 1961-1980	14,1	13,4	8,1	16,2	15,1	9,4
in 1981-1995	15,7	15,1	8,3	13,1	12,6	7,0
in 1996-2006	12,7	11,9	6,9	7,2	6,5	4,0
after 2006	7,4	6,7	5,3	4,4	4,2	3,0
Living space [m^2]						
=<30	7,4	6,9	5,4	32,9	31,2	20,6
31-60	8,3	7,9	5,2	16,9	15,9	10,7
61-90	19,7	19	12	19,6	18,8	12,2
91-120	30,5	29,1	17,2	25,2	24,2	14,2
>120	21,9	20,9	11,7	7,8	7,4	4,0
Flat ownership						
private	18,8	18	11	18,1	17,2	10,8
housing association	2,5	2,5	1,4	13,7	13,1	7,4
municipality, State Treasury, employer	5,6	5,4	3,2	18,9	17,6	12,8
Towarzystwo Budownictwa Społecznego (Low Cost Social Housing)	4,1	3,5	1,8	10,1	9,5	5,5
other	9,6	9,6	8,5	18,4	15,7	7,8
no information	10,1	10,1	7,6	22,8	22,8	16,7

For the type of buildings, the detached, single family-houses are suspected to be the most affected by energy poverty as even over 1/3 of such buildings can be energy poor due to the presented calculations. Also in general the older the building the more likely its inhabitants are to be energy poor. The houses built before 1946

do not follow this estimation strictly, however the share of the inhabitants living in such buildings is still above the average for Poland. When considering the living space of the house, the differences in calculation approach are significant. For this factor it is important to define the calculation method so the supporting policies would properly address the issue of houses with large living space where multi-generational families used to live and due to the migration tendencies, only few inhabitants remained causing them difficulties in sufficient heating of their living space.
In general, the statistical analyses of the most vulnerable consumers indicate similar clusters of people that face energy poverty as the organizations supporting the fight with energy poverty (ref. [7] and [8]). Those specific groups can be defined as:

- people living in detached, single family-houses or in old buildings with bad energy characteristic;
- people living in villages or in the islands of poverty;
- people with difficult financial status especially pensioners.

8.3 Vulnerable consumers characterization

The statistical approach provides estimation of the scale and general picture of the problem, however the direct information from energy poor consumers should provide the most precise data and definition of the scale of the problems regarding energy poverty. This is way the surveys targeted the energy poor consumer were carried out.
In Poland the survey was conducted among 215 consumers from Warsaw and surrounding regions. Most of the responders receive social benefits and maintain in contact with social welfare centers, which helps to target the most vulnerable consumers with Assis actions. Detailed information from the questioners are presented below.
The second survey was conducted to examine the consultancy preferences of consumers. The questionnaire consisted of 10 main questions and was answered by a group of 28 consumers.

9.1.1 Socio-demographic and living conditions results

The results presented in this chapter come from the questioners prepared for T2.5 which provide basic information on energy poor consumers. Five preliminary questions were related to the characteristics of the household. There were 215 responses, however some questions were omitted by some responders.

a) Which is your nationality?

All 215 responders declared their nationality as Polish.

b) How many people live in your house (including yourself)?

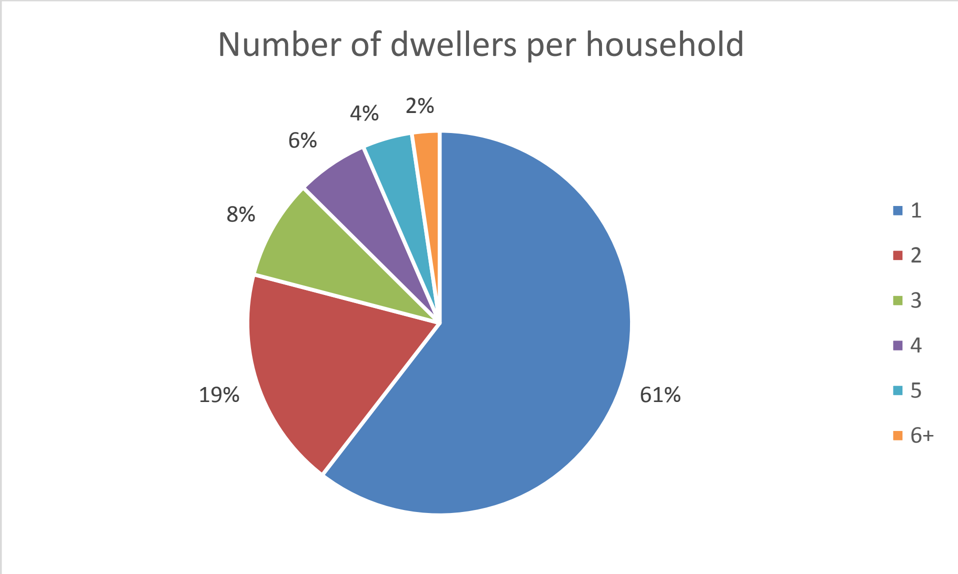


Figure 61 - Number of dwellers in the responders' households.

There was only 1 survey without the answer to this question. Among the rest the majority of responders live alone and only less than one fifth lives in the house of three or more tenants. The statistical analyses showed that people living in houses with multiple tenants are more likely to suffer from energy poverty, opposed to the results from the survey, however this fact is caused by specific large groups of vulnerable consumers answering the survey, which can be common in the city area. The more detailed analysis of these groups is presented below.

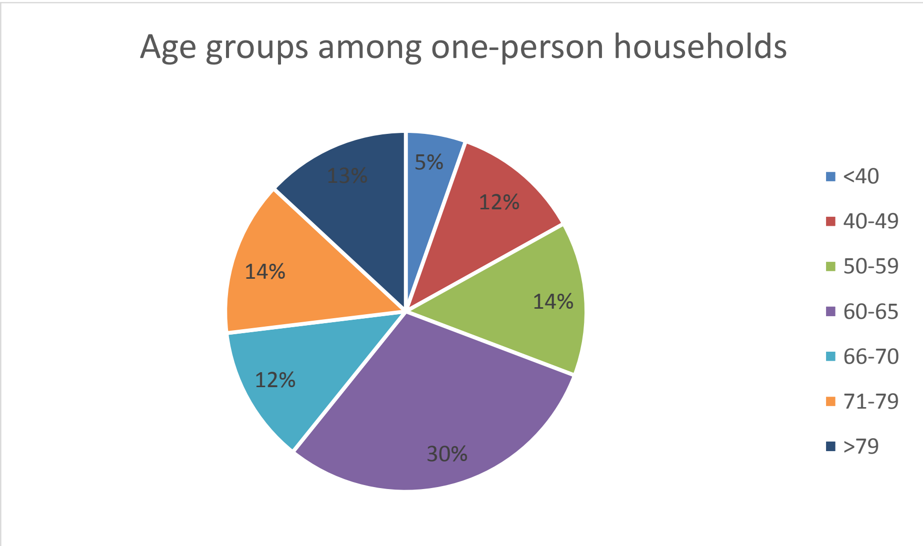


Figure 62 - Age groups of responders living alone.

The graph shows that over two thirds of people living alone are over 60 years old. This supports the fact that many people in the age for receiving a pension can be considered energy poor and many of those people live alone. This is fatherly shown on the graph below presenting the employment status of the people living alone.

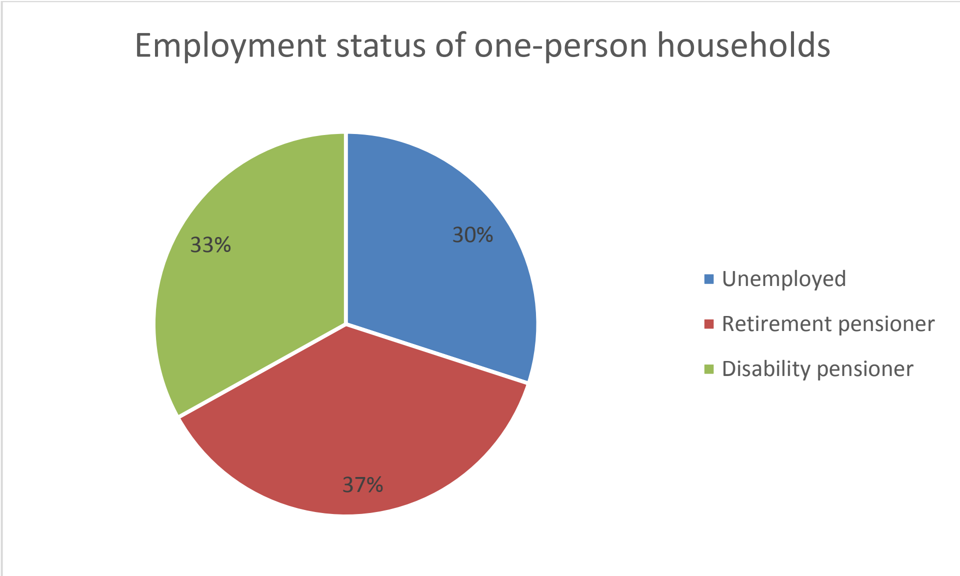


Figure 63 - The employment status of people living alone.

All people living in one-person households were either receiving some kind of pension or unemployed which shows that those common groups that can be affected by energy poverty.

c) What is the highest level of education you have completed?

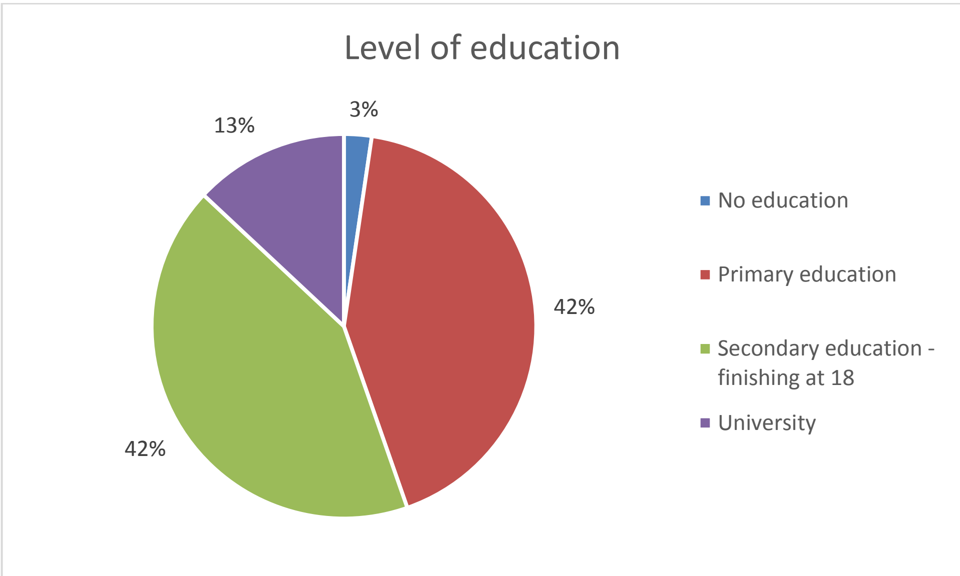


Figure 64 - Education level of the responders.

The vast majority of responders finished their education either at primary level (primary school) or at secondary level (high school, technical secondary school or vocational school). Five of the responders did not finish any of the education levels.

d) What is the surface of your house?

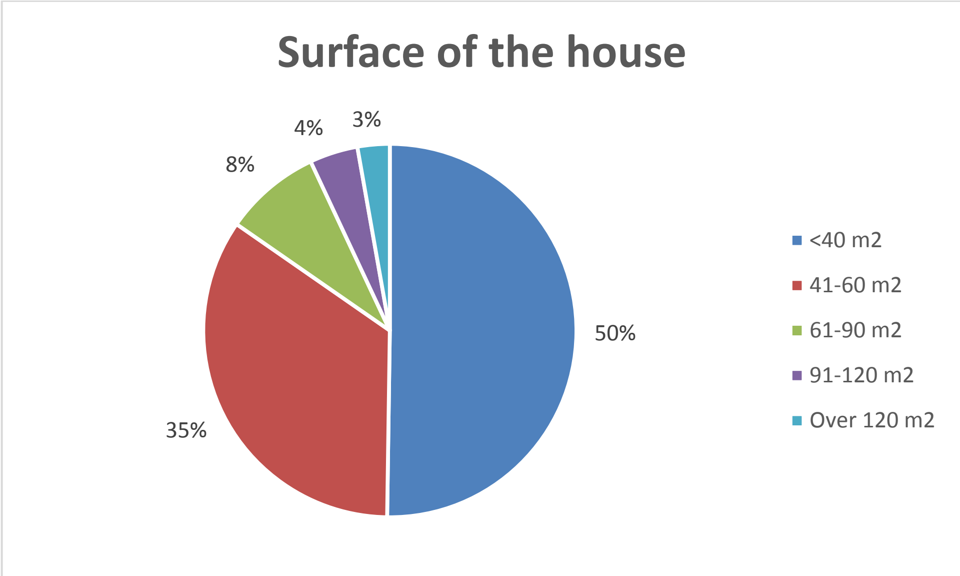


Figure 65 - Surface of the responders' house.

Half of the responders live in the flats below 40 m^2. Over one third lives in flats with surface between 41 and 60 m^2. These results correlate with the fact that most responses come from the city area.

9.1.2 Energy consultancy preferences

In this paragraph, the main questions and results about energy consultancy preferences are reported, together with some considerations about how the HEAs should interpret these results. The surveyed group for these consisted of 28 consumers.

1. "Have you ever heard of programs by professionals in the energy field (such as dedicated web sites, newsletters, social accounts,…) who provide you with external help to optimize your energy consumption?"

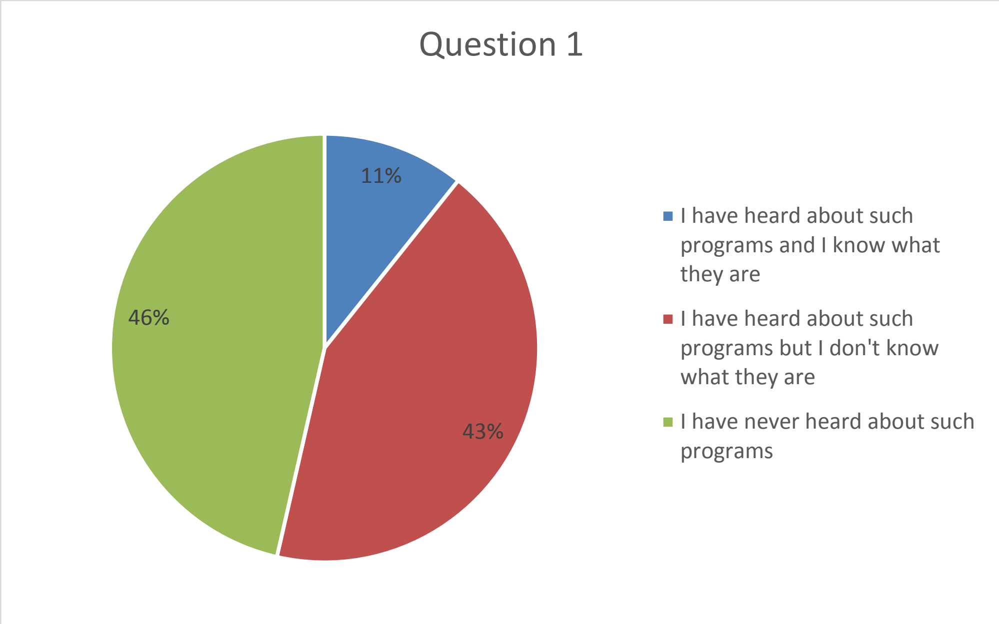


Figure 66 - Answers to question n.1.

Most of the interviewed consumers are not aware about the existence of energy consultancy services, or are not able to understand or know what they are offering. This means that the first step of the action performed by the HEA shall be to properly communicate with the target groups and to increase their awareness about energy saving programmes.

2. “Based on our knowledge about energy, do you think you would understand enough about the suggested actions to improve and optimize your energy consumption by the above mentioned programs?”

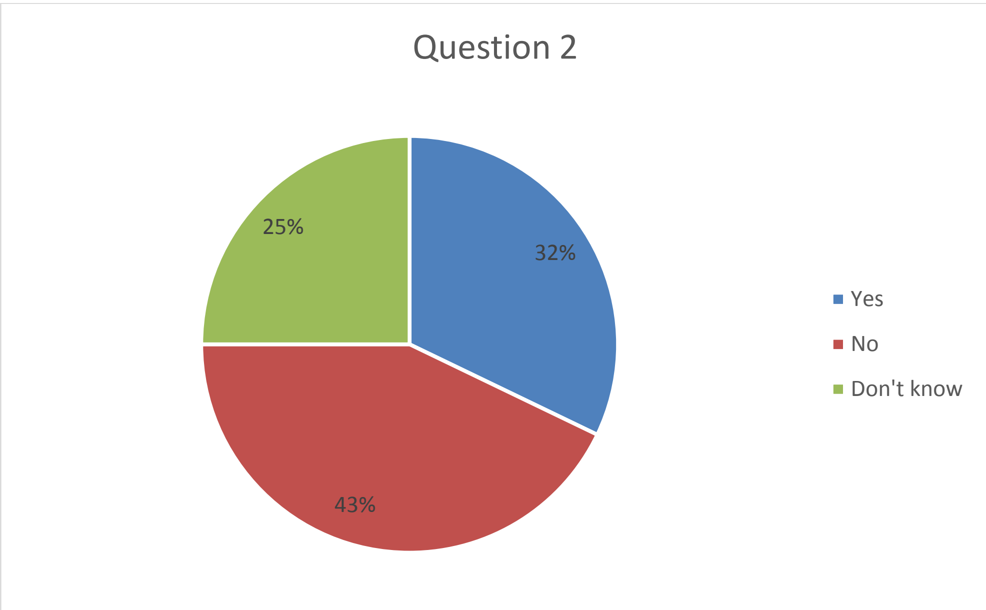


Figure 67 - Answers to question n.2.

Most people feel uncomfortable with their knowledge about energy sector and feel to be unable to understand the information provided by energy consultancy programs. It has to be remarked that the HEA will have to focus on their communication skills and to make the information they are passing as simple and complete as possible.

3. “Would you be willing to participate for free in programs which help you to optimize your energy consumption?”

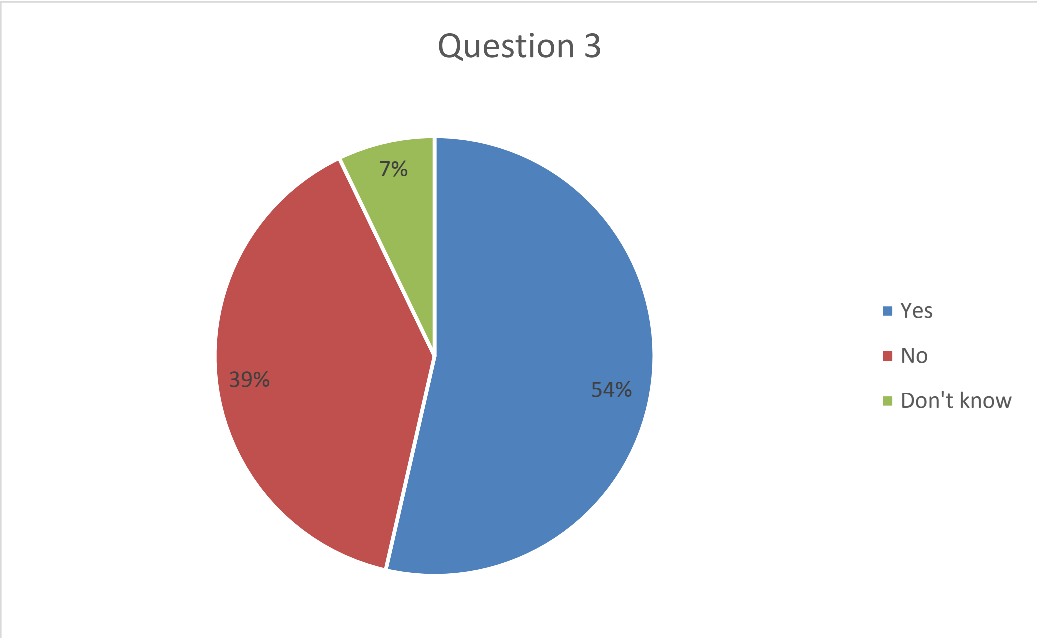


Figure 68 - Answers to question n.3.

Over one third of the interviewed persons seems not interested in participating in energy saving programs. In order to understand why, a further question has been proposed to those who answered “no” or “don’t know”: 3.1“If no or don’t know, why?”

The answers to this question have been grouped as per the following diagram (13 responses).

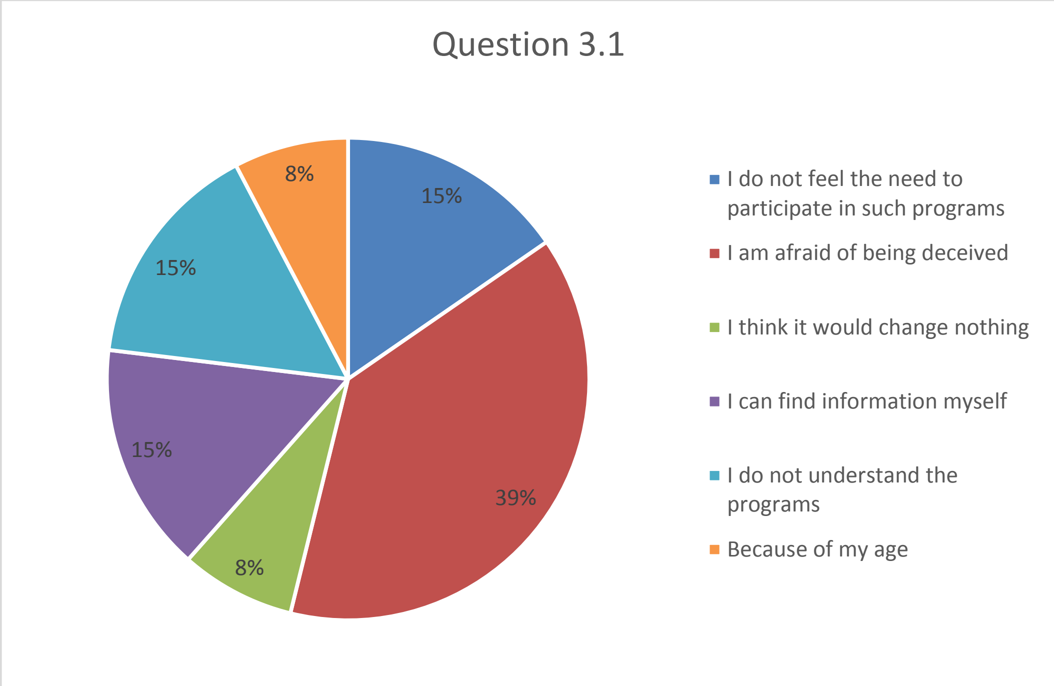


Figure 69 - Answers to question n.3.1.

Most people are afraid of being deceived (around 39% of the interviewed) the rest of the responders do not see the benefits of participating in such programs or does not now about them. One person answered that they do not participate due to their age.

This means that the HEA, with the most reluctant people, will have to be persuasive and earn their trust before proposing any privacy-related action (e.g. audit in the house or help in reading the energy bills).

4. If yes, who would you like to manage them?

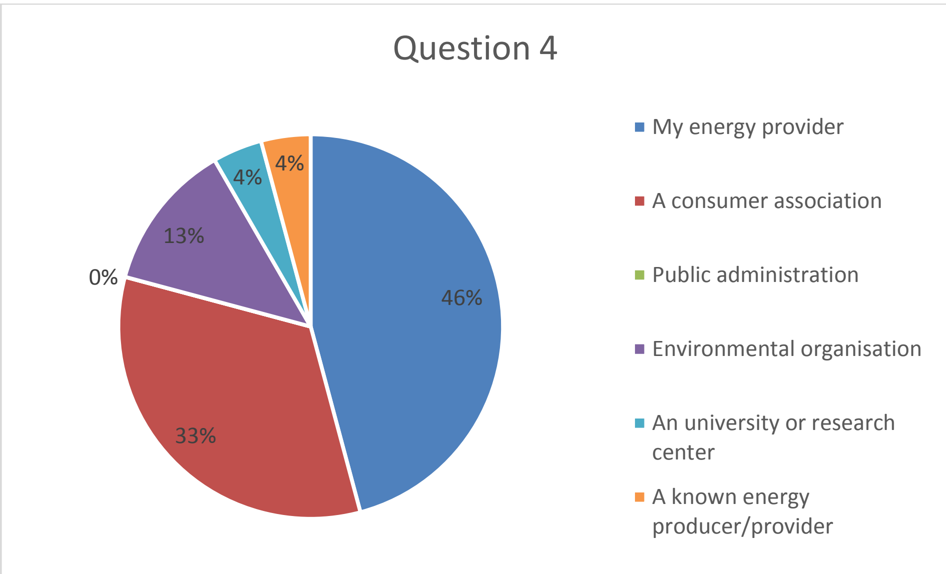


Figure 70 - Answers to question n.4 (23 responses).

Most people would feel comfortable if either their energy provider or a consumer association would manage the energy saving programs. This situation correlates with the risks mentioned in question 3.1 as the consumers association are perceived as trustworthy and the energy provider can already gather information about consumers' energy expenditure this way knows more about clients and can help with their experience as the specialist in the field.

Based on these results, it appears that the HEAs shall be viewed as trustworthy specialist and should represent a body of public trust.

5. Would you be willing to receive visits at your home by professionals during such programs to discuss your energy consumption and try to find a way to improve it?

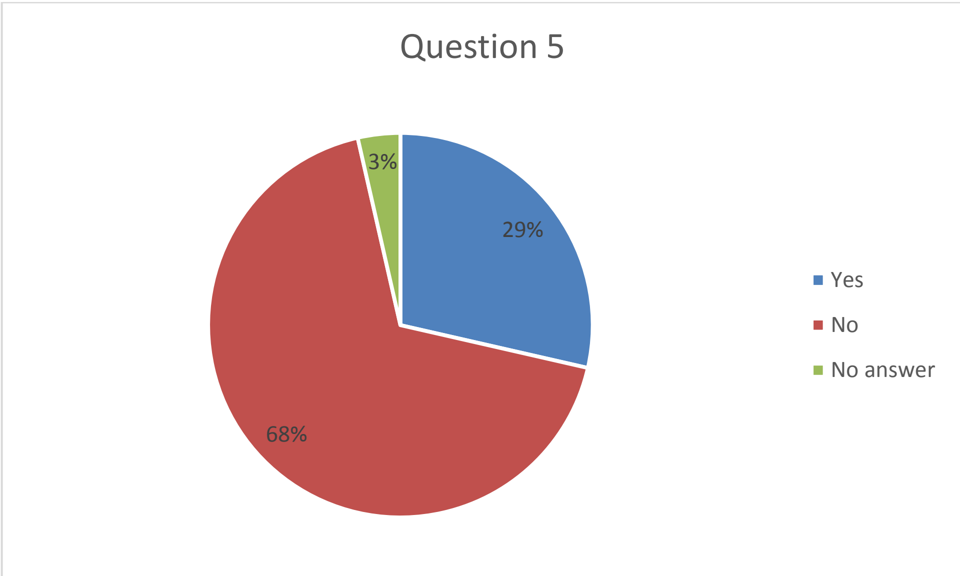


Figure 71 - Results from question n.5.

From this question, the already highlighted trust issue appears: only 29% of the interviewed are willing to let an energy consultant inside their house, while 68% indicated unwillingness to give access to their house. This can be a difficulty for the HEA to deliver sufficient help, especially when the situation of unfair market practises is significant and the consumers may be reluctant to letting anyone in.

6. What are your favourite channels through which participate and interact in such programs to help you to optimize your energy consumption?

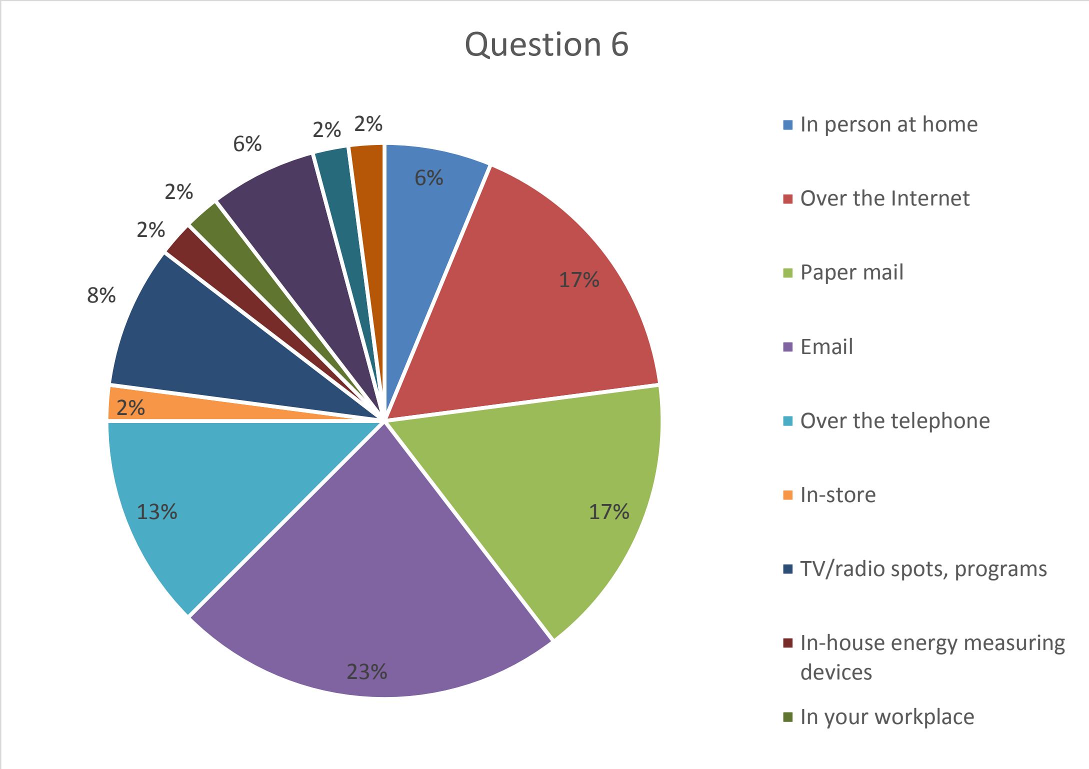


Figure 72 - Answers to question n.6.

From this question, the HEAs should take advice on how to communicate with their target groups. The interviewed could mark more than one answer for this questions and for the most common responses (over the internet, paper mail, email, over the phone) the responders chose multiple options. Only 6% of the interviewed wanted to receive information in person at their home as a preferred communication channel.

7. When is it most convenient time for you to learn more about the actions to optimize your energy consumption?

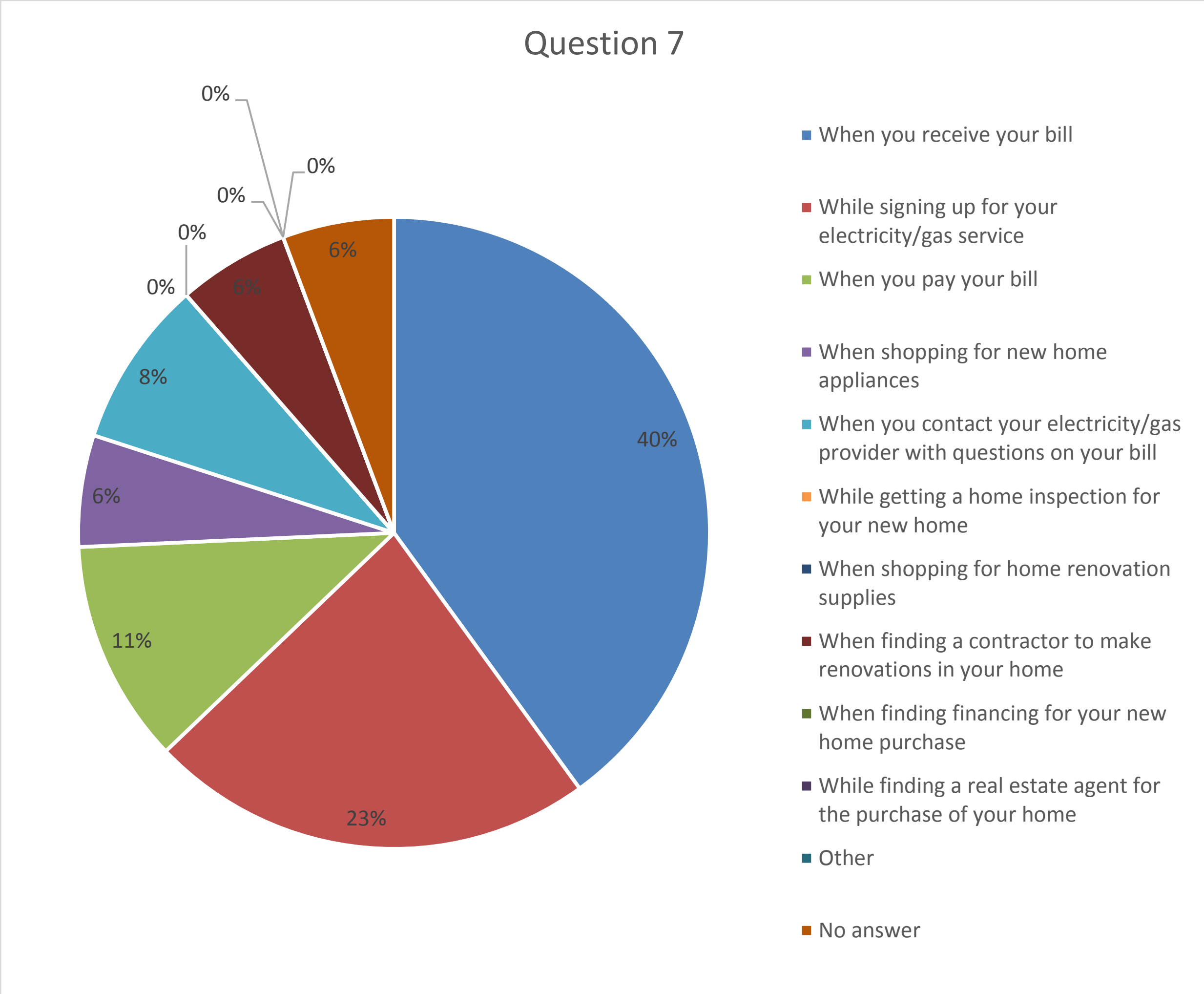


Figure 73 - Answers to question n.7.

Most of the interviewed persons prefer to receive the information along with their energy bills or when signing up for their energy service.

The HEAs should take into account this information, showing the importance to interact with energy providers to work in the direction of displaying the energy consumption information inside the energy bills. Moreover, they should be able to provide advice when the vulnerable consumers ask for support in reading their bills and understanding how they consume.

8. Which factors would most discourage you from joining such programs to help you to optimize your energy consumption (3 answers max)?

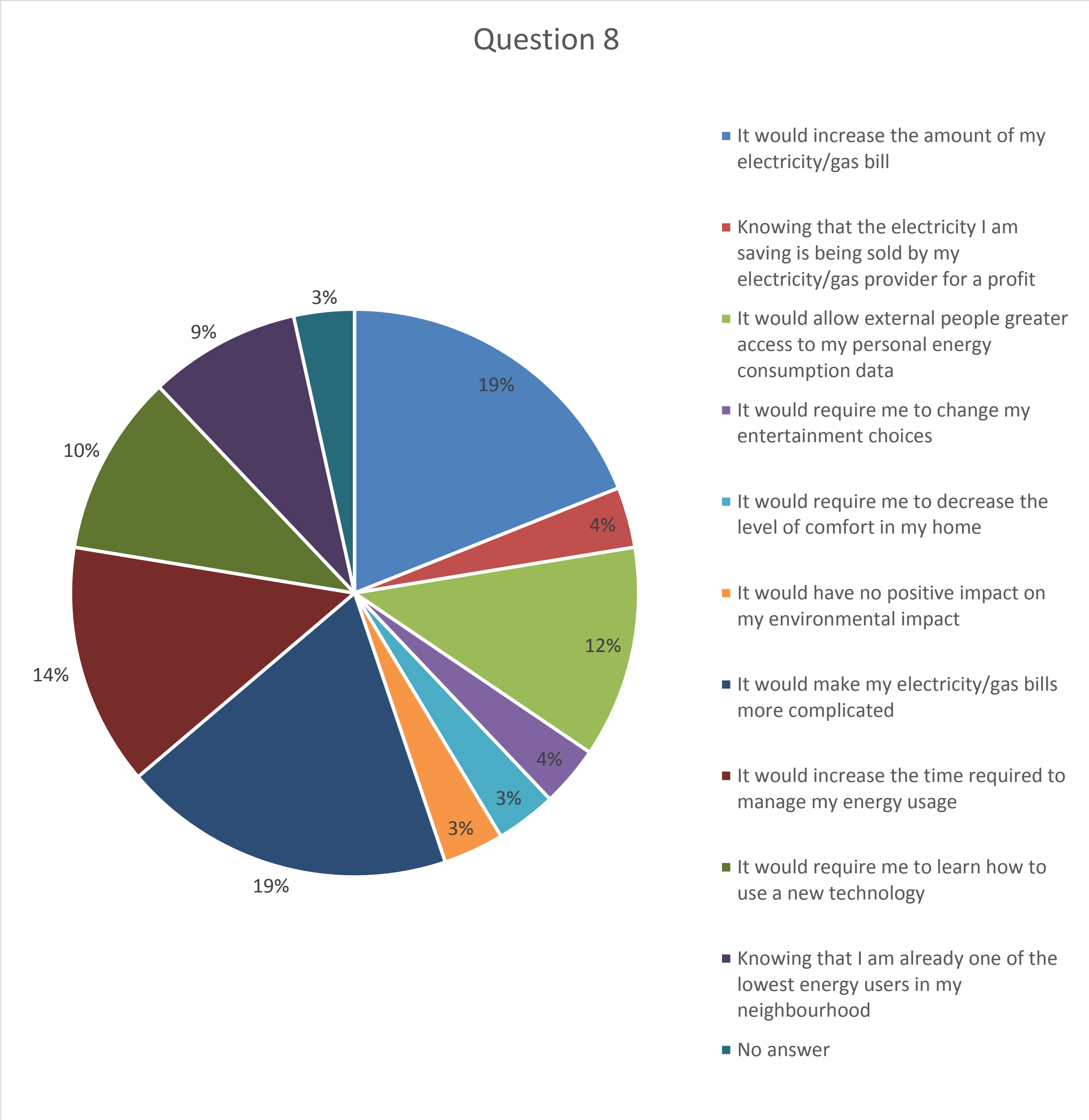


Figure 74 - Answers to question n.8.

From these answers, the concerns about the complexity of the energy bills and the time required to manage the energy use along with the amount on bills are of most importance to the consumers, however there is no singular threat that majority the respondents named thus the HEA should be prepared to confront most of them. These results suggest that HEAs, after earning the trust of the vulnerable consumers, will have to suggest solutions that each consumer can easily implement, without feeling to lose time or being afraid of the increase in bills. The proposed solutions should be easy to manage by the end users.

9. Which factors would most encourage you to join such programs to help you to optimize your energy consumption (3 answers max)?

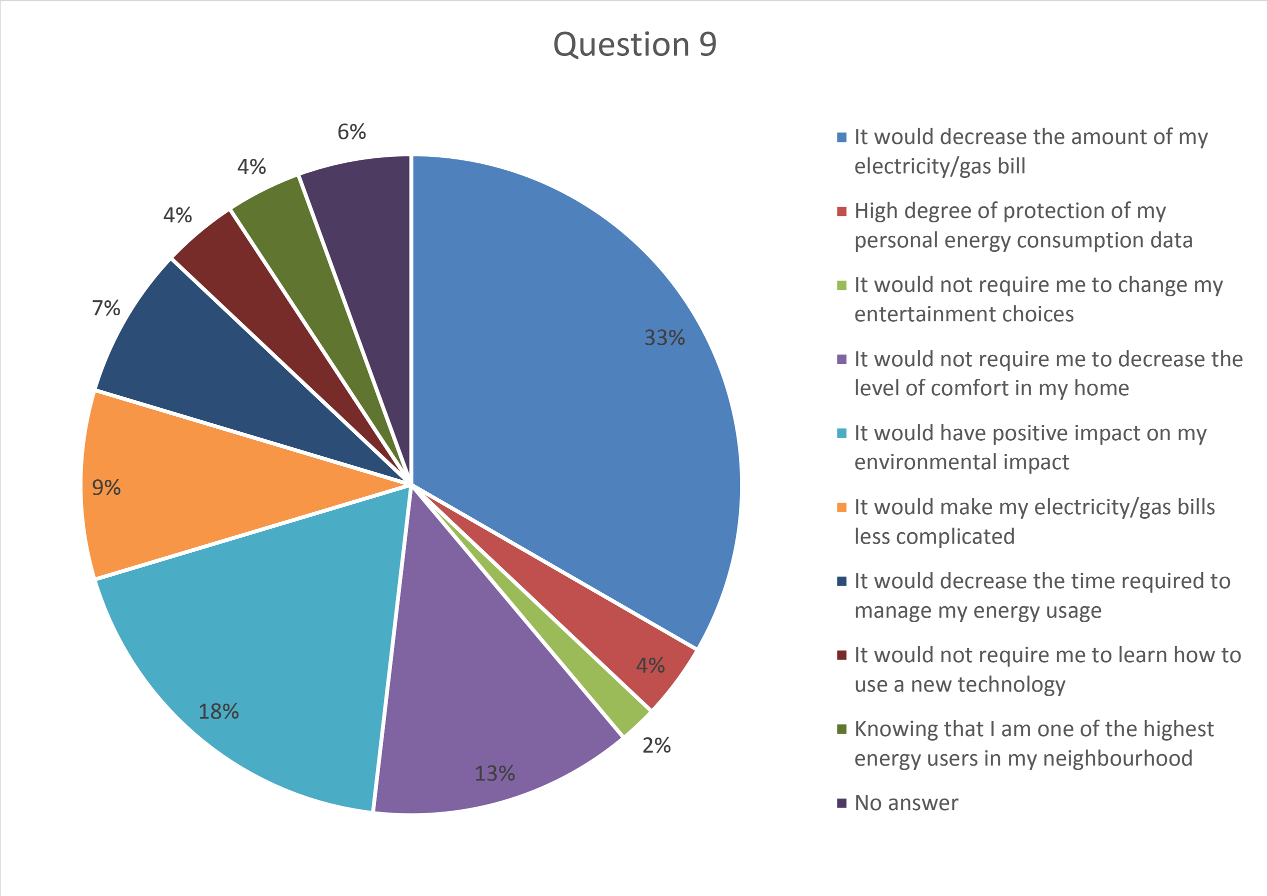


Figure 75 - Answers to question n.9.

From these results, it appears that the greatest benefit the users see in energy optimization programs would be related to economic savings and to environmental consciousness (almost one fifth marked both answers). Many of the consumers were also concerned about their comfort.

10. Which would be the most valuable output you expect to receive from such programs to help you to optimize your energy consumption?

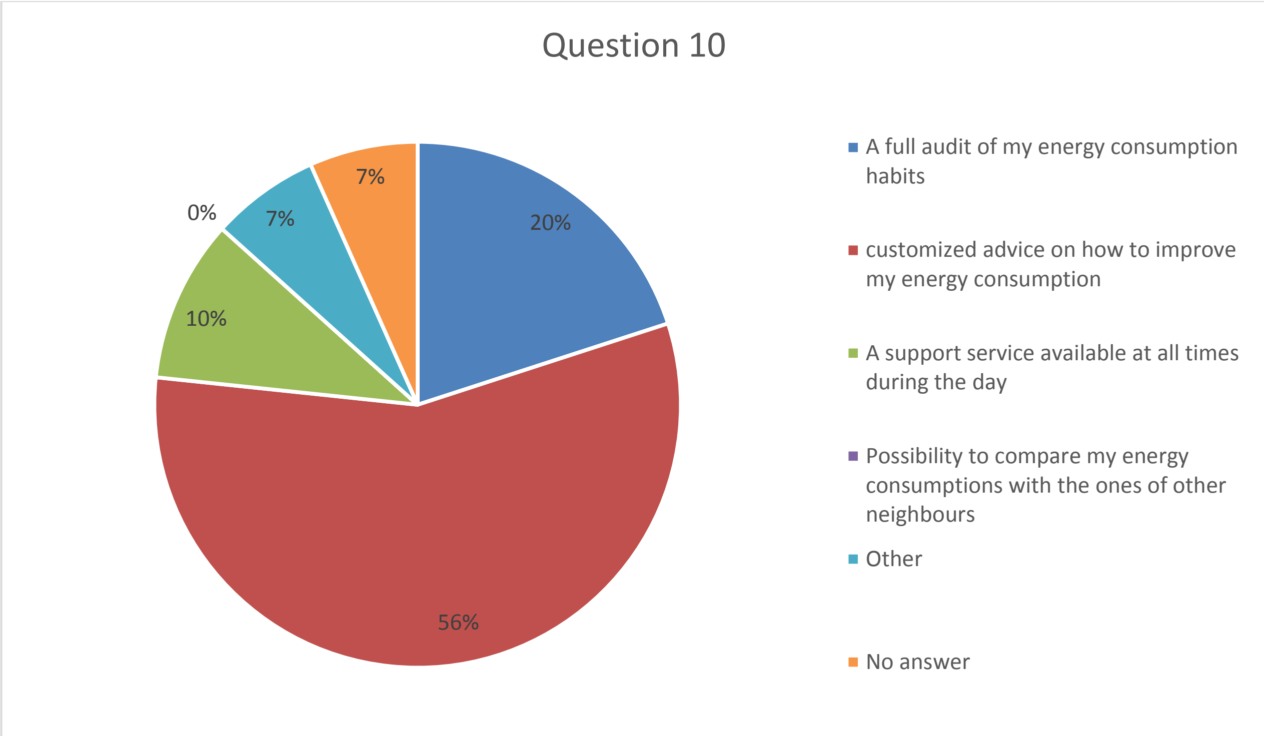


Figure 76 - Answers to question n.10.

In this case, most of the responders would prefer to receive advice on how to optimize or lower their energy. It is interesting to mention that more people chose to answer this question than declared willingness to participate in the program showing that maybe if the program supported the needs of the consumers well there may be more people willing to take part in them than declared at the beginning. The few answers related to the “Others” option showed the lack of need or even the strong reluctance against discussed programs.

8.4 Conclusions

As stated at the beginning of the chapter as for today there is no complex strategy to fight energy poverty in Poland. However, particularly at the local level, there are more and more initiatives related to air quality. These initiatives have not only environmental impact, but also help to mitigate the problem of energy poverty. The actions spread out over the whole country result in a synergy effect. As seen in the graph below, the number of people unable to keep their home decreases significantly.

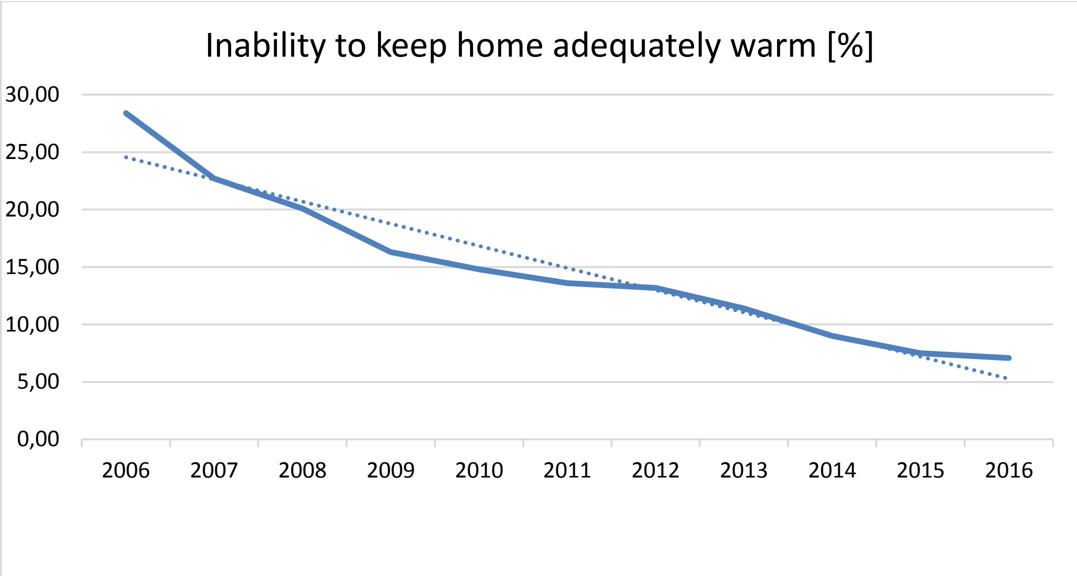


Figure 77 - Number of people unable to keep their home adequately warm.
Source: Eurostat Database

Within ten years from 2006, the rate went down from 28,4 to 7,1. This results places Poland below the EU average which in the recent years was about 9-10. The estimations of the whole energy poverty show that the total scale of the problem may be larger.

There is still a significant number of people facing the energy poverty. The most affected groups are people living in villages, possibly farmers, living in old, detached, single-family houses with poor energy characteristic. Where in cities the most affected are people from small flats, usually living alone, often old (pensioners) or receiving some social benefits. Many of the energy poor live in old buildings with bad energy characteristic.
The economic prosperity and low unemployment rate are good prognostics for further decrease or at least maintain of the present level, however still some additional supporting measures should be introduced and targeted actions should be taken in the following years.
The increasing interest of the society in the air quality and other environment related subjects and the raising interest of the policy makers in topics related the energy poverty create good bases for the ASSIST project actions, however the analysis of the questioners show that there are still many barriers for the HEA to overcome and the support of VSCS is needed to properly address them.

9. Spain

9.1 General Introduction

Nowadays in Spain, as well as in the rest of Europe, there is a high level of awareness of energy poverty being an issue of concern, especially among local and regional administrations. There is also a wide societal and institutional recognition of energy poverty as an issue of concern, which is still regarded as a direct consequence of the global financial and economic crisis on Spanish society. During the recent years there has been some remarkable advances in the local and regional administrations, in contrast with the lack of initiatives of the Spanish central government.

Energy poverty is understood as the result of a broader and less strictly defined condition referred to as energy vulnerability, described as the probability of a household to suffer from a lack of adequate energy services at home (Bouzarovski and Petrova, 2015). This new approach explains energy poverty as a dynamic condition caused by structural and short-term causes that go beyond the three main contributing factors traditionally considered (household income, energy prices and energy efficiency of the housing).

The report taken as a basis for this deliverable, tackles energy poverty in a novel way and presents an innovative analysis of results and indicators in terms of energy inequality, that allows for a comparison across groups with different economic levels.

More concretely, the official Low Income-High Costs indicator, used in the UK has been applied to the Spanish case, as well the Minimum Income Standard approach, which has been implemented assuming three different minimum income level thresholds, as there are no official estimations of this metric available for Spain.

Thanks to these methodological contributions six energy poverty indicators have been defined and presented in the report analysed.

The main objectives of the report analysed within the following sections are as follows:

- To provide updated estimates of the energy poverty indicators presented in the previous reports.
- To present new energy poverty indicators based on the income and expenditure approach, and in particular, on the Low Income-High Costs or indicator (LIHC) of the indicator based on the concept of Minimum Income Standard (MIS)

- To present disaggregated results of the different indicators used to measure the energy poverty based on location, socio-economic and housing variables.

9.2 Dataset

At this moment, there is no official methodology or specific information sources for measuring the energy poverty rates in Spain. However, there are several statistical sources freely available at the National Institute of statistics (INE) that can be used to calculate suitable indicators to estimate the intensity of the problem, their evolution over time and to produce valuable disaggregated results.

In particular, for the issue of this study analysed, two sources of key data have been used; i) micro-data from the survey on family budgets (Encuesta de Presupuestos Familiares; EPF) for the period 2006-2014 and ii) micro-data from the survey of Conditions of life (Encuesta de Condiciones de Vida, ECV) for the period 2006-2014. In addition to these data sources, the Eurostat aggregated results from the EU Survey on Income and Living Conditions (EU-SILC), of which CVD is part, have also been used to compare the incidence of indicators based on perceptions and statements from the home in Spain with other EU countries. Also the specific modules regarding the housing conditions of the ECV for 2007 and 2012 have been used to calculate the indicator of thermal comfort in summer. These modules are only available for the two years mentioned above.

In both cases (data from the CVD and from the EPF) the original microdata, i.e., the anonymous original answers of surveyed households, have been used. To obtain significant results from these sets of microdata, the results have been calculated taking into account elevation factors determined by the INE (Spanish Statistics National institute).These factors indicate the number of actual households representing each respondent home and determine the weight that each household in the sample should be taken to, for example, the calculation of an average.

Furthermore, for the analysis of the temporary factors, data from Eurostat (2007-2014) on the price of the electricity and gas have been used. In addition, data of interannual variation in gross domestic product, as well as the rate of unemployment and number of employed active population from the INE survey have been used.

9.3 Methodology for the energy poverty indicators calculation

As mentioned earlier, one of the objectives of the report analysed is to provide an estimation of the energy poverty rate in Spain (i.e. the percentage of houses or

people affected by this problem over the total population), according to different indicators.
According to the common literature, this can be done through three approaches or methodological alternatives (Healy, 2004):

- Temperature-based approach: This approach consists in measuring the internal temperature of the household order to check whether they comply with certain objective criteria (for example, falling temperatures between 18 ° C and 21 ° C) which serve to define what households are living in energy poverty. Although it has the advantage of being based on objective measures of satisfaction of the needs of a home, its application is difficult and, in practice, there are no estimates of energy poverty rates based on this approach. Another critical problem of this approach is that it only assesses energy poverty from the point of view of air conditioning, aside from many other household uses of energy (supply of domestic hot water, lighting, cooking food, entertainment and communication, etc.)
- Energy expenditure and household income-based approach: Developed initially in the United Kingdom, this approach has the advantage of considering all the uses of the domestic energy. The best known example of this approach is the official indicator of the United Kingdom between 2001 and 2012, based on the original proposal of Boardman (1991), according to which a home was in energy poverty if it had to spend 10% of your net income (approximately twice the median spending in energy per household in the United Kingdom at the time that this indicator was proposed) to maintain an adequate thermal comfort level. After the re-evaluation carried out by the so-called report Hills (Hills, 2012), this measurement methodology was replaced by the so-called indicator low income - high costs (LIHC), which currently is the official definition of energy poverty in England and Wales. Subsequently developed alternative approaches based on the principle that a certain level of energy expenditure at home is problematic if a) these sums are disproportionate with respect to their income, or b) if they use the family budget resources as to make their income fall below a certain threshold, i.e. the monetary poverty line or minimum income. Despite the objective nature of the baseline data, this approach leads to very different energy poverty rates estimations for the same data sets, depending on the thresholds applied. This implies a certain arbitrariness depending on the analyst decisions.
- Approach based on perceptions and statements of households: This approach was initially proposed by Irish researchers Jonathan Healy and John Clinch, and is based on the use of the results of the survey of conditions of life of Eurostat, and, specifically, on questions that seek to know if a home

is able to keep a suitable temperature, if they have delays in the payment of the energy bills, or if the household has any deficiency related to energy poverty (leaks, rot or humidity). Its subjective character is the main limitation of this approach, However, it is still the only approach that allows to make comparisons between countries of the European Union.

9.3.1 Approach based on perceptions and statements of households.

The European Survey on Income and Living Conditions (EU SILC) is the Encuesta de Condiciones de Vida (ECV - Living Conditions Survey). Its main objective is to have a reference source on comparative statistics about the distribution of income and social exclusion in Europe (INE, 2005, p.1). For that purpose, it annually collects data on the income and living conditions of households, as well as the composition of poverty and social exclusion, at national and EU level. It is therefore, carried out with harmonized criteria.

Within this survey, there are three questions that can be used to analyze the energy poverty:

- Can your household afford: […]
- Maintain your household at an appropriate temperature during winter?
- During the last 12 months: […]
- Did you have any delay in the payment of gas, water, heating, electricity community expenditures, etc..?
- Tell me if your household can afford: […]
- Maintain your household at an appropriate temperature during summer?

This estimation method is also called consensual approach (Healy, 2004, Healy and Clinch, 2004). It consists in a direct observation of the phenomenon based on the statements of the respondents about the state of their housing and their perceptions about their home and the living conditions. This is the main advantage and drawback of these indicators: their declared, subjective and dependent on the temporal and socio-cultural contexts, especially in the case of the first question, referred to the ability to maintain the house at an adequate temperature. The indicator about delays in the bill payment is also problematic because it includes water and community expenses, not just energy. However, given that there is no other methodology common to the entire EU, at present, this approach is the only one that allows comparing the incidence of energy poverty among Member States.

The results obtained from the survey are mainly presented in the form of a percentage of affected households for each indicator and disaggregation variable, although the figures for the number of people affected have been provided for some of those variables and indicators.

For 2014 the main results obtained from this indicator are as follows:

- 11,1% of the population (5.1 million people) could not maintain their household at an adequate temperature.
- 16.6% of the population (7.8 million people) were living in houses with some kind of imperfections associated to their wellbeing, such as leaks, dampness or rotting in their walls, roofs, floors or windows.
- 8% of households (3.2 million people) had delay in the payment of energy bills during the last 12 months.

9.3.2 Approach based on expenditures and incomes.

The main purpose of the Family Budget Survey (EPF – Encuesta de Presupuestos Familiares) is to obtain information on the nature and destination of consumption expenditures, as well as on several characteristics related to the living conditions of households. It has an annual periodicity and is carried out approximately on a sample of 24,000 households (INE, 2008).

The microdata of the EPF have been used to calculate three different types of indicators, as described below:

Disproportionate expenditures on domestic energy

This indicator classifies as being in energy poverty situation those households whose annual real energy expenditures represent more than a certain percentage of their annual net income. For the present study, four different thresholds have been defined 5, 10, 15 and 20% of annual income. These thresholds correspond, approximately, to one to four times the mean expenditure on energy over income per household in Spain.

According to this indicator, the situation in Spain during 2014 was as follows:

- 45% of households (equivalent to 20.7 million people) spent more than 5% of their income on energy;
- 15% of households (equivalent to 6.6 million people) spent more than 10% of their income on energy;
- 6% of households (equivalent to 2.6 million people) spent more than 15% of their income on energy;
- 3% of households (equivalent to 1.2 million people) spent more than 20% of their income on energy.

Low income – high cost (LIHC)

According to the definition of this indicator a household is energy poor if the equivalent cost necessary for paying the energy needed to maintain an adequate level of comfort is above the mean value, and if after discounting that expenditure from its equivalent income, the result is below the monetary poverty line (60% of the mean income per equivalent person after deducting the expenses associated with housing other than energy).

Within the report analyzed, for the calculation of 60% of the mean equivalent income, both the housing costs of the household and the average energy expenses have been discounted from the equivalent income, that is, according to this modification, the report considers that a household is in energy poverty when:

- The actual equivalent expenditure on energy is higher than the average of the real equivalent energy expenditure of the households surveyed
- The result of subtracting the equivalent energy expenditure from the equivalent household income is lower than 60% of the mean of the equivalent income with no housing cost or average energy costs.

According the LIHC indicator, as defined above, 10% of the households in Spain (equivalent to 4.9 million people) would be in situation of energy poverty during 2014.

Minimum Income Standard (MIS)

This indicator is based on a detailed investigation of the necessary income to reach a minimum standard of consumption of household goods and services according to size and composition of the home. According to this indicator, a household is energy poor if its income, after discounting the household and energy expenditures, are under the threshold of acceptable income minus the household and energy expenditures.

In Spain there is no official estimation about the acceptable income for a household. That's why three different income levels have been defined within the report analyzed, in order to be able to apply this indicator:

MIS1: The average of the minimum insertion incomes of the different Autonomous Communities in Spain, weighted by the population of those areas. For 2013 MIS1 is 415€/month.

MIS2: The highest minimum insertion income of the Autonomous Community in Spain. For 2013 MIS2 is 662€/month.

MIS3: The third threshold of 800 euros per month (for the first person in the household) in 2013, was defined arbitrarily in order to analyze how the percentage of households living in energy poverty would increase according to this indicator if the objective of the minimum income was to guarantee this level of income.

The values used for this indicator within the report can be found below:

	2014	2013	2012	2011	2010	2009	2008	2007	2006
MIS1	417	416	414	416	404	401	398	381	372
MIS2	666	663	641	659	639	635	630	603	589
MIS3	802	800	778	763	740	735	729	699	682

Figure 78 - MIS indicators for Spain.

According to the definition of MIS indicator, as defined above:

- 31% of households (equivalent to 17.1 million people) would be in energy poverty according to indicator MIS3 (minimum income threshold of 802 € / month in 2014 for the first person in the household);
- 21% of households (equivalent to 12.1 million people) would be in energy poverty according to indicator MIS2 (minimum income threshold of 666 € / month in 2014 for the first person in the household);
- 7% of households (equivalent to 4.7 million people) would be in energy poverty according to indicator MIS1 (minimum income threshold of 417 € / month in 2014 for the first person in the household)

9.4 Cluster analysis of the population

To represent in a synthetic way the values obtained in all energy poverty indicators in each Autonomous Community in Spain, the data has been organized in the form of a table (see Figure 79) corresponding to the 2014 results. To show the relative position of each territory with respect to the average for Spain, the following chromatic code is used: warm colors (red) indicate that the value of the indicator for the Region is above the national average; Cold colors (blue) indicate otherwise.

2014	5%	10%	15%	20%	LIHC	MIS1	MIS2	MIS3	ECV T°	ECV Ret	ECV got
Castilla – La Mancha	67,5%	36,4%	19,1%	9,0%	19,9%	8,2%	27,3%	41,3%	14,2%	8,1%	16,7%
Andalucía	44,3%	15,3%	5,8%	3,0%	9,2%	9,8%	28,9%	41,3%	15,8%	11,2%	23,4%
Extremadura	53,7%	19,5%	6,5%	2,5%	11,4%	7,4%	26,2%	40,6%	8,3%	3,8%	29,2%
Murcia, Región de	45,6%	15,5%	5,4%	2,1%	11,9%	9,7%	26,1%	37,2%	18,5%	11,1%	14,2%
Aragón	49,2%	19,2%	8,3%	3,6%	9,6%	4,6%	13,6%	23,2%	4,7%	4,7%	7,5%
Balears, Illes	46,4%	14,6%	5,9%	2,1%	9,9%	7,0%	21,9%	29,8%	6,6%	13,0%	26,7%
Canarias	27,2%	7,1%	2,6%	1,2%	4,6%	9,6%	29,8%	43,9%	0,2%	11,5%	35,3%
Cantabria	52,2%	15,8%	6,1%	2,2%	11,8%	5,9%	21,4%	30,7%	8,3%	4,5%	12,0%
Castilla y León	56,6%	23,1%	10,4%	4,4%	11,9%	4,6%	17,0%	27,3%	5,8%	2,4%	11,9%
Galicia	49,4%	16,8%	6,9%	2,3%	8,4%	5,8%	18,7%	30,7%	16,0%	6,9%	34,2%
Rioja, La	54,6%	21,7%	8,6%	4,3%	10,8%	4,5%	15,5%	24,3%	9,1%	5,5%	6,0%
Comunitat Valenciana	39,7%	10,7%	3,9%	1,8%	7,4%	7,9%	21,4%	31,3%	18,4%	11,7%	13,3%
Navarra, Comunidad Foral de	50,6%	17,9%	6,9%	3,9%	9,3%	4,1%	14,0%	21,8%	1,0%	3,9%	10,0%
Cataluña	44,7%	14,6%	6,1%	2,6%	9,6%	7,2%	19,0%	27,9%	8,7%	7,5%	7,5%
Madrid, Comunidad de	45,6%	12,5%	4,7%	1,9%	9,9%	6,0%	17,2%	24,5%	8,5%	5,7%	11,4%
Asturias, Principado de	41,1%	11,8%	4,0%	2,1%	7,6%	4,7%	16,3%	24,0%	13,7%	5,8%	15,5%
País Vasco	36,4%	8,2%	2,1%	0,8%	7,3%	3,8%	11,7%	19,0%	7,3%	3,5%	12,4%
España	45,5%	15,2%	6,1%	2,7%	9,6%	7,2%	21,3%	31,5%	11,1%	8,0%	16,6%

Figure 79 - Energy Poverty indicators for each Region in Spain.

As a summary, we can state that the four regions with the highest degree of energy poverty according to the indicators compiled are, Andalusia, Castilla La Mancha, Extremadura and Murcia. On the other hand, the three Regions with the least degree of affectation are the Basque Country, the Principality of Asturias and the Community of Madrid.

A comparative analysis of the energy poverty indicators disaggregated by the socio-demographic characteristics of the household indicates that households with children under 18 implies have greater incidence of delays in the payment of bills. On the other hand, only 3% of households with elderly people do not pay their bills on time, although they are more likely to suffer from inadequate temperatures inside the house.

On the other hand, there is a very noticeable difference for indicators based on expenditures and income, according to which the lower the educational level reached by the main person in the household, the greater the value of the energy poverty indicator. Furthermore, the working situation of the household is also shown as a factor of vulnerability. Unemployed households that depend on unemployment or other social benefits are more likely to be in energy poverty according to EPF and CVD indicators. In households with employment, there is also a greater incidence of energy poverty for people with temporary contracts compared to fixed contracts of indefinite duration.

Disaggregated by categories of marital status, the subgroup of households whose main person is widowed, separated or divorced have a higher incidence of energy poverty according to EPF indicators. Also large families (2 adults with 3 or more dependent children), single-parent families in which one adult is in charge, at least one dependent child or in households with older people, have higher rates of energy poverty. The same applies to households where the main person was born outside of Spain, especially if the origin country is a country that is not a Member State of the EU.

9.5 Conclusions

According to the report analyzed, the available indicators show that a significant part of Spanish households undergo conditions associated to energy poverty, that is, a signification proportion of households in Spain are considered to be in situation of energy poverty. More in concrete, results show that as of 2014:

- 11% of households (5.1 million citizens) stated to be unable to keep their home adequately warm during the cold season.
- 8% of households (4.2 million citizens) stated to be in arrears on utility bills including domestic energy.
- 15% of households (6.2 million citizens) devoted more than 10% of their annual income to domestic energy.
- 10% of Spanish households (4.9 million citizens) was in difficulties as per the official energy indicator of the UK. This means that, when discounted housing and energy costs, their income was below the monetary poverty line (60% of the equivalent median income); and that their equivalents energy expenditure was above the Spanish median.
- 21% of households (12.1 million citizens) was in difficulties according to the MIS indicator approach. According to this approach, their income after housing and domestic energy costs was below the highest Integration Minimum Income level of the country (corresponding to the Autonomous Community of Basque Country) less the average housing and energy costs of a Spanish household.

Spatially disaggregated results reveal that differences in climatic conditions across the country do not explain disparities in energy poverty levels in Spain, which are more associated to the level of studies of the household occupants, the type of family (Sigle-parental, large family, single person households have higher risk of being energy poor), or the labor situation of the household.
The ASSIST action in Spain will be basically focused in two different HEA profiles, one of them directly related with the energy companies front office and another one focused on a more social profile. The intention is to tackle well defined target groups of energy consumers already identified as potentially vulnerable, either by

the energy companies (aware of the difficulty in the payments of the bills by the end users) or by the social frontend (to which the vulnerable end users might have already gone for social aid).

The detailed planning of the ASSIST actions in Spain and the further obtained results will be reported in deliverable D5.2.

10. United Kingdom

10.1 Methodology

10.1.1 General

The source of data for housing and the household members, essential in modelling fuel requirement, is the English Housing Survey (EHS). The EHS is currently an annual survey, commissioned by the Department of Communities and Local Government (DCLG). For the purposes of producing the fuel poverty statistics, two years of the survey are combined. The EHS covers all tenures and includes a household interview and a physical inspection of properties by a surveyor. The information obtained through the survey provides an accurate picture of the type and condition of housing in England, the people living there, and their views on housing and their neighbourhoods. The survey is a random sample of housing and householders in England. The sample is clustered, with half of England being sampled each survey year, meaning that two consecutive years of the survey provide a national sample, which is what is used to derive the fuel poverty statistics. The two key components of the EHS used in the estimation of fuel poverty are:

- Interview Survey: An interview is conducted with the householder. The interview covers a wide range of topics that include: household characteristics, satisfaction with the home and the area, disability and adaptations to the home, and income details; and
- Physical Survey: The interview is followed by a visual inspection of the property, both internally and externally, by a surveyor. Data collected includes the number and type of rooms and facilities contained in the property, the condition of a wide range of aspects of the physical structure, details of the heating systems, approximate age of the property, and assessment of neighbourhood quality.

Currently, each year around 13,300 interviews are conducted with householders, and around 6,200 dwellings (approximately 6,000 households) have a follow up physical survey of their dwelling[7].

---

7 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/639120/Fuel_Poverty_Methodology_Handbook.pdf

10.1.2 Headline Figures[8]

In 2015 (latest data available), the average fuel poverty gap (the amount needed to meet the fuel poverty threshold – further details in par. 10.3) in England was estimated at £353, which is a decrease of 5.6 per cent in real terms from 2014 and continues the steady downward trend since 2013.

- The aggregate fuel poverty gap for England also continued to decrease in 2015 (by 0.5 per cent in real terms) to £884 million.
- In 2015, the proportion of households in fuel poverty in England was estimated at 11.0 per cent (approximately 2.50 million households). This is an increase of 0.4 per cent from 2014.
- In 2015, further progress was made towards the interim 2020 fuel poverty target, with 89.7 per cent of all fuel poor households living in a property with a fuel poverty energy efficiency rating of Band E or above.

Table 38 - Fuel poverty targets in UK for different Bands.

Year	Fuel Poverty Target	2010 (%)	2015 (%)	Percentage point change
2020	Band *E* or above	78.7	89.7	11.0
2025	Band *D* or above	29.4	63.1	33.7
2030	Band *C* or above	1.5	7.8	6.2

The relative nature of the fuel poverty indicator makes it difficult to isolate accurately absolute reason for change. The fuel poverty status of a household depends on the interaction between three key drivers; household incomes, fuel poverty energy efficiency ratings (FPEER) and required fuel costs. These are summarised below for 2015:

[8] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/639118/Fuel_Poverty_Statistics_Report_2017_revised_August.pdf

	Fuel poor	Non-fuel poor	All households
Median income	£10,118	£23,147	£21,333
Median FPEER	58.1	65.5	64.6
Median fuel costs	£1,430	£1,197	£1,236

Figure 80 - Fuel poverty status of a household in UK.

10.1.3 *Energy efficiency, dwelling and household characteristics*

- Households with insulated cavity walls are least likely to be in fuel poverty (6.2 per cent of households with an average gap of £185), compared to households with uninsulated solid walls (18.1 per cent with an average fuel poverty gap of £482).
- Older dwellings have a higher proportion of households in fuel poverty (18.0 per cent) compared to newer dwellings (4.2 per cent). They also have a much larger fuel poverty gap. The oldest dwellings (pre-1850) have an average fuel poverty gap of £899 compared to £182 for the newest dwellings (post-1990).
- The level of fuel poverty is highest in the private rented sector (21.3 per cent of households) compared to those in owner occupied properties (7.4 per cent). Those in the private rented sector also tend to be deeper in fuel poverty, with an average fuel poverty gap of £410, compared to £175 for those in local authority housing.
- When considering household composition, those living in 'multi-person (adult) households' are deepest in fuel poverty with an average fuel poverty gap of £493 compared to a single person under 60 (£227). However, the highest prevalence of fuel poverty is seen for lone parents with dependent child(ren) (23.6 per cent).

10.2 Cluster analysis of the population

Regional differences in the levels of fuel poverty are illustrated in the following figure:

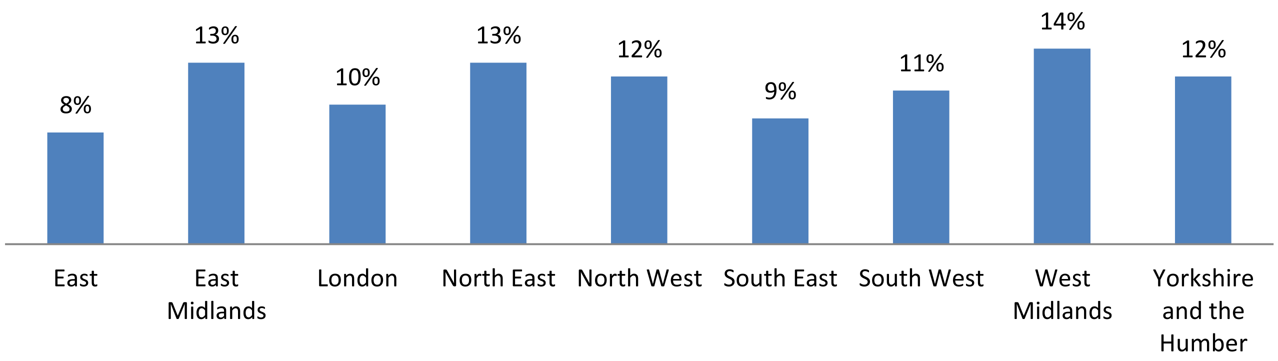


Figure 81 - % of households in fuel poverty split regionally[9].

There are clear regional differences when it comes to fuel poverty. The regions with the lowest levels of energy poverty are London, the South-East and the East. The regions with the highest levels of energy poverty are the West Midlands, the North-East and the East Midlands. These results align closely with general income level statistics.

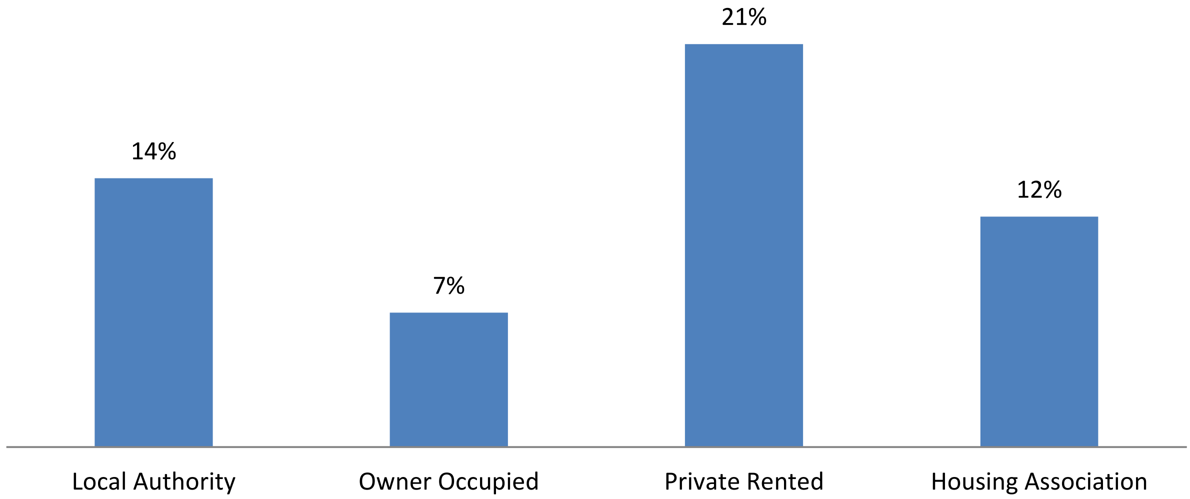


Figure 82 - % of households in fuel poverty split by tenure[10].

9 https://www.gov.uk/government/statistics/gdp-deflators-at-market-prices-and-money-gdp-march-2017-quarterly-national-accounts-march-2017

10 https://www.gov.uk/government/statistics/gdp-deflators-at-market-prices-and-money-gdp-march-2017-quarterly-national-accounts-march-2017

Figure 82 clearly shows that energy poverty is most prevalent in properties that are rented from private landlords. It would be interesting to see how this compares to results from other partner countries throughout the EU as a potential target sector.

10.3 Vulnerable consumers characterization

Fuel poverty in England[11] is measured using the Low Income High Costs (LIHC) indicator. Under the LIHC indicator, a household is considered to be fuel poor if:

- they have required fuel costs that are above average (the national median level)
- were they to spend that amount, they would be left with a residual income below the official poverty line

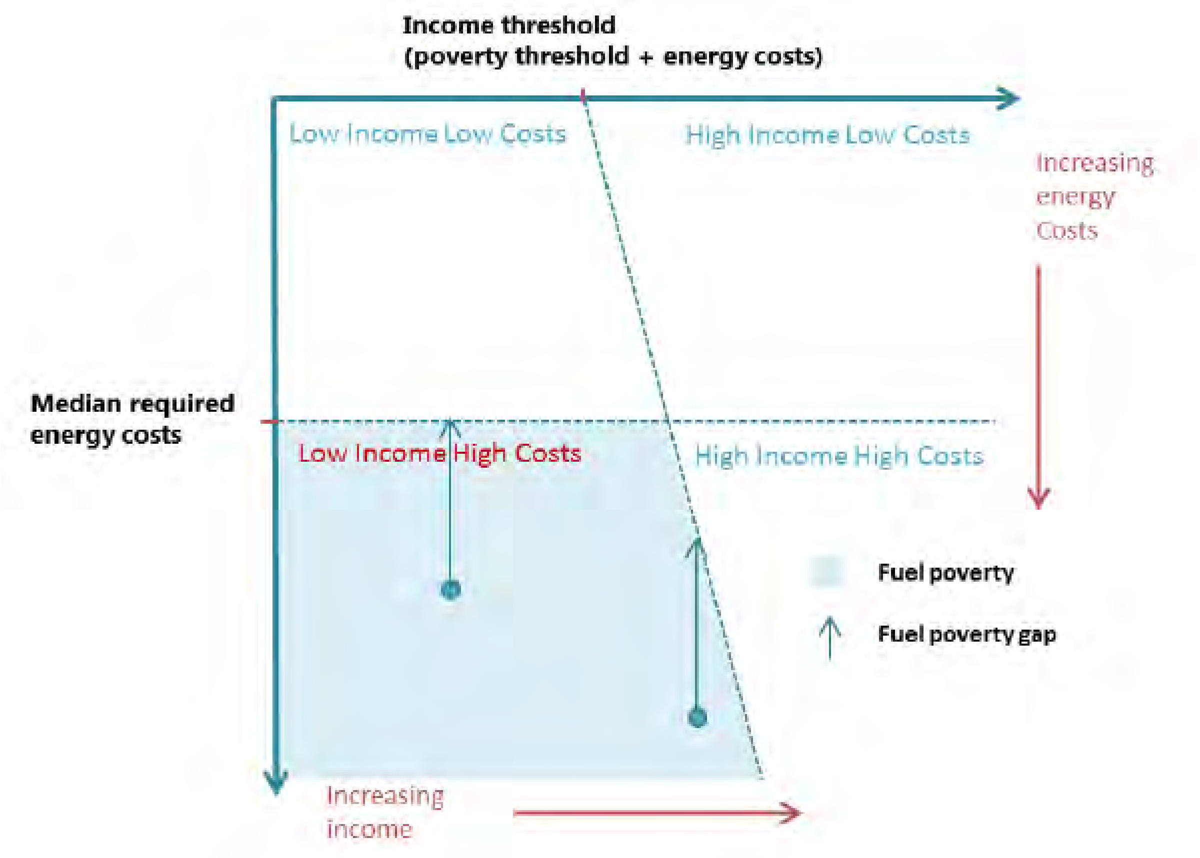


Figure 83 - Fuel poverty under the Low Income High Costs indicator.

Low Income High Costs is a dual indicator, which allows us to measure not only the extent of the problem (how many fuel poor households there are), but also the depth of the problem (how badly affected each fuel poor household is). The depth

11 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/639118/Fuel_Poverty_Statistics_Report_2017_revised_August.pdf.

of fuel poverty is calculated by taking account of the fuel poverty gap. This is a measure of the additional fuel costs (in pounds) faced by fuel poor households to meet the non-fuel poor household threshold. This is illustrated in Figure 1, where the indicator consists of:

- the number of households that have both low incomes and high fuel costs (shown by the shaded area in the bottom left hand quadrant in Figure 1.1); and
- the depth of fuel poverty among these fuel poor households. This is measured through a fuel poverty gap (shown by the vertical arrows in Figure 1.1), which represents the difference between the required energy costs for each household and the nearest fuel poverty threshold.

To get a sense of the depth of fuel poverty at a national level, the fuel poverty gap for each individual household is aggregated across all fuel poor households to produce an overall aggregate fuel poverty gap.

The fuel poverty indicator is a relative measure, as it compares households to national income thresholds and national median energy costs. A change in income will only have an impact on fuel poor households if they see relatively larger income changes (increase or decrease) than the overall population; the same is true for household energy costs. As a result, the proportion of households in fuel poverty remains, on the whole, stable over time, whereas the fuel poverty gap (which is measured in pounds) is more closely linked to changes in energy prices and the economy and therefore, a more informative measure when looking at the direct impacts of fuel poverty over time.

In December 2014, the Government introduced a new statutory fuel poverty target for England[12]. The target is to ensure that as many fuel poor homes as reasonably practicable achieve a minimum energy efficiency rating of a Band C[13], by 2030. To support the implementation of this target, the Government published 'Cutting the cost of keeping warm: a fuel poverty strategy for England[14], in March 2015. The strategy also set out interim milestones to lift as many fuel poor homes in England as is reasonably practicable to Band E by 2020; and Band D by 2025, alongside a strategic approach to developing policy to make progress towards these targets.

10.4 Conclusions

[12] Fuel poverty is a devolved matter, with each nation in the UK having its own policy target, measurement and outputs.

[13] Household energy efficiency ratings are banded from G (lowest) to A (highest).

[14] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/408644/cutting_the_cost_of_keeping_warm.pdf.

From the research shown above and the information gathered through stakeholder and consumer surveys, as well as analysis of the various schemes already running to alleviate energy poverty in the UK, action is being planned to work with partners already established in the local area.

Severn Wye are planning to work with a local organisation that already has volunteers working with vulnerable clients. The volunteers will then be trained through ASSIST to be able to recognise the signs of energy poverty and what actions can be taken to alleviate the issues.

The detailed planning of the ASSIST actions in the UK will be reported in deliverable D5.2.

11. General conclusions

The main conclusion that can be drawn from this work is that there is no uniform approach to the definition of energy poverty and to the segmentation of the end users in such a condition in different countries. However, even if there are significant differences in the approach of the different countries, some final considerations can be drawn.

It is impossible to use the same methodology for the segmentation in all the analysed countries. For some countries, a detailed segmentation was already available, performed by other bodies. This is the case of Belgium, Spain and the UK. For other countries, it was necessary to use internal resources in order to perform statistical analyses of data, collected either by the project partners or by other bodies. This is mostly the case of Finland and Italy. For Poland, a combination of different studies and statistical data analysis has been used. The last three countries presented also the results of a dedicated survey, that is supposed to be later proposed to the identified energy vulnerable consumers during ASSIST actions.

The indicators are reflecting the data availability in the single countries. For some countries, considering also the used external studies, it was possible to define common indicators, that are the most widely used at the moment in the EU. The more widespread indicator is LIHC: Low Income High Costs. For Belgium, Poland, Spain and the UK this indicator was already available in the existing studies. For Finland, it was calculated from the around 4.000 data collected through surveys, that were covering both income and energy expenses of the interviewed households. In Italy, there are statistics about income and statistics about energy expenses, but at the moment it is impossible to correlate them, so different indicators have been used. In particular, a “modified” LIHC indicator has been proposed (ref. [29] and [30]) based on the correlation between energy expenses, thermal confort and overall expenses of the household.

The amount of people affected by energy poverty, in all countries, is very high. For example, in Belgium, with the least inclusive indicator, at least 3.9% of the population is energy poor, and considering all the indicators and the possible overlaps the number increases to 21%; in Finland, for sure at least 3% of the investigated people were energy poor and at least 12% were at risk of energy poverty. In Italy, the situation is more complicated, however it is significant to highlight that at least 4% of the population has a high probability that the yearly energy expense is much higher than the food expense, while 8% has high probability of being totally unable to reach 75% of the minimum comfort. Moreover, a further 21% has a medium probability of having an energy expense higher than food expense and of being unable to reach the minimum level of comfort. In Poland, it is estimated that around 12% of the population is in energy poverty, while in Spain, according to the least inclusive indicator, the minimum number of energy poor is 8% of the population, while considering the most inclusive the amount increases to 17%. In the UK, where the issue has been taken care for a longe time, than in the other countries, a surprising result is that in 2015 the number of fuel poors increased of 0.4% with respect to 2014, reaching 11% of the population.

The categories most affected by energy poverty are different in each country, however some common features can be identified. First of all, people living in rented houses are more at risk than those who own the house where they live. In most countries, single parents and unoccupied people are more at risk than other categories, while old people are less affected than these categories. In all the countries, there are significant differences among the different regions, so there is no uniform distribution of energy poverty on a country scale. Finally, some recommendations can be drawn:

- There is a widespread need of aligned and complete data about households situation in all the EU MS: these data should comprise information about the structure of the house (e.g. building year, renovation works, adoptedenergy efficiency measures), about income, about energy expenses (both in € and kWh or equivalent energy units), about family habits and behaviours;
- There is the need of common indicators that cover all the possible issues related to energy poverty; however, this can be accomplished only with the availability of data that are the same for each MS;
- In some countries (e.g. Italy, Finland), higher attention shall be put to the issue of energy poverty, from policy makers and central governments, in order to have a full mapping of the phenomenon and to solve the lack of data and indicators mentioned above.

12. References

[1] A. Węglarz, G. Kubalski and D. Owczarek, "Propozycje mechanizmów wsparcia procesu przeciwdziałania zjawisku ubóstwa energetycznego w Polsce," Fundacja Instytut na rzecz Ekorozwoju, Warszawa, 2014.

[2] Urząd Regulacji Energetyki, "Mapa drogowa uwolnienia cen dla wszystkich odbiorców energii elektrycznej," 2008.

[3] D. Owczarek and A. Miazga, "Ubóstwo energetyczne w Polsce – definicja i charakterystyka społeczna grupy," Fundacja Instytut na rzecz Ekorozwoju, Warszawa, 2015.

[4] Ustawa z dnia 12 marca 2004 r. o pomocy społecznej (Dz.U. 2004 nr 64 pag. 593).

[5] Główny Urząd Statystyczny, "Badanie Budżetów Gospodarstw Domowych," Warszawa, 2012, 2013, 2014, 2015 and 2016.

[6] K. Sałach and P. Lewandowski, "Pomiar ubóstwa energetycznego na podstawie danych BBGD - metodologia i zastosowanie," Instytut Badań Strukturalnych, 2018.

[7] Fundacja Habitat for Humanity Poland, 2015. [Online]. Available: http://habitat.pl/ubostwo-energetyczne/. [Retrieved: February 9, 2018].

[8] D. Owczarek, "Fundacja Blisko," 2016. [Online]. Available: http://fundacjablisko.pl/index.php/owczarek-ubostwo-energetyczne-nowe-wyzwanie-nowe-rozwiazania/. [Retrieved: February 9, 2018].

[9] Eurostat Database, "Population unable to keep home adequately warm by poverty status," 2017. [Online]. Available: http://ec.europa.eu/eurostat/web/products-datasets/-/sdg_07_60. [Retrieved: February 1, 2018].

[10] ACA (2016). Pobreza, vulnerabilidad y desigualdad energética. Nuevos enfoques de análisis. España (2006 – 2016). (Madrid, Spain: Asociación de Ciencias Ambientales).

[11] Encuesta de Presupuestos Familiares; EPF: INE (2008). Encuesta de Presupuestos Familiares. Base 2006. Principales características. (Madrid, Spain: Instituto Nacional de Estadística).

[12] Encuesta de Condiciones de Vida, ECV: INE (2010). Encuesta de Condiciones de Vida ECV (7). Cuestionario de Hogar (Madrid, Spain: Instituto Nacional de Estadística).

[13] Boardman (1991): Boardman, B. (1991). Fuel poverty: from cold homes to affordable warmth (London ; New York: Belhaven Press).

[14] Hills, 2012: Hills, J. (2012). Getting the measure of fuel poverty. Final Report of the Fuel Poverty Review. (London, UK: Centre for Analysis of Social Exclusion. The London School of Economics and Political Science).

[15] INE, 2005, p.1: INE (2005). Encuesta de Condiciones de Vida. Metodología. (Madrid, Spain: Instituto Nacional de Estadística).

[16] Healy 2004: Healy, J.D. (2004). Housing, fuel poverty, and health: a pan-European analysis (Aldershot, England ; Burlington, VT: Ashgate Pub).

[17] Clinch: Clinch, J.P., and Healy, J.D. (2000). Cost-benefit analysis of domestic energy efficiency. Energy Policy 29, 113–124.

[18] INE, 2008: INE (2008). Encuesta de Presupuestos Familiares. Base 2006. Principales características. (Madrid, Spain: Instituto Nacional de Estadística).

[19] Koning Boudewijnstichting (King Baudouin Foundation), 2017. "Barometer energiearmoede (2009-2015)". Available: https://www.kbs-frb.be/nl/Activities/Publications/2017/20170313NT1 [Retrieved: February 20, 2018].

[20] UKDepartment for Business, Energy & Industrial Strategy, June 2017. "Fuel Poverty Methodology Handbook". Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/639120/Fuel_Poverty_Methodology_Handbook.pdf.

[21] Department for Business, Energy & Industrial Strategy, June 2017. "ANNUAL FUEL POVERTY STATISTICS REPORT, 2017 (2015 DATA) England, Statistical Release: National Statistics". Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/639118/Fuel_Poverty_Statistics_Report_2017_revised_August.pdf.

[22] HM Treasury, April 2017. "GDP deflators at market prices, and money GDP March 2017 (Quarterly National Accounts, March 2017)". Available: https://www.gov.uk/government/statistics/gdp-deflators-at-market-prices-and-money-gdp-march-2017-quarterly-national-accounts-march-2017.

[23] HM Government, March 2015. "Cutting the cost of keeping warm – a fuel poverty strategy for England" Presented to Parliament by the Secretary of State for Energy and Climate Change by Command of Her Majesty. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/408644/cutting_the_cost_of_keeping_warm.pdf.

[24] G. Huebner, D. Shipworth, I. Hamilton, Z. Chalabi and T. Oreszczyn, "Understanding electricity consumption: A comparative contribution of building factors, socio-demographics, appliances, behaviours and attitudes," Applied Energy, vol. 177, pp. 692-702, 2016.

[25] Istat, "Spese per consumi delle famiglie: anno 2016," Istat, 2017 [Micro-Data for Research use provided to RSE Spa by ISTAT]: https://www.istat.it/it/archivio/202093.

[26] E. R. Mansfield and B. P. Helms, “Detecting Multicollinearity," The American Statistician, vol. 36, pp. 158-160, 1982.

[27] G. James, D. Witten, T. Hastie and R. Tibshirani, “An Introduction to Statistical Learning, Springer New York, 2014.

[28] L. Breiman, J. Friedman, C. J. Stone and R. A. Olshen, “Classification and regression trees", CRC press, 1984.

[29] I. Faiella and L. Lavecchia, “Questioni di Economia e Finanza – La povertà Energetica in Italia”, Banca d’Italia, October 2014.

[30] I. Faiella, L. Lavecchia and M. Borgarello, “Una nuova misura della povertà energetica delle famiglie”, Banca d’Italia, October 2017.

[31] L. Oja, A. Vaahtera, I. Vehviläinen, S. Ahvenharju and L. Hakala, “Selvitys energiaköyhyydestä”, Ympäristöministeriön raportteja 21/2013.

[32] S. Runsten, K. Berninger, J. Heljo, J. Sorvali, P. Kasanen, J. Vihola and U. Uotila, “Pienituloisen omistusasujan energiaköyhyys”, Ympäristöministeriön raportteja 6/2015.

[33] Statistics Finland, Regional clasifications, https://www.stat.fi/meta/luokitukset/index_alue.html [Retrieved: January 18, 2018]

[34] Statistics Finland, NUTS regions, https://www.stat.fi/meta/luokitukset/nuts/002-2012/index.html [Retrieved: January 18, 2018]

[35] Statistics Finland, Regional State Admistrative Agencies, http://www.stat.fi/meta/luokitukset/avi/001-2013/index.html [Retrieved: January 18, 2018]

ASSIST 2GETHER

www.assist2gether.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 754051