**Hard to Count index for the 2021 Census**

**Project**: Hard to Count index for the 2021 Census

**Project SRO**: Iain Bell

**Project Executive**: Shaun Garvey

**Document Control**

|  |  |
| --- | --- |
| **Product Reference** | To be assigned by the Census Programme Management Office when final version is available |
| **Description** | Hard to Count index for the 2021 Census, developed by MDR/Methodology for the 2021 Census |
| **Owner** | Ercilia Dini |
| **Signed off** | Owen Abbott |
| **Consulted** | * PPP Transformation/Census Statistical and Outputs Design (Garnett Compton/Cal Ghee/Orlaith Fraser) * PPP Operations/Data Collection Services (Henry Brown) * PPP Operations/Census Field Operations (Lindsay Cracknell, Tansy Arthur) * Policy and Communication/Stakeholder Engagement (Lara Phelan, Neil Townsend) * Methods Data and Research/Methodology (Marie Cruddas) * Census Research Assurance Group (CRAG) |
| **Informed** | * PPP Transformation/Service Design and Management (Andy Teague) * MDR/Data Architecture/Address Register (Alister Calder, Mike James, Neil Hopper) * MDR/Methodology/Big Data (Owen Abbott) |
| **Related products**  **(input to projects)** | * 2021 Census Coverage Survey Design (Adriana Castaldo) * 2021 Census Coverage Assessment (Viktor Racinskij) * 2021 Census Coverage Adjustment (Alison Whitworth) * Field Operations Simulation (Peter Barber) * Census Statistical and Outputs Design (Orlaith Fraser) |

**Version Control**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Version** | **Created By** | **Date** | **Status** | **Changes Made** | **Written by** |
| 1.0 | Ercilia Dini | 4 April 2018 | V 1.0 | First draft paper sent for comments | Ercilia Dini |
| 2.0 | Ercilia Dini | 26 June 2018 | V 2.0 | Paper revised based on comments from Paul Smith, Marie Cruddas, Orlaith Fraser and Peter Barber | Ercilia Dini |
| 3.0 | Ercilia Dini | 17 July | V 3.0 | Paper revised based on comments from Salah Merad | Ercilia Dini |
| 4.0 | Ercilia Dini | 17 August 2018 | V 4.0 | Paper revised based on comments from CRAG members (meeting held on 8 August 2018) | Ercilia Dini |
| 5.0 | Ercilia Dini | 14 Sep 2018 | V 5.0 | Paper revised based on comments from Owen Abbott | Ercilia Dini |

**Hard to Count index for the 2021 Census**

­­­­­­

Ercilia Dini, Methodology, Office for National Statistics

September 2018

**Summary**

The 2021 Census, unlike previous censuses, will undertake the collection using an online questionnaire as the primary response mode. This change in the basic collection mode means that the non-response patterns observed in certain population groups in the 2001 and 2011 censuses may be different in the 2021 Census.

In addition to ‘willingness’ – as an innate propensity to respond to the census there will be an additional component which is the ability to respond, driven by access and use of digital technology. People who are ‘digitally excluded’ will require digital assistance or a way to respond to the census that may not be via the primary mode.

The ONS is conducting research to develop a Hard-to-Count (HtC) index to identify sub-populations/geographical (Lower Super Output) areas at risk of census non-response. This will be used as a tool in the 2021 Census to support pre-planning of field follow-up. It will also be used as a stratification variable in the sample design of the 2021 Census Coverage Survey and as a covariate in the Census estimation and adjustment methods. The HtC index will be key to achieving high quality census estimates.

The HtC index is composed of two domains: the *digital* domain and the *willingness to self-respond* domain. The digital domain is built as an indicator to measure infra-structures available which enable households to respond to an online census. The willingness domain is constructed using an area level (Lower Super Output) model that predicts non-response by day 10 after census day. The covariates used to build the model parameters are from previous census and updateable administrative data sources.

This paper describes the methodology used to develop the two HtC domains, results obtained and recommendations for further research.

**Table of Contents**

[1. Introduction 3](#_Toc517783026)

[2. Background 3](#_Toc517783027)

[3. Method 5](#_Toc517783028)

[3.1. Criteria to select domains and indicators 5](#_Toc517783029)

[3.2. Building the domains: data sources and method used 6](#_Toc517783030)

[3.2.1. Digital domain 6](#_Toc517783031)

[3.2.2. Willingness (self-response) domain 7](#_Toc517783032)

[3.3. Predicted willingness to self-respond versus Predicted final response rates 9](#_Toc517783033)

[4. Assigning hard-to-count categories to domains 10](#_Toc517783034)

[5. HtC index results 10](#_Toc517783035)

[6. 2017 Census test results 13](#_Toc517783036)

[7. Next steps 14](#_Toc517783067)

[8. References 15](#_Toc517783071)

[Annexes 16](#_Toc517783072)

[Annex A 16](#_Toc517783073)

[Annex B 16](#_Toc517783074)

[Annex C 23](#_Toc517783075)

1. **Introduction**

The provision of a Hard-to-Count (HtC) index is part of the Maximising Response Design product, which aims to support the strategic goals of maximising response and minimising the variability of response across areas in the 2021 Census (Abbott, 2015).

The HtC index will be used to address the following 2021 Census requirements:

1. To predict area self-return patterns by day 10 after the census (willingness to self-respond or not to a census).
2. To support planning of follow-up resources, and their allocation.
3. To support planning where paper questionnaires are to be sent as first option.
4. To identify where help or digital assistance will be required.
5. To be as a stratification variable for the Census Coverage Survey (CCS) sampling.
6. To be used in the census estimation and adjustment methods.
7. **Background**

At the time of the 2001 Census there was some information from the census validation survey (Heady et al, 1994) and from the ‘estimating with confidence’ project (Simpson et al, 1997) that non-response was associated with certain households and persons characteristics. These studies provided evidence for the need to develop a HtC index for the 2001 Census.

For the 2001 Census a HtC index was designed and used as the primary stratifying variable within both the design of the CCS and the One Number Census (ONC) estimation process (ONS, 2000).

To build the HtC index for the 2001 Census, ONS used information from the 1991 Census percentages of:

* Multi-occupied households;
* Private rented households;
* Unemployed Persons;
* Persons whose first language was not English (country of birth was used as proxy for this information); and
* Imputed households.

The rationale was that it had been observed from previous surveys that the higher these percentages the higher the levels of non-response. These percentages were summed and ranked, and a split was applied to census Enumeration Districts**[[1]](#footnote-1)**. Simulation studies were undertaken to evaluate the impact of four different distributions on the precision of population estimates. A split of 40%, 40% and 20% was chosen with the 20% being the hardest to count.

The decision on the split was based on the results of the simulation studies. They indicated that the distribution itself was not particularly important, provided that the HtC score was a good predictor of undercount. With the evidence available at the time and with uncertainty that the scores would be highly correlated with undercount, this coarse division was robust to misspecification. The index was applied as part of the ONC methodology, which was broadly successful in measuring non-response in the 2001 Census.

For developing the methods to be used in the 2011 Census there was more evidence, obtained from the 2001 Census Coverage Survey, on how census non-response was associated with certain characteristics of households and the population (Rahman and Goldring, 2006).

This allowed more detailed studies into census non-response. To develop the HtC categorisation for the 2011 Census, firstly, a model identifying factors associated with census non-response was fitted using record level data from matched 2001 Census and the CCS (Rahman and Goldring, 2006). To ensure subsequent predictions were based on more up to date sources, updateable proxies were obtained for the factors identified by this first model.

A second model was then constructed, this time using the proxies as area level covariates and the 2001 Census area level imputation rates as the outcome variable. The key assumption was that the structural relationship between non-response propensity and the key factors remained constant between the 2001 and 2011 censuses. This second model predicted area level non-response rates for the 2011 Census. The model predictions were used to rank Lower Super Output Areas (LSOAs) in England and Wales from the lowest predicted non-response rates to the highest. Based on the rank a categorisation was assigned splitting the distribution into 5 levels, the highest being the hardest to count. The percentages of LSOAs in each category were: 40% in HtC1, 40% in HtC2, 10% in HtC3, 8% in HtC4 and 2% in HtC5 (Hopper, 2011).

The design of the 40/40/10/8/2 categorisation was driven by consistency with the stratification for the 2001 Census coverage assessment and adjustment process which used a 40/40/20 split. The case for splitting those LSOAs in the 20% hardest to count group was driven by the observation that the largest variation in non-response rates in the 2001 Census was within this group. By splitting these 20% we could design the CCS sample more efficiently. There was relatively little difference in final non-response rates between LSOAs within the easiest 80% of LSOAs in 2001 and the ability of the model to differentiate between them was weak. However, there was a fairly steep gradient of differences between LSOAs in HTC categories 3, 4 and 5 (Hopper, 2011).

The primary aim of the 2011 HtC categorisation was to be used as a stratifying variable in the sample design of the CCS (ONS, 2012a). However, an innovation was to use the HtC information in the field work to allocate field staff resources across England and Wales, as well as for targeting census publicity. This approach integrated the 2011 Census collection and household and population estimates (ONS, 2012b), with the aim of maximising response rates and minimising variability. The strategy was successful (Abbott and Compton, 2014).

Return rates were monitored live during the 2011 Census. Mean return rates by HtC observed on the 7th April 2011, which related to the beginning of the field operation to follow up non-responding households, were in most areas equal to or higher than those predicted by the model. The variation in return rates observed by LSOA within each HtC category at the beginning of follow up was in line with that expected given the fit of the model to 2001 data. This provided evidence that the HtC categorisation worked well, both as a relative measure and as a specific measure of household non-response in the 2011 Census.

For the 2021 Census, a Hard-to-Count index is once again required as part of the strategy for delivering a high-quality set of census statistics. The overarching strategy for the field operation is to maximise response rates and minimise variability. Achieving this will require prior information about likely non-response patterns both in terms of self-response, and during field follow up. In addition, there will once again be a Census Coverage Survey, designed specifically to measure census non-response. This will require stratification and allocation methods which ensure that the sample is spread with respect to the likely levels of census non-response (after the census field follow-up has completed). Lastly, the estimation methods employed to derive the critically important census-based population estimates will be made more efficient with variables which explain the variation in census non-response.

However, the challenge for the 2021 Census Hard-to-Count index is that, unlike the 2001 and 2011 Censuses, the 2021 Census will use an online questionnaire as the primary response mode. This change in the basic collection mode means that the non-response patterns observed in certain population groups in the 2001 and 2011 Censuses may be different in the 2021 Census. In addition to ‘willingness’ – as an innate propensity to respond to the census there will be an additional component which is the ability to respond, driven by access and use of digital technology.

The main population groups or areas that will be harder-to-count in the 2021 Census will include people who are digitally excluded; such people will require help or a way to be able to respond to the census that may not be the primary mode (online self-completion).

Several options for the development of a HtC index for the 2021 Census were presented to the Census Research Assurance Group (CRAG) meeting held in May 2016 (Dini and Abbott, 2016).

The method endorsed by the members gives a set of relative measures of small geographical areas’ difficulty to be counted in a census, based on different domains related to this difficulty. Each of these domains can include one or more indicators**[[2]](#footnote-2)**. This is based on the idea that separate domains of areas’ difficulty to be counted in the census can be recognised, measured and used separately. The method takes inspiration from the widely used index of multiple deprivation (DCLG, 2015).

This paper presents the research conducted so far to develop two HtC domains for the 2021 Census: the digital and the predicted willingness to self-respond to the census. Each of these domains may be used as an index, per se.

Section 3 presents the method used to select these domains and indicators. It also presents the data sources and methodology used to build the indicators within each domain. Section 4 presents the method used to assign HtC categories into the digital and willingness to self-respond domains. Section 5 presents the results obtained so far for the Digital and Willingness domains of the HtC index. Section 6 presents results of self-response in the 2017 Census test with an analysis of the effect on self-response by sending a paper questionnaire or a Unique Access Code (UAC) as the first approach to households. Section 7 presents the next steps in the development of the HtC index. Section 8 contains the references used in this report and the Annexes provide more detail about the methodology and data sources used.

1. **Method**
   1. **Criteria to select domains and indicators**

The criteria to select domains were that they should:

* Be relevant to support the strategic goals of maximising response and minimising the variability of response across areas in the 2021 Census.
* Be robust to be used as a stratification variable for the CCS, and in the census estimation and adjustment methods.
* Take into account the census online mode of data collection.
* Take into account previous experience with surveys and censuses about population groups’ difficulty to be counted.

The criteria to select indicators within domains were that they should:

* Be ‘domain specific’ and appropriate to measure each domain.
* Be chosen to provide the best possible measure of each domain that contributes to the difficulty to be counted in an online census.
* Be derived using the most up-to-date data source and (as far as possible) be updatable at least yearly.
* Be statistically robust at the small area level. We propose Lower Super Output Area (LSOA) level as the highest level of geography.
* Be available for the whole of England and Wales at a small area level in a consistent form.
  1. **Building the domains: data sources and method used**

Two domains were chosen: a digital domain and a census willingness (self-response to census by day 10 after the census) domain.

The HtC index built for the 2011 Census was a ‘willingness to respond to a census’ and was derived from a model based on the 2001 final response rates. The index was used for both, to support planning of follow-up resources and their allocation, and to stratify the CCS sample.

This time information on day 10 return rates from the 2011 Census was available. Therefore, it was possible to build a model based on day 10 returns to be used to plan field follow-up. The rationale for this is that in the 2011 Census the field follow-up started after day 10. Returns before day 10 were therefore self-motivated.

As for the 2011 Census, this time it was also possible to build a model based on 2011 Census final response rates to be used as a stratifier for the CCS and in the census estimation and adjustment methods. However, having more than one HtC index may not be desirable.

This section presents the method used to build the digital and the willingness to self-respond domains. It also presents the method used to build a willingness domain based on a model using the 2011 Census final response rates. Finally, it compares the two models examining if they give different results that would justify the use of two willingness domains.

* + 1. ***Digital domain***

The 2021 Census, unlike the 2011 Census, will undertake the collection using an online questionnaire as the primary response mode. The main population groups or areas that will be harder to count in the 2021 Census will include people who are **digitally excluded[[3]](#footnote-3)** – such people will require digital assistance or a way to be able to respond to the census that may not be the primary mode (online self-completion), e.g. paper questionnaire may be the first option.

At present the digital domain measures one of the infra-structures available which enables households to respond to an online census. This is measured by the percentage of households in an area that have access to the internet (via fixed line broadband). The indicator is aggregated to LSOA level of geography. Annex A presents information on how the present digital indicator was built.

Responding to an online census is a specific process where people need to use the internet and more specifically, make use of the internet to interact with a government website to submit their personal information. At present, the digital domain gives a measure of geographical areas’ likelihood to use broadband internet. As it stands this is the best proxy we have to support where help or digital assistance will be required and to plan where questionnaires are to be sent as first option.

For the next census, we will need to use digital indicators to help us better understand patterns of internet use and people’s behaviour towards using the internet to interact with government websites.

In the future when fit for purpose data becomes available the digital domain will include an indicator of households and/or populations who use the internet to interact with government websites to submit forms.

Table 1 presents the data sources used to build the indicator in the digital domain. It also presents availability, periodicity for updates and the level of geography at which the data are available at present.

Table 2 presents potential data sources for other indicators that may be used in the digital domain when data becomes available.

Table 1: Indicator and data sources used in the digital domain, availability, periodicity and geography level.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Domain** | **Indicator** | **Data source** | **Availability**  (yes/no) | **Periodicity** | **Geography level** |
| **Digital**  **Infra-structure** | Percentage of households with internet access (fixed line broadband) by area and year  Base: Households | Ofcom, 2016 | Yes; Open data | Yearly | Postcode |

Table 2: Potential indicators and data sources for use.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Domain** | **Type of data** | **Indicator** | **Data source** | **Availability**  (yes/no) | **Geography level** |
| **Digital** | Digital infra-structure | Percentage of households with internet access (mobile network- 3G and 4G) by area and year.  Base: Households | Ofcom | No | Possibly postcode level |
| Use of internet to interact with government websites | Percentage of people renewing driving licence using online mode by area and year  Base: Persons | DVLA | Yes, in agreement with DVLA. | LSOA |
| Percentage of people sending online self-assessment tax returns by area and year  Base: Persons | HMRC | Yes, but not yet agreed with HMRC. | Not known |

* + 1. ***Willingness (self-response) domain***

The willingness domain gives a measure of households’ willingness to self-respond to a census by day 10 after the census day. The 2011 Census day 10 non-return rates (paper and online returns) were used as a proxy indicator for households’ willingness to self-respond to a census. The rationale for this is that in the 2011 Census the field follow up started after day 10. Returns before day 10 were therefore self-motivated.

Research work modelling ***record level*** non-response to the 2011 Census by day 10 after the census (record level model in CCS sampled areas) identified that region/country, type of accommodation, household tenure, household ethnicity, and age and structure of the household significantly affected households’ willingness to respond by day 10 after census day (Dini, 2017).

An area level (LSOA) model was used to build the ***parameters*** to predict response/ non-response to the 2021 Census. All households within an LSOA were considered equivalent, all experiencing the same conditions (area level covariates) and displaying the same probability of responding/not responding to the census, *p*. The dependent variable was the 2011 Census day 10 non-return rates. The covariates for the area model were proxy variables selected based on the variables that were statistically significant in the record level model in 2011 CCS sampled areas. These proxy covariates were from the 2011 Census data and, where available, from 2011 administrative data. In addition, English regions and Wales were also used as a covariate in the model.

Logistic regression was used to model this probability *p*. X describes the explanatory or covariates and the β’s are the coefficients used to convert the prevalence of the factor into an effect upon the response variable (propensity of household response/non-response). The model is given by:

Logit *p*= β0+ β1X1 + β2X2 + ...+ βnXn

where β0 isthe intercept and β1, β2…βn are the coefficients used to convert the prevalence of the covariates into an effect upon the outcome variable.

Table 3 presents a summary of the covariates and data sources used in the area level model to build the parameters. Annex B provides more information on data sources used in the model and the assumptions made where data for some LSOAs were not available.

The parameters given by the area model were applied (at present) to 2015 updated data sources (with exception of flats/maisonettes/caravans) to give the predicted non-response to the 2021 Census by day 10 after the census.

The built parameters will be applied yearly using updated data sources wherever possible. Therefore, updated predicted non-response to 2021 will be available yearly. The key assumption in the model is that the structural relationship between the 2011 Census day 10 non-return rate propensity and the key factors will remain constant between the 2011 and 2021 censuses despite the change in the mode of collection in the 2021 Census (the same assumption was applied between 2001 and 2011).

The predicted response/non-response values given by the model should not be taken as the actual response/non-response rates by day 10 after census for the 2021 Census. The predicted values are to be used only as the likelihood of LSOAs to self-respond to a census.

If more administrative data becomes available, they will be assessed for feasibility for use. Fit for use data will then be tested in the model. If proved to improve the fit of the model, they will be added to it. A potential administrative data source to be used is the Address Register. This would give an up to date percentage of flats/maisonettes/caravans to replace the 2011 Census data still in use.

Table 3: Covariates and data sources used to build the model parameters

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Domain** | **Covariate** | **Data source** | **Availability**  (yes/no) | **Year used in model parameter** | **Geography level** |
| **Willingness** | Percentage of flats/maisonettes: converted or shared (including bed-sits), in commercial building and of caravans/mobile or temporary structure by area and year  Base: Dwellings | 2011 Census | Yes | March 2011 | LSOA |
| Percentage of ‘Non White-British’ pupils by area and year  Base: Pupils | DfE: English and Welsh School censuses | Yes | 2011 | LSOA |
| Percentage of young adults aged 16-29 by area and year  Base: Persons | 2011 Census | Yes | March 2011 | LSOA |
| Percentage of claimants of Universal Credit and Job Seekers Allowance\* (Mar\_13) by area and year  Numerator: DWP Persons aged 16 and over \*\* claimants  Denominator: MYE\_2013 persons aged 16-65 | DWP | Yes | March 2013  MYE 2013 (reference period 30 June) | LSOA |
| Median price of residential property sales in E&W (Q1\_2011) by area and year  Base: Residential property | HM Land Registry data used in ONS publication | Yes | Quarter 1, 2011 | MSOA (assumption LSOAs within MSOAs with same values) |

\*Available from 2013 onwards; before this period information is available for claimants of Job Seekers Allowance only (ceased in April 2013).

\*\* Number of claimants aged 65 and over is very small and not published at LSOA level because of disclosure issues. DWP only releases a total count of claimants aged 65 and over.

* 1. **Predicted willingness to self-respond versus Predicted final response rates**

One of the objectives of developing a HtC index are to provide a stratification variable in the sample design for the 2021 CCS. The index will also be used in the 2021 Census estimation and adjustment methods.

Research work was undertaken to investigate if a model to predict final census non-response rather than day 10 non-response would be more suitable to build a HtC index to be used as a stratifying variable for the 2021 CCS sampling and for estimation and adjustment.

The method and data used in the model to predict final non-response was the same as in the model used the predict day 10 non-responses. The only difference was in the outcome variable used. Imputation rates from the 2011 Census were used as the outcome variable in this model. Annex C presents details of the research work conducted.

The results indicated that there was no strong evidence to justify the use of two models, one to build a HtC index to support planning of follow-up resources and allocation and another to build a HtC index to be used in the CCS sampling and census estimation and adjustment (more information in Annex C). The preferred choice would be to use one index for both, the planning of follow-up resources and as a stratifier for the CCS and census estimation and adjustment. This is best achieved by using the willingness to self-respond based on a model to predict day 10 non-responses.

1. **Assigning hard-to-count categories to domains**

The LSOA’s indicator in the Digital domain was ranked from the highest percentage of households’ internet connections to the lowest.

The LSOA’s predicted non-response to a census by day 10 after the census given by the model was ranked from the lowest to the highest.

The ranked indicators in each domain were then split into 5 categories, the highest – category 5 - being the hardest to be counted. The percentages of LSOAs used in the split were: 40% in HtC1, 40% in HtC2, 10% in HtC3, 8% in HtC4 and 2% in HtC5. More information on the split is given in Annex B.

At present the split uses the same 5 levels as for that used in the 2011 HtC. The split seems to work well for the Willingness domain (more information in Annex B).

In addition, if data on use of internet to interact with government becomes available and are fit for use the digital domain of the HtC index will need to be updated. More research may be needed to investigate the relevance of combining the two indicators of the digital domain, the access and use of internet.

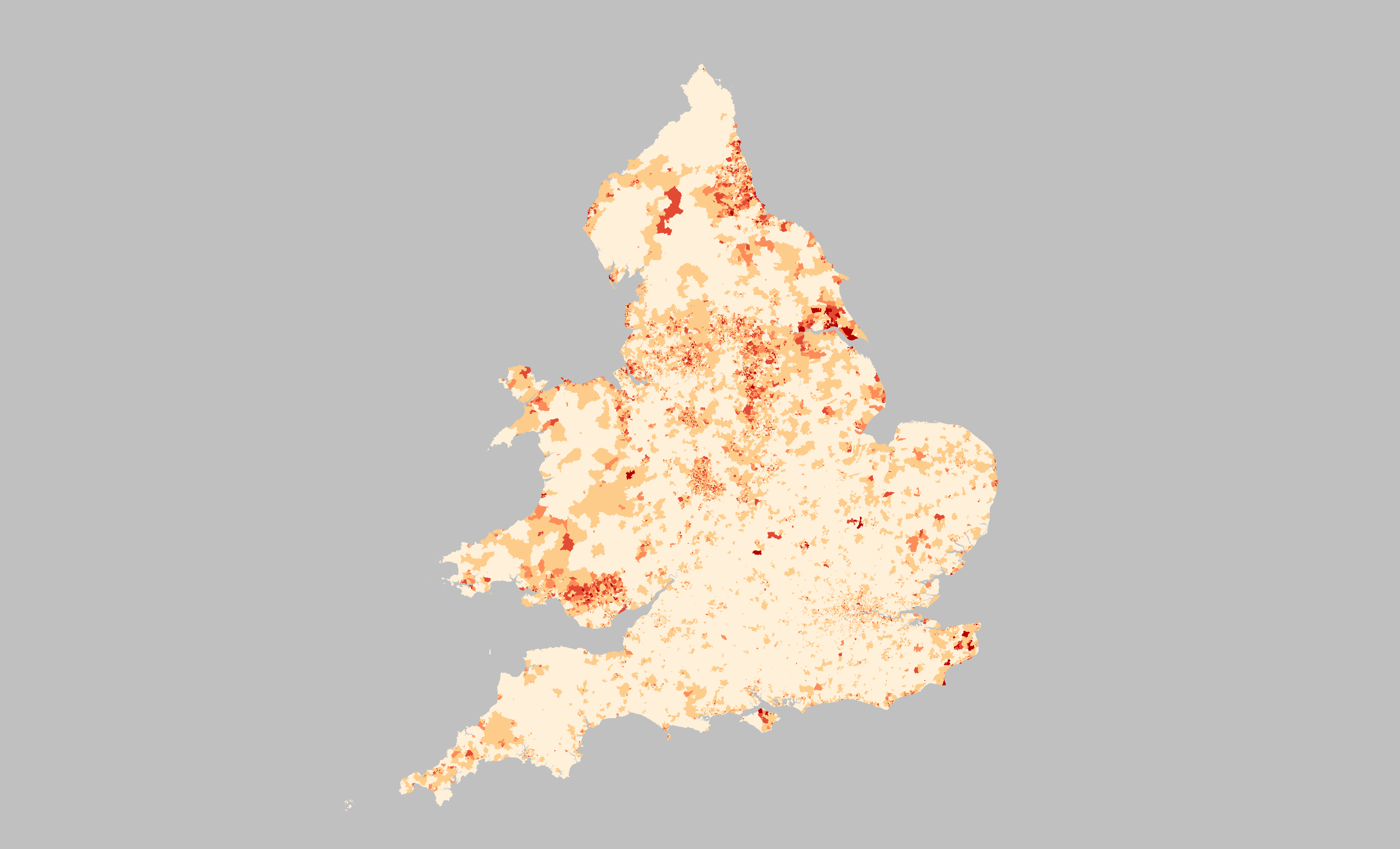
1. **HtC index results**

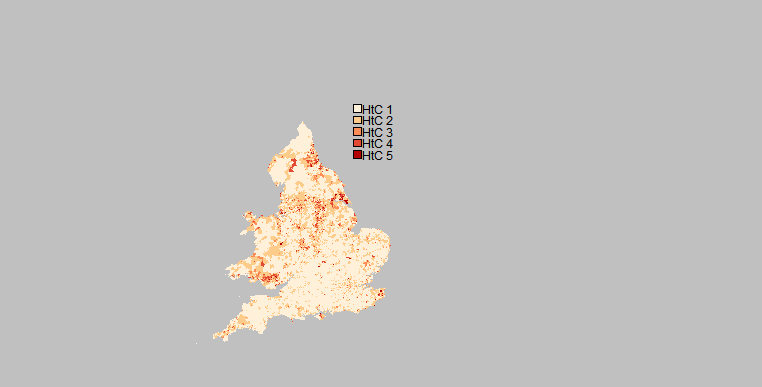
The results show the HtC categories attributed to the model predicted non-responses rates by LSOA using the split described in Section 4.

Maps 1 and 2 show the distribution of LSOAs in England and Wales according to the 5 categories of the HtC for the Digital and Willingness to census self-response domains, respectively.

Maps 3 and 4 show the distribution of LSOAs in Inner and Outer London according to the 5 categories of the HtC for the Digital and Willingness to census self-response domains, respectively.

Map 1: Digital domain: LSOAs by HtC category. England and Wales, 2016



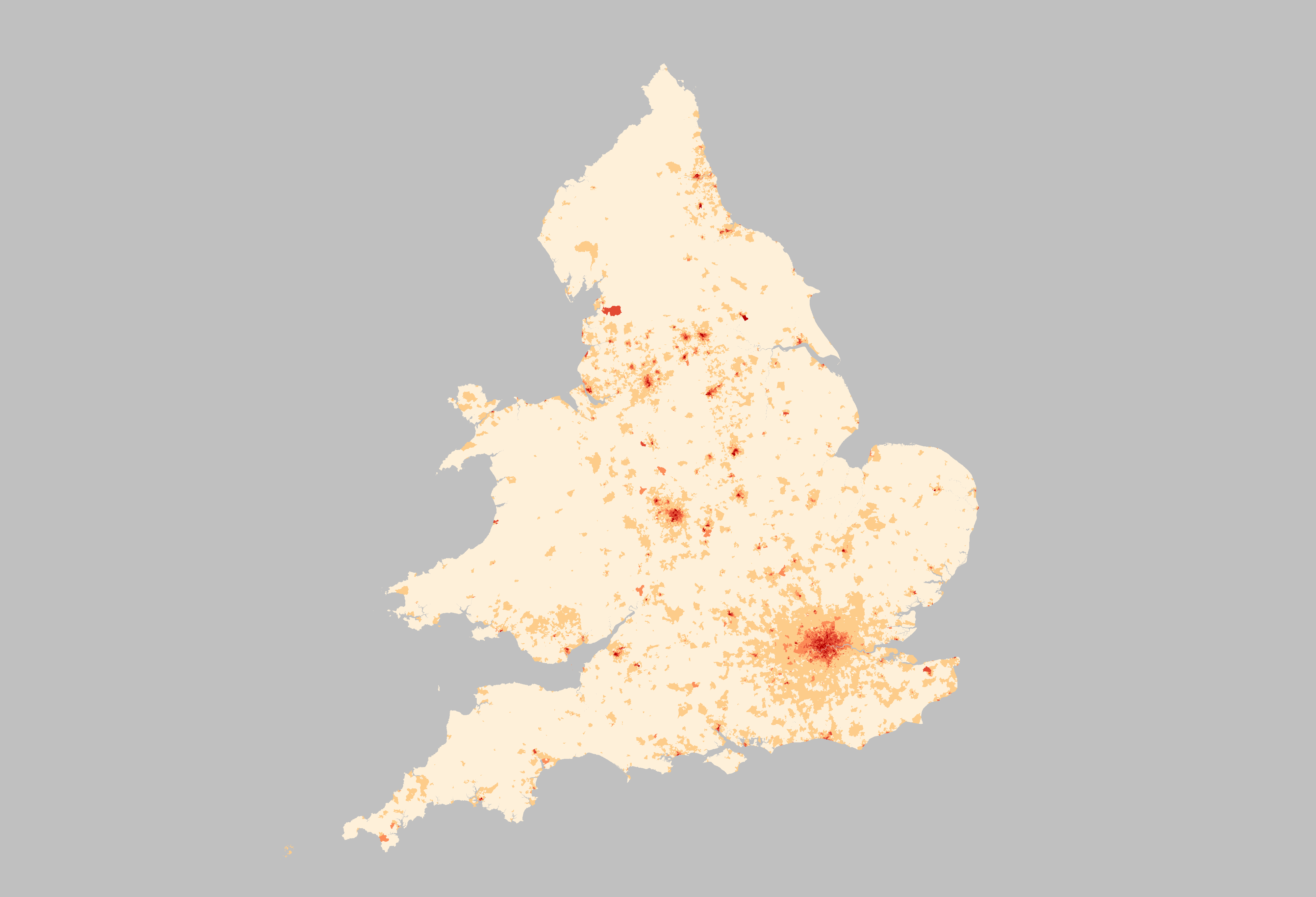


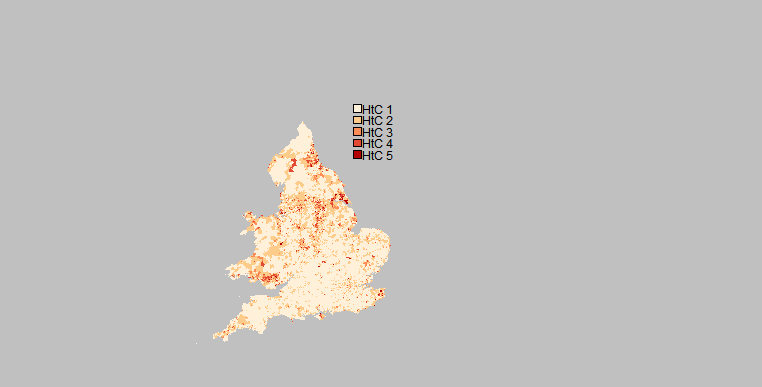
Note: HtC 5 = hardest areas to count

Map: Contains Ordnance Survey data © Crown copyright 2018

Source: Office for National Statistics licensed under the Open Government Licence v.3.0

Map 2: Predicted willingness to census self-response by day 10 after census day: LSOAs by HtC category. England and Wales.

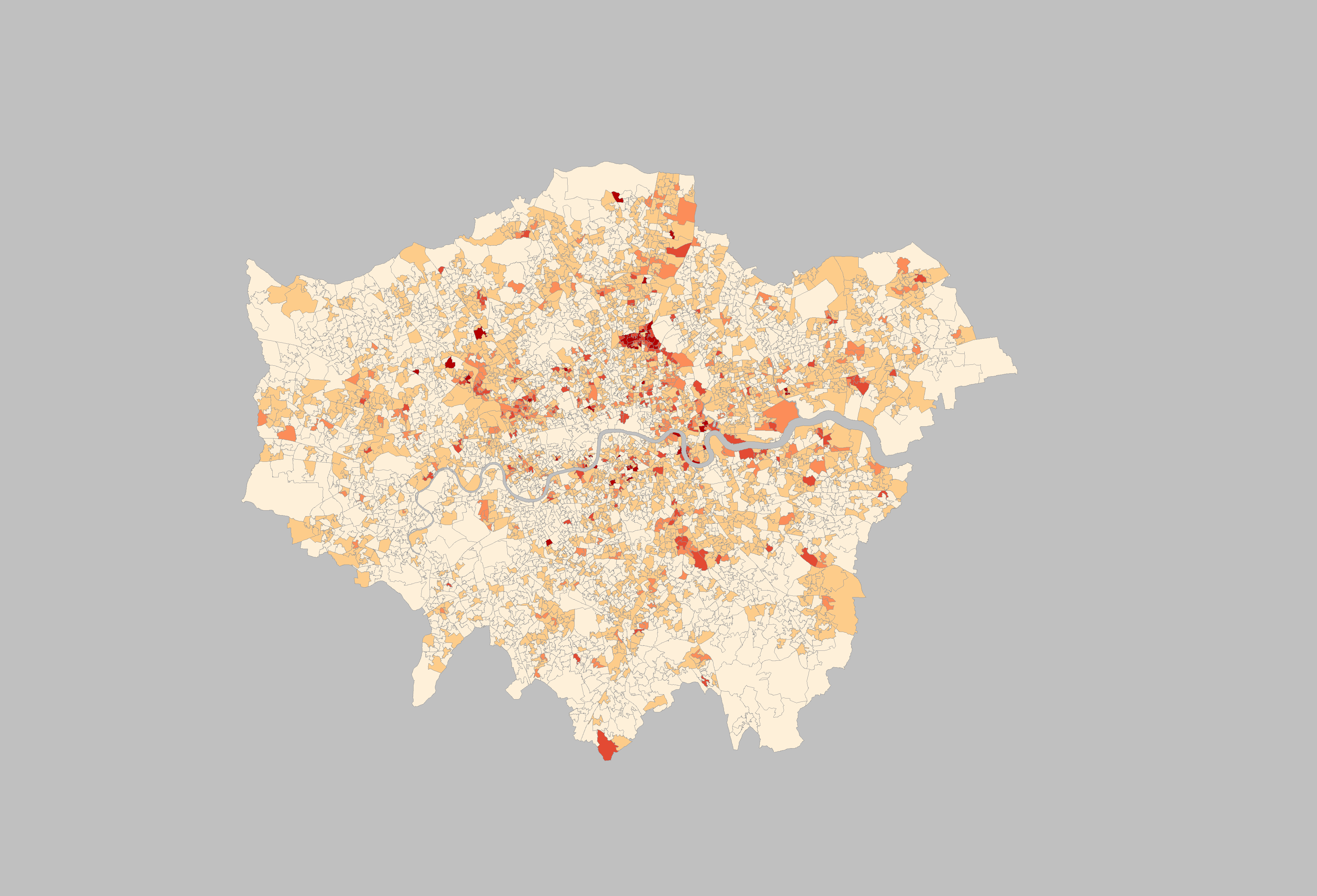


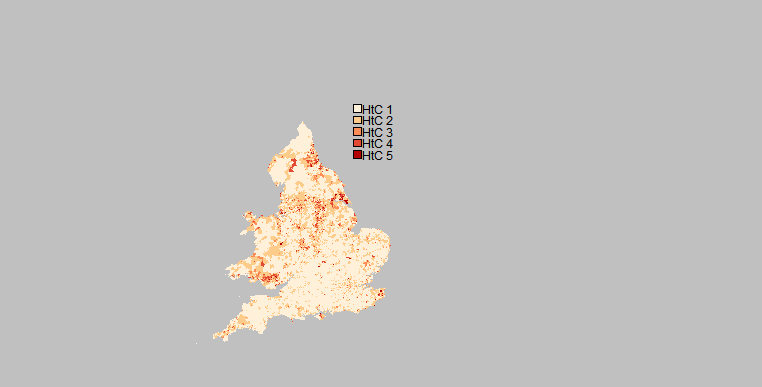


Map: Contains Ordnance Survey data © Crown copyright 2018

Source: Office for National Statistics licensed under the Open Government Licence v.3.0

Map 3: Digital domain: LSOAs by HtC category. London, 2016



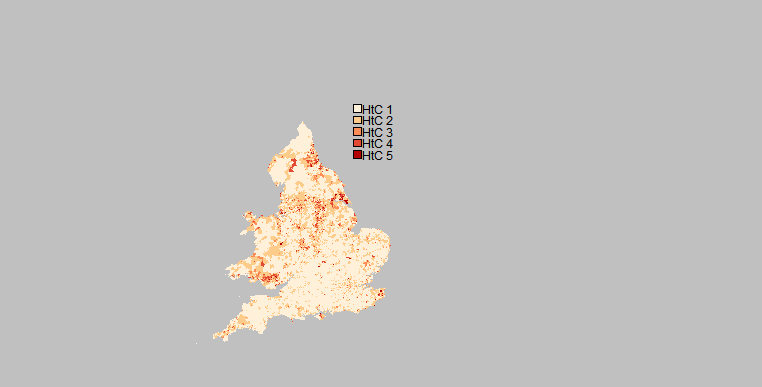


Map: Contains Ordnance Survey data © Crown copyright 2018

Source: Office for National Statistics licensed under the Open Government Licence v.3.0

Map 4: Predicted willingness to census self-response by day 10 after census day: LSOAs by HtC category. London.





Map: Contains Ordnance Survey data © Crown copyright 2018

Source: Office for National Statistics licensed under the Open Government Licence v.3.0

1. **2017 Census test results**

This section presents analysis using data from the designed experiments within the 2017 Census Test.

The sample in Component 2 Part 1 was designed specifically to test the impact of sending a paper questionnaire versus a letter containing a Unique Access Code (UAC) as the first approach to households. The total sample in Component 2 Part 1 was 60,000 addresses in England and Wales. Half of the sampled addresses were sent a paper questionnaire first and the other half received a UAC. Households in addresses that were sent a UAC first could request a paper questionnaire, but this was intentionally not made obvious. Households in addresses that were initially sent a paper questionnaire were also given a UAC, on the first page of the paper questionnaire, so it was clear that they could respond online if they chose to do so. Three reminders were sent to non-responding addresses in both groups, the first, 3 weeks after the initial questionnaire/UAC letter had been sent. All the reminders in both groups were via a letter containing the UAC only.

The 2017 Census Test sample was stratified using 9 categories constructed at LSOA level by combining 3 categories on likelihood to self-respond (based on return rate by day 10 in 2011 Census) and 3 digital categories using information on broadband internet availability (Ofcom, 2014). The cut points for the 3 categories for each domain were constructed using the Dalenius-Hodges method of stratification (Dalenius and Hodges, 1959) independently on each margin. The categories from each domain were then combined. For more information on the 2017 Census Test design and sample see ONS (2017).

Previous analysis of the test (Corps et al, 2017) showed that households were more likely to respond if they were sent a paper questionnaire first. This pattern of response was seen across the 9 categories of the sample stratification. The analysis also showed that the digital domain of the sample stratification used in the 2017 Census Test was a good predictor of online response and particularly by age group.

Further analysis was conducted to investigate the effect of paper versus UAC first on self-response i.e. without interference of reminder letters. The analysis considered all valid responses received by 7 April 2017 (i.e. before reminder letters were received by the households) amongst valid sampled addresses. The 5 HtC categories of the Willingness domain were assigned to the LSOAs sampled in the 2017 Census test and analysis of self-response was conducted by HtC categories. The rationale for using this post-stratification analysis was to check if levels of self-response in the census test were consistent with the HtC categories in the Willingness domain; that is the higher the HtC category the lower the self-response rate.

Overall self-response was very low. Approximately a quarter of the households that received paper first and only 15 per cent of the households that received UAC first self-responded to the census test. The percentage of self-response decreased as HtC categories of the Willingness domain increased. This pattern was seen for both online and paper responses but it was clearer in paper responses (Table 4).

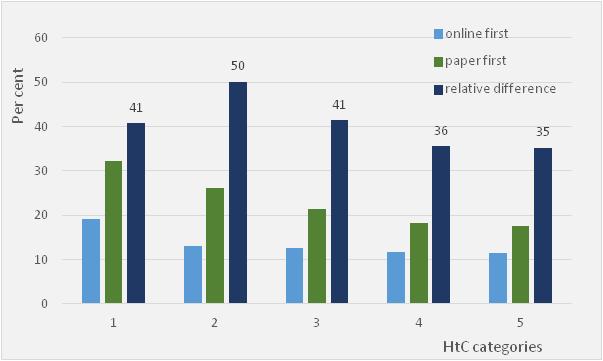
For the online first approach, the results showed that there were very small differences in self-response rates between categories 2 and 3 and categories 4 and 5 of the Willingness domain. For the paper first approach, the differences in self-response rate by categories of HtC were larger. The exception was for categories 4 and 5 where the differences in self-response rates were smaller than one percent (0.7%).

Table 4: Valid addresses, valid self-responses and percentages of self-responses by HtC categories of the Willingness domain.



Chart 1 presents the percentages of self-response by first mode of approach and the 5 HtC categories of the Willingness domain. It also presents the first mode of approach per cent relative difference in self-responses in HtC categories. The results showed that the relative difference of the effect paper/UAC first was slightly higher in category 2 of the HtC. It also showed that there was only one per cent relative difference between categories 4 and 5 and on difference between categories 1 and 3 of the HtC.

Chart 1: Percentage of self-response by first mode of approach and relative difference in self-response by HtC categories of Willingness domain. England and Wales.



This analysis used post stratification and the sample of the 2017 Census test had not been designed to evaluate the 5 HtC strata. Therefore, we need caution in interpreting these results.

In addition, we need to consider that the census test was a voluntary survey; self-response rates were very low and very likely to be from the most willing to participate. This consequently lead to response bias.

The 2021 Census will be mandatory, there will be media communication to gain people’s compliance with the census. Also, it is expected that the impact of the online first factor will be lower as people’s digital skills increase in the next few years. It is expected that household’s responses will be higher. The relative differences in approach mode seen by HtC categories of the Willingness domain (Chart 1) may be much smaller than those seen in the results of the census test.

Despite all these constraints and while the differences observed by HtC categories may be small, the results were reassuring. They showed the expected monotonically decreasing pattern in self response by HtC categories.

1. **Next steps**

* Investigate the feasibility of using DVLA data on people’s use of online mode to apply for or renew their driving licence.
* Update the digital domain to include digital information on people’s use of internet to interact with government websites (DVLA data) and people’s use of mobile network (Ofcom data) when these data become available and tested for feasibility of use.
* Continue updating the Willingness to census self-response domain as required when up to date data becomes available.

1. **References**

Abbott O (ONS, 2015). Hard-to-count strategy for the 2021 Census. ONS Internal report. Available on request.

Abbott, O. and Compton, G., 2014. Counting and estimating hard-to-survey populations in the 2011 Census. In: R. Tourangeau, B. Edwards, T. Johnson, K. Wolter and N. Bates, eds. 2014. *Hard-to-survey populations*. Cambridge: Cambridge University Press. Ch.4.

Corps D, Fraser O and Meirinhos V (ONS, 2017). Test Analysis – Analysis of online and paper respondents. ONS Internal report. Available at: <https://share.sp.ons.statistics.gov.uk/sites/cen/csod/CSOD_Stats_Design/Statistical_Design/2017%20Test/Reports/CRAG_paper_Nov17.docx>

Dalenius T and Hodges J L Jr (1959). Minimum variance stratification. Journal of the American Statistical Association, 54, 88–101.

Department for Communities and Local Government (DCLG, 2015). The English Indices of Deprivation 2015. Technical report. Available at: <https://www.gov.uk/government/publications/english-indices-of-deprivation-2015-technical-report>

Dini E and Abbott O (2016). Options for predicting non-response in the 2021 Census. ONS Internal report presented to the Census Research Assurance Group (CRAG). Available on request.

Dini E (2017). Modelling to predict 2011 Census household non-response by day 10. ONS Internal report. Available on request.

Heady P, Smith S and Avery P (1994). ‘1991 Census Validation Survey: coverage report’. ISBN: 9780116915917. Available at: <http://uroleaf.com/2001-census-statistical-metadata.pdf>

Hopper N (2011) Predicting patterns of household non-response in the 2011 Census. Survey Methodology Bulletin, vol 69, pp 9-22. Available at: <http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/method-quality/survey-methodology-bulletin/smb-69/index.html>

Office for National Statistics (2000). 2001 One Number Census Hard to Count Index. Paper presented to the One Number Census Steering Committee. Available at: <http://webarchive.nationalarchives.gov.uk/20160128201419/http://www.ons.gov.uk/ons/guide-method/census/census-2001/design-and-conduct/the-one-number-census/methodology/steering-committee/key-papers/index.html>

Office for National Statistics (2012a). Census Coverage Assessment and Adjustment Process: The 2011 Census Coverage Assessment and Adjustment Process. Available at:

<http://webarchive.nationalarchives.gov.uk/20160105160709/http://ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-user-guide/quality-and-methods/coverage-assessment-and-adjustment-methods/index.html>

Office for National Statistics (2012b). Coverage assessment and adjustment methods: Tuning the Coverage Estimation Process. Available at: <http://webarchive.nationalarchives.gov.uk/20160105160709/http://ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-user-guide/quality-and-methods/coverage-assessment-and-adjustment-methods/index.html>

Office for National Statistics (2017). ONS Census Transformation Programme. 2017 Census Test. Available at: <https://www.ons.gov.uk/census/censustransformationprogramme>

Office of Communications (Ofcom) (2014). Infrastructure Report 2014. Available at: <https://www.ofcom.org.uk/research-and-data/multi-sector-research/infrastructure-research/infrastructure-2014>

Rahman, N and Goldring, S (2006). Factors associated with household non-response in the Census 2001. Survey Methodology Bulletin, vol 59, pp 11-24. Available at: <http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/method-quality/survey-methodology-bulletin/smb-59/index.html>

Simpson, S, Cossey, R and Diamond, I (1997). 1991 population estimates for areas smaller than districts. Population Trends, vol 90, pp31-39.

# Annexes

## *Annex A*

Annex A presents a summary of the method used to build the indicator in the digital domain and the assumptions made where data was not available.

At present, the digital domain is composed of one indicator, the percentage of households with access to the internet via a fixed line broadband.

The percentage of households with access to the internet was estimated using the number of fixed line broadband connections to premises (households and small businesses) (Ofcom, 2016)[[4]](#footnote-4) and the number of occupied households given by the 2011 Census (NOMIS, 2011 Census)[[5]](#footnote-5). Ofcom and Census data were matched at postcode level and then aggregated to LSOA level of geography.

Number of internet connections at LSOA level

Percentage of households with access to internet = ------------------------------------------------------------- x 100

Number of occupied households at LSOA level

Ofcom publishes yearly the number of fixed line broadband connections at postcode level. It is Open data and available for free download. The number of connections refers to those from households and small businesses.

For disclosure reasons, Ofcom data is not available in a number of the postcodes. Ofcom 2016 data had 181,290 postcodes in England and Wales that fell in this category (14% of Ofcom and Census matched postcodes). Approximately 99% of these 181,290 postcodes had households with one or more usual residents counted by the 2011 Census. Given the large number of occupied households in these postcodes, an assumption was made that there were at least two broadband connections in each of these 181,290 postcodes.

Data were also checked for cases where the percentage of connections in the LSOAs were larger than 100. Using Ofcom 2016 data, approximately 2.5 per cent of the LSOAs (878 LSOAs) were in this category. One explanation is that these LSOAs had larger numbers of small businesses with internet connections (larger numerator) and the census only counted occupied households (smaller denominator). In these cases, an assumption was made that the percentage of households with access to the internet was 100%.

## *Annex B*

Annex B provides information on data sources used in the area level (LSOA) logistic regression model to predict non-response by day 10 after the census day. It provides information on the assumptions made where data for some LSOAs was not available. It also provides an analysis of the covariates used to build the parameters for the model.

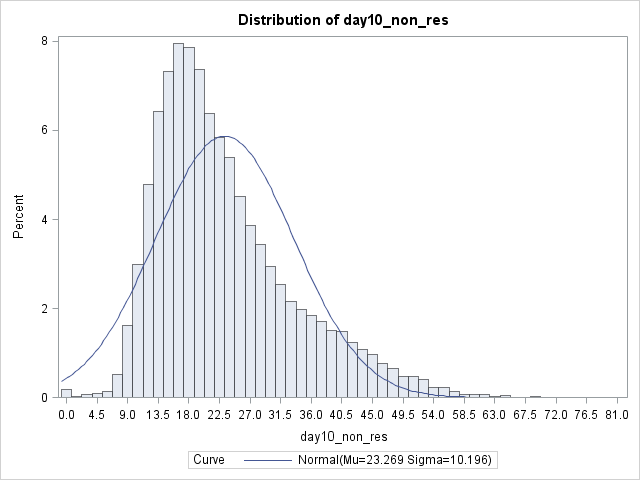
**Outcome variable**

The 2011 Census day 10 return rates (paper and online returns) were used as a proxy indicator for ‘willingness’ to self-respond to a census. Non-return rates were estimated as 100 minus return rates. The LSOAs percentage of non-returns was used as the outcome variable in the model. Figure B1 shows the distribution of LSOAs by the outcome variable. The mean day 10 non-return rate was approximately 23 per cent (+/- 10.2 per cent).

Information on census day 10 returns was available at LSOA level for the 2001 geography boundaries which were used during census day in March 2011. In December 2011 geography boundaries changed; some LSOAs split and some merged.

In cases where an LSOA split an assumption was made that the day 10 return rate in the ‘new’ LSOA was the same as in the ‘parent’ LSOA. In cases where LSOAs merged the counts of census returns (numerator) and the counts of addresses (denominator) for the merged LSOAs were summed up and divided by the number of these LSOAs. The assumptions made took into account the best-fit lookup published by ONS geo portal[[6]](#footnote-6).

Figure B1: Distribution of LSOAs by day 10 non-return rates. England and Wales, 2011 Census.



**Model covariates**

The covariates (explanatory variables) used to build the model parameters were from 2011 Census data and, where available, from 2011 administrative data.

The covariates and data sources used were:

* 2011 Census percentage of flats/maisonettes: converted or shared (including bed-sits), in commercial buildings and of caravans/mobile or temporary structure.
* 2011 Census percentage of young adults aged 16-29.
* Percentage of ‘Non White-British’ pupils from the 2011 English and the Welsh School censuses (Department for Education – DfE),
* Percentage of claimants (stocks) of Universal Credit and Job Seekers Allowance aged 16 to 65 (Department for Work and Pensions – DWP, 2013) [[7]](#footnote-7) and 2013 Mid-year-estimates (ONS, 2014)[[8]](#footnote-8).
* Median price of residential property sales in England and Wales, 2011 Land Registry data published by ONS (ONS, Housing, 2017) [[9]](#footnote-9) .
* Region (Inner London, Outer London and Wales and other English regions),

All covariates used data at LSOA level of geography, except the median price of residential property sales in England and Wales which is published by ONS at Middle Super Output Area (MSOA) level of geography. An assumption was made that all LSOAs within the respective MSOAs had the same median price of property sales.

The percentage of claimants of Universal Credit (UC) and Job Seekers Allowance (JSA) was given by the counts of claimants aged 16 and over (numerator) in 2013 (DWP) divided by the estimated population aged 16 to 65 (denominator) in 2013 (ONS) multiplied by 100. Data for 2013 was used because information for both claimants of UC and JSA has been published consistently only since this date. Published earlier counts (from October 1996 to April 2013) included counts of JSA only.

DWP counts of claimants aged 65 and over are very small and because of disclosure issues they are not published at LSOA level. DWP only releases a national count of claimants aged 65 and over. In 2013 there were 20 claimants aged 65 and over. No attempt was made to distribute these counts into LSOAs. The counts at age 65 and over by LSOA were assumed to be zero.

**Analysis of covariates**

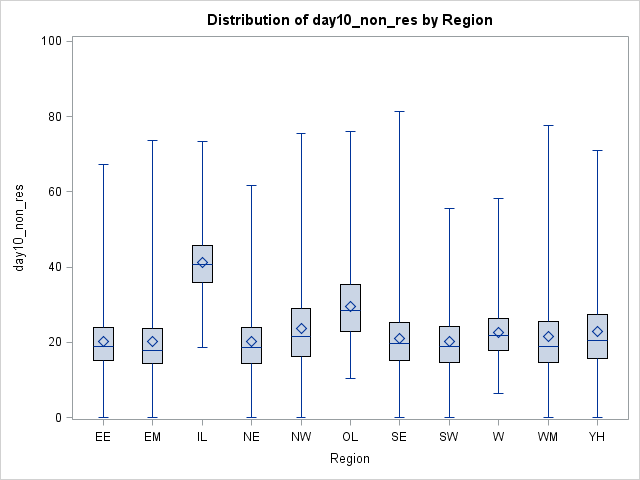
***Region/Country***

English regions and Wales were aggregated into 3 categories: Inner London, Outer London and other English regions and Wales. These 3 categories were used as a covariate in the model. The rationale for this was that:

1. This aggregated covariate showed a better fit in the initial record level model (Census matched to CCS) than using each region separately.
2. In the 2011 Census, non-returns in Inner London and Outer London were on average higher than non-returns in other English regions and Wales.

Figure B2 shows the distribution of LSOAs’ day 10 non-returns by region/country in the 2011 Census

Figure B2: Distribution of LSOAs’ day 10 non-returns by region/country. England and Wales. 2011 Census.



Region/Country

***School Census***

The percentage of ‘Non White-British’ was given by the counts of pupils from ‘Non White-British’ ethnic groups divided by the counts of all pupils multiplied by 100. The data used was from the 2011 English and Welsh School censuses (DfE).

Counts of pupils from the School censuses at LSOA level for 2011 used the 2001 geography boundaries. To attribute the percentages into LSOAs that merged or split we used the same approach as that used for the outcome variable of the model, and described previously in this Annex.

Because of disclosure issues, counts of pupils in ‘White-British’ and/or in ‘Non White-British’ ethnic groups were not provided for 1,772 (5 per cent) of the LSOAs in 2011. However, counts of pupils by these ethnic groups were available at Local Authority (LA) level.

The percentages of ‘Non White-British’ pupils in the LSOAs missing information were assumed to be the same as the percentage of ‘Non White-British’ pupils of the LAs these LSOAs belonged to.

As this approach could potentially overestimate the counts in these LSOAs, an adjustment to the estimates in LSOAs missing information was applied using the formulae shown below.

Adjusted percentage of ‘Non White-British’ in LSOAs missing information =



The results given by the prediction model using the adjusted for missed LSOAs was compared to the results given by the model when using the proxy LA values without adjustment. The results showed very small differences. Therefore, a decision was made to use the proxy values without adjustment.

*Validation of School Census data 2011 with 2011 Census*

The percentage of Non White-British pupils from the 2011 School censuses was validated against the percentage of Non White-British population given by the 2011 Census[[10]](#footnote-10). The per cent differences of ‘Non White-British’ given by both data sources were calculated by LSOA.

For England and Wales, 85 per cent of the LSOAs had per cent differences between +/- 10 per cent.

The lowest agreement was seen for Inner London; only 26 per cent of the LSOAs had per cent differences between +/- 10 per cent (Table B1).

The overall variance explained by the models using either ethnic groups given by the 2011 Census or given by the 2011 School census data was not affected. Therefore, the covariate ethnic group used in the model to build the parameters used ethnicity from the School Census data as the data is updatable.

Table B1: LSOA per cent differences between 2011 School Census and 2011 Census percentages of 'Non White-British'.



***Analysis of correlation between covariates***

Table B2 presents the results of the analysis of correlation between covariates. Pearson coefficient of correlation was very low or non-existent. This is a good indicator and it means that all the covariates used were contributing to the model. There was no collinearity effect between the covariates.

Table B2: Analysis of correlation between covariates.



**Model output**

**Model parameters and fit**

Table B3 shows the parameter estimates obtained by the model and Figure B3 shows the correlation between observed day 10 and predicted day 10 non-response rates. The model explained 70 per cent of the variance (deviance). In other words, the model ‘explained’ 70 per cent of the variation in the total LSOAs’ household non-response observed in 2011. It suggests a good fit to predict non-response.

Table B3: Model parameter estimates



Figure B3: Correlation between observed day 10 and predicted day 10 non-response rates.

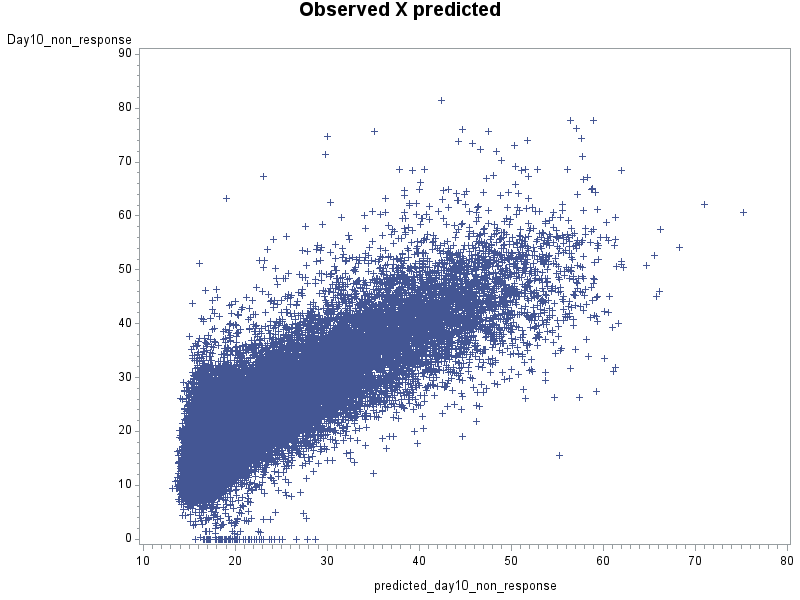
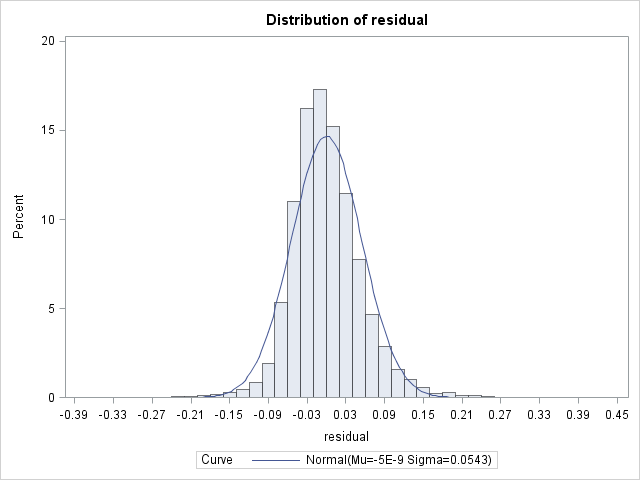


Figure B4 shows the distribution of residuals or unexplained part of the variance of the model at the LSOA level. The residual is given by the observed day 10 minus predicted day 10 non-response. The histogram shows the distribution of the residuals with a normal distribution overlaid. The mean distance of observed to predicted value for an LSOA is almost zero (-5E-9). This indicates that the model is not optimistic or pessimistic and predicts similar non-responses to those observed.

Figure B4: Distribution of residuals (observed minus predicted values).



**Applying the parameters to updatable data sources**

The parameters given by the area level model were applied to 2015 updated data sources to give the predicted non-response to the 2021 Census by day 10 after the census. The updated data sources used as covariates were:

* ONS 2015 Mid-year-estimates percentage of young adults aged 16-29 at LSOA level.
* The percentage of ‘Non White-British’ pupils from the 2015 English and the Welsh School censuses (DfE),
* Percentage of claimants (stocks) of Universal Credit and Job Seekers Allowance aged 16 to 65 (DWP, 2015) and 2015 Mid-year-estimates (ONS, 2016) [[11]](#footnote-11).
* The median price of residential property sales in England and Wales, 2015 from Land Registry data published by ONS.

In 2015, counts of pupils by ethnic groups were not provided for 1,389 (4 per cent) of the LSOAs. However, counts of pupils by ethnic groups were available at Local Authority (LA) level.

The percentages of ‘Non White-British’ in the LSOAs missing information were assumed to be the same as the percentage of ‘Non White-British’ of the LAs these LSOAs belonged to.

For median price of residential property sales which were published at MSOA level, an assumption was made that all LSOAs within their respectively MSOAs had the same median price values.

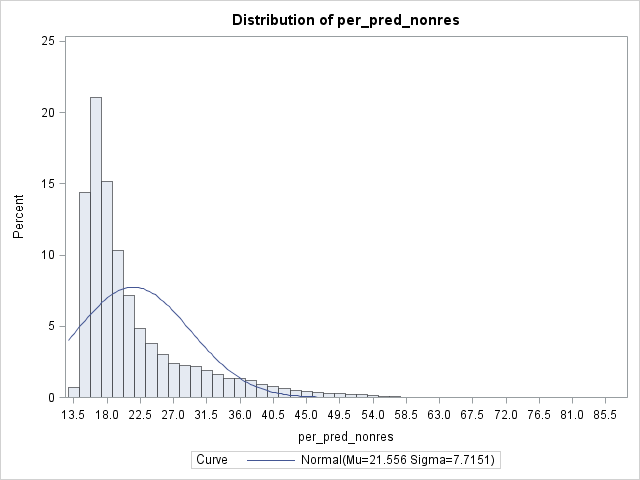
**Assigning Hard-to-Count categories**

The mean predicted day 10 non-response was approximately 22 per cent (+/- 7.7 per cent).

Figure B5 shows the distribution of LSOAs by predicted census day 10 non-response rates. The red bars in the figure are a rough illustration where in the predicted non-response rates the 40%, 40%, 10%, 8% and 2% cut off points fall in each of the 5 categories of the HtC. It shows that the split used worked well.

However, in using this prediction model, we have to bear in mind that the predicted values are still based on a primarily paper mode of census returns. Despite uncertainties to what extent this will be directly applicable to a primarily online census return strategy, this is the best prediction we have at the moment.

Figure B5: Distribution of LSOAs by predicted census day 10 non-response rates. England and Wales.



HtC1 HtC2 HtC3 HtC4 HtC5

The analysis of the 2017 Census test also provided evidence that the 5 level split used is working well. Despite the low response/high non-response rates observed in the 2017 Census test, and while the differences observed by HtC categories were small, the results were reassuring. They showed the expected monotonically decreasing pattern in self response by HtC categories, independently of the mode used to first approach the households (online or paper).

## *Annex C*

One of the objectives of developing a HtC index is to provide a stratification variable in the sample design for the 2021 CCS. Also, the index may be used as required in the 2021 Census estimation and adjustment.

Annex C describes the research work conducted to investigate if a model to predict final census non-response rather than day 10 non-response would be more suitable to build a HtC index to be used as a stratification variable for the CCS sampling and 2021 Census estimation. The rationale for using a prediction model based on the final non-response rates would be that the CCS takes place after the census follow up is completed.

The Annex describes the method used to model the predicted final non-response rates and compares the results of this model with the model to predict census day 10 non-response.

**Modelling to predict census final non-response rates**

The 2011 Census final imputation rates (or non-response rates) were used as the outcome variable in a model to predict the 2021 Census non-response.

The method and data used in the model to predict final non-response rates was the same as the model used the predict day 10 non-responses. The only difference was in the outcome variable used. A description of and analysis of the covariates used in the model was given in Annex B.

Figure C1 presents the distribution of 2011 Census non-response rates (imputation rates). Compared with the distribution of 2011 Census day 10 non-response (Figure B1) it gives evidence how well the last census follow up, estimation and adjustment worked to decrease the differences in non-response amongst LSOAs. The mean final non-response rate was approximately 5 per cent (+/- 3.5 per cent) compared with 23 per cent (+/-10.2 per cent) ten days after the census day.

Figure C1: Distribution of LSOAs by final non-response rates. England and Wales, 2011 Census.

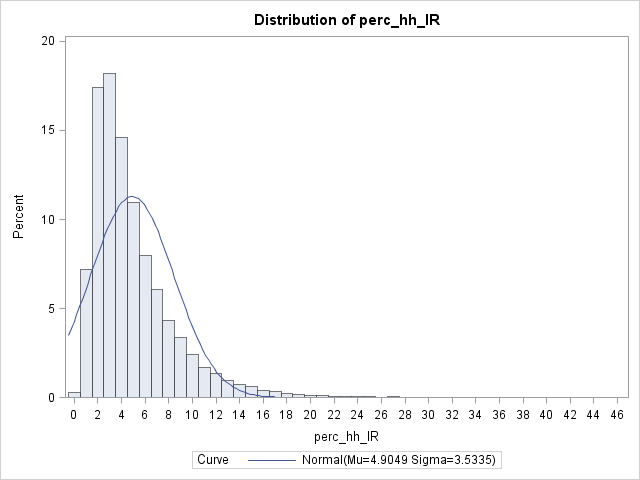


Table C1 presents the final non-response model parameter estimates. Figure C2 presents the distribution of LSOAs by the predicted values of non-response given by the day 10 non-response and the final non-response rates models. Figure C3 presents the correlation between the predicted values given by both models.

Table C1: Model parameter estimates



Figure C2: Distribution of LSOAs by predicted non-response rates given by the day 10 non-response and the final non-response rate models. England and Wales.

Predicted day 10 non-response Predicted final non-response rates

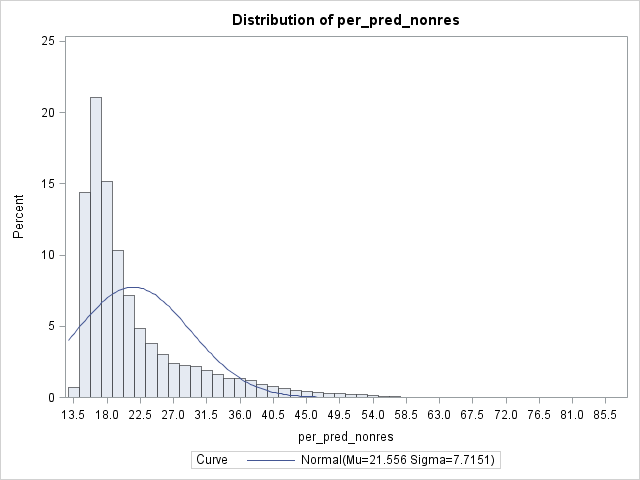
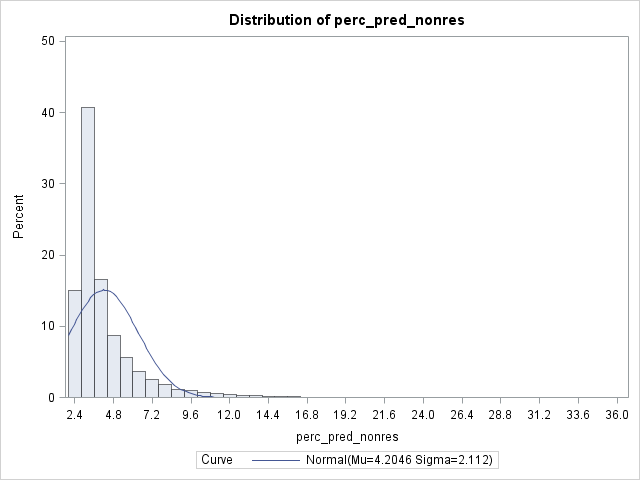


Figure C3: Correlation between LSOAs predicted day 10 non-response and predicted final non-response rates.





The correlation seen between LSOAs predicted day 10 non-response and predicted final non-response rates (Figure C3) is non-linear. This result suggests that the effect of the 2011 Census follow-up made less difference in reducing non-response in harder to count areas.

Figure C4: Distribution of LSOAs predicted day 10 non-response and predicted final non-response by HtC category. England and Wales.



Table C2a. Number of LSOAs in HtC categories 1 to 5 given by the predicted day 10 non-response and predicted final non-response rates models. England and Wales.



Table C2b: Percentage of LSOAs in HtC categories 1 to 5 based on predicted day 10 non-response where the predicted final non-response rates model categories agree. England and Wales.



Table C2c: Kappa statistical test of agreement in HtC categories attributed to LSOAs by each prediction model. LSOAs in England and Wales.



\* Kappa = 0.83 = Almost perfect agreement (Landis and Koch, 1977) [[12]](#footnote-12)

Table C3a: Number of LSOAs in HtC categories 1 to 5 given by the predicted day 10 non-response and predicted final non-response rates models. LSOAs in Inner and Outer London.



Table C3b: Percentage of LSOAs in HtC categories 1 to 5 based on predicted day 10 non-response where the predicted final non-response rates model categories agree. LSOAs in Inner and Outer London.



All LSOAs in Inner London were attributed to categories 4 or 5 of the HtC. In LSOAs in Outer London the lowest agreement was seen in category 1 of the HtC. Only 20 per cent of the LSOAs in category 1 of the day 10 nonresponse model were also attributed to category 1 of the HtC in the final nonresponse model. Most LSOAs attributed to category 2 in the predicted day 10 non-response model moved to category 1 in the predicted final non-response model. This is expected because the final non-response model uses imputation rates as the outcome variable which is post field follow up, estimation and adjustment.

As the Kappa statistic test is a weighted test and category 1 of the HtC had the lowest number of observations in London, the overall agreement is still substantial (Table C3c).

Table C3c: Kappa statistical test of agreement in HtC scores attributed to LSOAs by each prediction model. LSOAs in Inner and Outer London.



\* Kappa = 0.73 = Substantial agreement (Landis and Koch, 1977)

The main objective of these two prediction models was to attribute scores to the LSOAs according to their difficulty to be counted in a census.

The results given by the analysis comparing the two models and the attributed HtC categories using the split 40%, 40%, 10%, 8% and 2% indicate that there is no strong evidence to justify the use of two Willingness domains of the HtC index, one based on predicted day 10 non-response to support planning of follow-up resources, and their allocation and another based on predicted final non-response rates to be used in the CCS sampling and estimation.

1. ‘Enumeration Districts’: In the 1991 Census the country was divided into around 130,000 small areas (Enumeration Districts) and to each of these areas an enumerator was appointed. In England and

   Wales an average enumeration district (ED) comprised about 200 households. EDs were defined so that they generally did not cross the boundaries of administrative areas such as counties, local authority districts or wards. More information available at: <https://census.ukdataservice.ac.uk/media/51162/1991_defs.pdf> [↑](#footnote-ref-1)
2. ‘Indicators’: to be chosen to provide the best possible measure of each domain of areas’ difficulty to count.

   A statistical indicator is the representation of statistical data for a specified time, area or any other relevant characteristic.

   [↑](#footnote-ref-2)
3. ‘Digitally Excluded’ for the 2021 Census means those who are unable to use or access online services. [↑](#footnote-ref-3)
4. Ofcom (2016). Infrastructure report 2016: downloads. Available at:

   <https://www.ofcom.org.uk/research-and-data/multi-sector-research/infrastructure-research/connected-nations-2016/downloads> [↑](#footnote-ref-4)
5. ONS (2013). Postcode Headcounts and Household Estimates - 2011 Census. Available at:

   <https://www.nomisweb.co.uk/census/2011/postcode_headcounts_and_household_estimates> [↑](#footnote-ref-5)
6. ONS Geo Portal (2017). Lower Layer Super Output Area (2001) to Lower Layer Super Output Area (2011) to Local Authority District (2011) Lookup in England and Wales. Available at:

   <https://ons.maps.arcgis.com/home/item.html?id=6d3b1fc88b284a9bb9d4827530b16da4> [↑](#footnote-ref-6)
7. Department for Work and Pensions (2013). DWP Benefits.

   Available at: <https://www.nomisweb.co.uk/query/select/getdatasetbytheme.asp?theme=72> [↑](#footnote-ref-7)
8. Office for National Statistics (2013). Annual Small Area Population Estimates: 2013. Available at: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualsmallareapopulationestimates/2014-10-23> [↑](#footnote-ref-8)
9. Office for National Statistics (2017). Housing: Property price, private rent and household survey and census statistics, used by government and other organisations for the creation and fulfilment of housing policy in the UK. Available at: <https://www.ons.gov.uk/peoplepopulationandcommunity/housing#datasets> [↑](#footnote-ref-9)
10. Dini E (2017). LSOA per cent differences between 2011 School Census and 2011 Census’ percentages of 'Non White-British'. ONS Internal report. Available on request.

    [↑](#footnote-ref-10)
11. Office for National Statistics (2016). Annual Small Area Population Estimates: 2015. Available at: <https://www.ons.gov.uk/releases/smallareapopulationestimatesinenglandandwalesmid2015> [↑](#footnote-ref-11)
12. Landis, R and Kock, GG (1977). The Measurement of Observer Agreement for Categorical Data. Biometrics, Vol. 33, No. 1 (Mar., 1977), pp. 159-174. [↑](#footnote-ref-12)