# An Assessment of the Quality of the Matching Between the 2011 Census and the Census Coverage Survey

August 2013

This is one of a series of reports published to support the release of results from the 2011 Census. This series of methods and quality reports provides information on the different methods used to collect, process, clean, adjust and protect the census results. The series also reports on the quality assurance of the results and provides quality indicators.

Terms used in the series are explained in the 2011 Census glossary.

## Contents

# An Assessment of the Quality of the Matching Between the 2011 Census and the Census Coverage Survey

## 1. Summary

To measure coverage in the 2011 Census, the Office for National Statistics (ONS) conducted a Census Coverage Survey (CCS) across England and Wales six weeks after census day. In order to assess coverage in the census, the CCS was matched to the census records where the CCS was conducted.

This report gives estimates of the matching quality of the 2011 Census to the Census Coverage Survey matching process. The levels of error in the matching process had a direct impact on the quality of the population estimates from the coverage assessment and estimation process. It was therefore important to evaluate the quality of the matching, in order to give assurance that the population estimates were not affected significantly. This paper describes the methodology used to carry out this assessment of quality, through re-matching samples of data, and gives the results of this sampled re-matching by the different samples drawn, the hard-to-count (HtC) index and by shift worked by the matching staff.

The total number of postcodes in the CCS, and so available for re-matching, was 17,415. The sample contained a total of 1,084 postcodes (6.2%). The re-matching focused on hardest to count areas, which tended to have the lowest automatic match rates and so the largest workloads for the clerical matchers. The workload was defined as the number of potential matched pairs (candidate pairs) needing clerical resolution and the number of unmatched records, for which the matching staff searched the database to find matches. The sample contained a total of 83,356 original matching decisions; approximately 17% of the original matching decisions.

The results of this evaluation show that matching quality was very high. The overall precision of the matching in the sampled areas was 99.90% and overall sensitivity was 99.75%. It should be noted, however, that the sample was disproportionally skewed towards the ‘difficult to match’ areas, which had higher error rates. Therefore it is likely that the overall precision and sensitivity would be higher for matching the CCS to the census across the whole country.

Looking at the HtC index, the highest sensitivity was obtained in HtC 1 areas, with sensitivity of 99.92%. HtC 3 areas had the lowest sensitivity of 99.59%. Precision levels were more similar across the HtC index with the highest levels of 99.94% in HtC 2 and 3 and the lowest levels of 99.87% were found in HtC 4 and 5. These exceptionally high

quality standards were achieved through a mixture of automatic matching and a multi-stage hierarchical clerical matching process.
Despite there being a large team working over two shifts there was a consistency in the quality of the work carried out as the error rates for both the day and evening shifts were almost identical.

These accuracy rates are close to the matching targets and the small number of errors found will have had only a small impact on the quality of the population estimates. Users can therefore have confidence in the 2011 Census population estimates.

# 2. Background and introduction

The 2011 Census was designed to count everyone and produce population estimates by authority (LA). It was known that for a variety of reasons, some households and individuals would be missed from the census (undercount), while others would be counted more than once or in the wrong location (overcount).

To measure coverage in the 2011 Census, the Office for National Statistics conducted a Census Coverage Survey (CCS) across England and Wales. This was a voluntary survey carried out independently of the census, designed to measure coverage in the census. The CCS was used to estimate the population counted and missed by the census and adjust the census database for those estimated to have been missed (ONS, 2012a).[1]

Matching was a vital part of the coverage assessment and adjustment process, supplying the matched and unmatched data for the process of adjustment using Dual System Estimation (DSE). More information on DSE can be found in ONS (2010a).[2]

In each postcode where the CCS had been carried out, the following information was required:

i. the number of individuals in the census and in the CCS,
ii. the number in the census but not in the CCS, and
iii. the number in the CCS but not in the census.

| | | Counted by CCS | |
|---|---|---|---|
| | | Yes | No |
| Counted by census | Yes | (i) | (ii) |
| | No | (iii) | X |

To obtain this information it was necessary to match each census record within a CCS postcode, at both the individual and household level, with its CCS equivalent where one existed. Census and CCS households and people were either matched or declared to be unmatchable. The CCS sample was matched to the census in every LA in England and Wales between October 2011 and March 2012. Initially the matching was completed by one shift (day) but due to the time demands of the task, the matching team worked across a 2 shift operation (day and evening).

# 3. Process and methodology

Following the completion of the matching work at the end of March 2012 a sample of postcodes were selected for re-matching in order to assess the quality of the matching process and the decisions made, providing a measure of robustness for this element of the coverage assessment and adjustment methodology. The re-matching was performed by the more experienced matchers, most of whom were also more senior matchers. To ensure independence, no-one re-matched an area that they had worked on originally.

A comparison of the matching runs was made to identify where different decisions were made between the original matching process and the re-matching and to verify which of these decisions were correct. This information helped to establish the effect that matching accuracy had on the population estimate.

As in any matching project, it was impossible to eliminate matching errors entirely, as there were errors and inconsistencies in the census and CCS data. However, the use of clerical matching staff in the multi-stage hierarchical matching process was designed to minimise errors. Details of the matching strategy are given in ONS (2012b).[3] The original matching success criteria were:

- Low false negative match rates (i.e. not finding a match when one exists).
    - o False negative match rate < 0.1% of Census Coverage Survey (CCS) and census data.
- Minimal false positives (i.e. incorrectly matching two records).

The target false negative rate was exceptionally stringent and was chosen to prevent incorrect inflation of population estimates in the coverage assessment and adjustment process.

When evaluating the accuracy of record linkage projects, the usual quality measures used are precision and recall; recall is also known as sensitivity. Therefore, in order to align with best practice, sensitivity and precision have been computed for this evaluation of matching quality. Precision relates to the first success criterion above and sensitivity relates to the second success criterion above; they are defined as follows:

**Precision**
Precision measures the quality of the matches made (i.e. the proportion of matches made that were true matches) and is defined as:

$$Precision = \frac{\#\, true\ positives}{\#\, true\ positives + \#\, false\ postives} = 1 - \frac{\#\, false\ positives}{\#\, links\ made} = 1 - false\ positive\ rate$$

Where: # = ‘number of’

true positives = matches made that were actually true matches

**Sensitivity**
Sensitivity (recall) measures the proportion of all true matches that were actually found (made) and is defined as:

$$Sensitivity = \frac{\#\, true\, positives}{\#\, true\, positives + \#\, false\, negatives}$$

The denominator in the calculation of sensitivity was calculated as:

(# matches made - # false positives + # false negatives)

## 3.1 Sampling

It was not practical to re-match all of the postcodes and so a sampling exercise was carried out. The sample size was 1,000 postcodes. This was derived from the time available for the matching team to carry out the task and the throughputs achieved in the original matching process, whilst also ensuring sufficient sample to draw robust conclusions. A stratified random sample was taken using the following stratification factors:

a) The hard-to-count (HtC) index for that postcode (an integer score 1 – 5, where 5 is the hardest to enumerate).
b) The shift that performed the original matching work in the postcode (day or evening).
c) Whether the results from the Estimation Area (EA) looked unexpected within the estimation process, based on feedback from the QA team and panels (yes or no).

The primary stratification factor used was (c); the areas in the 'yes' category had the potential to have specific problems or barriers to estimation which may have been influenced by the matching. Previous experience has shown that some of the London areas are more difficult to enumerate and the coverage processes more difficult to complete. The two London regions (inner and outer) have the two highest mean HtC indexes, 4.4 and 3.2 respectively. (See ONS (2010b)[4] for more information on HtC). These areas were further evaluated in order to provide additional assurance that the original matching was of a high quality, and hence enable the later coverage processes to provide good quality census estimates in these areas. There were nine local authorities in these EAs and it was decided that half of the sample (500) would be postcodes within these areas and half would be drawn from the remaining LAs across the country.

**Sample 1**
For the EAs requiring further investigation from a total number of 1,120 postcodes, 500 (45%) were randomly selected proportionally across the EAs and HtC groups. For these EAs there were no HtC postcodes with an index of 1 so only 2, 3, 4 and 5 were used. The breakdown is given in Annex A, table A1.

**Sample 2**

The remaining sample of 500 postcodes was selected from EAs not requiring further investigation. Stratification factors were shift (day or evening) and HtC index (1 – 5). Approximately 3% of postcodes were randomly selected within each stratum; there was no requirement for sample to be drawn from every local authority. A subsequent request was made for additional quality assurance of results in Brent, Newham and Manchester and so a further sample of postcodes was made in these areas (sample 3). Although some postcodes in Brent, Newham and Manchester were originally included in sample 2, they were then all moved into the 3rd sample for consistency; thus the final number of postcodes in the second sample was less than 500 (16 moved to the 3rd sample and 19 were completely matched via the automatic matching algorithm and as such were never matched manually). Table A2 in Annex A gives the distribution of postcodes in the second sample, by HtC stratum.

**Sample 3**
The initial sample from LAs requiring further investigation was subsequently extended to include three additional areas that required further quality assurance; these areas were Brent, Newham and Manchester. Approximately 40 postcodes were randomly selected in each of the areas, stratified by HtC index. This sample size was based on time constraints and resource available, ensuring that sufficient sample was taken to give robust conclusions. In instances where these three areas had postcodes selected in the second sample (see above), they were then moved into the third sample; this included 5 in both Brent and Manchester and 6 in Newham. Table A3 in Annex A shows the number of postcodes in the third sample by LA and HtC stratum.

In total, across the three samples, 1,084 postcodes were selected for re-matching. The distribution of all the postcodes selected for re-matching by HtC stratum is given in table 1. The total number of postcodes in the CCS was 17,415.

**Table 1: Overall distribution of all postcodes sampled for re-matching by HtC stratum**

| | N in sample | Available | % sampled |
|---|---|---|---|
| HtC 1 | 173 | 5,906 | 2.9% |
| HtC 2 | 232 | 6,287 | 3.7% |
| HtC 3 | 142 | 2,222 | 6.4% |
| HtC 4 | 268 | 2,181 | 12.3% |
| HtC 5 | 269 | 819 | 32.8% |
| Total | 1,084 | 17, 415 | 6.2% |

## 3.2 Evaluation of match decisions

The sampled information was collated and showed lists of CCS and census questionnaire ID numbers and the matches made and not made by both the original and re-matching processes. It also showed whether the **'same'** decision had been made in both processes, or whether the decisions were **'different'**, e.g. census and CCS matched originally but not matched in re-matching.

To assess validity of the decisions that were the **same**, the data from Manchester, one of the priority areas, was checked to confirm if both decisions were correct. For sets of decisions that were **different**, the sample data was checked to identify if the correct decision was made originally or at re-matching.

A good original decision was where the household or person had been correctly matched between census and CCS forms, or there was no match to be made as there was no corresponding household or person. An incorrect original decision was where a match should have been made but wasn't (**missed match**), or a match had been made that shouldn't have been made (**incorrect match**). A missed match contributed to the false negative match rate whereas an incorrect match contributed to the false positive match rate.

It must be emphasised that changes to the matching process between the original matching and re-matching will have impacted on decisions made, and conflicting decisions may have been correct at the time each process was completed. Because of this, incorrect original decisions were categorised as shown in table 2 below; although not all were necessarily defined as errors, as the following reasons explain.

**Table 2: Categorisation of incorrect original decisions**

| Household | Person |
|---|---|
| 1 At least 1 dummy questionnaire | 1 In/Out mover |
| 2 Subjective decision | 2 Subjective decision |
| 3 Duplicate matched originally | **3 Incorrect match** |
| **4 Incorrect match** | **4 Missed match** |
| **5 Missed match** | |

Note: only the reasons in bold type (not greyed out) were considered as errors for the purposes of this investigation.

**Dummy questionnaires** - an instruction was given at the re-matching process that dummy questionnaires were not to be matched. However this match may have been made originally. This would show up on the database as two different decisions but both were the correct decision at the time. A dummy questionnaire was completed during the field operation for each valid household that did not return a census questionnaire. The reasons for dummy questionnaire completion include: non-contact with resident; refusal of resident to complete a questionnaire; an absent household; a second home; a holiday home; or a vacant property. This information was used to distinguish between occupied and unoccupied households for use during coverage assessment and adjustment.

**In/Out movers** – persons who appeared at a different address in the census and CCS. It was not considered essential, during the original matching process, to search extensively for in/out movers, if a household match had been made and local searches carried out.

These were routinely designated unmatchable originally but this practice was not followed in re-matching so many of these matches were subsequently made.
**Subjective decisions** – these were defined as any decision on which two matchers/experts were equally likely to disagree as to agree. Consequently there would be as much scope for variation within each matching process, as between the two matching processes (and between subsequent reviews). It was therefore not appropriate to consider these as errors.

**Duplicate matches** – questionnaires for which there was more than one possible correct match, and which have been matched at least once. It is likely in these cases that the original matcher/expert was aware of the duplicate, and elected to declare the other as unmatchable. This is not the same as a missed match. Duplicates were difficult to identify in the quality evaluation as the range of information was not available for this re-matching process.

Because of the explanations provided, it was only the reasons highlighted in bold type above that were considered as errors and it was these numbers that were used to calculate the quality metrics in this report.

In carrying out this comparison exercise, the range of original information was limited. For example contextual information (such as deferral notes, or an overview of the other data in a particular postcode) was not available which may have given mitigating reasons why matches were not made.

The original matching process included full access to the matching system and included cluster resolution as part of finalisation of LAs. With the comparison of the original and re-matching decisions, the matching system and contextual information was not available, and cluster resolution was not part of this evaluation process. Cluster resolution was performed at LA level at the end of the matching process, once the matching for a whole LA was completed. It involved quality assurance of multiple matches (clusters), where a record was matched to more than one different record, and had the ability to break or approve these matches. The remaining multiple matches were resolved by the estimation team. With the re-matching, cluster resolution was not part of the process, as the re-matching was not performed for whole LAs.

In examining all of these decisions, the starting point was the result. For example a match that had been made/not made originally and a different decision that had been made in the re-matching process. The only information that was available to be looked at was the census/CCS identification number which was used to look at the scanned image of the associated questionnaire to decide whether a match was or was not appropriate. This means we were underestimating our error rates where a) matches were missed in both passes and b) matches were incorrectly made in both passes. However results of the analysis of decisions that were the ‘same’ identified that any underestimation would be minimal. Additionally the census matching process was designed to minimise the possibility of missed matches by ensuring at least two people examined each non-match decision.

## 3.3 Comparison of matching decisions

The data identified decisions that had the **same** outcome both originally and at re-matching and also decisions that had **different** outcomes. The sample contained a total of 83,356 original matching decisions and 83,425 re-matching decisions; this represents approximately 17% of the original matching decisions. The number of decisions at re-matching was larger as there was no cluster resolution as part of this process and so the re-matched sample contained many ‘multiple match’ clusters.

In order to assess validity of decisions that had the **same** outcome from the original and re-matching processes the Manchester data was used. The Manchester data consisted of 3,882 sets of decisions that were looked at and results showed that two were not good decisions (0.05%), consisting of two separate occasions where two people were matched incorrectly both originally and at re-matching (due to localised duplication). It was therefore decided that the ‘**same’** decisions were of sufficient quality; that is, we can be confident that where decisions were the same, they were accurate.

For sets of decisions that had **different** outcomes, these were checked to identify whether the correct decision was made during the original matching or at re-matching.

When carrying out this exercise it was discovered that a number of discrepancies had appeared due to systematic differences between how the original and re-matching processes were carried out. Others were the result of individuals drawing different conclusions from subjective information. A total of 73 decisions were identified as subjective from the **different** outcomes, representing 0.09% of the total (83,356 decisions).These subjective differences of opinion were not included in the error counts. The remaining errors (as categorised in table 2), were used to calculate the quality metrics used in this report.

From a national total of 83,356 original decisions examined, 229 errors in the original matching were identified (0.27% of total decisions). These errors have been split into 63 incorrect matches, and 166 missed matches.

It is important to note that the ‘errors’, as defined, were a product of this comparison process, and that limitations of, or changes to, the process inevitably affected the number of errors. The apparent errors in matching were classified as either:

- Missed matches – matches which in retrospect it was decided should have been made, or
- Incorrect matches – matches which were made originally but it was decided should not have been made.

The missed matches would have been left as ‘unmatchable’ at the end of the original matching process, and so would have been confirmed by at least two people (matcher and expert/supervisor). The incorrect matches would have in most cases only been viewed by one person, and may not have been picked up by the expert or supervisor as only a percentage of these matches were presented for checking. These could have occurred at either the clerical or manual matching stage.

# 4. Results

Note that the following apply to all the tables in this report:

- 95% confidence intervals were calculated using a binomial approximation, and they have been calculated for the sensitivity, since the key aim was to minimise false negative matches (which would result in an over-estimate of the population).
- Percentages have (generally) been rounded to 2 decimal places and therefore may not add to the totals given.

## 4.1 Matching quality by sample

Tables 3 and 4 below summarise the results from the areas requiring further investigation, the three priority EAs (Brent, Newham and Manchester), and the rest of the country, including the sample size and the mean HtC index for the same EAs. The full results and analysis for each EA are available to download.[5]

**Table 3**

| Areas | No. of CCS Postcodes | No. of Postcodes in QA Sample | QA Sample Size | Mean HtC Index | Total Original Decisions | Total Errors |
|---|---|---|---|---|---|---|
| Areas for further investigation | 1,120 | 500 | 44.6% | 4.2 | 38,292 | 98 |
| Brent, Newham & Manchester | 466 | 119 | 25.5% | 3.6 | 11,236 | 53 |
| Rest of Country | 15,798 | 465 | 2.9% | 2.0 | 33,828 | 78 |

**Table 4**

| Areas | No. of Incorrect Matches | No. of Missed Matches | No. of original matches | Precision | Sensitivity | 95% CI for sensitivity |
|---|---|---|---|---|---|---|
| Areas for further investigation | 29 | 69 | 28,014 | 99.90% | 99.75% | (99.70%, 99.81%) |
| Brent, Newham & Manchester | 14 | 39 | 8,483 | 99.83% | 99.54% | (99.40%, 99.69%) |
| Rest of Country | 20 | 58 | 28,967 | 99.93% | 99.80% | (99.75%, 99.85%) |

Precision was very high across all three samples. Matching in Brent, Newham and Manchester had the lowest precision at 99.83% and the 'rest of country' had the highest precision of 99.93%.

Sensitivity was slightly higher for the 'rest of the country' than for the areas in samples one and three. The matching in Brent, Newham and Manchester (sample 3) had significantly lower sensitivity than the 'rest of country' and the 'areas for further investigation'. This was not unexpected, since the areas in sample 3 were chosen because additional quality assurance of their population estimates was required.

## 4.2 Matching quality by HtC

It is useful to look at the breakdown by the difficulty in matching the different areas. A proxy for this is the HtC index which rates postcodes with a 1 through to 5 where 1 represents the easiest to enumerate areas and 5 the hardest. The areas with a HtC index of 5 tended to have the lowest census response rates and so the match rates were also expected to be lowest in these areas. From table 3, it can be seen that the areas for further investigation carried a higher mean HtC index than the rest of the country.

**Table 5**

| HtC Index | No. of CCS Postcodes | No. of Postcodes in QA Sample | QA Sample Size | Total Original Decisions | Total Errors |
|---|---|---|---|---|---|
| HtC 1 | 5,900 | 173 | 2.9% | 10,043 | 13 |
| HtC 2 | 6,278 | 232 | 3.7% | 17,079 | 29 |
| HtC 3 | 2,220 | 142 | 6.4% | 11,586 | 45 |
| HtC 4 | 2,176 | 268 | 12.3% | 25,948 | 82 |
| HtC 5 | 810 | 269 | 33.2% | 18,700 | 60 |

As shown in table 6, precision was high, quite consistently across the HtC index. The highest precision of 99.94% was obtained in areas with HtC index of 2 and 3, and the lowest precision levels of 99.87% were found in HtC 4 and 5. Areas with HtC 3 had the lowest sensitivity (99.59%), and the highest sensitivity (99.92%) was in HtC 1 areas. Areas with a HtC index of 4 and 5 tended to be in London and large urban areas. In these areas matchers were aware that matching was very 'difficult' and time-consuming and high levels of concentration were required. It is likely that matching was easiest in areas with a HtC index of 1 and 2, leading to the lower error rates.

**Table 6**

| HtC Index | No. of Incorrect Matches | No. of Missed Matches | No. of original matches | Precision | Sensitivity | 95% CI for sensitivity |
|---|---|---|---|---|---|---|
| HtC 1 | 6 | 7 | 9,043 | 99.93% | 99.92% | (99.87%, 99.98%) |
| HtC 2 | 9 | 20 | 14,865 | 99.94% | 99.87% | (99.81%, 99.92%) |
| HtC 3 | 6 | 39 | 9,379 | 99.94% | 99.59% | (99.46%, 99.72%) |
| HtC 4 | 25 | 57 | 19,142 | 99.87% | 99.70% | (99.63%, 99.78%) |
| HtC 5 | 17 | 43 | 13,035 | 99.87% | 99.67% | (99.57%, 99.77%) |

## 4.3 Matching quality by shift

As an evening shift was introduced along with the extra staff required to man it, consistency of decision making across the shifts was considered. Tables 7 and 8 show the matching quality results for the two shifts separately.

**Table 7**

| Shift | No. of CCS Postcodes | No. of Postcodes in QA Sample | QA Sample Size | Mean HtC Index | Total Original Decisions | Total Errors |
|---|---|---|---|---|---|---|
| Day | 9,213 | 751 | 8.15% | 2.3 | 58,270 | 162 |
| Evening | 8,171 | 333 | 4.08% | 2.1 | 25,086 | 67 |

**Table 8**

| Shift | No. of Incorrect Matches | No. of Missed Matches | No. of original matches | Precision | Sensitivity | 95% CI for sensitivity |
|---|---|---|---|---|---|---|
| Day | 46 | 116 | 45,017 | 99.90% | 99.74% | (99.70%, 99.79%) |
| Evening | 17 | 50 | 20,447 | 99.92% | 99.76% | (99.69%, 99.82%) |

The precision and sensitivity across the two shifts were very similar, which would indicate that there was consistency in training and working practices across both shifts. Nearly 70% of the sample for re-matching was from the day shift because they matched most of the London EAs whilst the evening shift was trained. The similarity in the mean HtC index shows that both shifts on average worked on similar postcodes in terms of complexity, although in general the evening shift tended to focus on areas with HtC of 2 and 3, whereas the day shift tended to work on the more extreme areas with HtC index of 1, 4 and 5.

# 5. Conclusion and discussion

The quality of the matching between the 2011 Census and the Census Coverage Survey was very high. The overall precision of the matching in the sampled areas was 99.9% and overall sensitivity was 99.75%. These exceptionally high quality standards were achieved through a mixture of automatic matching and a multi-stage hierarchical clerical matching process.

There was consistency in the matching quality across the two shifts, indicating uniformity in training and working practices.

The sample represents only 6.2% of the CCS postcodes nationally and it should be noted that the sample was substantially skewed towards the more difficult areas that required further investigation or that have a HtC index of 4 and 5. The mean HtC index for the sample was 3.3 but the mean for all of the CCS postcodes was 2.2. Therefore it is likely that the overall precision and sensitivity for matching the CCS to the census across the whole country would be higher.

These high accuracy rates are close to the matching targets and the small number of errors found will have had only a small impact on the quality of the population estimates. Users can therefore have confidence in the 2011 Census population estimates.

The matching errors discussed in this report were identified through examination of the data at the end of the matching phase of the coverage assessment and adjustment process. At the next stage, the estimation team were able to look at results for a whole estimation area, as opposed to just a local authority or postcode. In addition to the individual matches, non-matches and the notes that the matching supervisors made they could also get an overview of the match rates for an area (something that the matching team were unable to see). This meant they were able to focus in on areas where there appeared to be something unexpected to check the results. As a result they were able to identify and correct 876 matching errors across all of the CCS postcodes. This will have improved the quality of the census results as the error rates were reduced further, therefore providing additional confidence in the final census population estimates.

# 6. References


[1] Office for National Statistics (2012a) 2011 Census Coverage Assessment and Adjustment Process. Available to download.

[2] Office for National Statistics (2010a) Trout, Catfish and Roach. Office for National Statistics. Available to download.

[3] Office for National Statistics (2012b) Automatic match rates for the 2011 Census to the Census Coverage Survey. Available to download.

[4] Office for National Statistics (2010b) Predicting patterns of household non-response. Available to download.

[5] 2011 Census: Quality of the Census Coverage Survey matching by estimation area and region. Available to download.

# 7. Annex A

**Table A1: Distribution of postcodes in the first sample (EAs requiring further investigation) by HtC stratum**

| | Camden | Croydon | Hackney | K & C | Lambeth | Richmond, Hounslow & Kingston | Westminster | Total |
|---|---|---|---|---|---|---|---|---|
| **Sampled Postcodes** (Available Postcodes) | **50** (112) | **67** (151) | **59** (133) | **51** (118) | **64** (145) | **86** (191) | **123** (276) | **500** (1,126) |
| HtC2 | **0** (0) | **10** (23) | **0** (0) | **0** (0) | **0** (0) | **34** (75) | **0** (0) | **44** (98) |
| HtC3 | **0** (0) | **23** (52) | **0** (0) | **0** (0) | **0** (0) | **28** (63) | **0** (0) | **51** (115) |
| HtC4 | **20** (45) | **32** (72) | **21** (47) | **4** (9) | **49** (110) | **24** (53) | **0** (0) | **150** (336) |
| HtC5 | **30** (67) | **2** (4) | **38** (86) | **47** (109) | **15** (35) | **0** (0) | **123** (276) | **255** (577) |

***Note:*** *Bold numbers indicate number of postcodes in the sample. Numbers in brackets indicate the number of postcodes available to draw the sample from.*

**Table A2: Distribution of postcodes in the second sample, by HtC stratum**

| HtC Index | Day Shift | | Evening Shift | |
|---|---|---|---|---|
| HtC 1 | **94** | (3,240) | **79** | (2,666) |
| HtC 2 | **87** | (2,892) | **96** | (3,265) |
| HtC 3 | **22** | (784) | **37** | (1,176) |
| HtC 4 | **22** | (900) | **21** | (687) |
| HtC 5 | **6** | (169) | **1** | (38) |
| Total | **231** | (7,985) | **234** | (7,832) |

***Note:*** *Bold numbers indicate number of postcodes in the sample. Numbers in brackets indicate the number of postcodes available (i.e. not in LAs covered by samples 1 and 3) to draw the sample from.*

Office for National Statistics

**Table A3: Distribution of postcodes in the third sample by LA and HtC stratum.**

| HtC Index | Brent | | Newham | | Manchester | |
|---|---|---|---|---|---|---|
| HtC 1 | **0** | (0) | **0** | (0) | **0** | (0) |
| HtC 2 | **0** | (0) | **0** | (0) | **5** | (32) |
| HtC 3 | **8** | (27) | **0** | (0) | **24** | (120) |
| HtC 4 | **28** | (95) | **37** | (112) | **10** | (51) |
| HtC 5 | **3** | (15) | **3** | (8) | **1** | (12) |
| Total | **39** | (137) | **40** | (120) | **40** | (215) |

***Note:*** *Bold numbers indicate number of postcodes in the sample. Numbers in brackets indicate the number of postcodes available to draw the sample from.*