

PAPER • OPEN ACCESS

## Use of Beta Regression to investigate the link between home air infiltration rate and self-reported health

To cite this article: S. Lu *et al* 2021 *J. Phys.: Conf. Ser.* **2069** 012178

View the [article online](#) for updates and enhancements.

You may also like

- [Automated algorithms for detecting sleep period time using a multi-sensor pattern-recognition activity monitor from 24 h free-living data in older adults](#)  
Verónica Cabanas-Sánchez, Sara Higuera-Fresnillo, Miguel Ángel De la Cámara et al.
- [Executive dysfunction after multiple concussions is not related to cerebrovascular dysfunction](#)  
Erin D Ozturk, Mary Alexis Iaccarino, Jason W Hamner et al.
- [A review of physiological and behavioral monitoring with digital sensors for neuropsychiatric illnesses](#)  
Erik Reinertsen and Gari D Clifford



**IOP | ebooks™**

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

# Use of Beta Regression to investigate the link between home air infiltration rate and self-reported health

Lu S.<sup>1</sup>, Symonds P.<sup>1</sup>, Verschoor N.<sup>2</sup>, Chalabi Z.<sup>1</sup>, Taylor J.<sup>3</sup>, Davies M.<sup>1</sup>

<sup>1</sup>Institute of Environmental Design and Engineering, UCL, London, UK

<sup>2</sup>CFP Green Buildings, Apeldoorn, Netherlands

<sup>3</sup>Department of Civil Engineering, Tampere University, Finland

**Abstract.** The UK has introduced ambitious legislation for reaching net zero greenhouse gas (GHG) emissions by 2050. Improving the energy efficiency of homes is a key priority in achieving this target and solutions include minimising unwanted heat losses and decarbonising heating and cooling. Making a dwelling more airtight and applying insulation can result in a lower energy demand by reducing unwanted heat loss through fabric and openings. However, the supply of sufficient outdoor air is required to dilute indoor airborne pollutants. This research investigates the relationship between dwelling air infiltration and self-reported health at population neighbourhood level for Greater London. This paper links data from a variety of sources including Energy Performance Certificates (EPCs), the Greater London Authorities' Large Super Output Area (LSOA) Atlas and the Access to Healthy Assets and Hazards (AHAH) database at LSOA level. Beta regression has been performed to assess the influence of air infiltration rate on self-reported health, whilst controlling for other socioeconomic factors. All factors have been ranked in order of their association with self-reported health. Findings indicate that air infiltration rate has a positive association with the percentage of people reporting themselves to be in "good or very good" health.

## 1. Introduction

The UK government revealed its carbon emissions plan in 2019 that the country will target net zero emissions by 2050 [1]. In 2016, approximately 500,000 GWh of energy was consumed by the UK domestic sector, accounting for 30% of total consumption [2]. Space heating is responsible for roughly 60% of home energy consumption. Given that most dwellings that will exist in 2050 have already been built, improving home energy efficiency is a key priority. The UK Government's efforts to promote home energy efficiency began towards the end of 20<sup>th</sup> century [3]. The Decent Home scheme aimed at providing minimum standards of housing conditions for those who are housed in the public sector [4]. The Warm Front scheme helped the poorest UK households to install heating measures and insulate their homes [5]. The Green Deal assisted occupiers to pay for energy efficient home improvements ranging from building envelope to renewable energy [6].

Human health and wellbeing and the influence of the built environment is not a new topic. The first schemes in the UK date back to the 19<sup>th</sup> century when Edwin Chadwick in his report 'The Sanitary Conditions of the Labouring Population' revealed that diseases could be eradicated by improving the ventilation of London homes [7]. In the 20<sup>th</sup> century, Ebenezer Howard envisioned the garden city concept which aimed at providing citizens with a healthy living environment [8].



The direct impact of housing on human health has been acknowledged in many studies where poor housing increased risk for depression, anxiety, respiratory and cardiovascular diseases [9, 10]. Cold, damp and mould are some of the highest risk factors on occupant health. Cold is strongly associated with winter deaths, however better heating and higher SAP ratings can help alleviate this issue [11, 12]. Emissions from indoor combustion, volatile organic compounds (VOCs) and toxic metals are also found to be damaging to occupant health [13-15]. Summer overheating within homes also affects occupant thermal comfort [16]. These issues rely heavily on ventilation to help regulate the indoor environment, promoting human health and wellbeing.

This research extends recent work by the authors investigating the link between home energy efficiency and self-reported health in Greater London [17]. Here we link to additional neighborhood data in the form of the Access to Healthy Assets and Hazards (AHAH) version 2 to draw on wider environmental, social and demographic factors as opposed to using the index of multiple deprivation as in Symonds et al. [17]. This paper focuses on the link between air infiltration and self-reported health and ranks this variable in terms of its association with self-reported health at Large Super Output Area (LSOA) level to other environmental, social and demographic factors.

## 2. Methods

This study is comprised of secondary data analysis using data derived from LSOA Atlas, Access to Healthy Assets and Hazards (AHAH) v2, Strategic noise mapping and Energy Performance certificates (EPCs) databases. Data were filtered, sorted and combined at Lower Super Output Area (LSOA) level for London area (N=4,835). Variance inflation factors were calculated to avoid multicollinearity between independent variables in the regression. Statistical analyses were then carried out to investigate association between self-reported health and median air infiltration rate and other confounding factors at LSOA level. Figure 1 illustrates the methodology schematically.

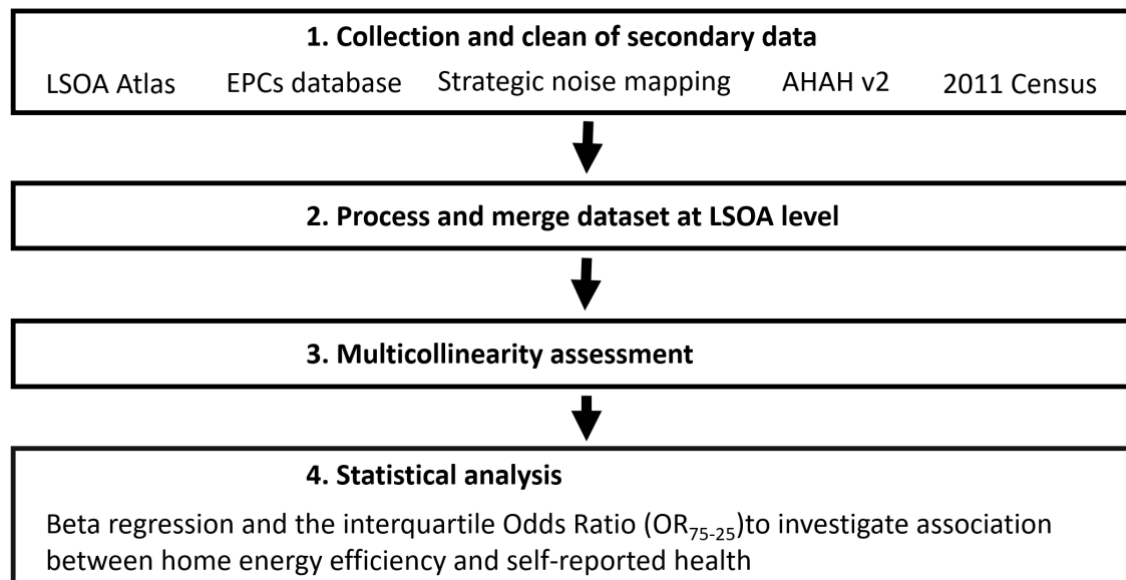


Figure 1. Research methodology.

### 2.1. Data and sources

Data was obtained primarily from four large datasets: LSOA Atlas, AHAH v2, EPCs and the 2011 census [18-21], with additional data sourced from other sources such as strategic noise mapping and Greater London Authority [22, 23]. Table 1 summarises each variable included in this study, the publication year and the data source.

Table 1: Variables included in the analysis.

Variable	Unit	Year(s)	Sources
Good or very good health	%	2011	2011 Census
Air infiltration	ach	2017	EPC database
Domestic energy consumption	kWh/m <sup>2</sup>	2011	2011 Census
One-person household	%	2011	2011 Census
Rail and road noise	dBA	2012	DEFRA
Air quality index	Index	2008	LAEI
Road casualties	Cases/ thousand	2011	Department for Transport
Distance to nearest GP practices	km	2017	AHAH
Distance to A&E hospitals	km	2017	AHAH
Distance to dentists	km	2017	AHAH
Distance to passive green spaces	km	2017	AHAH
Distance to active green spaces	km	2017	AHAH
Distance to blue spaces	km	2017	AHAH
Female fraction	0-100 (%)	2011	ONS: population data
Age	0-100	2011	
Owned outright housing	%	2011	2011 Census
Employment rate	%	2011	2011 Census
Annual household income	£	2011	GLA
Household without car access	%	2011	2011 Census
Average GCSE and equivalent point score per pupil	Score	2011/ 2012	2011 Census

2.1.1. *The Dependent variable* in this study is self-reported health. This data was collected as part of the 2011 census [20]. Residents were asked to rate their general state of health based on a five-point scale: ‘very good’, ‘good’, ‘fair’, ‘bad’ or ‘very bad’. In this study, subjective health is defined by the proportion of respondents who rated their personal health status as ‘good’ or ‘very good’. Therefore, subjective health is used as a proxy for health and wellbeing among residents within the local neighbourhood (LSOA). It should also be noted that self-reported ‘good’ or ‘very good’ health may not reflect the true clinical health status of individuals.

2.1.2. *The main independent variable* considered is the median air infiltration rate of homes within LSOAs and is expressed in air changes per hour (ach). Data relating to the energy efficiency characteristics of homes was extracted from the EPC database. The air infiltration rate at individual dwelling level is calculated based on the SAP methodology [21] and includes uncontrolled ventilation relating to:

- Chimneys, flues, fans and passive vents
- The number of stories in the dwelling
- Wall construction depending on steel, timber or masonry types
- Unsealed suspended floors or draught lobbies
- Unsealed windows and doors

LSOA level air infiltration is then obtained by calculating the median value for all dwellings inside an LSOA.

2.1.3. *Other independent variables considered* include built environment, demographic, social and economic factors. For built environment factors, the percentage of one-person households and domestic energy consumption which combines both domestic electricity and gas consumption were derived from 2011 census [20]. Road casualties per thousand persons were sourced from the Department for Transport [24]. Data on road and rail noise were derived from DEFRA's strategic noise mapping was averaged at LSOA level using GIS [19]. Air quality was measured using the Air Quality Index (AQI) which includes PM10, NO<sub>x</sub> and NO<sub>2</sub> [21]. Distance to GP practice, A&E hospitals, dentists, blue spaces, passive and active green spaces for each LSOA were all derived from AHAAH v2 database [18].

Demographic and economic factors, employment rates and households without car access were extracted from the 2011 census [20]. Female fraction and the median age of each LSOA were derived from ONS data [20]. Owned outright housing describes the percentage of households where the owner-occupier owns a house outright. Furthermore, estimated median annual household income was provided by the GLA [23]. In England, the General Certificate of Secondary Education (GCSE) is the qualification that marks the graduation of compulsory education [25]. The average GCSE point scores were taken from the department for education via neighbourhood statistics [20].

## 2.2. Statistical analysis

Python and R were used for data processing, analysis and visualization including calculating variance inflation factors, performing beta regression analysis, and calculating interquartile odds ratios and p-values [26, 27].

2.2.1. *Variance inflation factors* (VIFs) measure the amount of multicollinearity in a set of multiple regression variables [28]. In this study, VIFs were used to identify the existence and severity of multicollinearity between independent and confounding variables. Variables with VIF >5 were subsequently excluded from the regression analysis.

2.2.2. *Beta regression* was used to examine the relationship between self-reported health and the independent variables (covariates). Proposed by Ferrari and Cribari-Neto, beta regression model is designed for continuous variates where values are assumed to reside in the standard unit interval. This method assumes that the dependent variable is beta-distributed with its mean related linearly to independent variables via coefficients and a link function [26]. Beta regression is therefore appropriate for this analysis as the self-reported health metric is constrained to within 0-100% (or 0-1). In R, the *betareg* R package was used to implement maximum likelihood regression and establish a logit link function [29]. The dependant variable was hereby mapped from the original 0-1 to the range of real values with beta regression coefficients  $\beta_k$  identified for each independent k<sup>th</sup> covariate.

2.2.3. *The interquartile odds ratios* (ORs) have been used to quantify the strength of the association between covariates and self-reported health. ORs represent the odds for outcome A to occur given exposure B exists, compared to the odds for outcome A to occur given exposure B's absence [30]. The OR may be calculated for each k<sup>th</sup> covariate as:  $e^{c\beta_k}$ , where c represents one unit increase in the covariate [31]. In this analysis we calculate the interquartile OR, which is the ratio of the odds for each covariate at its 75<sup>th</sup> vs the 25<sup>th</sup> centile:

$$OR_{75-25} = e^{(X_k^{75} - X_k^{25})\beta_k} = \frac{e^{\beta_k X_k^{75}}}{e^{\beta_k X_k^{25}}}$$

Here  $\beta_k$  represents the fitted beta regression coefficient and  $(X_k^{75} - X_k^{25})$  is the absolute difference between the 25<sup>th</sup> and 75<sup>th</sup> centile for each covariate. This helps compare associations between different covariates since they all have different units, distributions and scales.

2.2.4. *Probability values (p-values)* are defined as the largest probability under the null hypothesis about the unknown distribution of the test statistic to have observed a value as or more extreme than the actual observed value [32]. A small p-value (less than 0.05) means the observed outcome is possible but not very likely under the null hypothesis. A p-value of less than 0.05 is said to be statistically significant at the 95% confidence level.

### 3. Results

#### 3.1. Descriptive statistics

Table 2 summarises the characteristics of all variables in this study at Greater London LSOA Level. On average, 84% percent of people within an LSOA reported ‘good or very good’ health. For London dwellings, the median air infiltration rate is around 0.6 ach.

Table 2: Summary statistics of dependent and independent variables for Greater London LSOAs

Variable	Percentiles			
	Mean	25%	50%	75%
Self-reported health (%)	83.8	81.1	83.8	86.5
One person household (%)	30.5	24.1	30.0	36.2
Air infiltration (ach)	0.61	0.56	0.61	0.66
Domestic energy consumption (kWh/m <sup>2</sup> )	447.2	352.9	391.2	461.3
Rail and road noise (dBA)	82.4	72.5	75.0	135.0
Distance to nearest GP practices (km)	0.72	0.44	0.63	0.89
Distance to A&E hospitals (km)	4.46	2.80	4.19	5.81
Distance to dentists (km)	0.78	0.46	0.67	0.97
Distance to blue spaces (km)	2.17	1.10	1.88	2.89
Distance to passive green spaces (km)	0.74	0.25	0.47	0.85
Distance to active green spaces (km)	0.41	0.24	0.36	0.52
Air quality index	100.2	89.6	97.2	107.9
Road casualties (/thousand)	8.77	1.77	5.53	11.71
Female fraction	0.51	0.49	0.51	0.52
Median age (years)	35.23	31.75	34.20	37.57
Owned outright housing (%)	21.96	11.60	20.20	30.70
Employment rate (%)	62.39	57.04	63.13	68.22
Annual household income (£)	35829.5	27046.5	32663.8	42305.9
Household without car access (%)	37.83	22.40	36.90	53.65
Average GCSE Point Score	479.87	440.35	477.10	517.00

#### 3.2. Beta regression

Regression results which yield interquartile ORs for air infiltration and covariates with self-reported health are presented in figure 2. Results indicate that air infiltration has a positive and statistically significant association with self-reported health. Air infiltration shows positive association with self-reported health. It had the 8<sup>th</sup> highest interquartile OR amongst the variables which had positive associations.

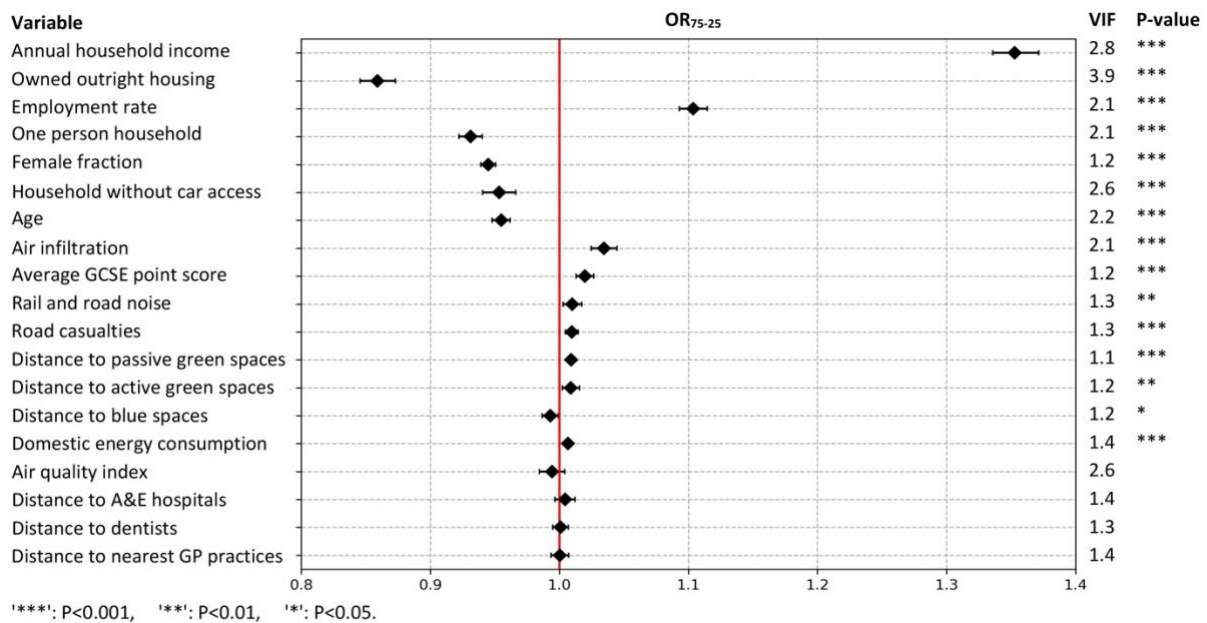


Figure 2: ORs by association strength of covariate for Greater London LSOAs

#### 4. Discussion

This research aims to analyze the relationship between home air infiltration rate and self-reported health in Greater London. Results suggest a clear link exists between air infiltration rate and self-reported health at population level. In terms of significance and positive association, air infiltration ranks lower in terms of its interquartile OR than some of the key socio-demographic factors such as household income, employment rate and age, but is above the education variable (Average GCSE point score) and all environmental variables including distance to green and blue spaces, road and rail noise and the air quality index.

Increasing the air tightness of dwellings is widely accepted as a means of cutting down carbon emissions from the buildings sector, by reducing heat losses through bulk air movement. However, the statistically significant positive association between air infiltration and self-reported health observed in this study suggests higher air infiltration rates may positively contribute to subjective health. This could be due to higher uncontrolled ventilation allowing indoor moisture, pollutants, VOCs and odors to be diluted with fresh outdoor air [13-16].

The observed relationship between air infiltration rate and self-reported health is consistent with previous studies. Hamilton et al.'s modeling study revealed that compensatory ventilation is required in airtight dwellings to help reduce indoor pollutants, and hence improve health [12]. Hence, it is critical to maintain appropriate levels of background ventilation whilst improving air tightness when it comes to promoting occupant health and wellbeing.

#### 5. Limitations

The study has several limitations. The first limitation relates to the variables included in this analysis. While this study adopted secondary data for regression analysis, variable collinearity is an important issue in regression analysis. Due to high collinearity with other variables, some predictors were removed. Although only variables with a VIF of less than 5 were included, some degree of collinearity between covariates still exists.

Another limitation concerns the estimation of air infiltration rates using EPC data based on SAP calculations. Despite the large uncertainties in the SAP methodology [33], air infiltration rate only describes unintended ventilation and intended ventilation (e.g. window opening, extract fans, etc...) is excluded.

Variation in the time horizon of the used data is a third limiting factor. Air quality data was collected in 2008. However, health service, green and blue space locations were gathered in 2017. The data collection period spans over a ten-year period, which inevitably leads to inconsistencies. Also, since the self-reported health survey was carried out in spring, seasonal variations in mood for subjective wellbeing are also not accounted for [34].

The fourth limitation relates to the studied area and generalisability. This research only considers homes within Greater London which has a unique housing stock and climate. Finding may not be applicable to stocks which rely more on mechanical ventilation in different climate zones. Future research could examine whether these associations are applicable to other cities and climates.

A final limitation is that results are aggregated at LSOA level (1,000-3,000) people which only allows inference at neighbourhood level. Data at individual level would be preferable, but it is much more difficult to match health records and personal data at this resolution.

## 6. Conclusion

This study investigated the relationship between dwelling infiltration rate and self-reported health at population (LSOA) level for Greater London. Whilst annual household income and housing tenure appear to be the main indicators of subjective wellbeing, results show a positive and statistically significant relationship between air infiltration and self-reported health. These findings imply the importance of appropriate compensatory ventilation in homes which have undergone air tightening. It is thus crucial for policy developers and engineers to consider indoor environmental impacts while retrofitting the existing building stock.

## References

- [1] BEIS 2019 UK becomes first major economy to pass net zero emissions law. [Online]. Available at: [www.gov.uk/government/news/uk-becomes-first-major-economy-to-pass-net-zero-emissions-law](http://www.gov.uk/government/news/uk-becomes-first-major-economy-to-pass-net-zero-emissions-law). [Accessed 1 February 2021].
- [2] Department of Business Energy & Industrial Strategy. Energy Consumption in the UK (ECUK) 1970 to 2019, Department of Business Energy & Industrial Strategy, London, 2020
- [3] Mavrogianni A, Davies M, Kolokotroni M, Hamilton I, editors. A GIS-based bottom-up space heating demand model of the London domestic stock. Proceedings 11th International IBPSA Conference, *Building Simulation*; 2009.
- [4] National Audit Office, “*The Decent Homes Programme - National Audit Office*,” The Stationery Office, London, 2010
- [5] National Audit Office, “*The Warm Front Scheme - National Audit Office*,” National Audit Office, London, 2009.
- [6] Department of Energy and Climate Change, “*The Green Deal*” Department of Energy and Climate Change, London, 2010.
- [7] Chadwick E. *The sanitary conditions of the labouring population*. Edited by Flinn MW. Edinburgh: Edinburgh University Press; 1842.
- [8] Howard E. *Garden cities of to-morrow*: Routledge; 2013.
- [9] Marmot M, Allen J, Goldblatt P, Boyce T, McNeish D, Grady M. *Fair society, healthy lives. Strategic review of health inequalities in England post-2010*. 2011.
- [10] Ige J, Pilkington P, Orme J, Williams B, Prestwood E, Black D, et al. The relationship between buildings and health: a systematic review. *Journal of public health*. 2019;**41**(2):e121-e32.
- [11] Packer CN, Stewart-Brown S, Fowle SE. Damp housing and adult health: results from a lifestyle study in Worcester, England. *Journal of Epidemiology & Community Health*. 1994;**48**(6):555-9.
- [12] Hamilton I, Milner J, Chalabi Z, Das P, Jones B, Shrubsole C, et al. Health effects of home energy efficiency interventions in England: a modelling study. *BMJ open*. 2015;**5**(4):e007298.
- [13] Burr ML. Health effects of indoor combustion products. *Journal of the Royal society of Health*. 1997;**117**(6):348-50.



- [14] Jones AP. Indoor air quality and health. *Atmospheric environment*. 1999;**33**(28):4535-64.
- [15] Tan SY, Praveena SM, Abidin EZ, Cheema MS. A review of heavy metals in indoor dust and its human health-risk implications. *Reviews on environmental health*. 2016;**31**(4):447-56.
- [16] Mavrogianni A, Wilkinson P, Davies M, Biddulph P, Oikonomou E. Building characteristics as determinants of propensity to high indoor summer temperatures in London dwellings. *Building and Environment*. 2012;**55**:117-30.
- [17] Symonds P, Verschuur N, Chalabi Z, Taylor J, Davies M. Home Energy Efficiency and Subjective Health in Greater London. *J Urban Health*. 2021 Apr 23. doi: 10.1007/s11524-021-00513-6. Epub ahead of print. PMID: 33893624.
- [18] CDRC. 2020. Access to Healthy Assets & Hazards. [Online]. Available at: [data.cdrc.ac.uk/dataset/access-healthy-assets-hazards-ahah](https://data.cdrc.ac.uk/dataset/access-healthy-assets-hazards-ahah) [Accessed 01 February 2021].
- [19] MHCLG. 2019. Ministry of Housing, Communities & Local Government EPC Registers. [Online]. Available at: [www.epcregister.com](http://www.epcregister.com). [Accessed 01 February 2021].
- [20] ONS. 2011. 2011 Census. [Online]. Available at: [/www.ons.gov.uk/census/2011census](http://www.ons.gov.uk/census/2011census). [Accessed 01 February 2021].
- [21] Greater London Authority. 2014. LSOA Atlas. [Online]. Available at: [data.london.gov.uk/dataset/lsoa-atlas](https://data.london.gov.uk/dataset/lsoa-atlas) [Accessed 01 February 2021].
- [22] DEFRA. 2015. Strategic noise mapping Explaining which noise sources were included in 2012 noise maps. [Online]. Available at: [assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/453620/noise-mapping-source-guidance.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/453620/noise-mapping-source-guidance.pdf) [Accessed 01 February 2021].
- [23] Greater London Authority. 2019. Household Income Estimates for Small Areas. [Online]. Available at: [data.london.gov.uk/dataset/household-income-estimates-small-areas](https://data.london.gov.uk/dataset/household-income-estimates-small-areas) [Accessed 18 May 2020].
- [24] Department for Transport. 2015. Road Casualties by Severity. [Online]. Available at: [data.london.gov.uk/dataset/road-casualties-severity-borough](https://data.london.gov.uk/dataset/road-casualties-severity-borough) [Accessed 18 May 2020].
- [25] Ofqual. 2015. Completing GCSE, AS and A level Reform. [online]. Available at: [www.gov.uk/government/consultations/gcse-as-and-a-levels-new-subjects-to-be-taught-in-2015](http://www.gov.uk/government/consultations/gcse-as-and-a-levels-new-subjects-to-be-taught-in-2015) [Accessed 01 February 2021].
- [26] Zeileis A, Cribari-Neto F, Grün B, Kosmidis I. Beta regression in R. *Journal of statistical software*. 2010;**34**(2):1-24.
- [27] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*. 2011;**12**:2825-30.
- [28] Thompson CG, Kim RS, Aloe AM, Becker BJ. Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. *Basic and Applied Social Psychology*. 2017;**39**(2):81-90.
- [29] Zeileis A, Cribari-Neto F, Gruen B, Kosmidis I, Simas AB, Rocha AV, et al. Package ‘betareg’. R package. 2016;**3**:2.
- [30] Bland JM, Altman DG. The odds ratio. *Bmj*. 2000;**320**(7247):1468.
- [31] Szumilas M. Explaining odds ratios. *Journal of the Canadian academy of child and adolescent psychiatry*. 2010;**19**(3):227
- [32] Mahuli A, Mahuli S. Significance of statistical significance. *Journal of Dental Research and Review*. 2015;**2**(3):106-.
- [33] Jenkins D, Simpson S, Peacock A. Investigating the consistency and quality of EPC ratings and assessments. *Energy*. 2017;**138**:480-9.
- [34] Murray G, Allen NB, Trinder J. A longitudinal investigation of seasonal variation in mood. *Chronobiology international*. 2001;**18**(5):875-91.