A Novel Statistical Method to Improve Energy Efficiency of Housing Stock in the Eastern Mediterranean Climate

Bertug Ozarisoy[1] and Hasim Altan[2]

1 Sustainable Environment and Energy Systems (SEES) Graduate Program, Middle East Technical University, Northern Cyprus Campus, Kalkanli, Guzelyurt, North Cyprus, via Mersin 10, Turkey

2 Department of Architecture, College of Architecture and Design, Prince Mohammad bin Fahd University, Khobar, Dhahran, Saudi Arabia

Abstract

This paper presents a novel methodology that was developed according to the *in-situ* measurements of the building-fabric thermal performance to assess as-built energy models of case-study residential tower blocks. It seeks to corroborate the empirical model by integrating the findings from the questionnaire survey and energy-assessment analyses to feed building-energy simulation (BES) procedures. It discusses the uncertain input parameters for the BES that quantitative modelling adopted for the energy assessment of dynamic thermal simulation findings in conjunction with occupants' socio-demographic characteristics, occupancy patterns, household size and environmental conditions recorded. The present study revealed that weekday cooling consumption patterns were significantly and strongly related to the weekend heating consumption patterns ($\chi^2 = 54{,}590$, $p < 0{,}001$, Cramer's $V = 0{,}522$). Specifically, longer duration of heating consumption was related to longer duration of cooling consumption. The developed framework provides a scientific background that can be included in the BES platform to serve as a set of guidelines for the European Performance of Buildings Directive scientific committee for energy policy design and retrofitting strategies.

Keywords: Energy efficiency, Energy use, Overheating risk, Energy retrofit, Thermal comfort

1. Introduction

The conventional terminology of energy efficiency gap (EEG) was defined by Jaffe and Stavins as "the energy efficiency gap is described as the gap that exists between the current or expected future energy use of homes and the optimal current or future energy use" (Ozarisoy & Altan, 2022). Jaffe and Stavins then further describe the EEG in order to provide a foundation of theoretical information to guide policymakers in the implementation of energy policy design in the residential sector. However, the starting point of a well-established meaning for EEG dates to the 1970s. A pilot study project entitled "Drivers and Barriers to Improving Energy Efficiency and Reducing CO_2 Emissions in the Private Housing Sector" states that in the UK little action had been taken towards energy conservation through the method of considering existing housing stock that was built in the UK pre-1973 (Altan, 2004). This pilot study demonstrated the relevance of the rapid increase in energy prices in the 1970s. The present study is the first to target and conduct building-energy simulation (BES) procedures on existing Cypriot social-housing stock. A BES analysis was integrated into the implementation of energy performance certificates (EPCs) because of the reliable assumptions thereof to assess the energy performance of case-study residential tower blocks (RTBs). The energy simulation inputs seek to identify the impact of household occupancy patterns and habitual adaptive behaviour on home-energy performance to provide a basis for the information that is needed to properly calibrate the building-energy performance of targeted households. It also envisages demonstrating that occupants' real-life energy-use experiences have had a significant impact on calibrating domestic-energy use to simultaneously identify discrepancies between the actual and predicted energy use on the dynamic energy-simulation platform. As such, the present study was developed according to human-based BES input parameters obtained from the questionnaire survey, *in-situ* measurements, the infrared-radiometer thermography (IRT) survey and environmental monitoring of the project site to demonstrate real data for energy-use policy making decisions. This conceptual framework can be applied to efforts to implement the Energy Performance of Buildings Directives (EPBD) mandates and to demonstrate the exemplar development framework, policy and regulations as it relates to the social-housing stock of the Republic of Cyprus (RoC) and other EU-27 member states.

doi:10.18086/eurosun.2022.02.06 Available at http://proceedings.ises.org

2. Method

2.1. The Concept of Statistical Representativeness

In order to provide a background analysis for developing the concept of statistical representativeness, this section presents a review of selected theoretical information and exemplar pilot projects and their applicability. Table 1 delineates the original research articles that were reviewed to identify the most appropriate concept of statistical representativeness for the present study.

Tab. 1: Reviewing the concept of representativeness in statistics.

References	Concept(s)	Method(s)	Outcome(s)
Chasalow and Levy (2021)	Law, social and behavioural sciences	- Stratification of population sample size - Generalisability of research findings	Equal representation of each subgroup of the user population was recommended
Hirsch and O'Donnell (2017)	Education, social and behavioural sciences	- A multiple-choice test-based survey was distributed - Descriptive statistics - 4 x 2 chi-square test of independence	A unique set of test questions to identify students who hold common representativeness was developed
Schmill *et al.* (2014)	Global change sciences (i.e., human factors, climate, remote sensing)	- Pearson's chi-square tests - Heat mapping - Histograms - Kullback-Leibler *f*-divergence test - Multivariate analysis	Implementing a variety of methods for making assessments about the representativeness of a collection of case studies across the globe
Schouten *et al.* (2009)	Social and behavioural sciences	- Population R-indicators - Chi-square statistics used to test independence and goodness-of-fit - Logistic regression models	A mathematically rigorous definition and perception of representative response was developed

Ozarisoy & Altan (2021) and Hu and Kohler-Hausmann (2020) highlight the importance of the integration of the STS approach into using the conceptual analysis of sampling size to test associations gathered through various experimental statistical analyses. Chasalow and Levy (2021) discuss the STS approach providing a multidisciplinary integrated conceptual framework that enables both statisticians and engineers to interrogate taken-for-granted terms and categories while developing benchmarks in energy-policy design. These studies indicate that representativeness of sampling size shows differences between one pilot study and another, due to demographic structures, geography and the political conditions of each research context. Chasalow and Levy (2021) explain their own representativeness concerns while developing a novel methodological framework for representativeness in statistics, politics and machine learning. In this theoretical study, these scholars contribute a sense of the variety of meaning and values associated with representativeness in order to prove the validity of chosen sampling sizes. Chasalow and Levy indicate that they did not select a large sampling to develop their own representative study. In their pilot research project, these scholars predominantly focused on sampling in Europe and the United States more than other parts of the world. According to Chasalow and Levy, it is difficult to identify the limitations of their own selected sampling size representativeness because of the geographical extent of their research context. To avoid a research bias and provide a generalisation of their research findings, their study set out to offer reliable and statistically representative sampling criteria that would allow the targeted reader group to understand their research outcomes.

Garett (1942), Jensen (1926), Kurksal and Mosteller (1980) and McNemar (1940) all discuss the issue of how to check a particular sample for representativeness. Kiaer (1976) states that "the representative method can be applied in several ways". To explain this claim, Kiaer develops two different sample types. His first method is an arbitrary filtering of the variables gathered from the sampling. Kiaer explains that this selection should be done in a "haphazard or random way" to avoid giving preference to subject respondents in certain occupations or belonging to particular social strata. This method of selection highlights that the representativeness of sampling size results from the absence of selective discretion. His second method involves allocating representativeness in a mechanical procedure that can provide a feasible method of design to undertake statistical analysis faster and provide an opportunity to detect discrepancies in the sample.

In previous studies that develop the concept of statistical representativeness of sampling size, one of the main strengths of the selection of a random sampling method is that it enables the use of all available data to ensure a proportionate match on known relevant variables. This was proven by Kiaer in 1976 in a population study that consisted of surveyors in rural areas and used census data to allocate counts per country and then selected districts within the countries to "represent the main industry groups within the country as well as its various geographic conditions" (Kiaer, 1976). This method of design demonstrates that it can be applicable to choose a single geographic domain to develop the concept of statistical analysis by integrating census data and applying outcomes to other geographic domains that have been shown to have similar demographic structures, political assets and cultural norms.

Hirsch and O'Donnell (2001) have developed a scientific method of design that measures the reliability of datasets and the applicability of concepts of statistical design to provide an understanding of conceptual change that can have a long-term impact on conceptual-level analysis. Schmill *et al.* (2014) developed a theory of analytics for assessing global representativeness in social science studies to guide future scholars that was aimed at addressing sampling bias and providing a public domain for similar pilot projects. This pilot study aimed to reduce the gap between local and global researchers in providing analytical methods of design that could help scholars assess the representativeness of a sampling size and assist in correcting any bias for the interpretation of statistical findings. Schmill *et al.* (2014) recommends the chi-square test as the most appropriate statistical method that allows scholars to apply to their own hypothesis testing. In the Schmill study, Pearson's chi-square tests were used for testing the independence of two samples using a model function defined over a contingency table of observed versus expected values. To test the reliability of their statistical analysis, Schouten *et al.* (2007) conducted logistic regression models to predict the type of responses expected. In a further study conducted by Schouten *et al.* (2009), the researchers developed an advanced indicator to identify the concept of statistical representativeness of sample sizes. In this pilot study, the researchers found that the field survey approach gave more accurate results regardless of the limitation of only being able to recruit relatively small sample sizes. As an outcome of this pilot study, Schouten *et al.* (2009) recommend that in multinominal logistic regression models, variables give a significant contribution at the 5% level, and where this cannot be done, these variables should be excluded from the sampling size.

The study adopts the Cohen's statistical convention to identify the significance of occupants' TSVs within *in-situ* recorded environmental parameters. On further questioning the identified 9.46% margin of error, a priori power analysis using G*Power 3.1.9 was conducted to determine the minimum sample size. With the power set at 0.08, alpha level set at 0.05, a moderate effect size of 3.5 (odds ratio) and with the proportion of the control group at 0.5, results indicate that a total of 104 participants would be needed in order to reach an adequate sample size for the study. While this section discusses theoretical information around the concept of statistical representativeness, further exemplary pilot studies are reviewed and discussed in Section 2.2.

2.2. Equation model

Semi- structured interviews and participant feedback were transcribed and translated. The Statistical Package for Social Sciences (SPSS) Version 28.0 software (IBM: Armonk, NY, U.S.) was utilised to conduct the quantitative analysis; and tests-of-associations were conducted between the numeric factors and the questionnaire responses to join the questionnaire results with the statistical analysis. Little and Rubin (2002) suggest that the pattern of missing data is a more important factor than the actual amount of missing data. Missing values that are randomly scattered throughout the data matrix (i.e., data is missing completely at random; MCAR) is a less serious problem than missing values that are not random. If the missing values on a variable are not random but are associated with other variables in the data, this is referred to as missing at random (MAR) and can be alleviated with imputation methods. The three pie charts below summarise the frequency and percentage of missing data in the dataset by variable, case/observation and individual values. The third pie chart represents the full data matrix and was used to evaluate the 5% threshold of the proportion of missing values in the data matrix that was discussed above. If the missing values on a variable are associated with the variable itself (i.e., missing not at random; MNAR), this has serious implications for inferences, regardless of imputation procedures (Dong *et al*., 2021). Recently, Schouten *et al.* (2009) proposed two definitions of representativeness with respect to survey response: strong (given in Definition 1.1) and weak (given in Definition 1.2). Definition 1.1 (strong). A response subset is representative with respect to the sample if response propensities are the same for all units in the population:

$$\forall_i \ E(R_i) = \rho_i = P(R_i = 1 \mid I_i = 1) = \rho \qquad \ldots \qquad (1)$$

The response of a unit is independent of the response of all other units, which denotes the response of unit *i* and is an indicator showing whether a unit took part in the survey. Schouten *et al.* (2009) notes that strong representativeness corresponds to the Missing Completely at Random (MCAR) pattern for every target variable *y*. This means that non-response does not cause estimators to be biased. Although this definition is appealing, its validity can never be tested in practice. To solve this problem, a weaker definition of representativeness was introduced by Schouten *et al.* (2009).

$$\bar{\rho}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} \rho_{ih} = \rho, for\ h = 1, 2, \ldots, H, \qquad (2)$$

Definition 1.2 (weak). A response subset is representative of categorical variable x with H categories if the average response propensity over the categories is constant, where is the population size of category *h,* is the response propensity of unit i in category *h*, and summation is over all units in this category.

2.3. Assumption tests

A general rule for sample sizes is that group sizes are approximately equal if *n* of the largest group is no more than about twice *n* of the smallest group. Another general rule for sample sizes is at least 10% of the sample should be in each group. Categorical variables with very uneven splits between categories present problems for several multivariate analyses (Tabachnick & Fidell, 2007). The following variables were recoded to reflect the conceptualisation of statistically representative findings in accordance with the research hypotheses, as shown in Table 2.

Tab.2: Coded Variables During the Data Preparation Stage.

Age band	Cooling energy consumption in August 2015	Floor level
Cooling consumption on weekdays	Cooling energy consumption in summer 2015	Health condition
Clothing insulation levels of participants	Heating energy consumption in winter 2015	Occupation
Type of cooling control in home	Cooling energy consumption in August 2016	Heating consumption on the weekend
Ethnicity	Cooling energy consumption in summer 2016	Household density
Thermal preference	Heating energy consumption in winter 2016	Income
Interviewed room condition	Metabolic rates of participants	Length of residency
Orientation	Reasons for thermal discomfort	Space conditioning
Overall thermal satisfaction in summer	Thermal sensation in bedrooms 1, 2, 3 and living room	Type of cooling system
Window closing reasons	Window opening patterns in winter	Type of heating system
Note: Additionally, all categorical variables were recoded in ordinal sequence where possible (e.g., metabolic rate) Variables related to occupants' thermal preferences were recoded from very cold to very hot All variables were recoded from smallest value to largest value All dichotomous variables were recoded to 1 = yes, 0 = no		

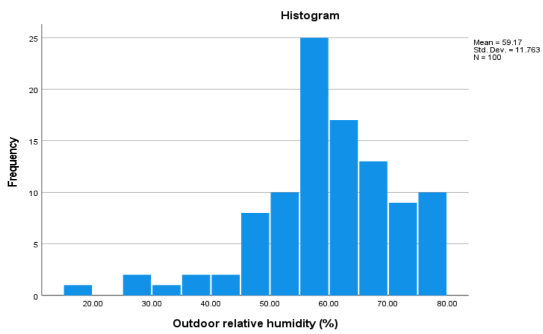
Virtually all parametric statistics have an assumption that the data comes from a population that follows a known distribution. This assumption of normality is often erroneously applied, however, because many populations are not normally distributed. Therefore, researchers need to understand what their samples consist of. It is standard practice to assume that the sample mean from a random sample is normal because of the central-limit theorem. However, almost all variables have a slight departure from normality. If researchers have a large enough sample, then any statistical test will reject the null hypothesis. In other words, the data will never be normally distributed if the sample size is large enough.

It was found that the data was missing completely at random (MCAR). After preparing the data for analysis, it was observed that out of 100 recorded cases, 98 cases contained missing data (98.0%) and out of 53 variables, 2 variables contained missing data (2.8%), which amounted to a total of 0.04% missing information in the dataset. To assess whether the pattern of missing values was MCAR, Little's MCAR test (Little, 1988) was conducted. The null hypothesis of Little's MCAR test is that the pattern of the data is MCAR and follows a chi-square distribution. Using an expectation-maximisation algorithm, the MCAR test estimates the univariate means and correlations for each of the variables. The results revealed that the pattern of missing values in the data was MCAR: χ^2 (104) = 121,645, p = 0,114. Even though the proportion of the total missing data is less than 5% and the data is MCAR, the final sample size may still be affected by listwise or pairwise deletion when the analysis is run. Listwise deletion removes a case if a case has any missing value for any of the variables used in an analysis.

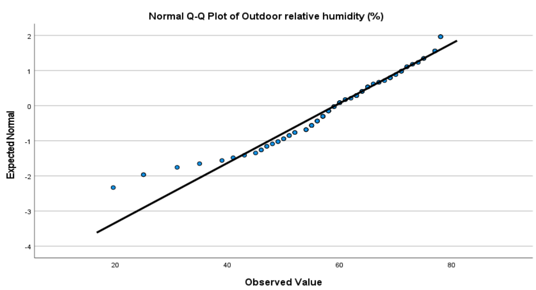
To assess normality, skewness and kurtosis statistics are assessed. Skewness refers to the symmetry of the distribution and kurtosis refers to the peakedness. Variables that have distributions that are very asymmetrical, flat, or peaked could bias any test that assumes a normal (i.e., bell-shaped) distribution. Generally, skewness and kurtosis values (converted as z-scores) that fall outside ±4 should be further inspected for potential outlier removal, nonparametric testing, or transformation. However, researchers may have flexibility in larger samples (Field, 2013). Some normality tests are done for sample sizes smaller than 100 (i.e., Shapiro–Wilks and Kolmogrov–Smirnov tests). If these tests are significant beyond $p < 0.001$, these variables should be further inspected (Gamst, Meyers & Guarino, 2008). Graphing methods are also employed for assessing normality. These graphs include histograms, normal quantile–quantile (Q-Q) plots and box plots, as shown in Figures 1 (a)-(d).

Statistics			
Outdoor relative humidity (%)			
N	Valid	100	
	Missing	0	
Mean		59.166	
Median		59	
Std. Deviation		11.76264	
Skewness		-0.853	-3.5394191
Std. Error of Skewness		0.241	
Kurtosis		1.39	2.90794979
Std. Error of Kurtosis		0.478	
Minimum		19.6	
Maximum		78	

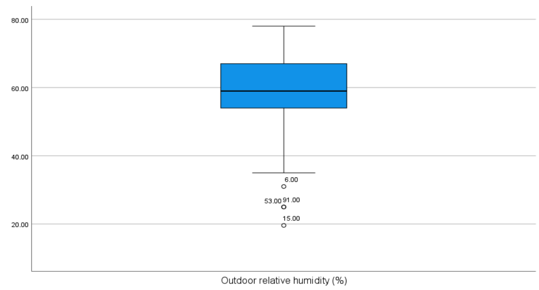
(a)



(b)



(c)



(d)

Fig.1: (a) Skewness and Kurtosis of the *on-site* recorded outdoor relative humidity (RH); (b) Histogram of the outdoor Relative Humidity (RH); (c) Normality analysis of the outdoor RH; (d) Whisker graph of the outdoor RH.

Histograms should look fairly bell-shaped. Q-Q plots should follow a straight line when plotting the expected values against the observed values. Box plots show the overall interquartile range and whether extreme values exist in the variable (see section on outliers below). If the data contains outliers, graphic displays both with and without the outliers should be examined to see how the graphs changed. If a continuous variable has serious deviations from normality, it must be addressed through transformation (log, inverse, Box-Cox, etc.), recoding into an ordinal variable, or assessment for whether nonparametric analysis needs to be conducted. Regardless, researchers should run analysis both with and without outliers to see whether the pattern of results changes. Univariate and multivariate outliers are also known as extreme values and can significantly bias any parametric test. We have checked our variables for univariate outliers. SPSS identifies these values as being three times the interquartile range beyond the 25th and 75th percentile values. In the present study, multivariate outliers were tested before the primary analyses were conducted where appropriate, as shown in Table 3.

Tab.3: Descriptive analysis of the variables related to identification of 'neutral' adaptive thermal comfort thresholds for benchmarking.

Variable Name	Mean	Median	Std. Deviation	Percentiles		
				25th	50th	75th
Overall thermal satisfaction in summer [0 to 6] - Discrete	1.64	2.00	1.453	1.00	2.00	3.00
Overall thermal satisfaction in summer [-3, +3] - Continuous	1.68	2.00	1.36241	1.00	2.00	3.00
Thermal sensation in bedroom 1 [0 to 6] - Discrete	0.52	1.00	1.396	0.00	1.00	2.00
Thermal sensation in bedroom 2 [0 to 6] – Discrete	0.46	1.00	1.290	0.00	1.00	1.00
Thermal sensation in bedroom 3 [0 to 6] – Discrete	0.55	1.00	1.234	0.00	1.00	1.00
Thermal sensation in kitchen [0 to 6] - Discrete	-0.35	-1.00	1.533	-2.00	-1.00	1.00
Thermal sensation in living room [0 to 6] - Discrete	0.20	0.00	1.595	-1.00	0.00	1.75
Thermal sensation in bedroom 1 [-3, +3] - Continuous	0.51	1.00	1.30651	0.00	1.00	2.00
Thermal sensation in bedroom 2 [-3, +3] - Continuous	0.45	1.00	1.23399	0.00	1.00	1.00
Thermal sensation in bedroom 3 [-3, +3] - Continuous	0.53	1.00	1.20147	0.00	1.00	1.00
Thermal sensation in kitchen [-3, +3] - Continuous	-0.42	-1.00	1.39393	-2.00	-1.00	1.00
Thermal preference	1.50	1.00	1.07778	1.00	1.00	3.00

In this study, a dataset related to occupants' thermal sensation votes (TSVs) was designed in accordance with the thermal comfort assessment convention recommended by Wang *et al.* (2018). As presented in Table 3, in accordance with Wang *et al.* (2018), the thermal sensation scale was set out in two conceptual assessment criteria in order to undertake the statistical analysis accurately. First, Wang *et al.* (2018) recommend a 7-point discrete thermal sensation scale that can be applied to assess occupants' TSVs. In this case, the TSV is set as an ordinal variable, thus enabling researchers to undertake Cramér's *V* test for the statistical analysis and apply the statistical findings whenever it is appropriate at the time of developing an evidence-based energy policy design. In the present study, the dataset was coded as follows: 0 = -3, 1 = -2, 2 = -1, 3 = 0, 4 = +1, 5 =+2, 6 = +3. Notably, the [-3, +3] scale band represents the outcome of occupants' thermal sensation as an ordinal measure used to accurately conduct the Cramér's *V* test. This type of coding was applied to determine households' TSVs gathered through a questionnaire survey. In the questionnaire survey, questions related to households' thermal sensation were ranked on a 7-point Likert scale that could be used as an ordinal measure. In this dataset, to provide consistency of the interpretation of households' TSVs [0 to 6], a coding range representation of thermal sensation scale band at [-3, +3] was used, which was developed by Fanger in the 1970s and was commonly used by thermal comfort scholars between 1990 and 2000.

Second, Wang *et al*. (2018) recommend a 7-point continuous thermal sensation scale that can be applied to assess occupants' TSVs. With the TSV set as a continuous variable, researchers are able to undertake Pearson's correlation analysis. This method of design is commonly applied by thermal comfort researchers to identify “neutral” adaptive thermal comfort thresholds. Using occupants’ TSVs as continuous variables is the most well-known method for reporting field survey findings concurrently with *in-situ* measurements or *on-site* environmental monitoring findings. This is an essential method of design that was developed by a team of experts at the University of California at Berkeley to contribute to the ASHRAE Global Thermal Comfort Database II (see at - https://repository.uel.ac.uk/item/89zv1). Further to this on-going method of analysis in thermal comfort studies, in the present study, the dataset was coded as follows: -3 = Cold, -2 = Cool, -1 = Slightly cool, 0 = Comfortable, +1 = Slightly warm, +2 = Warm, +3 = Hot. Notably, the [-3, +3] scale band enables thermal comfort researchers to identify “neutral” adaptive thermal comfort thresholds for benchmarking. In this dataset, to provide consistency of the interpretation of households’ TSVs, the [-3, +3] coding range represents the [Cold to Hot] thermal sensation scale, which was recommended by Fanger in the 1970s and further developed by de Dear in 1998 and 2001.

3. Results and Discussions

3.1. Assessing households’ socio-demographic characteristics

This section explores the relationships among the household socio-demographic characteristics, energy use and other variables collected from the respondents during the field-study period; and a set of questions developed to undertake statistical tests was utilised to determine the correlations between different socio-demographic characteristics based on the survey findings. The relationships between occupant age, economic status, education level, occupation, income, and health conditions were examined using cross tabulations with chi-square tests, as shown in Table 4.

Tab.4: Relationships Between Age Band, Economic Status, Education, Occupation, Income, and Health Conditions.

Research Questions	Age Band	Economic Status	Education Level	Occupation	Income	Health
Q 1.1: What is your age?	1	0,201	0,114	0,399***	0,213	0,496**
	—	0,447	0,989	0,000	0,324	0,000
Q 1.2: What is your economic status?	0,201	1	0,416**	0,171	0,136	0,178
	0,447	—	0,000	0,763	0,936	0,661
Q 1.3: What is your highest level of education?	0,114	0,416**	1	0,162	0,190	0,288
	0,989	0,000	—	0,648	0,333	0,196
Q 1.4: What is your occupation?	0,399***	0,171	0,162	1	0,275	0,342**
	0,000	0,763	0,648	—	0,092	0,000
Q 29: What is your monthly income?	0,213	0,136	0,190	0,275	1	0,174
	0,324	0,936	0,333	0,092	—	0,480
Q 28: How is your health in general?	0,496**	0,178	0,196	0,342**	0,174	1
	0,000	0,661	0,288	0,000	0,480	—

As shown in Table 4, age bands were significantly related to the health conditions, and this relationship was strong (χ^2 = 73,739, p < 0,001, Cramer’s V = 0,496). Younger age appeared to report better health conditions (good or very good) than older age. Household occupation was significantly related to health conditions, and the relationship was moderate (χ^2 = 33,071, p < 0,001, Cramer’s V = 0,342). A greater proportion of participants with very good condition worked outside the home than those with mediocre health conditions, whereas none of the participants who retired had very good conditions. Household occupations were also significantly associated with age with a moderate-strong relationship (χ^2 = 44,810, p < 0,001, Cramer’s V = 0,399). The results revealed that economic status was significantly related to educational level (χ^2 = 48,808, p < 0,001, Cramer’s V = 0,416). A greater proportion of participants with high school had full time jobs than those with only elementary or secondary school, whereas most participants who were pensioners graduated from elementary or secondary school, and none of undergraduate/postgraduate students were pensioners. Income was not significantly related to any variables.

Tab.4: Relationships Between Age Band, Economic Status, Education, Occupation, Income, and Health Conditions (Continued).

Health condition - Age, χ^2 (12) = 73,74, p = 0,000, *Cramer*'s V = 0,496
Health condition – Employment status, χ^2 (12) = 9,48, p = 0,661, *Cramer*'s V = 0,178
Health condition – Education level, χ^2 (9) = 10,83, p = 0,288, *Cramer*'s V = 0,196
Health condition – Occupation, χ^2 (9) = 33,07, p = 0,000, *Cramer*'s V = 0,342
Health condition – Income, χ^2 (9) = 8,63, p = 0,472, *Cramer*'s V = 0,170
Income - Age, χ^2 (12) = 8,63, p = 0,324, *Cramer*'s V = 0,213
Income – Employment status, χ^2 (12) = 5,58, p = 0,936, *Cramer*'s V = 0,136
Income – Education level, χ^2 (9) = 10,22, p = 0,333, *Cramer*'s V = 0,190
Income – Occupation, χ^2 (9) = 8,55, p = 0,480, *Cramer*'s V = 0,174
Occupation – Age, χ^2 (12) = 44,81, p = 0,000, *Cramer*'s V = 0,399
Occupation – Employment status, χ^2 (12) = 48,81, p = 0,000, *Cramer*'s V = 0,416
Occupation – Education, χ^2 (9) = 6,89, p = 0,648, *Cramer*'s V = 0,162
Age – Education, χ^2 (12) = 3.68, p = 0,989, *Cramer*'s V = 0,114
Age – Employment status, χ^2 (16) = 16.09, p = 0,447, *Cramer*'s V = 0,201
Employment status – Education, χ^2 (12) = 48.81, p = 0,000, *Cramer*'s V = 0,416

As shown in Table 4, feed-forward interviews of the residents of 100 flats revealed that age bands were significantly related to the health conditions, and this relationship was strong (χ^2 = 73,739, p < 0,001, Cramer's V = 0,496). Younger age appeared to report better health conditions (good or very good) than older age. Household occupations were also significantly associated with age with a moderate-strong relationship (χ^2 = 44,810, p < 0,001, Cramer's V = 0,399). Of the surveyed flats, 73% had owner-occupiers whose ages ranged between 55 and 65 or were 65 and older; these age bands were in the high-income group, and the energy consumption of these households was higher than the national average, all of which demonstrates an association between age and level-of-income factors, which suggests that household socio-demographic characteristics should be evaluated before any type of building retrofitting is developed.

3.2. Assessing occupants' thermal comfort

In Table 5 (a) Pearson's correlation analysis was undertaken to assess occupants' TSVs. To provide an accurate conventional method of design in thermal comfort studies, households' TSVs [-3, +3] coding range represents the [Cold to Hot] thermal sensation scale. Wang *et al.* (2018) recommended 7-point continuous thermal sensation scale which allows researchers to conduct Pearson's correlations at the time of identifying 'neutral' adaptive thermal comfort in longitudinal field studies. In Table 5 (b) Cramer's V test was undertaken to demonstrate relationships between households' TSVs, orientation factor of each RTB in the post-war social housing estate and floor level differences of each apartment in the RTB. To provide an accurate conventional method of design in thermal comfort studies, households' TSVs [0 to 6] coding range represents [-3, +3] thermal sensation scale band. Wang *et al.* (2018) recommended 7-point discrete thermal sensation scale could be applied to assess occupants' TSVs. This means that the variables related to occupants' TSVs could be used as ordinal variables to conduct the relevant Cramer's V test. It should be noted that in the questionnaire survey pro-forma, the survey was set on a 7-point Likert scale to assess occupants' TSVs.

In this respect, Cramer's V test was conducted to demonstrate the appropriateness of the chosen statistical analysis. Hence, according to adaptive thermal comfort theory which was developed by Haghighat and Donnini (1998) and Haldi and Robinsion (2008) recommended that [-3, +3] coding should be applied as a continuous variable to conduct Pearson's correlations while interpreting the households' TSVs. These scholars highlighted the interpretation of households' TSVs by selecting these variables which could not have a significant effect on the outcome to identify 'neutral' adaptive thermal comfort thresholds in a field study. In the present study, both Pearson's correlations and Cramer's V test were conducted and only Pearson's correlation findings were reported according to the statistical convention. Additionally, Cramer's V findings were presented to respect the convention in statistics. In this present study, Fisher's exact test was applied before undertaking Cramer's V test to avoid any research bias and demonstrate the statistical representation of this method of design. However, in general many thermal comfort studies recommend Pearson's correlation outputs could provide a reliable guidance to the researchers to measure the effect of households' TSVs.

Tab.5 (a): Relationships Between Occupant TSVs for Each Occupied Space in the Summer: Living Room, Kitchen, Bedroom 1, Bedroom 2, Bedroom 3.

Thermal sensation votes (TSV) for each occupied space		Living Room	Kitchen	Bedroom 1	Bedroom 2	Bedroom 3
Living Room	Pearson's correlation	1	0,462**	0,302**	0,146	0,200*
	Significance	—	< 0,001	0,002	0,147	0,046
Kitchen	Pearson's correlation	0,462**	1	0,133	0,205*	0,220*
	Significance	< 0,001	—	0187	0,041	0,028
Bedroom 1	Pearson's correlation	0,302**	0,133	1	0,763**	0,724**
	Significance	0,002	0187	—	< 0,001	< 0,001
Bedroom 2	Pearson's correlation	0,146	0,205*	0,763**	1	0,829**
	Significance	0,147	0,041	< 0,001	—	< 0,001
Bedroom 3	Pearson's correlation	0,200*	0,220*	0,724**	0,829**	1
	Significance	0,046	0,028	< 0,001	< 0,001	—
** Correlation is significant at the 0,01 level (two-tailed) *Correlation is significant at the 0,05 level (two-tailed)						

As shown in Table 5 (a), several strong and moderate positive correlations related to the occupants' decisions on TSVs in the summer were detected. TSVs in bedroom 1, bedroom 2, and bedroom 3 were strongly and positively correlated with each other (*r*s= 0,724 - 0,829, *p*s < 0,001). A moderate positive correlation was noted between the TSVs in the living room and kitchen spaces (r = 0,462, p < 0,001). TSVs in living room was significantly but weakly correlated with TSVs in bedroom 1 (r = 0,302, p = 0,002) and bedroom 3 (r = 0,200, p = 0,046). TSVs in kitchen was significantly but weakly related to TSVs in bedroom 2 (r = 0,205, p = 0,041) and bedroom 3 (r = 0,220, p = 0,028), which indicates that the position of the rooms in the flats should be taken into account to assess the occupants' thermal comfort and provide a basis for an ordinal logistic regression analysis; this is presented in Figures 2 (a)-(d). It is important to know occupants' TSVs in their bedrooms because of the significance of night-time sleep is strongly correlated with the hot and humid climate characteristics of this Eastern-Mediterranean region. Table 5 (b) demonstrates the Fisher's Exact tests if over 25% of cells had less than 5 expected counts that revealed relationships between occupant TSVs for each occupied space in the summer, orientation and floor level.

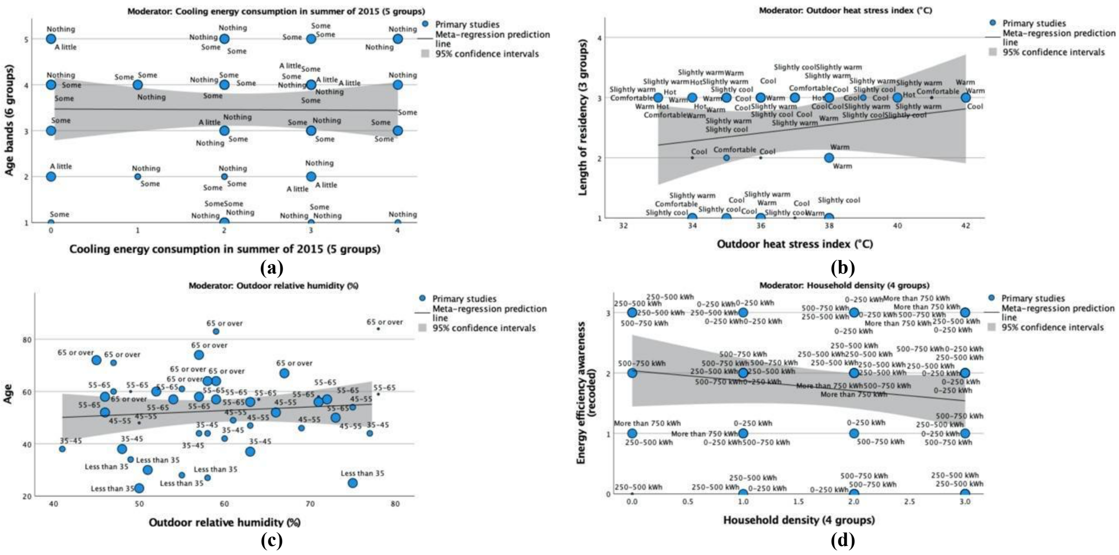


Fig.2: (a) Meta-regression analysis between age bands and cooling consumption in summer by exploring energy efficiency awareness; correlations between households' length of residency and outdoor heat stress index factor by thermal sensation votes integration; (c) regression analysis between households' age and outdoor relative humidity; (d) correlations between households' energy efficiency awareness and households' density by exploring energy consumption.

Tab. 5 (b): Relationships Between Occupant TSVs for Each Occupied Space in the Summer: Living Room, Kitchen, Bedroom 1, Bedroom 2, Bedroom 3, RTB Orientation and Floor Level.

Thermal sensation votes (TSV) for each occupied space		**Orientation**	**Floor Level**
Living Room	Cramer's V	0,226	0,232
	Significance	0,379	0,220
Kitchen	Cramer's V	0,279	0,222
	Significance	0,118	0,384
Bedroom 1	Cramer's V	0,274	0,177
	Significance	0,176	0,952
Bedroom 2	Cramer's V	0,272	0,194
	Significance	0,121	0,891
Bedroom 3	Cramer's V	0,263	0,221
	Significance	0,094	0,489
Orientation	Cramer's V	1	0,197
	Significance	—	0,188
Floor Level	Cramer's V	0,197	1
	Significance	0,188	—

Living room TSV – Orientation, *Fisher's exact* = 15,40, p = 0,379, *Cramer*'s V = 0,226
Kitchen TSV – Orientation, *Fisher's exact* = 19,72, p = 0,118, *Cramer*'s V = 0,279
Bedroom 1 TSV – Orientation, *Fisher's exact* =20,81, p = 0,176, *Cramer*'s V = 0,274
Bedroom 2 TSV – Orientation, *Fisher's exact* = 22,54, p = 0,121, *Cramer*'s V = 0,272
Bedroom 3 TSV – Orientation, *Fisher's exact* = 20,19, p = 0,094, *Cramer*'s V = 0,263
Floor level - Orientation, *Fisher's exact* = 12,11, p = 0,188, *Cramer*'s V = 0,197
Living room TSV – Floor level, *Fisher's exact* = 18,15, p = 0,220, *Cramer*'s V = 0,232
Kitchen TSV – Floor level, *Fisher's exact* = 15.35, p = 0,384, *Cramer*'s V = 0,222
Bedroom 1 TSV – Floor level, *Fisher's exact* = 10,23, p = 0,952, *Cramer*'s V = 0,177
Bedroom 2 TSV – Floor level, *Fisher's exact* = 10,09, p = 0,891, *Cramer*'s V = 0,194
Bedroom 3 TSV – Floor level, *Fisher's exact* = 13,86, p = 0,489, *Cramer*'s V = 0,221

Occupant TSVs for living room, kitchen and bedrooms 1, 2 and 3 in the summer: (0) to (6)
RTB orientation: 0 (north-east), 1 (south), 2 (north-west), 3 (south-west) and 4 (south-east)
Different floor levels: 0 (ground), 1 (first), 2 (second), 3 (third), 4 (fourth) and 5 (fifth

The results revealed that orientation and floor level were not significantly related to any TSVs. This was probably due to the small floor area of these spaces, which means the physical condition of the RTBs can lead to thermally uncomfortable indoor-air temperatures due to the poor window design in the interviewed flats. Data related to the occupants' adaptation to slightly warmer indoor-environment conditions and outdoor-air temperatures could be seen as a significant contribution to the ASHRAE Global Thermal Comfort Database II in terms of the delineation of a specific method to conduct a longitudinal field survey in this particular south-eastern Mediterranean climate and the prediction of neutral adaptive thermal comfort levels with the use of an ordinal logistic regression analysis (see at - https://doi.org/10.6078/D1F671). The present study also provides a roadmap to the EN 15251 thermal-comfort assessment criteria in the event that industry-based temperature design criteria are unable to comply with the ASHRAE Global Thermal Comfort Database II because they conflict with the occupants' adaptive comfort temperatures, as shown in Figure 3.

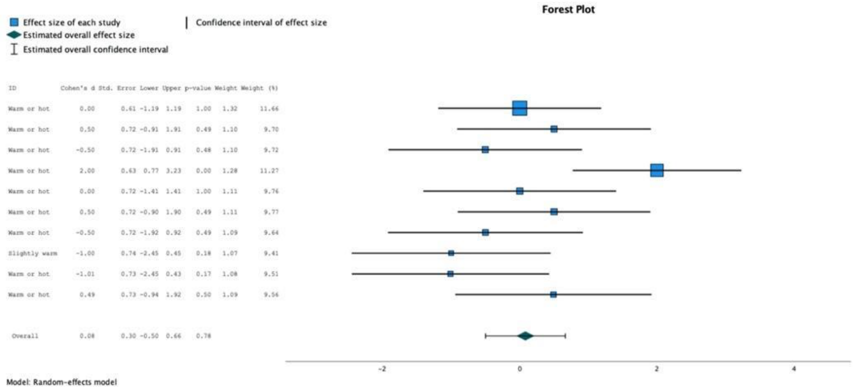


Fig.3: Households' thermal sensation votes by using Forest Plot analysis.

According to Figure 3, the recorded temperatures were above the acceptable benchmark of 25°C that was determined to maintain the occupants' thermal comfort (BSI, 2005; CIBSE, 2017; CEN, 2007). Additionally, the average mean temperatures that were recorded across the indoor measurement results and the outdoor monitoring results were ranged from 30,59–32,12°C, which is above the recommended thermal-comfort level of 23–25°C indicated by the CIBSE TM52 Overheating Task Force. It is worth noting that recorded daily running mean outdoor temperatures reflected the thermal experiences of the occupants more accurately than the monthly mean temperatures, because the outdoor mean temperatures sometimes changed in significantly shorter intervals (Nicol *et al*., 2012). Even though the monthly mean temperature was taken as an average temperature of the month as a whole in the present study, the occupants' TSVs were found to be correlated with their thermal experiences and their ability to adapt their physiological body temperatures to changing summer climate conditions (Ozarisoy & Altan, 2023).

4. Conclusions

The present study detailed a model that can be employed to investigate a range of potential retrofitting interventions related to housing-energy consumption and to address issues associated with overheating risks experienced during the summer; as such, the research design and methodology of the present study aims to enhance policymakers' understandings of the complex nature of energy consumption and occupant thermal comfort. The developed framework provides a scientific background that can be included in the building energy simulation (BES) platform to serve as a set of guidelines for the EPBD scientific committee for energy policy design and retrofitting strategies. The study findings can be extrapolated by current industry benchmarks or assessment criteria as a new European Norm (EN) that can be adopted by other EU countries. The findings of this study enhanced the overall understanding of the complex interrelationships between household socio-demographic characteristics, building thermal properties and occupants' habitual adaptive behaviour related to thermal comfort in heat-vulnerable MFHs. It was found that TSVs in living room was significantly but weakly correlated with TSVs in bedroom 1 (r = 0,302, p = 0,002) and bedroom 3 (r = 0,200, p = 0,046). TSVs in kitchen was significantly but weakly related to TSVs in bedroom 2 (r = 0,205, p = 0,041) and bedroom 3 (r = 0,220, p = 0,028). The present study is the first to follow the recommended methodology laid out in *EN 15251 – Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting and acoustics*. To date, no other studies have focused on the development of this particular *EN 15251* standard; as such, the present study developed a socio-technical-systems (STS) conceptual framework to provide a significant contribution to the body of knowledge related to a novel methodological framework for building performance evaluation studies.

5. Acknowledgments

The authors would like to acknowledge the Graduate School, School of Architecture, Computing & Engineering (ACE), London, United Kingdom. The lead author (Dr. Bertug Ozarisoy) would also like to acknowledge the Sustainable Environment and Energy Systems (SEES) Graduate Program at the Middle East Technical University Northern Cyprus Campus.
Data Access Statement: The datasets generated during and/or analysed during the current study are publicly available at - https://doi.org/10.15123/uel.8q713

6. References

Altan, H. (2004). Energy efficiency in housing: drivers and barriers to improving energy efficiency and reducing carbon dioxide emissions in private sector housing. PhD Thesis. University of Sheffield. ISNI: http://isni.org/isni/0000000134185603 (accessed on 10/05/2019)

Chasalow, K., & Levy, K. (2021). Representativeness in statistics, politics, and machine learning. In *FAccT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 77–89). Association for Computing Machinery, Inc. https://doi.org/10.1145/3442188.3445872

British Standards Institution (BSI). (2007). BS EN 15251: 2007: Indoor Environmental Input Parameters for Design and Assessment of Energy Performance of Buildings Addressing Indoor Air Quality, Thermal Environment, Lighting and Acoustics; BSI: London, UK, 2007.

CEN. (2007). Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting and acoustics. EN 15251. Technical report, European Committee for Standardization, Bruxelles. [accessed on 11/05/2018]

CIBSE. (2017). CIBSE TM59 - Design methodology for the assessment of overheating risk in homes. London: The Chartered Institution of Building Services Engineers.

Dong, B., Liu, Y., Fontenot, H., Ouf, M., Osman, M., Chong, A., … Carlucci, S. (2021). Occupant behavior modeling methods for resilient building design, operation and policy at urban scale: A review. Applied Energy, 293. https://doi.org/10.1016/j.apenergy.2021.116856

Field, A. (2013). Discovering statistics using IBM SPSS statistics (4th ed.). Los Angeles, CA: Sage.

Gamst, G., Meyers, L. S., & Guarino, A. J. (2008). Analysis of variance designs: A conceptual and computational approach with SPSS and SAS. New York: Cambridge University Press.

Garrett., H. E. (1942). The Representativeness of a Sample. The American Journal of Psychology 55, 4 (1942), 580–581. http://www.jstor.org/stable/1417131

Haghighat, F., & Donnini, G. (1999). Impact of psycho-social factors on perception of the indoor air environment studies in 12 office buildings. *Building and Environment*, *34*(4), 479–503. https://doi.org/10.1016/S0360-1323(98)00034-1

Haldi, F., & Robinson, D. (2008). On the behaviour and adaptation of office occupants. *Building and Environment*, *43*(12), 2163–2177. https://doi.org/10.1016/j.buildenv.2008.01.003

Hirsch, L. S., & O'Donnell, A. M. (2001). Representativeness in Statistical Reasoning: Identifying and Assessing Misconceptions. *Journal of Statistics Education*, *9*(2). https://doi.org/10.1080/10691898.2001.11910655

Hu, J. and Kohler-Hausmann, I. (2020). What's Sex Got to Do with Machine Learning?. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (Barcelona, Spain) (FAT* '20). Association for Computing Machinery, New York, NY, USA, 513. https://doi.org/10.1145/3351095.3375674

Jensen., A. (1926). Report on the Representative Method in Statistics. Bulletin de l'Institut international de statistique 22, 1 (1926), 359–438.

Kiaer., A. N. (1976). Den repræsentative undersøgelsesmethode = The representative method of statistical surveys. Oslo: Aschehoug.

Kruskal, W., and Mosteller, F. (1980). Representative Sampling, IV: The History of the Concept in Statistics, 1895-1939. International Statistical Review / Revue Internationale de Statistique 48, 2 (1980), 169–195. http://www. jstor.org/stable/1403151

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. Journal of the American Statistical Association, 83, 1198–1202.

Little, R. J. A., & Rubin, D. B. (2002). Statistical analysis with missing data (2nd ed.). New York: Wiley.

McNemar, Q. (1940). Sampling in psychological research. Psychological Bulletin 37, 6 (1940), 331–365. https://doi.org/10.1037/h0063476

Nicol F, Humphreys, M. and Roaf, S. (2012). Adaptive Thermal Comfort, Principles and practice, London Routledge.

Ozarisoy, B., & Altan, H. (2021). Developing an evidence-based energy-policy framework to assess robust energy-performance evaluation and certification schemes in the South-eastern Mediterranean countries. *Energy for Sustainable Development*, *64*, 65–102. https://doi.org/10.1016/j.esd.2021.08.001

Ozarisoy, B., & Altan, H. (2022). Significance of occupancy patterns and habitual household adaptive behaviour on home-energy performance of post-war social-housing estate in the South-eastern Mediterranean climate: Energy policy design. *Energy*, *244*. https://doi.org/10.1016/j.energy.2021.122904

Ozarisoy, B., & Altan, H. (2023). Energy Policy Design in the Eastern Mediterranean Basin: A Roadmap to Energy Efficiency, Taylor & Francis Group, Routledge Explorations in Energy Studies, ISBN 9781032493046, https://www.routledge.com/Energy-Policy-Design-in-the-South-Eastern-Mediterranean-Basin-A-Roadmap/Ozarisoy-Altan/p/book/9781032493046

Schmill, M. D., Gordon, L. M., Magliocca, N. R., Ellis, E. C., & Oates, T. (2014). GLOBE: Analytics for assessing global representativeness. In *Proceedings - 5th International Conference on Computing for Geospatial Research and Application, COM.Geo 2014* (pp. 25–32). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/COM.Geo.2014.21

Schouten, B., Cobben, F., & Bethlehem, J. (2009). Indicators for the representativeness of survey response. *Survey Methodology*, *35*(1), 101–113.

Tabachnick, B. G., & Fidell, L. S. (2007). Using multivariate statistics (5th ed.). Boston, MA: Pearson.

Wang, Z., de Dear, R., Luo, M., Lin, B., He, Y., Ghahramani, A., & Zhu, Y. (2018). Individual difference in thermal comfort: A literature review. Building and Environment. Elsevier Ltd. https://doi.org/10.1016/j.buildenv.2018.04.040