Predicting patterns of household non-response in the 2011 Census

1. Summary

The census produces population estimates based on the direct census count adjusted for under-coverage estimated from the Census Coverage Survey (CCS). The accuracy of these population estimates is dependent upon both the response rate and its variation within estimation strata. Thus, there are two key aims for the 2011 Census field operation: to minimise differential non-response and maximise response (Cabinet Office, 2008). In order to facilitate the achievement of these aims a hard to count categorisation (HTC) has been produced. The HTC classifies lower layer super output areas (LSOAs) from 1 to 5, in advance of the census, according to their expected relative difficulty of enumeration. The HTC will form part of the machinery that drives the allocation of enumerators to the field operation. It is through this allocation that differential non-response will be minimised and overall response maximised: more resources will be placed into those areas expected to be harder to enumerate than in those areas expected to be easier. In addition, the same HTC will be used to stratify the sample in the Census Coverage Survey (CCS). The key assumption is that we can relatively accurately predict patterns of non-response in the 2011 Census.

This paper describes the method to produce the hard to count categorisation. The final categorisation has yet to be completed and will incorporate more timely data. Some analysis of the accuracy of the HTC categorisation produced by the current methodology is presented for indicative purposes. In addition, details of the distribution of HTC categories by Local Authority using data currently available in January 2010 are given. This distribution may change as the data is updated later in 2010.

2. Introduction

One of the key objectives of Field Force Planning for 2011 Census is to minimise differential under coverage: adjustments for under-coverage are made with greater certainty if under-coverage is uniform (Abbott, 2009a). In order to reduce differential under coverage, more resources will be provided to enumerate areas where response rates are likely to be low. In addition, approximately three times more total resource will be available for follow-up in 2011 than in 2001. As the time available for following up unreturned census questionnaires is relatively short (even though the period is longer than in 2001), there will be little opportunity to increase recruitment in response to an unexpectedly low questionnaire return rate. Therefore, it is desirable that the initial resource allocation matches the amount of work to be done as closely as possible.

A field allocation algorithm that estimates the amount of work required to achieve a high response rate has been developed for the 2011 Census. Key to this algorithm is the estimation of the proportion of households that will require follow up in the field in order to elicit a response. This paper describes the construction of a model that uses area level information to predict final household non-response propensities. The model extends previous work by Rahman and Goldring (2006), who explored household factors associated with non-response in the 2001 Census. Under the assumption that the model of the relationship between area level information and household non-response observed in the

2001 Census is a good proxy for a 2011 model, predictions for 2011 are obtained by applying up to date area level information through the model fitted to 2001 data. As these predictions are for final household non-response propensities based on the design of the 2001 Census, adjustments are required to produce non-response rate estimates for the proportion of households requiring follow up in 2011. These adjustments account for the effectiveness of follow-up, measured from the 2001 Census, and for the impact of postal delivery, measured from the 2007 Census Test. Model based predictions are used to construct the hard to count categorisation (HTC) before these adjustments are made.

The HTC is a national (England and Wales) categorisation that assigns LSOAs into one of five classes according to the predicted household non-response rate. The HTC categorisation will be used as a general indicator of the expected level of non-response. In addition to being a parameter in the field allocation algorithm, the HTC categorisation is also used as a stratifying variable in the sample design of Census Coverage Survey (as outlined in Abbott, 2009b).

This paper firstly reviews previous work to identify household factors that associate with household non-response in the 2001 Census. These household factors are used as a guide to find area level information from updatable sources that are used to produce an equivalent model. Section 4 describes how these factors are used in the current statistical model that combines the prevalence of these factors through their estimated coefficients to explain 2001 household non-response propensity. Section 5 explores how well the current model was able to explain 2001 household non-response and section 6 describes the formation of the HTC categorisation and provides information on the 2009 version of the HTC used in the 2009 Census Rehearsal.

3. Factors associated with household non-response in the 2001 Census

In order to predict patterns of household non-response in the 2011 Census, it is necessary to understand the causes of household non-response. The most relevant data source to analyse factors that may affect response rates in the 2011 Census is the 2001 Census. The 2001 Census Coverage Survey (CCS) was an independent survey that visited every household in a 1% sample of postcodes to assess census participation. The CCS collected household data which could be compared between households that participated in the 2001 Census with those that did not. Rahman and Goldring (2006) constructed a multilevel logistic regression model of household non-response using household level data from the 2001 CCS. They found that household tenure and location and householder age, ethnicity and family status all strongly affected the likelihood of participation (Rahman and Goldring, 2006). Similar household factors have been consistently identified as determinants of non-response in other studies (e.g. 1991 Census (Heady *et al.,* 1994); 1991 and 2001 Census linked studies of social survey non-response (Foster, 1998; Freeth and Sparks, 2003; Durrant and Steele, 2009)). This suggests that similar factors will also be determinants of household non-response in 2011.

In order to make robust predictions for the 2011 Census, it is necessary that the prevalence of factors affecting household non-response is measurable from current data sources. Thus, individual household level variables are not optimal for use in a model constructed to predict household non-response in 2011. An alternative is to build a household model using variables available for small areas that correlate with those identified Rahman and Goldring study but are derived from data sources updated on a yearly basis. The Rahman and Goldring study was therefore used as a guide to find relevant area level variables that could be derived from administrative data sources that existed in 2001 and continue to exist consistently today for use to model household non-response in a census. According to the

Rahman and Goldring study, explanatory factors for household non-response are likely to include measures of ethnicity, urbanity, housing deprivation, young adults, economic deprivation and societal engagement. To allow household non-response rates to be produced at a high resolution, it is necessary that predictions are made at a reasonably low level of geography. As a large amount of high quality administrative data is published at the lower layer super output area level (LSOA), this level of geography was selected to be used as the building block of the model. Six updatable variables from administrative data sources were identified as correlates of household non-response (appendix 1) and are included in the area level model:

The proportion of individuals claiming Jobseeker's Allowance (DWP)
The proportion of pupils in state schools that are non-'White British' (DCSF)
The relative median house price (Land Registry)
The dwelling density (council tax data from the DCLG)
The proportion of people 16-29 years of age (ONS mid-year population estimates)
The standardised rate of notable offences recorded (Home Office)

In addition, variables derived from the 2001 Census were also considered for inclusion. Though these variables are not updatable, three are included as they significantly increase the ability of the model to explain household non-response in the 2001 Census, particularly where the updatable sources do not fully reflect all the factors identified by Rahman and Goldring (2006). These are:

The proportion of households paying some rent (private or social)
The proportion of households with single occupiers
The proportion of households in flats, apartments or maisonettes

A full description of these variables, including scatter plots showing their relationship to household non-response rates, is given in appendix 1. Additional variables that did not make it into the final model are also described in appendix 1.

4. Modelling household non-response

This section describes how a model explaining the non-response propensity of household using area level covariates was developed.

4.1. The dependent variable

The 2001 One Number Census project produced accurate population counts for each Local Authority by making an adjustment to the direct census count to correct for under-coverage. Households representing the estimated under-coverage were imputed either into a physical location where no response was received (e.g. refusals, non-contacts, which were recorded by enumerators who produced 'dummy forms' for such households) or into areas where similar households already existed (see ONS, 2005). Though the design of the 2001 One Number Census was to provide accuracy at the Local Authority level, there was a high correspondence between the numbers of 'dummy forms' produced for each Local Authority and the numbers of households to be imputed. This suggests that the household imputation rate by LSOA in 2001, though having greater error than the local authority rates, is an accurate reflection of the household under-coverage rate in the 2001 Census for local areas. Therefore, for the area level model described here, the imputation rates were used as the dependent variable. An important advantage is that these rates are available for the whole country, and are not restricted to the CCS sample. However, they do have some error associated with them. Though modelling is expected to smooth away error, there is the possibility that consistent error may be captured and incorporated into the model. Thus, model based predictions should be treated with some caution. In addition, it should be borne in mind that, though we largely refer to the model as a model of non-response, it is actually a model of under-coverage.

4.2. Model form

The aim of the following modelling exercise was to explain household response behaviour. All households within an LSOA are considered equivalent, all experiencing the same conditions (area level explanatory variables) and displaying the same probability of not responding, *p*. Logistic regression was used to model this probability *p*. X describes the explanatory, or independent, variables 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 non-response):

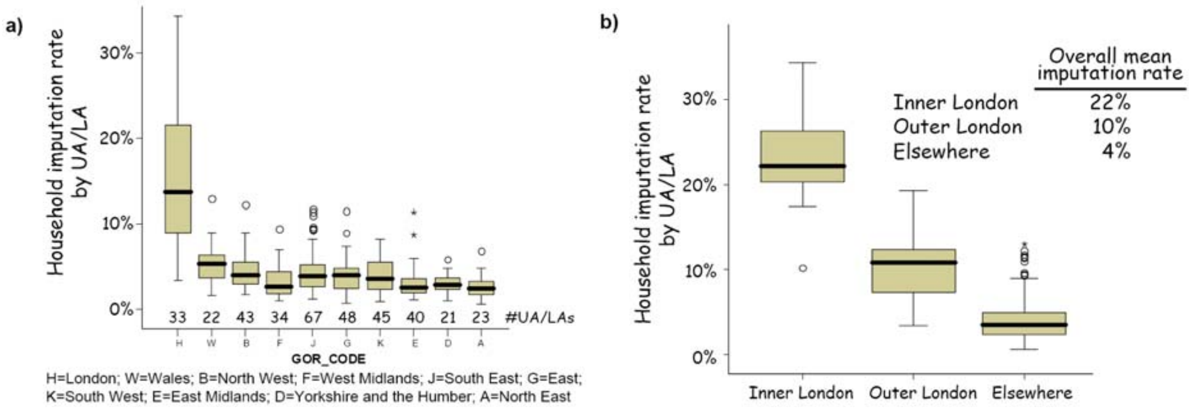
$$\text{Logit } p = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$$

Model fit is assessed using the improvement in the -2 log likelihood of the model compared to the null model (e.g. intercept only).

4.3. Regional effects

In 2001 the overall household imputation rate in London was 15%. This was much greater than for the other Government Office Regions (GORs) where overall household imputation rate for each GOR ranged from 3.3% to 5.4%. Comparing the spread of household imputation rate by Local Authority also suggests that London is distinct. This distinction becomes clearer by splitting London into Inner London and Outer London and combining all other GORs together to produce three regions (figure 1). These three regions are called 'Inner London', 'Outer London' and 'Elsewhere'. The three region geography has a better fit than GOR geography in models of non-response and was used for modelling.

Figure 1: Box plots of GORs (a) and Regions (b) showing the spread of household non-response rates across Local Authorities in the 2001 Census. The ordering of GORs in a) is driven by the overall household imputation rate.



4.4. Explanatory variables

In order to capture as much societal change as possible in our model, its construction was originally limited to those variables derived from administrative data sources that were updated at least yearly. Model version 1 was thus composed of the categorical region variable and five continuous variables that interacted with the region variable: pcJSA (the percentage of individuals claiming Jobseeker's Allowance by LSOA); pcNWB (the percentage of pupils in state schools that are non-'White British' by LSOA); region_rel_hp (the median house price of residential properties within an MSOA relative to the regional mean); dw_density (the dwelling density of residential properties within an LSOA); and pcAge (the percentage of people between 16-29 years of age by LSOA). This resulted in a 7.8% improvement in the -2 log likelihood from 9728924 for the intercept only null model to 8965840 for model version 1.

Adding a categorical variable for each local authority to model version 1 significantly increases fit (-2 log likelihood of 8806824, improvement over null model =9.5%). This suggests that there are differences between Local authorities that are not captured in the regional model (version 1) which relies on differences between LSOAs to explain differences between local authorities within each region. Some of these differences are likely to be real; others may be due to factors that are unlikely to be repeated in a future census. Without identifying their causes, it is difficult to foresee how these differences between local authorities will be maintained in the future. Though inclusion of a local authority categorical variable in the model increases fit, this needs to be balanced against the function of the model to capture change and predict future non-response rates. Thus, a local authority categorical variable was not included in the final model. However, a number of measurable Local authority level covariates were added to the model to produce model version 2. These covariates are the percentage of people within the local authority on jobseeker's allowance (LA_pcJSA), the percentage of pupils resident by local authority who are non-'white British'(LA_pcNWB), the median dwelling density of an LSOA within each Local Authority (LA_median_dd), the percentage of people in a Local Authority between 16 and 29 years of age (LA_pcAge) and the standardised local authority count of notable offences (LA_offences). Inclusion of these variables improved model version 2 fit significantly over model version 1 (-2 log likelihood of 8955858, improvement over null model =7.9%). Though this improvement in fit is much smaller than that seen by including a local authority categorical variable, all these local authority level covariates are updatable.

It is known that household tenure and single occupation were important explanatory factors in household non-response in the 2001 Census (Rahman and Goldring, 2006). However, administrative data sources capturing these variables were not identified (see appendix 1). In order to assess the impact of these variables in fit of our model we derived a number of LSOA level variables from the 2001 census and assessed their impact in the context of model version 2. Three such census variables significantly improved the fit of the model. These are pcSingOccs (the percentage of households by LSOA that had a single occupier in 2001), pcRent (the percentage of households by LSOA that paid some rent (private or social)) in 2001, and pcFMA (the percentage of households by LSOA that were within flats, apartments or maisonettes in 2001). Inclusion of these three census variables created model version 3, which has a -2 log likelihood of 8918382 (improvement over null model =8.3%). The local authority variable LA_pcJSA was no longer significant when these census variables are added and was dropped from the model.

Though adding these three 2001 Census variables increases fit, it was possible that they could weaken the contribution to the model of the updatable variables. Therefore their inclusion may decrease the amount of change predicted when the model is used to predict non-response rates using current data for updatable variables. However, this may also be considered a good feature of the model, as it means that we are taking a conservative approach that balances between the patterns seen in 2001 and differences in variables that are changing over time. In other words, we have some protection against making extreme estimates where the updatable sources have changed significantly. Analysis of the predictions made in 2009 by both model version 2 and 3, also contributed to the selection of model version 3 as the version used in the 2009 Census Rehearsal.

The models presented here are household level models of non-response that use area level factors to explain differences in household non-response propensities. Thus, the models make no attempt to explain the differences between households that respond and households that do not respond within the same area. As a consequence, when assessing model fit at the household level, the improvement over the null model appears modest: Model version 3 represents an 8.3% improvement.

Model of household non-response version 3:

$$\begin{aligned}\text{Logit } p_{hijk} = \; & \text{region}_k * \{\beta_{ohijk} + \beta_{1k}(\text{pcJSA})_{hi} + \beta_{2k}(\text{pcNWB})_{hi} + \beta_{3k}(\text{Region\_rel\_hp})_{hi} \\ & + \beta_{4k}(\text{dwelling\_density})_{hi} + \beta_{5k}(\text{pcAge16-29})_{hi} + \beta_{11k}(\text{pcSglOccs})_{hi} \\ & + \beta_{12k}(\text{pcFMA})_{hi} + \beta_{13k}(\text{pcRent})_{hi}\} + \beta_{7}(\text{LA\_pcNWB})_{hij} \\ & + \beta_{8}(\text{LA\_median\_dd})_{hij} + \beta_{9}(\text{LA\_pcAge})_{hij} + \beta_{10}(\text{LA\_Offences})_{hij}\end{aligned}$$

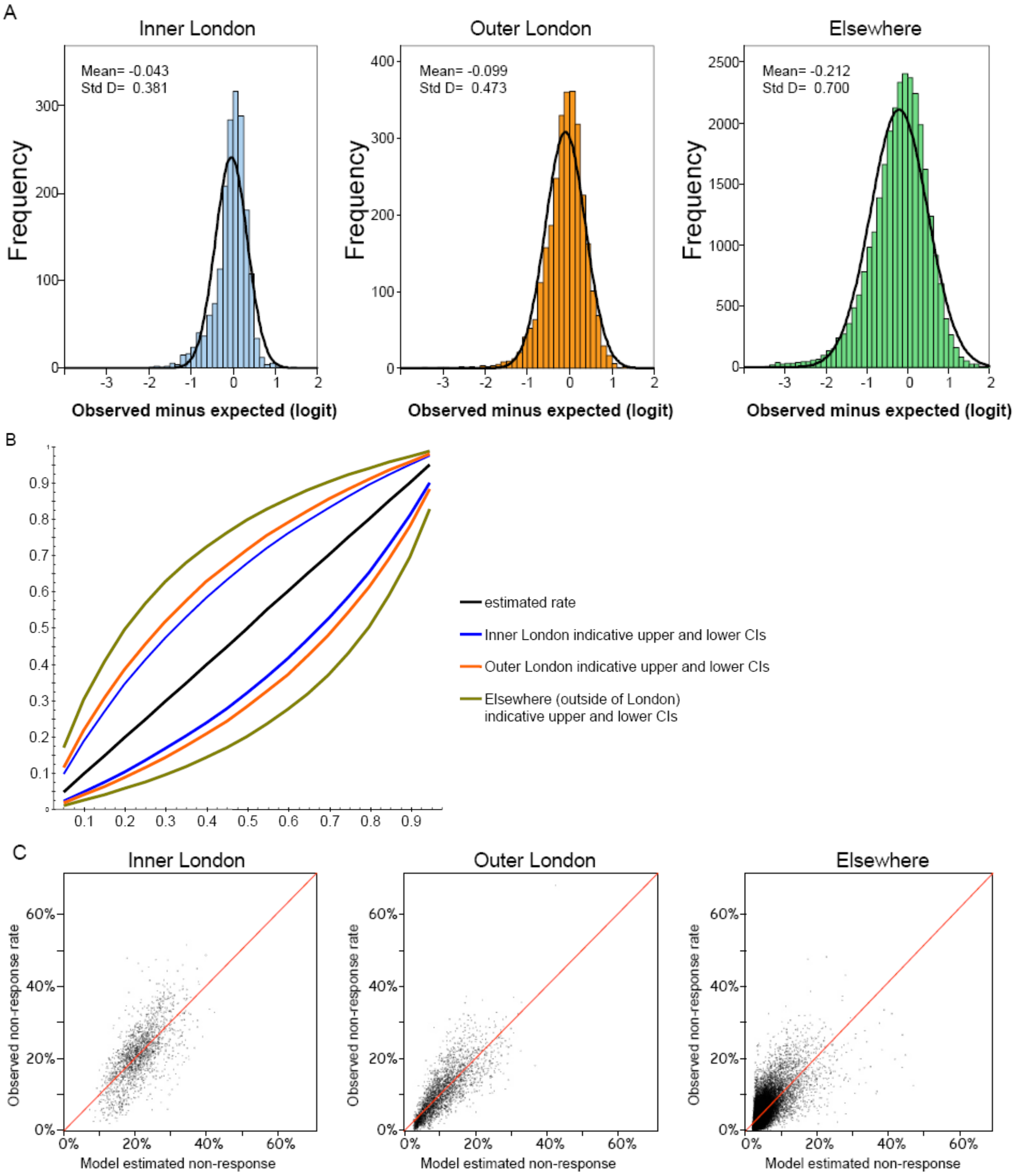
Where p_{hijk} is the probability of household(h) non-response in LSOA(i), within Local Authority(j) and region(k). A description of the variables is given in appendix 1. Estimated values of the β's are given in appendix 2.

5. Empirical measurement of model fit

In principle, the model of household non-response propensity can be used to derive estimates for 2011 Census household non-response rates. To do this the 2001 values for the covariates are replaced with 2011 (or most timely) values. Since the relationship between the explanatory factors and household non-response may change since 2001, model based predictions can only be indicative. In other words, there is unquantifiable uncertainty in the ability of the model to explain future household non-response rates. If the assumption that the relationships modelled using 2001 remains constant is violated significantly, then our estimates of non-response will have some bias: localities having similar characteristics will either be consistently over-estimated or consistently under area.

This is additional to the uncertainty due to the limited ability of the model to explain household non-response rates in 2001. In order to empirically quantify the fit of the model, the observed household non-response logit *p* in 2001 was compared to the model expected logit *p* for every LSOA. This allows us to analyse what proportion of the variation in logit *p* response rates by LSOA was explained by the model. The model predicted LSOA variation was 47% of the total variation in LSOA level household non-response rates. This suggests that the model 'explains' 47% of the LSOA variation observed in 2001. The residual, or unexplained part of the variation at the LSOA level, is given by the observed minus expected logit *p* (where *p* is proportion of households not responding) measured for each LSOA. A histogram was produced for each region in the model and a normal distribution overlaid (figure 2a). In each of the three regions, the mean distance from the model predicted to the observed value for an LSOA is negative (the model predicted slightly more undercount than observed). As expected, the distribution of LSOA residuals is approximately normal. Under the assumption that the inter-relationships captured by the model continue to exist, the standard deviations of the LSOA level residuals given in figure 2a can be used to provide an indication of the smallest confidence interval that we can place any estimate of final response rate for any LSOA based on the 2001 Census field design (design changes since 2001, most notably postal delivery, are not modelled directly). An indication of the size of these confidence intervals (1.96*standard deviation for indicative 95% level of confidence) on the probability scale is given in figure 2b. This shows that, for instance, if we predict a non-response rate of 50% in Inner London, the indicative confidence interval is about 32-68%. It is stressed that this is indicative only, based on empirical fit and assuming that the 2001 model is valid and stable. These indicative confidence intervals apply to final response rates under the operational conditions of the 2001 Census. A comparison of model predicted non-response rates plotted against observed household non-response rates by LSOA and region is shown in figure 2c.

Figure 2: A) Empirical representation of model fit by region on the logit scale. Frequency refers to the number of LSOAs having an observed minus expected logit value in the ranges shown. B) Indicative confidence intervals at the LSOA level based on empirical measurement of the fit of the model to 2001 Census data. C) Scatter plot of model fit by region on the probability scale. Each data point is an LSOA.



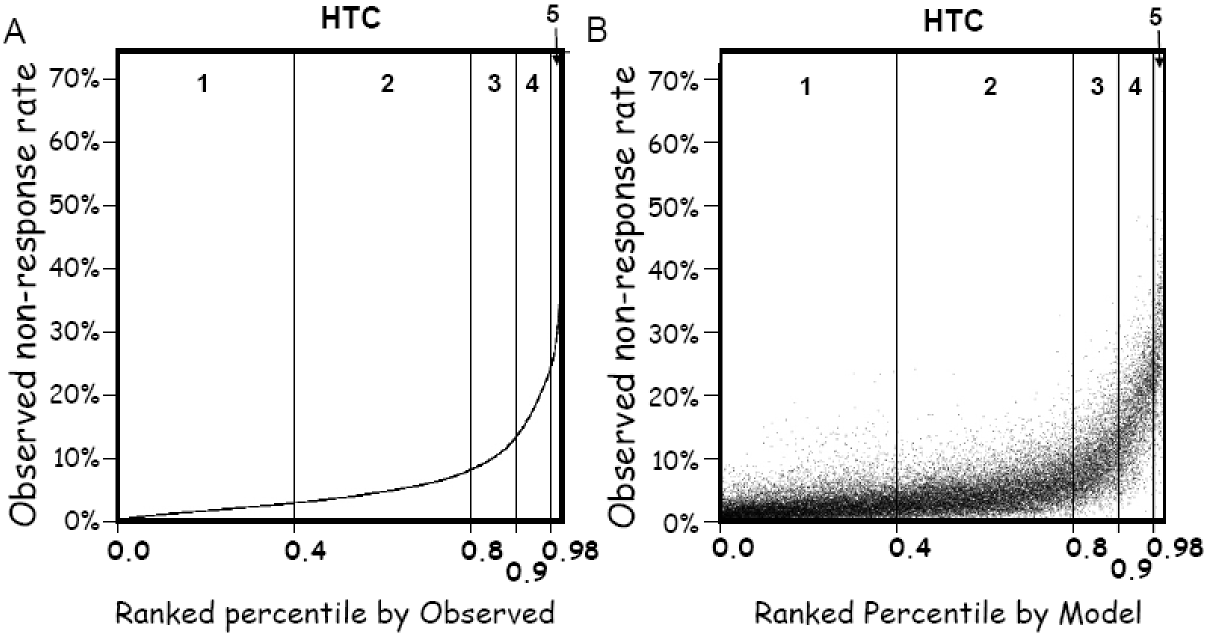
6. The hard to count categorisation

This section describes how predictions of household non-response by LSOA are used to produce the hard to count (HTC) categorisation. The HTC will be used to stratify LSOAs for the census coverage assessment and adjustment process. In addition, the HTC will be used to categorise estimates of household non-response rates for workload planning. Workload planning requires estimates of day 10 non-response rates. It is on census day plus 10 (day 10) that enumerators are deployed to follow up non-responding households to elicit a response. As the model of household non-response estimates final non-response rates under the operational conditions of the 2001 Census, further adjustments are required to account for the effectiveness of follow up in 2001 and for the operational changes since 2001, most notably postal delivery.

6.1. Construction and likely performance of the HTC categorisation

The hard to count categorisation (HTC) is constructed from the model based predictions of household non-response described above. They are used to rank LSOAs in England and Wales from lowest predicted non-response rates to highest. From this ranking, those LSOAs that are expected to comprise those areas easiest to enumerate are classed as HTC-1. LSOAs are thus assigned a class from HTC-1 to HTC-5 by increasing expected difficulty of enumeration. LSOAs classed as HTC-5 are expected to have the lowest response rates and so be the hardest to enumerate. 40% of LSOAs are categorised as HTC-1, 40% are categorised as HTC-2, 10% as HTC-3, 8% as HTC-4 and 2% of LSOAs are categorised as HTC-5. The design of the 40/40/10/8/2 categorisation is partly 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% of hardest to enumerate group is driven by the observation that the largest variation in non-response rates is within this group when the HTC is constructed using 2001 data (figure 3). There is relatively little difference in non-response rates between LSOAs within the easiest 80th percentile and our ability to differentiate between them is weak (figure 3). However, there is a fairly steep gradient of differences between LSOAs in HTC categories 3, 4 and 5.

Figure 3: Scatter plot of a) observed rank percentile of non-response by LSOA and b) ranked percentile of model based predictions for non-response, in the 2001 Census against the observed household non-response rates. Vertical lines represent HTC boundaries.



The ability of the area level household non-response model to differentiate between relative ranks of LSOA household non-response rate is not perfect (compare figure 3 part a) with part b)). If we define error as a mis-categorisation occurring when the observed non-response rate is higher than the expected value for the category above, or lower than the expected value for the category below, then in 2001 there would have been an error rate of up to 20%. If a distance of 2 categories is used to define error, then an error rate of approximately 5% is observed using 2001 data. This suggests that the HTC is an appropriate tool to predict patterns of non-response: the HTC generally describes the relative difficulty of enumeration of LSOAs (also see table 2). However, the HTC can not be used to precisely predict the household non-response rate for a given LSOA. Thus, it is likely that there will be some areas where the actual household non-response rate is substantially different from that predicted. In addition, there may be clustering of mismatches due to non-random effects not captured in the model. Thus, contingency plans are being developed to deal with unforeseen difficulties in enumeration where ever they may arise.

6.1. The 2009 version of the HTC categorisation

A hard to count categorisation was created for the 2009 Census Rehearsal using the most currently data available for variables in the model (figure 4). This categorisation and the rates shown were as though we were performing a Census in 2009: no account of the voluntary nature of the rehearsal was applied to the data presented here (therefore this data should not be used to judge the success of the 2009 Census Rehearsal directly). Due to increased prevalence of factors since 2001 is clear that household non-response will be a bigger challenge in the 2011 Census (table 1). Indeed, the model suggests that this increased challenge will be most strongly felt in areas that proved difficult in 2001. Though the predicted final response rate is higher in 2009 than in 2001, this reflects the operational conditions in the 2001 Census. More than three fold more resource will be committed to follow-up in the 2011 Census. The HTC categorisation forms part of the method to judge where that extra resource ought to be targeted.

Table 1: Model predicted final household non-response rates by HTC in 2001 and 2009. The model makes no account of operational changes since 2001 that are designed to reduce non-response, especially in hard to enumerate areas. Overall non-response rate is calculated from weighted average of HTC categorisation.

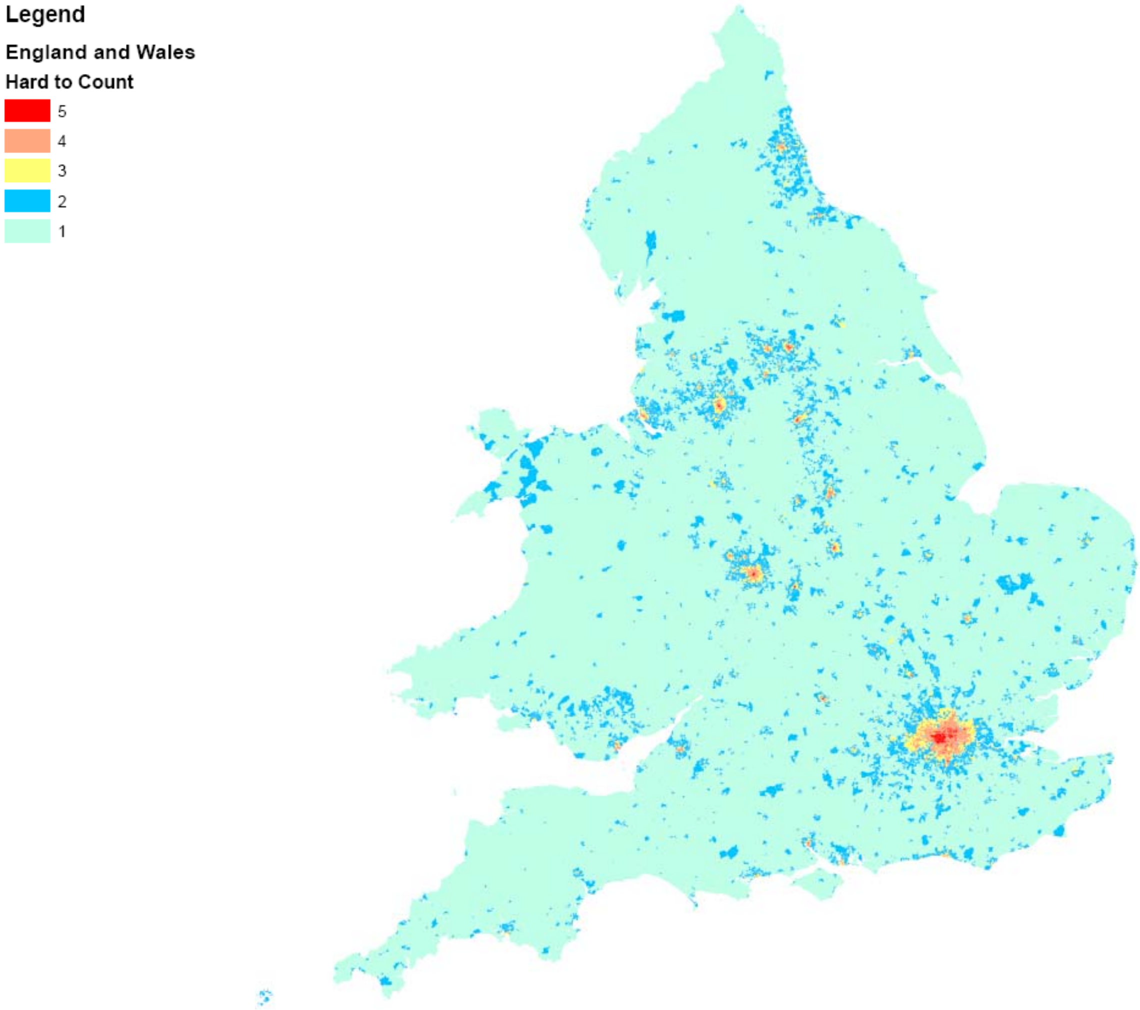
	Model predicted final household non-response rate in 2001	Model predicted final household non-response rate in 2009
HTC-1	2.7%	2.8%
HTC-2	4.6%	5.1%
HTC-3	9.1%	10.5%
HTC-4	17.4%	19.6%
HTC-5	27.7%	31.4%
Overall	5.4%	5.9%

In order to validate the HTC produced in 2009, it has been used to assess non-response rates for social surveys in 2006. This analysis confirmed that the HTC is a strong predictor of current household non-response (table 2).

Table 2: Non-response rates by HTC in social surveys performed by the ONS in 2006. Eligible cases. (GHS and LFS wave 1 only). Overall non-response rate is calculated from weighted average of HTC categorisation.

	General Household Survey	Labour Force Survey	Family Resources Survey	Expenditure and Food Survey
HTC-1	19.6%	27.1%	31.8%	37.3%
HTC-2	22.3%	29.8%	36.9%	42.0%
HTC-3	28.7%	33.0%	41.0%	47.1%
HTC-4	30.1%	39.9%	47.1%	56.4%
HTC-5	39.3%	41.4%	55.0%	62.8%
Overall	22.3%	30.0%	35.4%	41.3%

Figure 4: The distribution of the HTC categories from HTC2009 across England and Wales.



7. References

Abbott, O. (2009a). ONS report presented at RSS meeting in June 2009
'Precision of Census estimates for different levels and patterns of Census response'.
http://www.rss.org.uk/rssadmin/uploads/1348720_Precision%20for%20different%20census%20patterns%20for%20RSS%20v0.2.pdf

Abbott, O. (2009b). Population Trends, 137, pp25-32.
'2011 UK Coverage Assessment and Adjustment methodology.'
http://www.palgrave-journals.com/pt/journal/v137/n1/pdf/pt200931a.pdf

Cabinet Office (2008). The Stationery Office, London. ISBN: 9780101751322
'Helping to shape tomorrow – The 2011 Census of Population and Housing in England and Wales.' CM 7513.

Durrant, G.B. and Steele, F. (2009). J. R. Statist. Soc. A, 172, 2, pp. 1-21.
'Multilevel modelling of refusal and non-contact in household surveys: evidence from six UK government surveys.'

Foster, K. (1998). GSS Methodology Series, No.8. ISBN: 1857742729
'Evaluating non-response on household surveys: Report of a study linked to the 1991 Census.'

Freeth, S. and Sparks, J. (2003). Internal ONS document
'Summary of the 2001 Census-linked studies of survey non-response.'

Office for National Statistics (2005). Office for National Statistics, London
'One Number Census Evaluation Report'.
http://www.statistics.gov.uk/census2001/pdfs/onc_evr_rep.pdf

Office for National Statistics (2006). Office for National Statistics, London
'Enumeration Targeting categorisation to be used in the 2007 Census test'.
http://www.statistics.gov.uk/census/pdfs/EnumerationTargetingCategorisation.pdf

Rahman, N. and Goldring, S. (2006). Survey Methodology Bulletin, vol 59, pp 11-24.
'Factors associated with household non-response in the Census 2001.'
http://www.s3ri.soton.ac.uk/isi2007/papers/Paper13.pdf

Appendix 1.

This appendix provides full details of variables in the area level model. Details of additional variables considered, but not ultimately used in the model are also provided. Criteria for model inclusion include: the availability of data for 2001 and currently in a consistent and timely fashion; lack of collinearity with other variables; and impact on model fit. When non-updatable data sources are considered, the impact on their inclusion upon the coefficients for updatable variables was also considered (i.e. did their inclusion reduce the ability of the model to capture change?).

A1.1: The percentage of individuals claiming Jobseeker's Allowance
Previous studies have shown that householder unemployment is a determinant of household non-response in the census and social surveys (Foster, 1998; Freeth and Sparks, 2003; Rahman and Goldring, 2006; Durrant and Steele, 2009). In the area model of household non-response this variable was used at the LSOA level (pcJSA) and the Local Authority level (LA_pcJSA). These variables were derived from jobseeker's allowance claimant counts available from the Department of Works and Pensions at the NOMIS website (www.nomisweb.co.uk). Scatter plots of the relationship between these variables and household non-response rates are shown in figure A1.

A1.2: The proportion of pupils that are non-'White British'
Ethnicity of householder has previously been shown to be a determinant of household non-response in both the 2001 Census and in social surveys (Foster, 1998; Freeth and Sparks, 2003; Rahman and Goldring, 2006). In particular, Rahman and Goldring reported that the 'proportion born outside of the UK' was a determinant of 2001 census non-response in their area model (Rahman and Goldring, 2006). The school census (also known as PLASC, for pupil level annual school census) was initiated in England in 2002. It is a statutory requirement for all schools in England receiving state funding to fill in a return, which is collected by the department of children, schools and families (DCSF). Though the DCSF school census only covers England, a similar process of data collection occurs in Wales. The Welsh school census began in 2004 and its data is maintained on the Welsh National Pupil Database. Though these data sources do not go back to 2001, they have the widest coverage of any continuously updated data source collecting ethnicity data. The 2002 DCSF school census data contains records of over 6.5 million pupils, 97.1% of which contain both pupil residence postcode and ethnicity data. This dataset was used to produce variables measuring the proportion non-'White British' at the LSOA level (pcNWB) and LA level (LA_pcNWB). Less than 1.5% of LSOAs have less than 50 pupils resident. In these cases, the appropriate MSOA values are exchanged. Scatter plots of the relationship between these variables and household non-response rates are shown in figure A1.

A1.3: The relative median house price
The Rahman-Goldring model of census non-response uses a housing deprivation variable that is derived from the Indices of Multiple Deprivation (IMD: Housing Wider Barriers) (Rahman and Goldring, 2006). The indices of multiple deprivation (IMD) are not comparable over time and do not cover Wales equivalently. They are therefore inappropriate variables to consider in the model. In addition, studies of both social survey and census household non-response identified that householders living in a converted/shared house, those paying part rent/part mortgage, those who are single occupiers or are young are consistently less likely to participate (Foster, 1999; Freeth and Sparks, 2003; Rahman and Goldring, 2006). Householders in these groups tend to live where accommodation is relatively less expensive. As the house price market in actual values has been very volatile since 2001, actual house prices are not suitable, but rather an internally relative measure is required. For this purpose, a measure relative to the local authority average may be more appropriate than

the national average, as an individual's place of residence is often restricted geographically (e.g. by place of employment or family considerations etc.). However, creating a house price measure relative to the local authority median limits its use to models for which individual local authorities are identified. To overcome this limitation, the MSOA level variable 'reg_rel_hp' was produced. This variable is derived from Land Registry data relating to median house price by MSOA at point of sale and is relative to the regional mean of MSOA median house prices. To construct this variable, the natural log of the ratio between the MSOA median and the mean of MSOA medians within the region is taken. Scatter plots of the relationship between this variable and household non-response are shown in figure A1.

A1.4: Dwelling density
It is known that those living in cities tend to be less likely to participate in the census or social surveys (Foster, 1998; Freeth and Sparks, 2003; Rahman and Goldring, 2006; Durrant and Steele, 2009). In busy cities such as London, people may feel that they are one of a crowd and so become anonymous. Such people may feel that their individual participation may not be important as they feel they are represented by their neighbours. A measure of the number of households per LSOA is provided yearly from council tax records and is published on NeSS (neighbourhood.statistics.gov.uk). To convert this number to a dwelling density, it was divided by the land-based area in hectares of each LSOA (provided by ONS Geography) to produce the variable dwelling_density (dw_den in column 2 of table A2). In addition to capturing social anonymity, it is likely that this dwelling density measure also captures those areas were there are many flats and/or converted houses. Householders who live in such dwellings tend to not respond in social surveys or the census (Foster, 1998; Freeth and Sparks, 2003; Rahman and Goldring, 2006; Durrant and Steele, 2009). A further dwelling density measure was created at the Local Authority level. The variable LA_median_dd is the median dwelling density by LSOA within a Local Authority. Scatter plots of the relationship between these variables and household non-response rates are shown in figure A1.

A1.5: The proportion of people between 16 and 29 years of age
It is known that young adults are less likely to participate in the Census and are less likely to be contacted in social surveys (Freeth and Sparks, 2003; Rahman and Goldring, 2006; Durrant and Steele, 2009). In particular, Rahman and Goldring reported that the 16-29 age group was more than 1.7 times more likely not to participate in the 2001 Census than the 40-59 age group (Rahman and Goldring, 2006). Thus, it was felt that the proportion of the population between the ages of 16 and 29 may be an important determinant of household non-response in our model. Mid-year population estimates for each are published yearly by ONS. These estimates are Census anchored and based on births and deaths and estimates for migration and are produced by Local Authority. The LSOA figures are constrained to the Local Authority value and produced by a ratio change method based on patient record and child benefit data. A variable was produced at the LSOA level (pcAge16_29) and at the Local Authority level (LA_pcAge). Scatter plots of the relationship between these variables and household non-response rates are shown in figure A1.

A1.6: The proportion of households with single occupiers in 2001
Rahman and Goldring (2006) showed that households containing single occupiers were twice as likely not to respond in the 2001 Census as households containing families. No centrally collected administrative data source that captured this was identified, though households obtaining a single occupier discount could potentially be identified from Council Tax records. pcSglOccs is a variable at the LSOA level that gives the percentage of households in the post-imputed 2001 Census that have a single occupier. Scatter plots of the relationship between this variables and household non-response is are shown in figure A1.

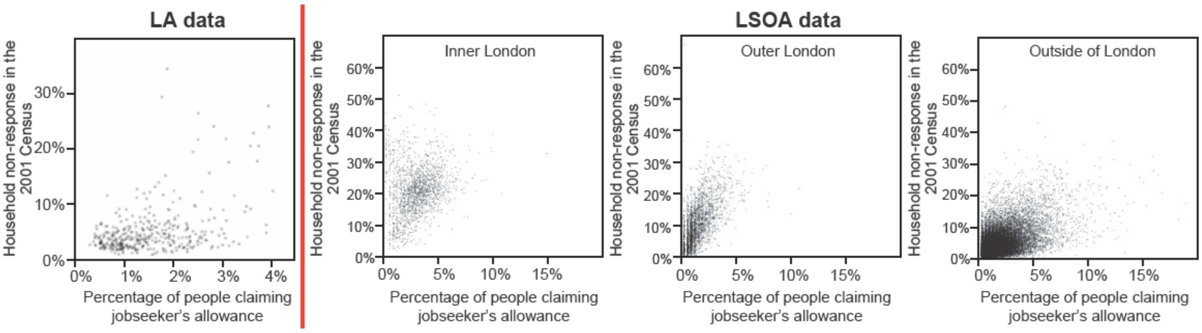
A1.7: The proportion of households that resided in flats in 2001
Rahman and Goldring (2006) showed that household accommodation was a determinant of household non-response in the 2001 Census. It was felt that the relative house price and the dwelling density variables may not completely capture this. pcFMA is a variable at the LSOA level that gives the percentage of households in the post-imputed 2001 Census that are within flats, apartments or maisonettes. Scatter plots of the relationship between this variables and household non-response is are shown in figure A1.

A1.8: The proportion of households that paid some rent in 2001
Rahman and Goldring (2006) showed that households paying rent more than twice as likely not to respond in the 2001 Census than owner occupiers. An administrative data source capturing those paying rent was not identified. pcRent is a variable at the LSOA level that gives the percentage of households in the post-imputed 2001 Census that pay some rent (public or privately). Scatter plots of the relationship between this variables and household non-response is are shown in figure A1.

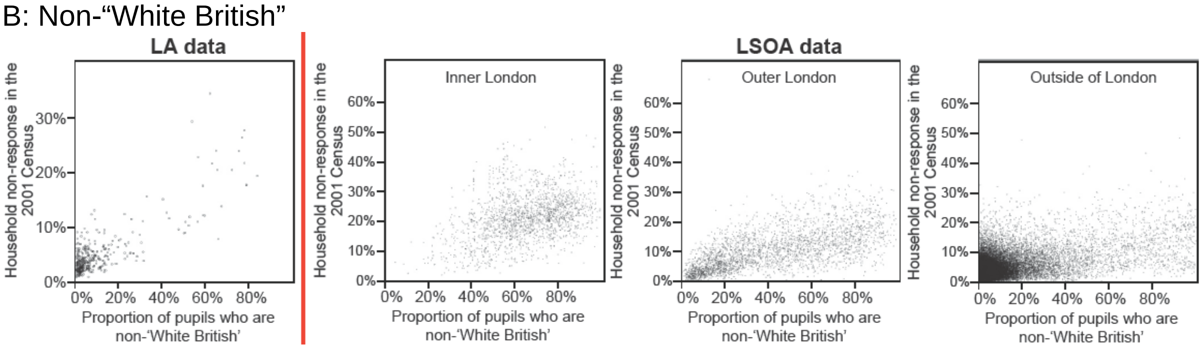
A1.9: The number of notable offences committed within a local authority
There is no clear picture of whether crime rate is associated with Census non-response from other studies identified. However, it seemed likely that crime rates could indicate life dissatisfaction, which may be associated with non-response. As this data is available as a number for local authority only, it was converted into a standardised form by dividing by the number of persons within the Local Authority and multiplying by 100 (LA_Offences). In order to examine the nature and strength of the relationship between number of notable offences and household non-response rates a number of simple models were built that showed that household non-response did tend to increase as the standardised proportion of notable offences increases. As for other variables considered, there is evidence that this relationship varies by region. A Scatter plot of the relationship between this variable and household non-response is shown in figure A1.

Figure A1: Scatter plots showing the relationship between variables in the model and household non-response in 2001.

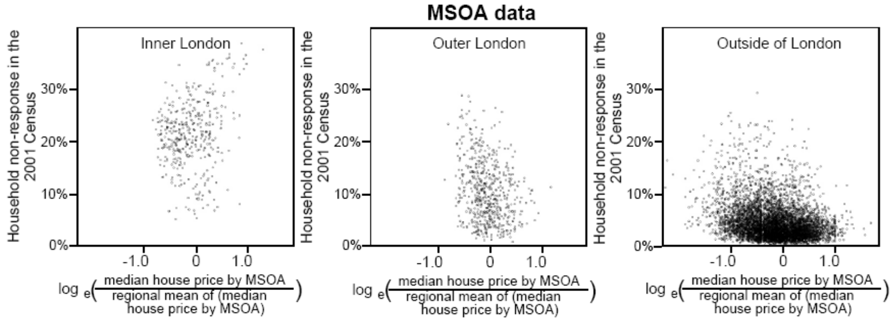
A: Jobseeker's Allowance



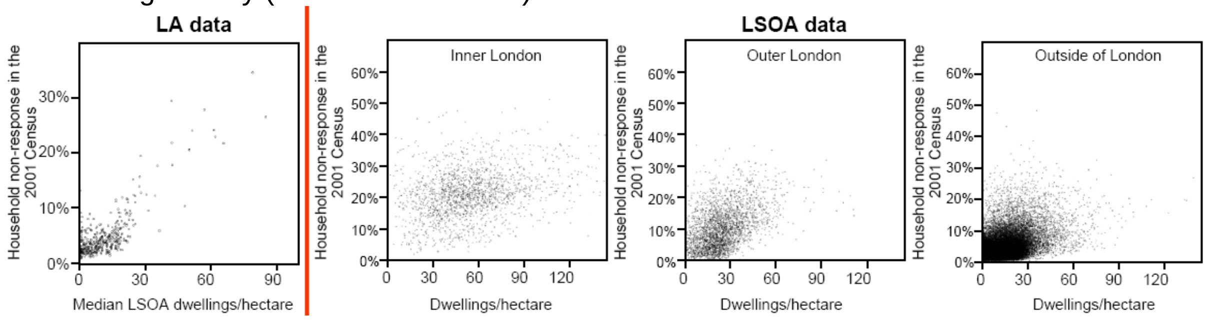
B: Non-"White British"



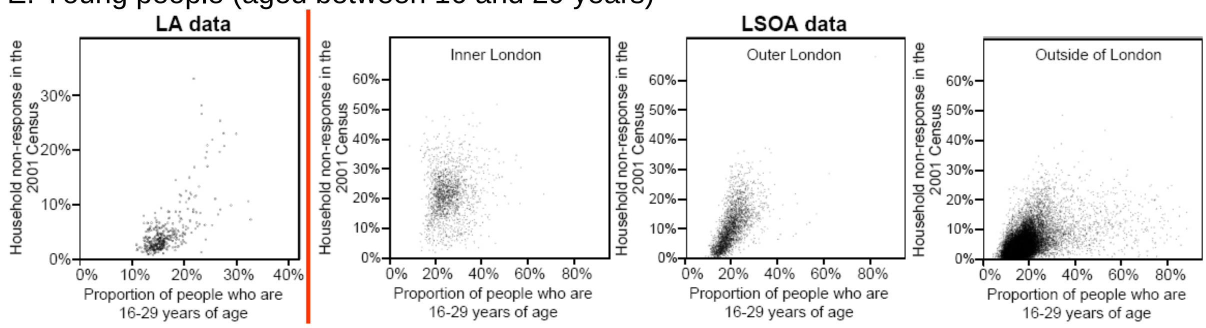
C: Relative house price



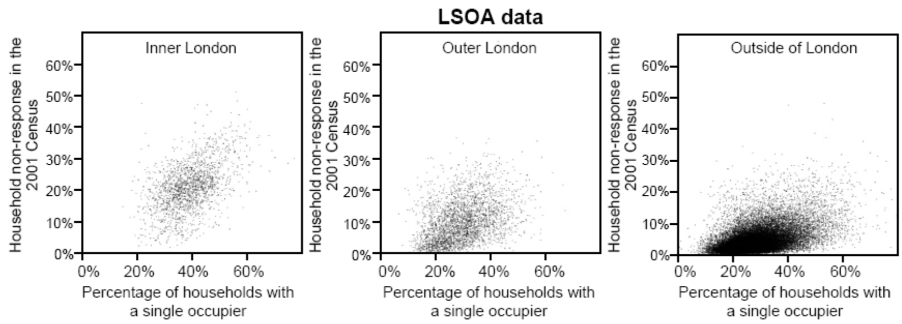
D: Dwelling density (residences/hectare)



E: Young people (aged between 16 and 29 years)

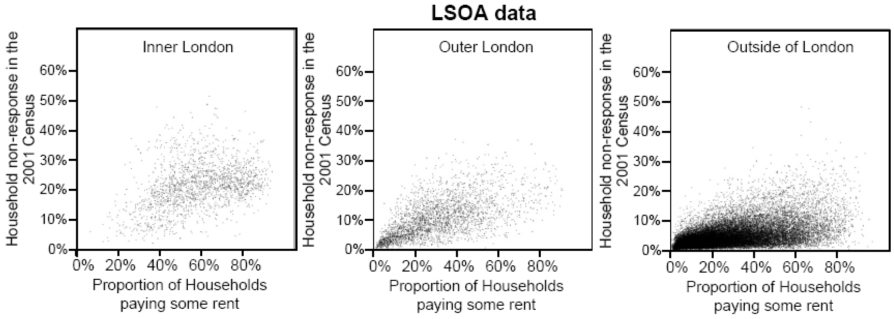


F: Single Occupiers

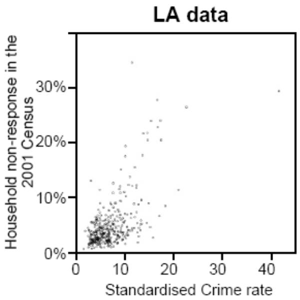


G: Resident in flats, apartments or maisonettes

H: Pay some rent



I: Standardised crime rate



A1.10: Other variables considered, but not used

In addition to the variables described above, a number of other variables have been considered. These are described briefly below.

A1.10.1. Council Tax Band. Studies of both social survey and Census non-response indicate that householders living in flats are more likely not to respond than others. The proportion of dwellings in Council Tax bands A (or in A and B) was therefore examined for its ability to capture 2001 Census non-response. The idea here was that flats would be more likely to banded A (or B) than many other forms of residence. However, this approach proved fruitless for 2 main reasons:

a. There was not a uniform distribution of housing type for each council tax band across England and Wales. This is compounded by the failure of a presence of a proportion of a particular council tax band to predict the make-up of dwellings in the remainder of the LSOA. The result of which was the failure to find any clear relationship between the proportion of dwellings of council tax A (or bands A and B) within an LSOA in 2001 and 2001 Census response rates (not shown).
b. Council tax bands were recalculated in Wales in 2006 and it is not simple to make an adjustment to match the conditions for banding in place in 2001.

The non-updatable Census variable pcFMA (proportion of households that are flats, apartments or maisonettes by LSOA) has been used.

A1.10.2. Proportion of sales that are flats. The land registry data collects property type. Thus it is possible to determine the proportion of sales in an area that are of flats. This could be used as a measure of the actual proportion of flats in the area. Though this approach was feasible, it was dropped because it was found that the proportion of property sales that were flats showed a very strong relationship with dwelling density throughout England and Wales (not shown).

A1.10.3. Commercial geo-demographic designations. A number of commercial companies produce categorisations used to segment the UK market based on address. There is evidence that variation in 2001 Census response rates can be partially explained by the certain geo-demographic tools. However, these market segmentation tools were not designed to measure Census non-response. The clustering algorithms used are poorly suited to non-response resulting in poor consistency between hierarchical levels of geo-demographic designation and Census non-response. Furthermore, no evidence could be found to suggest that commercial geo-demographic tools used data sources that were superior to the publicly available data sources described above in their relevance to Census non-response.

A1.10.4. Proportion of households living in privately rented accommodation. Studies of both social survey and Census non-response indicate that householders living in privately rented accommodation are more likely not to respond than others. Data to capture this consistently at the LSOA level is unavailable. The non-updatable Census variable pcRent (proportion of households that rent either privately or in social housing) has been used.

A1.10.5. Proportion of persons seeking housing assistance. A measure of this was included in the IMD: Housing wider barriers sub-domain, used in a model produced by Rahman and Goldring (2006). It was possible that this variable captures some housing deprivation relevant to census non-response that is distinct from the dwelling density and relative house price measures described above such as a measure of the demand for social housing. However, we failed to find a source of data that captured this at the LSOA level and would go back to 2001. We considered the variable **pcOverCwd** (percentage of households with more than half a person per room recorded in the 2001 Census) as a measure of housing deprivation. In the context of the other variables in the model, no improvement of fit was found, and the variable was dropped.

A1.10.6. Indices of Multiple Deprivation. These variables were considered, but dropped because they were not updated using a consistent methodology and they do not equivalently cover Wales.

A1.10.7 Single Occupiers from Council tax data. It is possible that an updatable variable on single occupiers could be obtained from Council Tax records, as a single person discount is usually offered. However, this information is not gathered centrally.

A1.11: Additional relevant information. It has been suggested that response rate data from the electoral commission or Local Authority place survey could be used as an indicator of Census response rates. In order to evaluate this, data is required for 2001. This is to demonstrate that there is a relationship between response rates in these processes and the Census and to quantify its strength. This criterion is not met for electoral roll or LA place survey data. These data sources have been used to validate model predictions. The strongest validation of the model predictions is social survey non-response rates shown in table 2.

Appendix 2

This appendix provides the full details of model version 3. This model was used to predict household non-response by LSOA and produce the HTC2009 presented in section 6 and appendix 3.

Table A2: Area level model of household non-response, including estimated β's.

Model	Estimated βs		Significance
logit p=	β1 (pcJSA) Inner London	0.0123	P<0.0001
α	β1 (pcJSA) outer London	0.0377	P<0.0001
+β1(pcJSA)	β1 (pcJSA) Elsewhere	0.0392	*
+β2(pcNWB)	β2 (pcNWB) Inner London	0.0090	P<0.0001
+β3(region_rel_hp)	β2 (pcNWB) outer London	0.0095	**
+β4(dwelling_density)	β2 (pcNWB) Elsewhere	0.0085	***
+β5(pcAGE16_29)	β3 (rel_hp) Inner London	0.0560	P<0.0001
+β11(pcSglOccs)	β3 (rel_hp) outer London	-0.1597	P<0.0001
+β12(pcFMA)	β3 (rel_hp) Elsewhere	-0.0177	P<0.0001
+β13(pcRent)	β4 (Dw_den) Inner London	-0.0005	P<0.0001
+Region	β4 (Dw_den) outer London	0.0010	P<0.0001
+β1(pcJSA)*Region	β4 (Dw_den) Elsewhere	0.0038	P<0.0001
+β2(pcNWB)*Region	β5 (pcAge) Inner London	0.0019	P<0.0001
+β3(reg_rel_hp)*Region	β5 (pcAge) outer London	0.0296	P<0.0001
+β4(dw_den)*Region	β5 (pcAge) Elsewhere	0.0150	P<0.0001
+β5(pcAGE)*Region	β7 (LA_pcNWB)	0.0035	P<0.0001
+β11(pcSglOccs)*Region	β8 (LA_median_dd)	0.0117	P<0.0001
+β12(pcFMA)*Region	β9 (LA_pcAge)	0.0043	P<0.0001
+β13(pcRent)*Region	β10 (LA_Offences)	-0.0138	P<0.0001
+β7(LA_pcNWB)	β11 (pcSgl) Inner London	0.0203	P<0.0001
+β8(LA_median_dd)	β11 (pcSgl) outer London	0.0187	****
+β9(LA_pcAGE)	β11 (pcSgl) Elsewhere	0.0140	P<0.0001
+β10(LA_Offences)	β12 (pcFMA) Inner London	-0.0020	P<0.0001
	β12 (pcFMA) outer London	-0.0031	P<0.0001
	β12 (pcFMA) Elsewhere	0.0016	P<0.0001
	β13 (pcRent) Inner London	0.0043	P<0.0001
	β13 (pcRent) outer London	0.0090	P<0.0001
	β13 (pcRent) Elsewhere	0.0067	P<0.0001
	Inner London intercept	-3.6775	P<0.0001
	Outer London intercept	-4.4874	P<0.0001
	Elsewhere intercept	-4.3845	P<0.0001

*β1(pcJSA) outer London not significantly different from β1(pcJSA) Elsewhere (p=0.62).
**β2(pcNWB) outer London not significantly different from β2(pcNWB) Inner London (p=0.03).
***β2(pcNWB) Elsewhere is significantly different from β2(pcNWB) Inner and outer London (p=0.002).
****β11(pcSgl) outer London is significantly different from β11(pcSgl) Inner London (p=0.005).

β6 is the coefficient for LA_pcJSA, which was removed from the model when it became insignificant.

Appendix 3

Table A3: Distribution of HTC categories across Local Authorities.
(2010v1 release, subject to change)

	HTC-1	HTC-2	HTC-3	HTC-4	HTC-5
Adur	36%	64%	0%	0%	0%
Allerdale	72%	28%	0%	0%	0%
Amber Valley	78%	22%	0%	0%	0%
Arun	56%	38%	5%	0%	0%
Ashfield	55%	45%	0%	0%	0%
Ashford	66%	34%	0%	0%	0%
Aylesbury Vale	71%	28%	2%	0%	0%
Babergh	87%	13%	0%	0%	0%
Barking and Dagenham	0%	4%	68%	28%	1%
Barnet	0%	30%	48%	22%	0%
Barnsley	50%	48%	1%	0%	0%
Barrow-in-Furness	50%	48%	2%	0%	0%
Basildon	38%	60%	2%	0%	0%
Basingstoke and Deane	57%	42%	1%	0%	0%
Bassetlaw	73%	27%	0%	0%	0%
Bath and North East Somerset	45%	46%	8%	1%	0%
Bedford	33%	46%	16%	5%	0%
Bexley	1%	75%	19%	4%	0%
Birmingham	1%	40%	36%	20%	2%
Blaby	75%	25%	0%	0%	0%
Blackburn with Darwen	21%	55%	20%	4%	0%
Blackpool	19%	68%	13%	0%	0%
Blaenau Gwent	34%	66%	0%	0%	0%
Bolsover	73%	27%	0%	0%	0%
Bolton	26%	58%	13%	3%	0%
Boston	58%	42%	0%	0%	0%
Bournemouth	12%	66%	12%	9%	0%
Bracknell Forest	38%	61%	1%	0%	0%
Bradford	16%	55%	24%	5%	1%
Braintree	56%	44%	0%	0%	0%
Breckland	65%	35%	0%	0%	0%
Brent	0%	0%	21%	72%	7%
Brentwood	67%	33%	0%	0%	0%
Bridgend	64%	36%	0%	0%	0%
Brighton and Hove	2%	52%	29%	16%	1%
Bristol, City of	6%	65%	19%	9%	0%
Broadland	93%	7%	0%	0%	0%
Bromley	18%	61%	18%	4%	0%
Bromsgrove	88%	12%	0%	0%	0%
Broxbourne	41%	57%	2%	0%	0%
Broxtowe	45%	48%	7%	0%	0%
Burnley	28%	63%	8%	0%	0%
Bury	40%	53%	7%	0%	0%
Caerphilly	63%	37%	0%	0%	0%
Calderdale	43%	49%	7%	2%	0%
Cambridge	0%	43%	49%	7%	1%
Camden	0%	0%	0%	45%	55%
Cannock Chase	52%	48%	0%	0%	0%

	HTC-1	HTC-2	HTC-3	HTC-4	HTC-5
Canterbury	43%	48%	9%	0%	0%
Cardiff	18%	58%	13%	11%	0%
Carlisle	43%	54%	3%	0%	0%
Carmarthenshire	70%	30%	0%	0%	0%
Castle Point	67%	33%	0%	0%	0%
Central Bedfordshire	69%	30%	1%	0%	0%
Ceredigion	60%	32%	6%	2%	0%
Charnwood	49%	39%	11%	1%	0%
Chelmsford	54%	43%	3%	0%	0%
Cheltenham	29%	52%	16%	3%	0%
Cherwell	64%	34%	2%	0%	0%
Cheshire East	66%	33%	1%	0%	0%
Cheshire West and Chester	66%	30%	4%	0%	0%
Chesterfield	46%	51%	3%	0%	0%
Chichester	76%	24%	0%	0%	0%
Chiltern	67%	33%	0%	0%	0%
Chorley	64%	36%	0%	0%	0%
Christchurch	57%	43%	0%	0%	0%
City of London	0%	0%	0%	100%	0%
Colchester	41%	54%	5%	0%	0%
Conwy	62%	38%	0%	0%	0%
Copeland	63%	35%	2%	0%	0%
Corby	24%	76%	0%	0%	0%
Cornwall	74%	26%	0%	0%	0%
Cotswold	76%	24%	0%	0%	0%
County Durham	50%	48%	1%	0%	0%
Coventry	5%	66%	23%	5%	1%
Craven	84%	16%	0%	0%	0%
Crawley	6%	89%	5%	0%	0%
Croydon	0%	30%	30%	38%	2%
Dacorum	38%	62%	0%	0%	0%
Darlington	43%	49%	8%	0%	0%
Dartford	41%	57%	2%	0%	0%
Daventry	80%	20%	0%	0%	0%
Denbighshire	67%	29%	3%	0%	0%
Derby	20%	59%	16%	5%	0%
Derbyshire Dales	95%	5%	0%	0%	0%
Doncaster	50%	47%	2%	1%	0%
Dover	58%	42%	0%	0%	0%
Dudley	36%	59%	4%	1%	0%
Ealing	0%	8%	39%	52%	2%
East Cambridgeshire	87%	13%	0%	0%	0%
East Devon	83%	17%	0%	0%	0%
East Dorset	91%	9%	0%	0%	0%
East Hampshire	79%	21%	0%	0%	0%
East Hertfordshire	71%	29%	0%	0%	0%
East Lindsey	80%	19%	1%	0%	0%
East Northamptonshire	73%	27%	0%	0%	0%
East Riding of Yorkshire	69%	30%	1%	0%	0%
East Staffordshire	57%	36%	7%	0%	0%
Eastbourne	17%	68%	12%	3%	0%

	HTC-1	HTC-2	HTC-3	HTC-4	HTC-5
Eastleigh	66%	34%	0%	0%	0%
Eden	86%	14%	0%	0%	0%
Elmbridge	38%	62%	0%	0%	0%
Enfield	0%	15%	46%	38%	1%
Epping Forest	44%	56%	0%	0%	0%
Epsom and Ewell	34%	61%	5%	0%	0%
Erewash	63%	37%	0%	0%	0%
Exeter	15%	66%	15%	4%	0%
Fareham	85%	15%	0%	0%	0%
Fenland	72%	28%	0%	0%	0%
Flintshire	76%	24%	0%	0%	0%
Forest Heath	47%	53%	0%	0%	0%
Forest of Dean	76%	24%	0%	0%	0%
Fylde	67%	33%	0%	0%	0%
Gateshead	24%	65%	10%	1%	0%
Gedling	38%	61%	1%	0%	0%
Gloucester	32%	54%	11%	3%	0%
Gosport	33%	65%	2%	0%	0%
Gravesham	29%	62%	6%	3%	0%
Great Yarmouth	49%	41%	10%	0%	0%
Greenwich	0%	12%	41%	43%	4%
Guildford	48%	48%	5%	0%	0%
Gwynedd	56%	40%	4%	0%	0%
Hackney	0%	0%	0%	45%	55%
Halton	38%	61%	1%	0%	0%
Hambleton	82%	18%	0%	0%	0%
Hammersmith and Fulham	0%	0%	1%	77%	22%
Harborough	89%	11%	0%	0%	0%
Haringey	0%	0%	6%	90%	4%
Harlow	7%	89%	4%	0%	0%
Harrogate	66%	32%	2%	0%	0%
Harrow	0%	21%	58%	21%	0%
Hart	85%	15%	0%	0%	0%
Hartlepool	33%	62%	5%	0%	0%
Hastings	25%	60%	11%	4%	0%
Havant	59%	41%	0%	0%	0%
Havering	7%	79%	12%	1%	0%
Herefordshire, County of	78%	22%	1%	0%	0%
Hertsmere	26%	71%	3%	0%	0%
High Peak	73%	25%	2%	0%	0%
Hillingdon	0%	46%	48%	6%	1%
Hinckley and Bosworth	83%	17%	0%	0%	0%
Horsham	72%	28%	0%	0%	0%
Hounslow	0%	9%	48%	42%	0%
Huntingdonshire	76%	24%	0%	0%	0%
Hyndburn	32%	62%	6%	0%	0%
Ipswich	11%	72%	11%	5%	0%
Isle of Anglesey	77%	23%	0%	0%	0%
Isle of Wight	64%	35%	1%	0%	0%
Isles of Scilly	0%	100%	0%	0%	0%
Islington	0%	0%	0%	66%	34%

	HTC-1	HTC-2	HTC-3	HTC-4	HTC-5
Kensington and Chelsea	0%	0%	0%	4%	96%
Kettering	64%	34%	2%	0%	0%
King's Lynn and West Norfolk	78%	18%	2%	1%	0%
Kingston upon Hull, City of	7%	71%	17%	6%	0%
Kingston upon Thames	0%	56%	31%	13%	0%
Kirklees	35%	47%	16%	2%	0%
Knowsley	23%	73%	4%	0%	0%
Lambeth	0%	0%	0%	82%	18%
Lancaster	42%	47%	11%	0%	0%
Leeds	24%	51%	13%	9%	2%
Leicester	0%	38%	46%	12%	4%
Lewes	66%	34%	0%	0%	0%
Lewisham	0%	0%	20%	80%	0%
Lichfield	75%	25%	0%	0%	0%
Lincoln	21%	65%	12%	2%	0%
Liverpool	3%	62%	23%	11%	1%
Luton	1%	48%	39%	12%	1%
Maidstone	63%	36%	1%	0%	0%
Maldon	88%	12%	0%	0%	0%
Malvern Hills	78%	22%	0%	0%	0%
Manchester	0%	26%	46%	23%	5%
Mansfield	55%	44%	2%	0%	0%
Medway	34%	59%	7%	0%	0%
Melton	87%	13%	0%	0%	0%
Mendip	76%	22%	1%	0%	0%
Merthyr Tydfil	56%	44%	0%	0%	0%
Merton	0%	20%	51%	29%	0%
Mid Devon	79%	21%	0%	0%	0%
Mid Suffolk	80%	20%	0%	0%	0%
Mid Sussex	69%	30%	1%	0%	0%
Middlesbrough	24%	58%	14%	5%	0%
Milton Keynes	21%	71%	8%	1%	0%
Mole Valley	74%	26%	0%	0%	0%
Monmouthshire	79%	21%	0%	0%	0%
Neath Port Talbot	51%	49%	0%	0%	0%
New Forest	80%	20%	0%	0%	0%
Newark and Sherwood	83%	17%	0%	0%	0%
Newcastle upon Tyne	9%	55%	21%	15%	1%
Newcastle-under-Lyme	54%	43%	2%	0%	0%
Newham	0%	0%	0%	94%	6%
Newport	35%	57%	5%	2%	0%
North Devon	74%	24%	2%	0%	0%
North Dorset	80%	20%	0%	0%	0%
North East Derbyshire	67%	33%	0%	0%	0%
North East Lincolnshire	50%	48%	3%	0%	0%
North Hertfordshire	46%	52%	3%	0%	0%
North Kesteven	85%	15%	0%	0%	0%
North Lincolnshire	67%	30%	3%	0%	0%
North Norfolk	84%	16%	0%	0%	0%
North Somerset	68%	30%	2%	0%	0%
North Tyneside	24%	71%	5%	0%	0%

	HTC-1	HTC-2	HTC-3	HTC-4	HTC-5
North Warwickshire	89%	11%	0%	0%	0%
North West Leicestershire	82%	18%	0%	0%	0%
Northampton	27%	55%	16%	2%	0%
Northumberland	59%	40%	1%	0%	0%
Norwich	0%	59%	35%	5%	0%
Nottingham	0%	40%	32%	25%	3%
Nuneaton and Bedworth	45%	51%	4%	0%	0%
Oadby and Wigston	17%	81%	3%	0%	0%
Oldham	28%	51%	17%	4%	0%
Oxford	0%	45%	41%	12%	2%
Pembrokeshire	75%	25%	0%	0%	0%
Pendle	35%	47%	18%	0%	0%
Peterborough	22%	60%	16%	2%	0%
Plymouth	31%	52%	11%	6%	0%
Poole	49%	49%	1%	0%	0%
Portsmouth	0%	61%	28%	11%	0%
Powys	76%	24%	0%	0%	0%
Preston	23%	56%	13%	8%	0%
Purbeck	86%	14%	0%	0%	0%
Reading	10%	53%	30%	8%	0%
Redbridge	0%	12%	53%	33%	2%
Redcar and Cleveland	50%	46%	4%	0%	0%
Redditch	38%	55%	7%	0%	0%
Reigate and Banstead	55%	44%	1%	0%	0%
Rhondda, Cynon, Taff	66%	34%	0%	0%	0%
Ribble Valley	90%	10%	0%	0%	0%
Richmond upon Thames	0%	72%	28%	0%	0%
Richmondshire	72%	28%	0%	0%	0%
Rochdale	30%	56%	12%	2%	1%
Rochford	81%	19%	0%	0%	0%
Rossendale	55%	45%	0%	0%	0%
Rother	81%	19%	0%	0%	0%
Rotherham	46%	49%	4%	1%	0%
Rugby	57%	40%	3%	0%	0%
Runnymede	37%	62%	2%	0%	0%
Rushcliffe	66%	29%	4%	0%	0%
Rushmoor	31%	63%	7%	0%	0%
Rutland	91%	9%	0%	0%	0%
Ryedale	90%	10%	0%	0%	0%
Salford	17%	64%	13%	6%	0%
Sandwell	2%	66%	28%	4%	0%
Scarborough	58%	35%	7%	0%	0%
Sedgemoor	69%	31%	0%	0%	0%
Sefton	45%	48%	6%	1%	0%
Selby	84%	16%	0%	0%	0%
Sevenoaks	73%	27%	0%	0%	0%
Sheffield	20%	54%	19%	6%	1%
Shepway	51%	42%	8%	0%	0%
Shropshire	78%	22%	1%	0%	0%
Slough	0%	42%	55%	3%	0%
Solihull	47%	50%	3%	0%	0%

	HTC-1	HTC-2	HTC-3	HTC-4	HTC-5
South Bucks	78%	23%	0%	0%	0%
South Cambridgeshire	82%	18%	0%	0%	0%
South Derbyshire	85%	15%	0%	0%	0%
South Gloucestershire	62%	37%	1%	0%	0%
South Hams	86%	14%	0%	0%	0%
South Holland	79%	21%	0%	0%	0%
South Kesteven	72%	28%	0%	0%	0%
South Lakeland	81%	19%	0%	0%	0%
South Norfolk	89%	11%	0%	0%	0%
South Northamptonshire	92%	8%	0%	0%	0%
South Oxfordshire	81%	19%	0%	0%	0%
South Ribble	71%	29%	0%	0%	0%
South Somerset	79%	20%	1%	0%	0%
South Staffordshire	87%	13%	0%	0%	0%
South Tyneside	17%	71%	12%	1%	0%
Southampton	0%	60%	23%	17%	1%
Southend-on-Sea	18%	67%	11%	4%	0%
Southwark	0%	0%	2%	88%	9%
Spelthorne	40%	57%	3%	0%	0%
St Albans	50%	47%	3%	0%	0%
St Edmundsbury	56%	44%	0%	0%	0%
St. Helens	45%	54%	1%	0%	0%
Stafford	70%	30%	0%	0%	0%
Staffordshire Moorlands	85%	15%	0%	0%	0%
Stevenage	6%	92%	2%	0%	0%
Stockport	41%	56%	3%	0%	0%
Stockton-on-Tees	46%	46%	6%	2%	0%
Stoke-on-Trent	27%	64%	7%	3%	0%
Stratford-on-Avon	87%	13%	0%	0%	0%
Stroud	72%	28%	0%	0%	0%
Suffolk Coastal	75%	25%	0%	0%	0%
Sunderland	27%	65%	6%	2%	0%
Surrey Heath	65%	35%	0%	0%	0%
Sutton	2%	60%	31%	7%	0%
Swale	61%	38%	1%	0%	0%
Swansea	43%	48%	8%	1%	0%
Swindon	37%	56%	6%	1%	0%
Tameside	23%	72%	4%	1%	0%
Tamworth	54%	44%	2%	0%	0%
Tandridge	78%	22%	0%	0%	0%
Taunton Deane	59%	39%	2%	0%	0%
Teignbridge	81%	19%	0%	0%	0%
Telford and Wrekin	41%	57%	2%	0%	0%
Tendring	64%	34%	1%	0%	0%
Test Valley	73%	27%	0%	0%	0%
Tewkesbury	75%	25%	0%	0%	0%
Thanet	32%	57%	8%	2%	0%
The Vale of Glamorgan	54%	46%	0%	0%	0%
Three Rivers	51%	47%	2%	0%	0%
Thurrock	31%	64%	4%	1%	0%
Tonbridge and Malling	71%	29%	0%	0%	0%

	HTC-1	HTC-2	HTC-3	HTC-4	HTC-5
Torbay	37%	58%	4%	0%	0%
Torfaen	50%	50%	0%	0%	0%
Torridge	76%	24%	0%	0%	0%
Tower Hamlets	0%	0%	0%	42%	58%
Trafford	24%	65%	9%	1%	1%
Tunbridge Wells	57%	38%	4%	0%	0%
Uttlesford	86%	14%	0%	0%	0%
Vale of White Horse	80%	20%	0%	0%	0%
Wakefield	47%	50%	2%	1%	0%
Walsall	16%	68%	12%	4%	0%
Waltham Forest	0%	7%	20%	66%	8%
Wandsworth	0%	0%	17%	82%	1%
Warrington	61%	38%	2%	0%	0%
Warwick	35%	58%	6%	1%	0%
Watford	2%	81%	17%	0%	0%
Waveney	58%	38%	4%	0%	0%
Waverley	73%	27%	0%	0%	0%
Wealden	86%	14%	0%	0%	0%
Wellingborough	49%	45%	6%	0%	0%
Welwyn Hatfield	25%	66%	9%	0%	0%
West Berkshire	76%	23%	1%	0%	0%
West Devon	94%	6%	0%	0%	0%
West Dorset	82%	18%	0%	0%	0%
West Lancashire	67%	33%	0%	0%	0%
West Lindsey	83%	17%	0%	0%	0%
West Oxfordshire	86%	14%	0%	0%	0%
West Somerset	83%	17%	0%	0%	0%
Westminster	0%	0%	0%	0%	100%
Weymouth and Portland	54%	44%	3%	0%	0%
Wigan	52%	47%	2%	0%	0%
Wiltshire	73%	27%	0%	0%	0%
Winchester	68%	29%	3%	0%	0%
Windsor and Maidenhead	49%	49%	2%	0%	0%
Wirral	38%	57%	5%	0%	0%
Woking	31%	56%	13%	0%	0%
Wokingham	59%	41%	0%	0%	0%
Wolverhampton	1%	66%	25%	7%	1%
Worcester	33%	56%	11%	0%	0%
Worthing	23%	68%	8%	2%	0%
Wrexham	54%	44%	2%	0%	0%
Wychavon	86%	14%	0%	0%	0%
Wycombe	51%	39%	9%	0%	0%
Wyre	72%	26%	1%	0%	0%
Wyre Forest	62%	38%	0%	0%	0%
York	42%	49%	8%	0%	0%