

2004 Local Authority studies

Analysis of data and evidence for Southwark

Contents

	Page
Executive summary	3
1 Introduction	4
2 Background information on Population Estimates and One Number Census	4
3 Analysis	5
3.1 Enumeration	5
a) Analysis of Council Tax data	6
b) Analysis of Enumerator Record Books (ERBs)	8
c) Analysis of Council Tax based response rates by 2001 Ward	8
d) Enumeration information	8
e) Estimated response compared to other areas	9
f) Qualitative information obtained about the Census enumeration	10
3.2 Census Coverage Survey	10
a) Missing sample	10
b) Sample balance	10
Dummy form distribution	10
Recalculated Hard to Count score	11
c) Localised undercount and the CCS	15
d) CCS fieldwork and response rates	19
3.3 One Number Census process	19
a) Response rates	19
b) Collapsing in the ONC process	20
c) Outliers	20
d) Stratification	20
e) Household and person imputation results	22
4 Population definitions	25
5 Processing	25
6 Other - communal establishments	26
7 Other - administrative sources	26
8 Conclusions and recommendations for adjustment	28
9 Adjustment and methods used	28
10 Summary of adjustments	28
11 Quality assurance	29

Executive summary

The Office for National Statistics (ONS) has now concluded a series of studies designed to improve population estimates in the areas that proved to be hardest to count in the 2001 Census in England and Wales. The results of this analysis, involving experts from local government and other bodies, has confirmed the findings contained in reports by the Statistics Commission and the Local Government Association. These reports concluded that the One Number Census (ONC) worked well in most areas but that there were a few cases where it was not able to sufficiently adjust for under-enumeration in exceptional circumstances.

More information can be found at www.statistics.gov.uk/pfdir/census0704.pdf

A report was produced for each Local Authority examined, and this report sets out the analysis and conclusions for the Local Authority of Southwark. It provides background information about population estimates and the Census, and describes analysis carried out in certain key areas - enumeration, Census Coverage Survey (CCS), One Number Census (ONC), population definitions and processing. A comparison with administrative sources is also shown. Conclusions and recommendations for adjustment are detailed at the end of the analysis, along with an explanation of the methods used for adjustment.

Southwark was selected for detailed analysis as part of the local authority population studies, as it was identified as an area where there was a significant risk of an under-estimate of the population by the 2001 Census.

Localised under-enumeration in three wards (Friary, Consort and Basset) was also identified, which the One Number Census was unable to adjust for. To resolve this problem these wards have been removed from the data and considered as a separate stratum. The population of this stratum was then re-estimated based on the Council Tax count of addresses and the household size from the existing Census Ward level results.

Furthermore, there were a large number of unprocessed forms in Southwark, and an assessment has been made on the number of forms that would not have been accounted for by the CCS, for which a revision is required.

After consideration of all the evidence, the population of Southwark in 2001 was re-estimated, and it was concluded that the Census had under-estimated the population by 6,552.

2004 Local Authority studies: Analysis of data and evidence for Southwark

1 Introduction

Southwark (Census 2001 population 244,866) is situated on the south side of the Thames in London, immediately opposite the City of London and Westminster.

In order to estimate census undercount across England and Wales contiguous Local Authorities (LAs) were grouped together to form Estimation Areas (EAs) which consisted of about half a million population. In most cases EAs consisted of several Local Authorities. Where an LA was sufficiently large the EA consisted only of that one LA, i.e. the LA was an EA in itself. LAs were divided into Enumeration Districts (EDs) which were pre-planned workload areas of around 200 households within which an individual enumerator worked. EDs did not cross Ward boundaries.

Southwark is part of a Census Estimation Area containing one other Local Authority (LA), Lambeth.

Southwark was selected for investigation for a number of reasons:

- Analysis of Council Tax data shows that the Census estimated 6,500 fewer households in Southwark, including 4,200 fewer occupied households than are recorded on Council Tax.
- The Census Coverage Survey (CCS) may not have covered the most difficult areas to enumerate, according to analysis of the Hard to Count (HtC) distribution, as there has been a large change between the 1991 and 2001 data used to calculate the Hard to Count scores.
- There was a large number of unprocessed forms in Southwark.

2 Background information on Population Estimates and One Number Census

Table 2.1 sets out the Census and mid year estimates (MYEs) for Southwark for the period 2000 – 2002. Note that the 2001 and 2002 mid year estimates shown here are those based on 2001 Census data, while the 2000 mid year

estimate is based on rolled forward data from the 1991 Census.

Table 2.1
Mid Year Estimate (MYE) series and revisions for 2000, 2001 and 2002

	Population	Change since 2000 MYE
2000 MYE	238,700	
Census	244,866	6,200
2001 MYE	245,400	6,700
Revised 2001 MYE	251,100	12,400
Revised 2002 MYE	251,300	12,600

NB: The numbers in the above table do not add up due to rounding.

Adjustments were made to the Mid Year Estimates in September 2003. More information can be found at www.statistics.gov.uk/about/Methodology_by_theme/Revisions_to_Population_Estimates/default.asp

Table 2.2 shows the Confidence Interval associated with the ONC population estimate

Table 2.2
95 per cent Confidence Interval for Southwark's One Number Census estimate

Relative C.I.	+/-	Confidence Interval Range	
3.1 %	7,591	237,275	252,457

Mid Year Population Estimates

The following tables provide information on Southwark's population between 1991 and 2000. These indicate whether the population has increased or decreased throughout the intercensal period, whether the area is one of large change and what effect migration has had on the area's population. The population of Southwark was steadily increasing between 1991 and 2000.

Table 2.3
Population profile for the MYE series up to 2000 (unrevised)

Year	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Mid Year Population Estimate (thousands)	227.2	227.4	229.4	228.8	232.0	229.9	230.5	232.0	235.8	238.7

NB: The mid-year population estimates in this table are those that existed before the 2001 Census, and have since been revised.

Table 2.4
Average annual changes in components of population estimates (unrevised)

Average annual change since mid-91 (thousands)	E & W	Southwark
Average annual change	0.5	1.3
Average annual natural change	0.3	2.0
Average annual change in migration	0.1	-2.4

NB: The numbers in the above table do not add up due to rounding.

Table 2.5
Migration profile for MYE series to 2000 (unrevised)

Migration	Mid-92	Mid-93	Mid-94	Mid-95	Mid-96	Mid-97	Mid-98	Mid-99	Mid-00
Net Internal							-2.1	-2.2	-2.3
Net International							0.2	1.1	0.1
Total Net	-3.0	-1.1	-4.3	-0.1	-5.6	-2.6	-1.9	-1.1	-2.2

Table 2.5 shows that migration was variable for Southwark over the intercensal decade, with net out migration in all years between 1992 and 2000. However, the level of natural change (excess births over deaths) shown in table 2.4 means Southwark was a growing area between 1991 and 2000.

One Number Census Quality Assurance information

The One Number Census Quality Assurance Information Pack for Southwark, published in 2003, can be found in the link below, and includes the following information:

- Population Estimates
- Confidence Intervals
- Diagnostic Ranges
- Dependency Ratios
- Age-Sex Profiles depicting the above figures
- Census Coverage Survey Maps including Hard to Count information

The map in the link below shows the location of the CCS Postcodes within the Local Authority. White circles show postcodes with a Hard to Count (HtC) index of 1 (the easiest areas to

enumerate), light green circles have a HtC index of 2 and dark green circles have a HtC index of 3 (the hardest areas to enumerate). Local Authority boundaries are marked in blue and the Estimation Area boundary in red.

www.statistics.gov.uk/census2001/onc_qa/pdfs/Southwark.pdf

Information provided by the Local Authority Southwark is not an area that has queried their Census results. ONS sent a letter to Southwark on 26th September 2003 regarding unprocessed forms.

3 Analysis

This section of the report covers the detailed analysis carried out for Southwark. More information on the overall approach to the LA studies can be found at www.statistics.gov.uk/downloads/theme_population/LAStudy_FullReport.pdf

3.1 Enumeration

When looking at possible ward level undercount, it is necessary to consider that one of the key assumptions underpinning the One Number Census (ONC) is that the undercount is

approximately randomly distributed - that is, the distribution of the undercount is mainly determined by the age and sex profile of an area and the distribution of hard to count areas as measured by the hard to count index. The assumption is that there are no other factors that have a large influence, such as the quality of the work carried out by the census enumerators. If this assumption is not true the resulting ONC estimates of population may not adequately reflect the true undercount. The Census Coverage Survey (CCS), although a large sample nationally, is highly unlikely to be able to measure this additional source of variability at low levels, i.e. Enumeration Districts or Wards. The Local Authority studies have made detailed comparisons with Council Tax data, and conducted in-depth analyses on enumerator record books (ERBs). The finding of these pieces of work are described below.

a) Analysis of Council Tax data

Southwark shows an overall difference between the Council Tax records and the Census results of 6,519 addresses, a 6.09 per cent difference. This is the 4th largest positive numerical difference, and the 2nd largest positive proportional difference for the 376 LAs in England and Wales.

The map on page 7 highlights the differences between Council Tax and Census by ward. CCS postcodes are indicated by the red dots.

The map shows that the 4 wards shaded darkest blue (Peckham, Nunhead, Village and Surrey Docks) are the ones with the largest percentage difference between Census and Council Tax figures and contain no CCS postcodes - which is a cause for concern as the ONC may therefore not have been a robust estimate.

The table below shows the differences between the Census and Council Tax for each ward, ordered by percentage difference, so that the ward at the top has the largest percentage difference. As shown in Table 3.1.1, Peckham ward has the largest difference between the 2 sources, with 12 other wards having a difference of over 300 addresses.

Difference between Census and Council Tax occupied and unoccupied dwelling counts

Southwark shows an overall difference between occupied dwellings according to the Council Tax records and the Census results of 4,223 dwellings, a 4.0 per cent difference. Amongst all English LAs, this is the 3rd largest positive

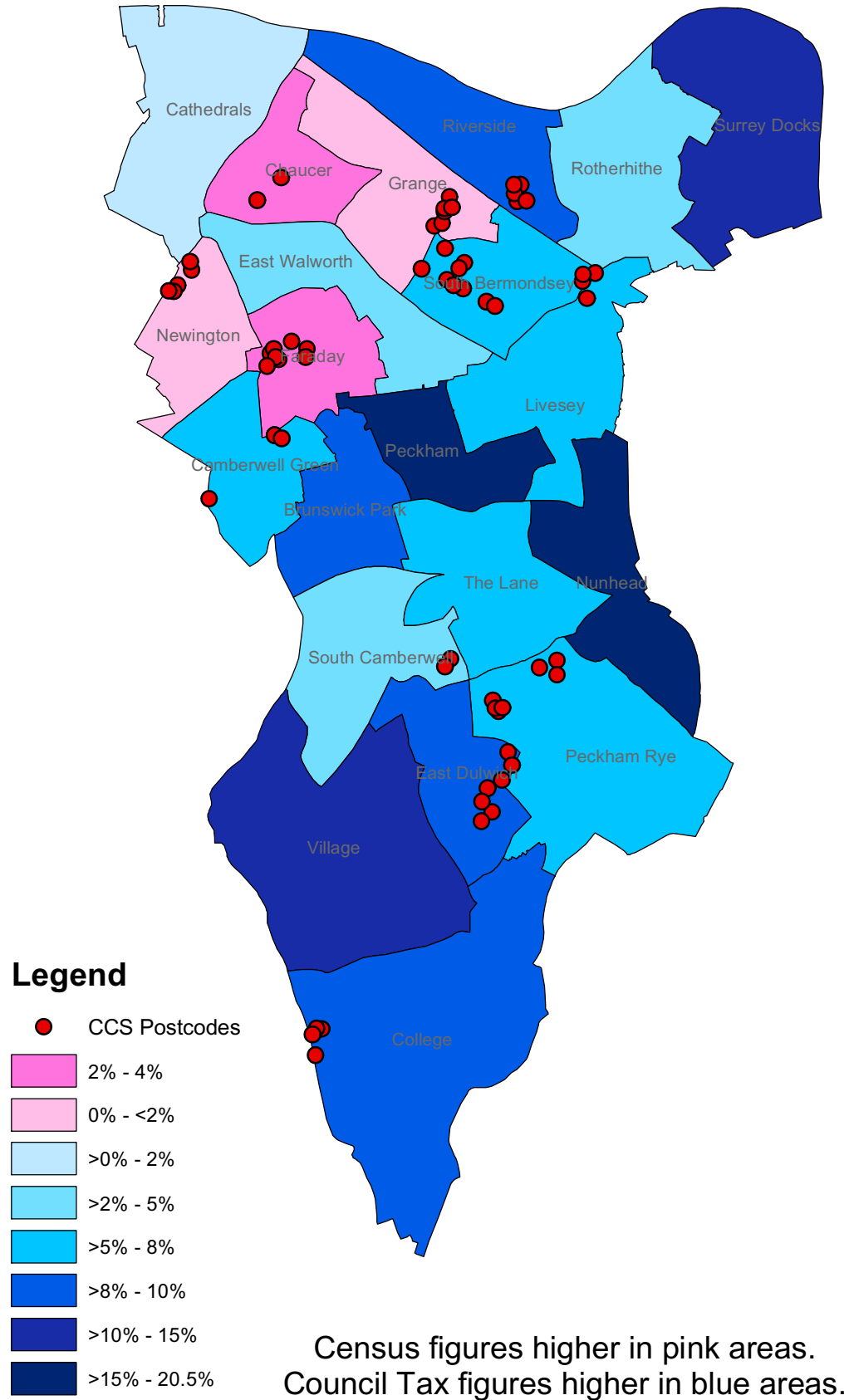
Table 3.1.1
Differences between Census and Council Tax dwelling counts by 2003 ward

2003 Ward	Census Dwellings	Council Tax Dwellings	Difference	Percentage Difference
Peckham	4,223	5,071	848	20.1%
Nunhead	4,575	5,452	877	19.1%
Village	4,101	4,710	609	14.9%
Surrey Docks	5,318	5,904	586	11.0%
Riverside	5,604	6,126	522	9.3%
East Dulwich	4,699	5,101	402	8.6%
Brunswick Park	4,701	5,096	395	8.4%
College	4,625	5,010	385	8.3%
Camberwell Green	5,580	5,976	396	7.1%
Livesey	5,185	5,531	346	6.7%
Peckham Rye	4,984	5,301	317	6.4%
South Bermondsey	5,258	5,570	312	5.9%
The Lane	5,360	5,663	303	5.7%
East Walworth	5,376	5,635	259	4.8%
Rotherhithe	5,176	5,388	212	4.1%
South Camberwell	4,859	5,040	181	3.7%
Cathedrals	5,466	5,518	52	1.0%
Grange	5,719	5,665	-54	-0.9%
Newington	5,775	5,714	-61	-1.0%
Chaucer	5,266	5,108	-158	-3.0%
Faraday	5,270	5,060	-210	-3.9%

NB: Due to boundary changes between Census day and 2003, some wards on the above table will be different to 2001 wards.

Since Census day 2001, populations from the following wards have been combined into a new one (Peckham ward): Brunswick, Friary and Liddle. Barse, Consort and Waverley have also been combined into Nunhead ward since Census day 2001.

Figure 3.1
Map showing the percentage difference between Council Tax and Census for Southwark



numerical difference, and the 4th largest positive proportional difference, and is a cause for concern.

Table 3.1.2
Differences between Census and Council Tax dwelling counts

	Census	CTB1	Difference	% Difference
Occupied	105,806	110,029	4,223	4.0%
Unoccupied/ Vacant	1,857	4,353	2,496	134.4%

b) Analysis of Enumerator Record Books (ERBs)
Investigations were carried out for Wards where a high discrepancy in either direction had been identified between the Census results and the Council Tax. This involved analysis of Enumerator Record Books (ERBs), investigating individual records to assess the quality of the enumeration in the area, to ensure that all addresses had been accounted for and establish whether the differences between the Census and Council Tax may be definitional (eg differences in recording of vacant properties). This analysis also allowed an estimate to be produced of the number of dummy forms that should have been completed by Enumerators.

Dummy forms are completed by census enumerators to account for census forms that either have not been returned (ie a non-response) or for which the enumerator has determined that they should not be returned (ie a valid non-response such as a vacant household). For 2001, the types of dummy form were:

- Refusal
- Absent Household
- No contact
- Vacant
- Second/Holiday Home

The first three of these can be considered as some form of non-response. In addition, some census forms are returned with no residents (and these are not all visitor households) which can also be considered as a type of dummy form (an 'empty' household).

Investigation of the ERBs in Peckham, Nunhead, Village and Surrey Docks wards (where there are the largest differences between Council Tax and Census figures) was carried out. This analysis

identified a large difference between the number of non-response dummies according to the ERBs and the Census database records. Analysis of the ERBs also reveals that the address lists did not represent these wards well. Therefore, there is evidence to suggest that significant numbers of households were missed in Peckham, Nunhead, Village and Surrey Docks wards due to enumeration problems. Furthermore, there were a large number of derelict and demolished properties in the Liddle ward of Southwark.

c) Analysis of Council Tax based response rates by 2001 Ward

Table 3.1.3 on page 9 shows a comparison of Council Tax Figures and households counted in the Census, broken down by ward. An implied Census response rate has been calculated by taking the number of counted Census households and dividing this by the number of assumed occupied Council Tax addresses (the total number of Council Tax addresses minus those counted by the Census as vacant, second homes and visitor only households). In wards where the Council Tax implied response rate is substantially lower than the ONC household response rate for the LA, it is possible that significant enumeration failings, for which the ONC may not have been able to make a robust adjustment, may have occurred. The table is ordered by the implied census response rate, with the wards with the lowest response rate at the top of the table.

The ONC household response rate for Southwark is 79.6 per cent. Looking at Table 3.1.3, Liddle, Friary, Consort and Barsest wards show quite a large difference between the ONC household response rate and the implied Census response rate.

d) Enumeration information

Census enumerators collected certain information on households where no contact was made. According to this data, there were 1,319 (1.4 per cent) vacant households and 298 (0.32 per cent) second homes in Southwark.

Census returns indicated that there were 152 (0.16 per cent) visitor only households, and were 1,059 (1.12 per cent) empty households (returned household forms which have no usual residents, or very little information on the form). These figures are not extreme and are consistent with expectations.

Table 3.1.3
Analysis of Council Tax based response rates by 2001 Ward

Ward Name	2001 Council Tax addresses	Census Vacants, 2nd homes and visitor households	Assumed occupied Council Tax addresses	Counted Census households	Absolute Difference	Implied Census Response Rate
Liddle	3,320	33	3,287	2,096	1,191	64%
Friary	3,916	4	3,912	2,530	1,382	65%
Consort	3,488	43	3,445	2,329	1,116	68%
Barset	3,106	48	3,058	2,085	973	68%
Brunswick	4,857	131	4,726	3,371	1,355	71%
Alley	3,226	29	3,197	2,283	914	71%
Rotherhithe	4,562	85	4,477	3,251	1,226	73%
Browning	4,999	8	4,991	3,627	1,364	73%
St. Giles	5,368	95	5,273	3,853	1,420	73%
Waverley	3,106	10	3,096	2,275	821	73%
Rye	3,982	62	3,920	2,885	1,035	74%
Riverside	6,384	209	6,175	4,568	1,607	74%
Dockyard	9,958	123	9,835	7,412	2,423	75%
Lyndhurst	5,139	43	5,096	3,844	1,252	75%
Burgess	2,772	48	2,724	2,055	669	75%
Bellenden	5,357	118	5,239	3,995	1,244	76%
Ruskin	3,929	3	3,926	3,005	921	77%
The Lane	3,561	47	3,514	2,711	803	77%
Cathedral	4,275	246	4,029	3,166	863	78%
College	3,450	23	3,427	2,718	709	79%
Bricklayers	4,680	32	4,648	3,690	958	79%
Abbey	3,834	74	3,760	3,053	707	81%
Newington	5,411	43	5,368	4,393	975	82%
Faraday	5,769	90	5,679	4,676	1,003	82%
Chaucer	5,190	167	5,023	4,243	780	84%
Totals	113,639	1,814	111,825	84,114	27,711	75%

NB: It should be noted that the above table shows a comparison with the wards as they were on Census day 2001. Since Census day 2001, populations from the following wards have been combined into a new one (Peckham ward): Brunswick, Friary and Liddle. Barset, Consort and Waverley have also been combined into Nunhead ward since Census day 2001.

e) Estimated response compared to other areas

The ONC contingency strategy used a classification of 'similar' areas (known as borrowing strength areas) which were to be used if it was judged the ONC process had failed. This classification is useful for comparing response rates across areas, as we would expect the response rates measured by the ONC to be similar for these LAs. Further information on the contingency and borrowing strength strategy can be found at www.statistics.gov.uk/census2001/pdfs/oncinfopaper.pdf, which includes the borrowing strength areas in Annex E.

The Borrowing Strength areas for Southwark and their associated ONC response rate figures are shown in Table 3.1.4 below. The response rate for Southwark is lower than the mean response rate for its borrowing strength areas.

Table 3.1.4
ONC Response rates for similar (borrowing strength) areas

Area	ONC Response Rate
Southwark	76.8%
Borrowing Strength Areas	
Islington	77.9%
Lambeth	78.8%
Hackney	72.1%
Lewisham	80.8%
Haringey	83.4%
Mean Response Rate for BS Areas	78.6%

The ONC response rate for the other local authority within the same Census estimation area (EA) as Southwark is in the table below. It shows that the response rates for Southwark are broadly consistent with the other LA in the EA.

Table 3.1.5
Response rates for all LADs in the Estimation Area

Local Authority	ONC Response Rate
Southwark	76.8%
Lambeth	78.8%

f) Qualitative information obtained about the Census enumeration

The analysis of fieldwork intelligence indicated many of the enumerators in Southwark had double workloads, which may have led to a reduction in the quality of enumeration. The qualitative information also indicated that the proportion of direct and late returns fell outside the expected range.

3.2 Census Coverage Survey

A key element of the One Number Census was the Census Coverage Survey (CCS) - a post enumeration survey that was designed to measure undercount in the Census. The aim was to survey a representative sample of postcodes across the country, which would allow differential undercount to be quantified. Because under-enumeration is disproportionately distributed across areas, the CCS was stratified according to a Hard to Count (HtC) index, constructed from the following 1991 Census variables which are associated with undercount:

- Multi -occupancy
- Unemployment
- Country of birth (which is associated with language difficulty)
- Private rented accommodation
- Number of households imputed in 1991

ONC imputation rates by key variables can be found at www.statistics.gov.uk/census2001/imputation_rates_by_variable.asp, and confirm that the above variables were associated with undercount in the 2001 Census.

The index categorises Enumeration Districts (EDs) into 3 groups representing the easiest 40 per cent, the next 40 per cent and the hardest 20 per cent of EDs nationally. The sample was then selected separately within each of these strata. This meant that the CCS sample had good coverage of areas with each HtC index value, based on 1991 Census data, which was the only information available at the time of the 2001 Census.

a) Missing sample

As part of the ONC process, investigations were carried out into sampled postcodes, with high levels of CCS only, or Census only counts. If it was found that the CCS or Census were enumerated so poorly as to be out of scope of the ONC (ie the ONC would not be able to compensate for the undercount), these postcodes were removed from the sample.

Four postcodes in Southwark were removed from the CCS sample, due to poor CCS coverage, and interviewers identifying too few households in one postcode.

ONC figures indicate that 6 of the 17 CCS workloads have coverage of less than 80 per cent. This is in line with other London areas but may indicate that the CCS is less robust.

b) Sample balance

This analysis assessed whether the selected sample was adequately balanced across indicators associated with undercount (both across the Estimation Area and also the constituent Local Authorities), and also whether the CCS was successful in achieving high response rates, and therefore measuring the undercount in the Census.

The analyses used 2001 Census data, in an attempt to examine the balance of the sample across a number of variables.

Dummy form distribution

As mentioned in section 3.1 (b), dummy forms are completed by census enumerators to account for census forms that either have not been returned (ie a non-response) or for which the enumerator has determined that they should not be returned (ie a valid non-response such as a vacant household).

The refusal, absent, no contact and empty households (which can be thought of as non-response dummy forms) provide an indication of response rates across an area, and can be used to assess whether there is any significant imbalance within the CCS sample selection by comparing the proportions of these dummy forms between the sample and the non-sampled areas. This would show whether the CCS is likely to provide a sample that underestimates or overestimates undercount.

The proportions of census households that were classified as refusals, no-contact, absent or empty households are examined. The data used are the

unadjusted census records - that is prior to the ONC adjustments. The CCS sampled postcodes within an Estimation Area (EA) are compared with the whole of the EA, and the same analysis done at LA level. This is also carried out within the Hard to Count (HtC) strata, since we would expect the proportions of dummy forms to be different across these strata. It is valuable to look at proportions of dummy forms at both LA and the EA level because the ONC estimation system produced estimates first at the EA level, then apportioned these estimates out to the LAs within that EA.

The map on page 12 shows the spread of dummy non-response forms by ward, with CCS postcodes indicated by red dots. Lyndhurst is the ward with the largest percentage of non-response dummy forms (ie the darkest wards on the map), and contains no CCS postcodes. However, most of the wards with a large percentage of non-response dummy forms do contain a number of CCS postcodes. The fact that the CCS visited these areas indicates that undercount in these areas is likely to have been identified and adjusted through the ONC process.

Table 3.2.1 below shows the proportion of Census dummy forms across the Estimation area and within the CCS sample for each Hard to Count stratum. The difference between proportions is not high so does not give cause for concern.

It should be noted that there are no HtC 1 areas, and that hard to count strata 2 and 3 were collapsed for this Estimation area, due to the small sample size of hard to count 2. See section 3.3 (b) for details.

Table 3.2.2 below shows the proportion of dummy forms for each Hard to Count stratum of each Local Authority in the Estimation area, both across the whole stratum and within the CCS sample. Again, the differences between proportions are not large.

To explore this further, the distribution of the proportions of dummy forms are shown below in figures 3.2.2 to 3.2.3 on pages 13 and 14, with the proportion for the Southwark as a whole shown in figure 3.2.2 and broken down by hard to count stratum for figure 3.2.3 (c).

There was one ED in the CCS sample with more than 25 per cent dummy forms.

The graphs below all indicate that the sample is balanced, which will have allowed the One Number Census to identify undercount across different hard to count strata.

Figure 3.2.3 (a) is not relevant, as there is no Hard to Count 1 stratum in Southwark.

Figure 3.2.3 (b) is not relevant, as the Hard to Count 2 stratum in Southwark is small and has been combined with Hard to Count 3.

Recalculated Hard to Count score

The Hard to Count score was derived using a number of 1991 Census variables which were associated with undercount. The score was used to determine the level of the Hard to Count index for each 1991 Enumeration District in England and Wales. The index was then used within the CCS sampling strategy as the primary stratifier within each Estimation Area, and as a stratum for estimation.

Table 3.2.1
Proportion of dummy forms across the Estimation area and within the CCS sample for each Hard to Count stratum

Estimation Area	Hard to Count Index stratum	Count of dummy forms in CCS	Proportion of dummy forms across all areas	Proportion of dummy forms in CCS	Difference between proportions
Lambeth and Southwark	3	523	13.6	15.9	-2.3%

Table 3.2.2
Analysis of proportions of dummy forms across the Local Authorities for each Hard to Count stratum

Local Authority District	Hard to Count Index stratum	Count of dummy forms in CCS	Count of households in CCS	Proportion of dummy forms across all areas	Proportion of dummy forms in CCS	Difference between proportions
Southwark	3	143	1,421	8.9	10.1	-1.2
Lambeth	3	380	1,870	17.5	20.3	-2.9

Figure 3.2.1
Map to show the percentage of non-response dummy forms for Southwark

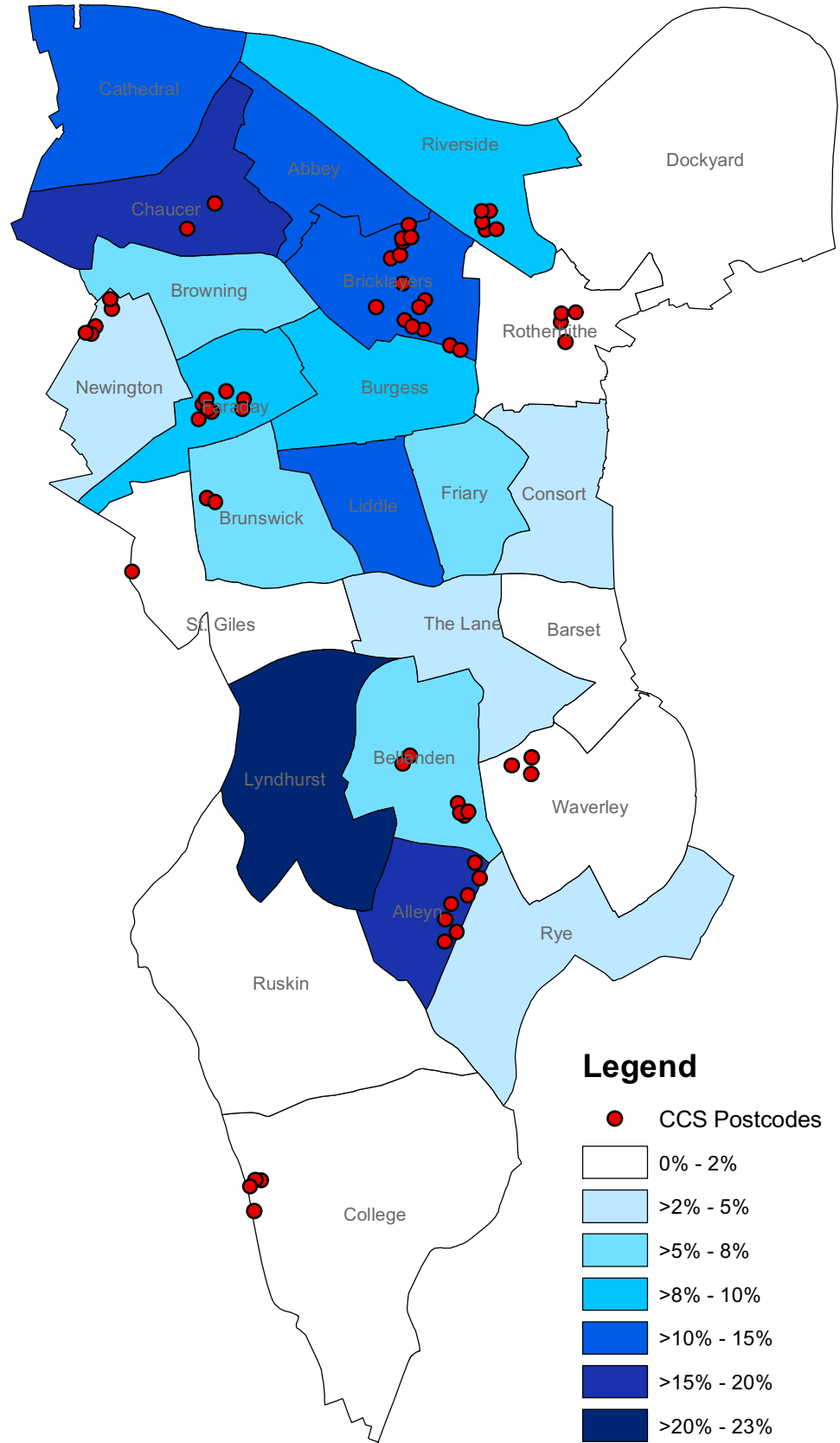


Figure 3.2.2
Distributions of the proportions of dummy forms in each 2001 Enumeration District

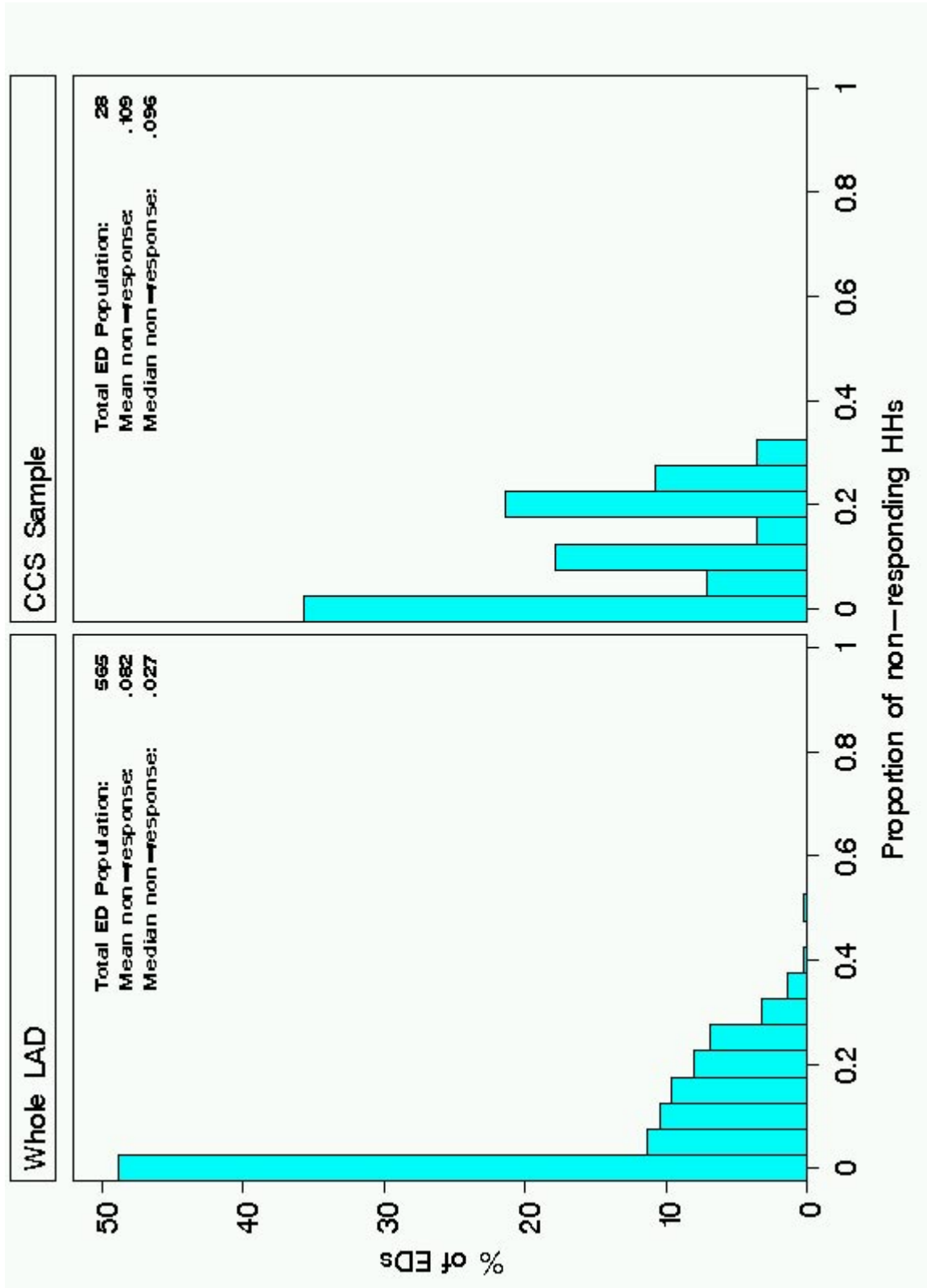
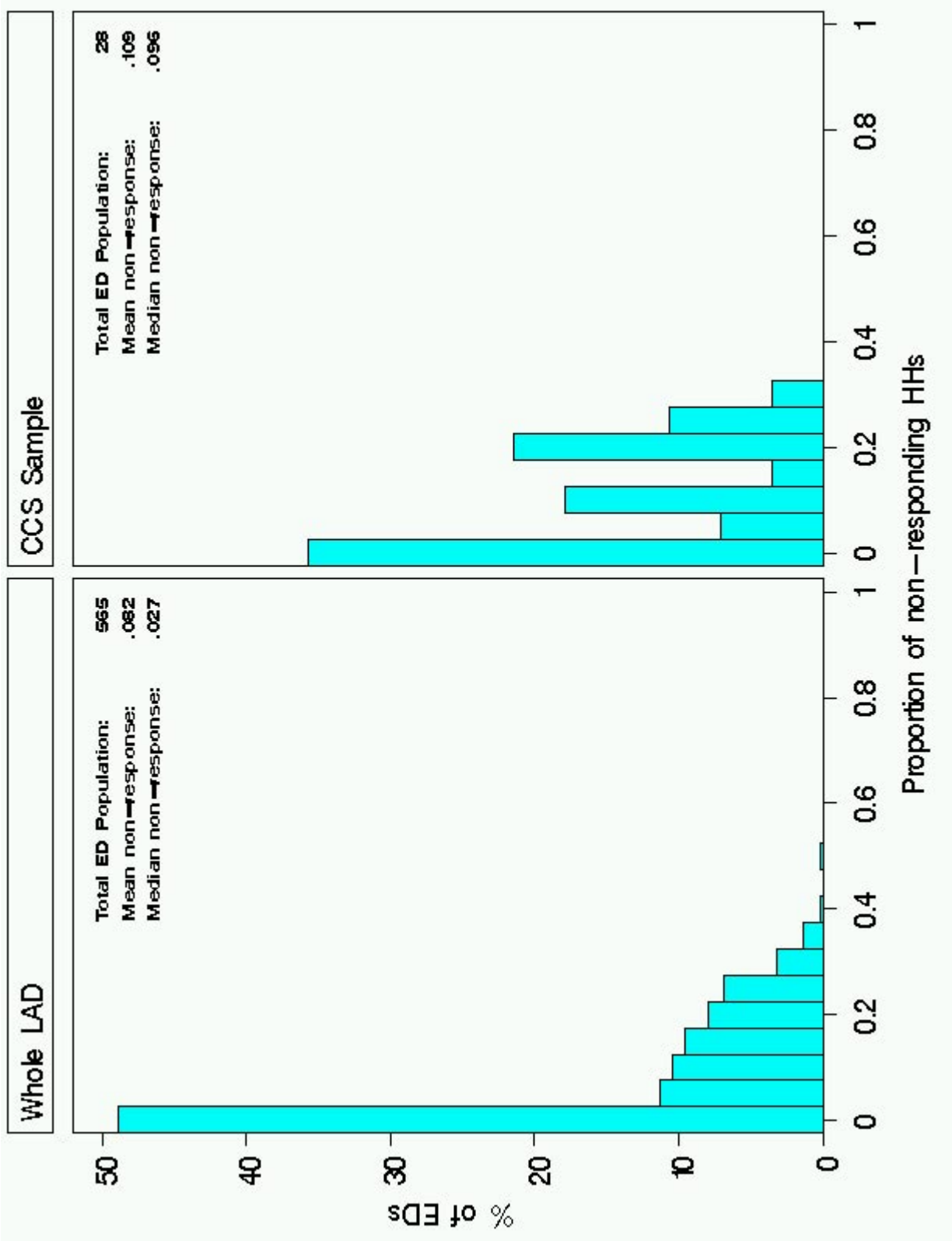


Figure 3.2.3(c)
Distributions of the proportions of dummy forms in each 2001 Enumeration District for the HtC 3 stratum



As the data used to derive the HtC index was from the 1991 Census, there was a risk of the sample being unbalanced with respect to the 'real' hard to count information. We can assess this by using a new hard to count score, derived from the 2001 Census data, at postcode level. The assumption underpinning the analysis is that the recalculated score is highly correlated with the real undercount - that is, that these variables are associated with undercount in 2001. From the patterns observed in the CCS, there is evidence to suggest that this assumption is reasonable. The score is calculated as the sum of the proportions of:

- Unemployed persons
- Persons whose country of birth was a non-English speaking nation
- Privately rented households
- Dummy form 'non-response' households
- Multi-occupied dwellings

The derived score is calculated for all postcodes across the Estimation Area, and comparisons can be drawn between the score distributions for the Estimation Area/Local Authority and the CCS sampled postcodes.

Section 3.3 (d) Stratification, also looks at the recalculated hard to count scores but provides a general analysis on how the area has changed in terms of hard to count between 1991 and 2001. This section focuses on comparing the proportions of hard to count areas in the CCS sample and the rest of the EA or LA

Figure 3.2.4 on page 16 shows the recalculated Hard to Count score distributions for the whole EA and for CCS sample postcodes. The distributions are fairly similar, indicating that the CCS sample was balanced across the EA as a whole.

Figure 3.2.5 (a) is not relevant, as there is no Hard to Count 1 stratum in Southwark

Figure 3.2.5 (b) is not relevant, as there is no Hard to Count 2 stratum in Southwark

Figure 3.2.5 (c) on page 17 shows the recalculated Hard to Count score distributions for the population and sample across the whole Estimation Area by 2001 hard to count groups. Distributions are fairly similar for the population and the sample for hard to count strata 3, which indicates that the CCS

was targeted in areas which allowed it to be a balanced sample.

Figure 3.2.6 on page 18 shows the recalculated Hard to Count score distributions for Southwark Local Authority and for the CCS sample areas within Southwark. Again, distributions are similar between the two, although the CCS sample has a higher proportion of postcodes between a Hard to Count score of 0.2 and 0.4.

c) Localised undercount and the CCS

Potential areas of localised undercount were identified in a number of ways including comparing the ONC household response rate with the response rate implied by Council Tax (see section 3.1 (c), table 3.1.3). Areas with high proportions of dummy forms were also investigated. Where the CCS had visited these wards, it was investigated to check the level of undercount that had been observed by the CCS, and whether this looked plausible when compared to the levels implied by the Council Tax.

The following table on page 19 shows the level of undercount identified by the CCS. The One Number Census was based on dual system estimation, which combines the numbers of people enumerated by the Census and/or CCS and estimates those people missed by both. The DSE (dual system estimate) column shows this figure for each ward. A measurement of the Census undercount implied by the CCS can therefore be obtained by dividing the DSE by the number of people counted in the Census, and using this figure as a multiplier. The final column in the table shows this figure. If the CCS had found a level of undercount that was similar to that implied by the Council Tax data, then there is evidence to suggest that the ONC has made a robust adjustment. If, however, the CCS had no sample in these wards or it had not found a reasonable level of undercount, then the ONC may not have been able to make a robust adjustment.

In Southwark, Peckham and Nunhead were the wards with the biggest differences with Council Tax records. Due to ward changes, since Census day 2001, populations from the following wards have been combined into a new ward (Peckham ward): Brunswick, Friary and Liddle. Barset, Consort and Waverley have also been combined into Nunhead ward since Census day 2001. The CCS did not sample Friary, Liddle, Barset, or Consort.

Figure 3.2.4
Distributions of the recalculated Hard to Count score for the Estimation Area

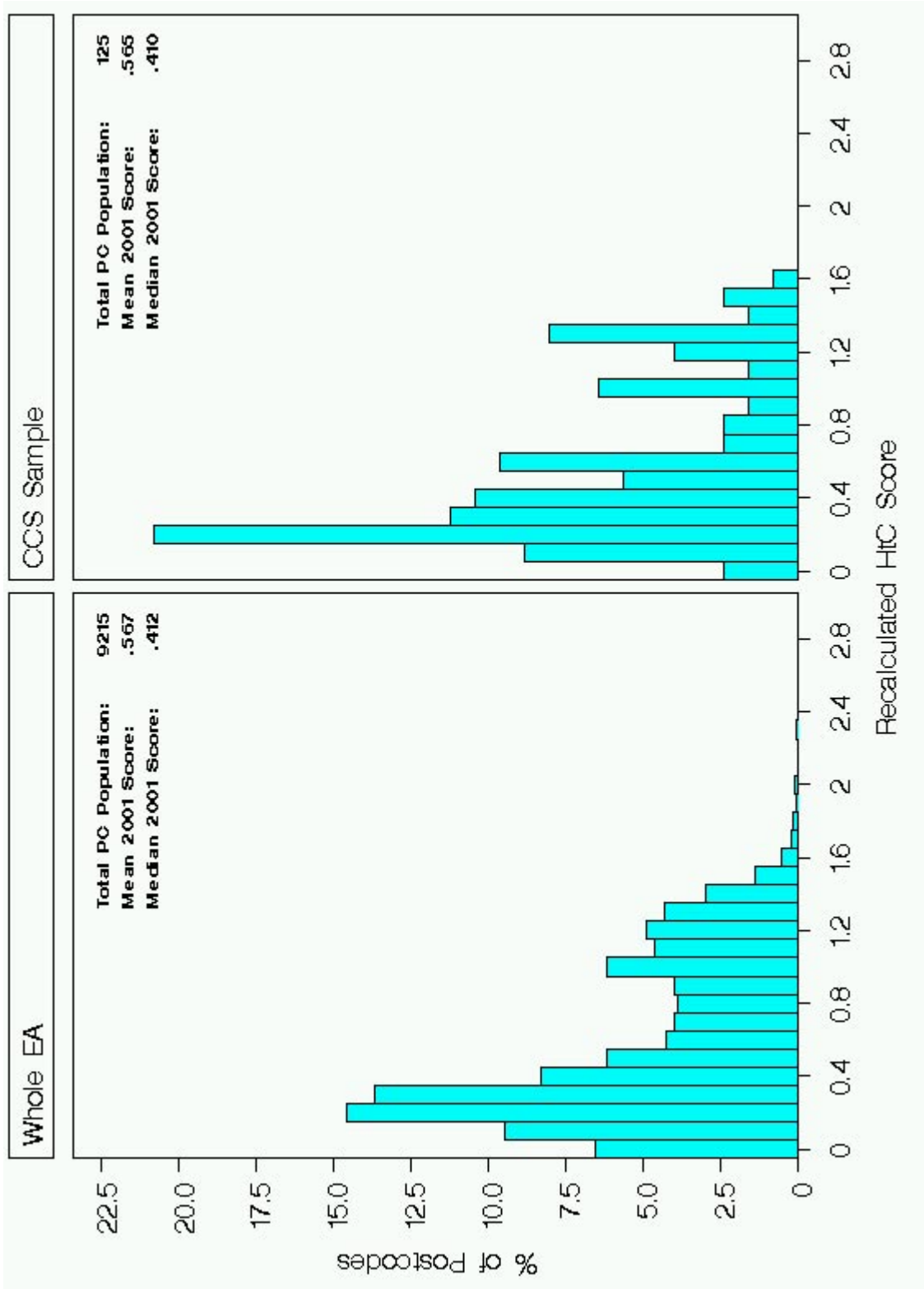


Figure 3.2.5(c)
Figure 3.2.5 (c) - Distribution of the recalculated Hard to Count score for the Estimation Area by Hard to Count group 3

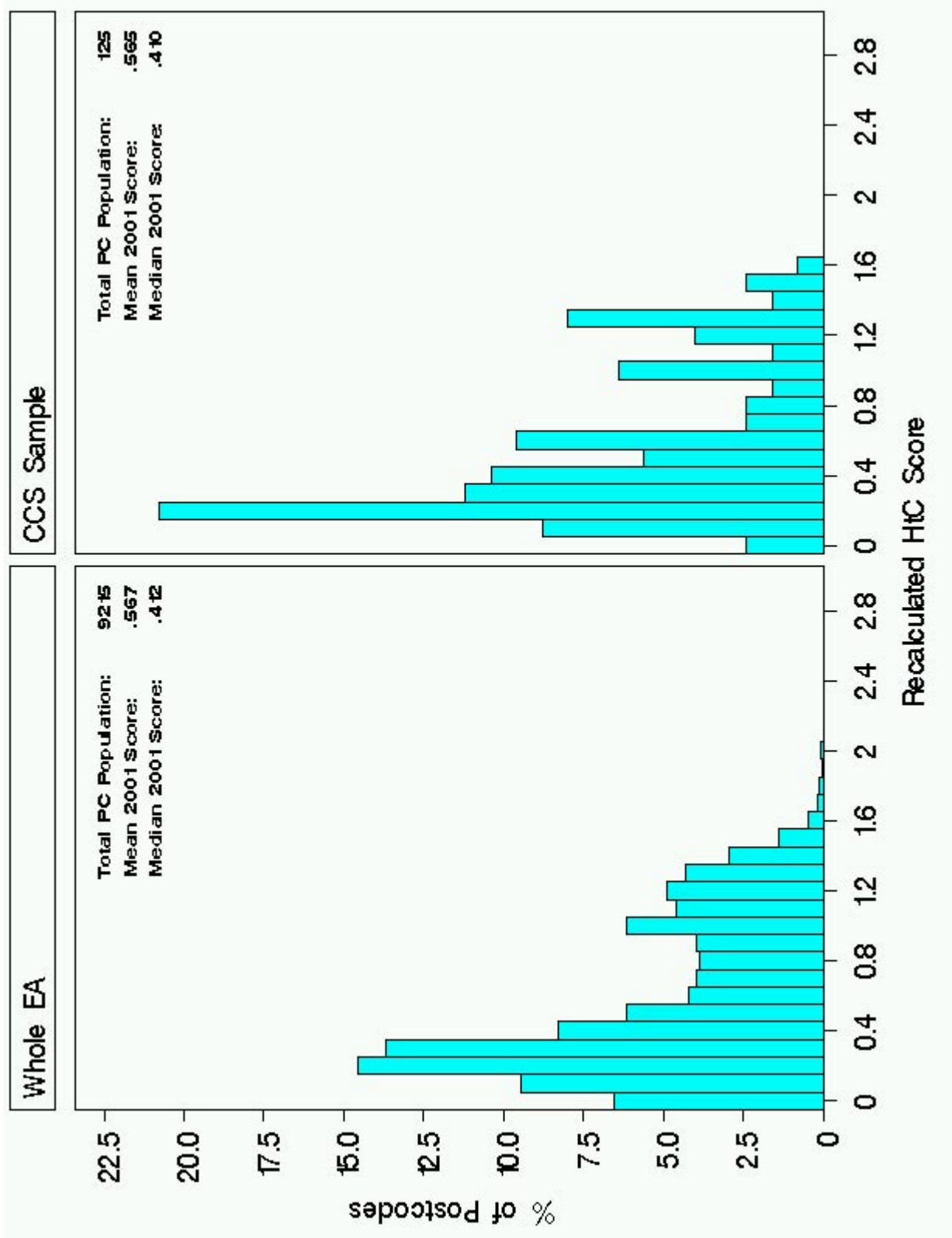


Figure 3.2.6
Distributions of the recalculated Hard to Count score for the Local Authority

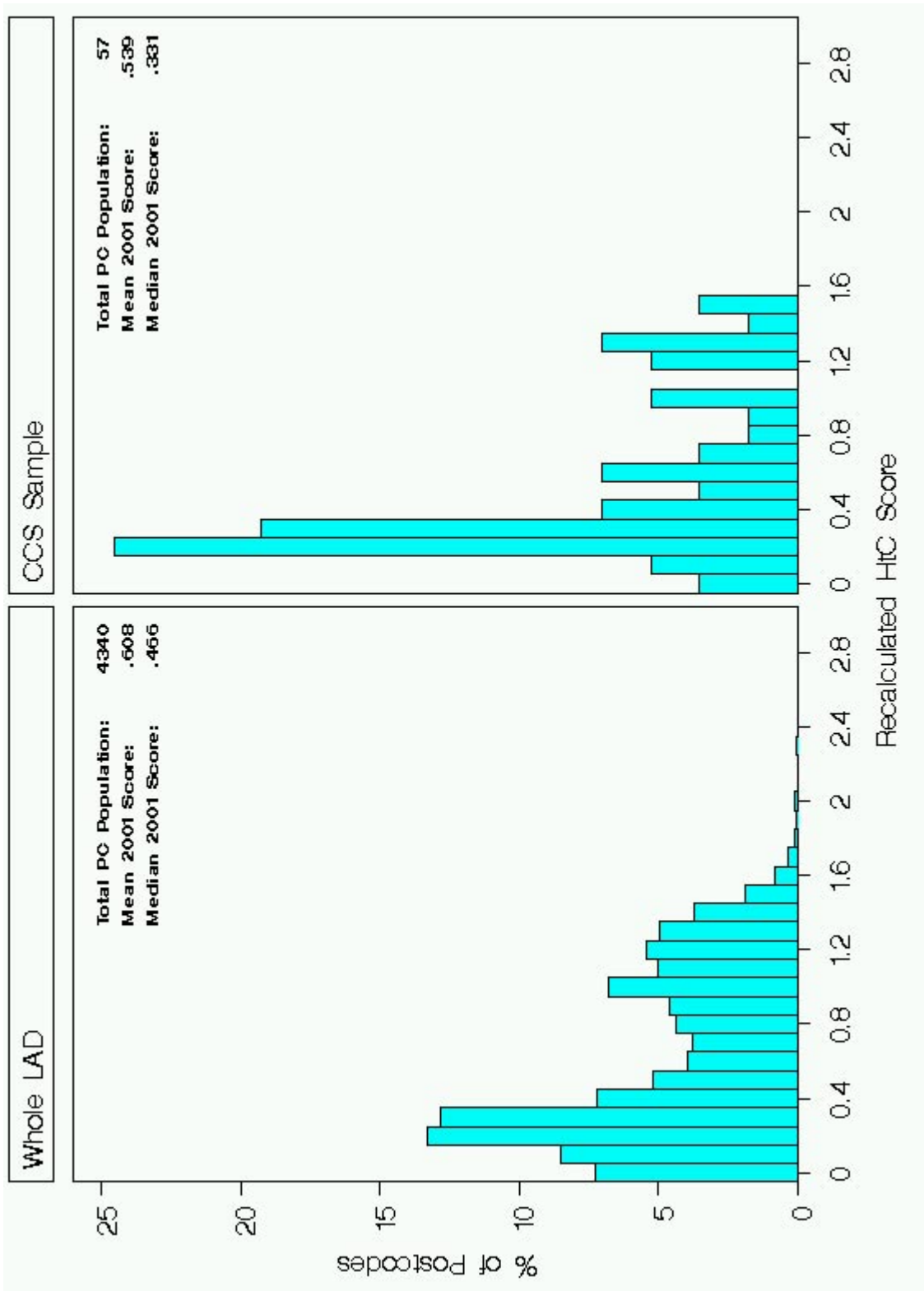


Table 3.2.3
Census Coverage Survey implied undercount by 2001 ward
HOUSEHOLDS

Ward code	No. CCS postcodes	Counted in Census	DSE	DSE/Counted in Census	CCS undercount
Brunswick	3	58	90.7	1.564	26.1%
Waverley	3	83	100.2	1.207	17.2%

PERSONS

Ward code	No. CCS postcodes	Counted in Census	DSE	DSE/Counted in Census	CCS undercount
Brunswick	3	125	215.2	1.722	41.9%
Waverley	3	201	251.8	1.253	20.2%

Table 3.2.4
Response rates, matching outcomes and sum of the missed person estimates by Hard to Count group for the CCS sample in the Estimation Area.

Local Authority	Hard to Count group	CCS person response rate	CCS only count	Census only count	Matched count	Estimate of persons missed in both
Southwark	3	79.6%	636	571	2,245	169
Lambeth	3	76.3%	646	733	2,405	215

The CCS has found a high level of undercount in parts of the Peckham ward, with a mean undercount of 26.1 per cent for households and 41.9 per cent for persons across the wards from parts of which it was formed. There is also a fairly high level of CCS undercount in Nunhead ward, with a mean undercount of 17.2 per cent for households and 20.2 per cent for persons across the wards from parts of which it was formed. Whilst these undercount rates are high and in line with the analysis of Council Tax data, the sample sizes are small which might mean that there is not enough data to make an overall robust adjustment in wards with no CCS sample.

d) CCS fieldwork and response rates

There were few problems with the CCS in Southwark, although some interviewers did have heavy workloads, and there was a high proportion of refusals of 10 per cent.

Table 3.2.4 above shows the response rates, matching outcomes and sum of the dual system estimates by Hard to Count group. The CCS has found many extra people compared with the census only counts. Hard to count strata 2 and 3 were collapsed due to the low sample size of hard to count 2 areas (see section 3.3 (b) Collapsing in the ONC process). The CCS has therefore performed in line with expectations.

3.3 One Number Census process

The One Number Census project (ONC) was set up by ONS for the 2001 Census to address

the fact that it is inevitable that some people and households will not be counted in any population census. By conducting a large post-enumeration survey (the Census Coverage Survey, CCS) and combining the results of both the Census and CCS in what is known as a dual system approach, the aim of the project was to estimate and adjust the Census database for undercount, and to ensure that robust results could be obtained for each local authority area. Detailed information on the One Number Census can be found in the following links:

A Guide to the One Number Census:
www.statistics.gov.uk/census2001/pdfs/oncguide.pdf

One Number Census methodology and Quality Assurance process report:
www.statistics.gov.uk/census2001/pdfs/onc_qa_process.pdf

Analyses into how well the ONC process worked for Southwark were carried out as detailed in the sections below.

a) Response rates

The estimated person level response rate for Southwark Local Authority is 76.8 per cent.

Across the Local Authority, the dummy forms suggested a household response rate of 91.1 per cent, whereas the ONC estimated 79.6 per cent (Ranked 1st across all LAs in terms

of the difference). There were 33 unfilled dummy forms after the imputation process. 13,418 households were imputed into random postcodes. As noted in the ERB analysis, there is evidence that enumerators failed to fill in dummy forms, which explains this discrepancy (see section 3.1 b).

Table 3.3.1 below shows the estimated response rates by Hard to Count group for this Local Authority.

Table 3.3.1
Estimated Response rates by Hard to Count group for Southwark.

Hard to Count group	Household Response rate	Person Response rate
3	79.6%	76.8%

b) Collapsing in the ONC process

For the purpose of One Number Census (ONC) estimation the population was divided into 37 age-sex groups. In addition, each postcode was classified into one of three Hard to Count (HtC) levels, 1 being the easiest and 3 the hardest. This means that there were 111 separate estimation strata in any given Estimation Area (EA). However, in some cases it was not possible to produce good quality estimates of these groups separately. In these cases groups were combined, referred to as “collapsing strata”. Because there was no HtC 1, and because HtC 2 and 3 were collapsed, there were only 37 strata for this EA. More details of how this was applied for Southwark can be at www.statistics.gov.uk/census2001/pdfs/collapsing_strata.pdf

c) Outliers

Each postcode in the CCS is used to represent undercount in a number of postcodes in the population. When an unusual observation occurs in a postcode it is likely that it is not representative of other postcodes. In order that the unusual postcode (referred to as an outlier) does not have an undue influence on the rest of the population, it is removed from the sample.

Within the ONC strategy, outliers were identified using pre-defined ‘ratio’ limits for individual observations. The ratio was defined by the ratio of the dual system estimate to the census count for each age-sex observation in each postcode. These limits were different for each Hard to Count stratum. The limits were a ratio of 3 for the Hard to Count 1 stratum, a ratio of 4 for the Hard to Count 2 stratum and a

ratio of 5 for the Hard to Count 3 stratum. If an observation was classified as an outlier through this method, it was not used in the calculation of any model parameters. For example, the CCS finds 6 people that the Census missed, and both the Census and CCS find 1 person. The Census will show 1 person, and the dual system estimate is 7 persons. The ratio for this observation is then 7, which is classified as an outlier. This methodology was based on the data from the simulation studies, which assumed an overall 95 per cent Census coverage.

The Lambeth and Southwark Estimation Area had 10 outliers, ranked 6th across all 101 Estimation Areas. The outliers within this Estimation Area contained 2.6 per cent of the persons found in the CCS that were missed by the Census, ranked equal 48th amongst all estimation areas. This is not high enough to be a cause for concern.

d) Stratification

The ONC estimation process used pre-defined stratification. These were the Hard to Count (HtC) index and 37 five-year age-sex groups. The HtC index was based on 1991 Census data, and was used to draw the sample and to form estimation groups. Whilst this was the best stratification that could be used at the time, there may be areas where the Hard to Count index gave a poor stratification. Analyses have been undertaken to examine the change between the 1991 HtC distribution and that implied by the 2001 Census data.

Figure 3.3.1 on page 21 shows how the HtC score distribution for Enumeration Districts (EDs) has changed from the 1991 Census to the 2001 Census for the whole Estimation Area. A score of greater than 0.43 on figure 3.3.1 indicates that the ED would be a HtC 3 area using 2001 data. It can be seen that the Estimation Area is harder to count in 2001 than it was in 1991, with more EDs in HtC 3. This is mainly a result of the number of privately rented homes, although there is also a relatively large proportion of people from non-English speaking countries

Note that this analysis looks at the overall change between 1991 and 2001 in the distribution of hard to count areas across the Estimation Area and LA as a whole. Part of section 3.2 (b) also looked at the recalculated hard to count distribution, but in terms of the CCS sample compared with the rest of the EA and LA.

Figure 3.3.1
Distribution of 1991 and 2001 Hard to Count scores for EDs within the Estimation Area

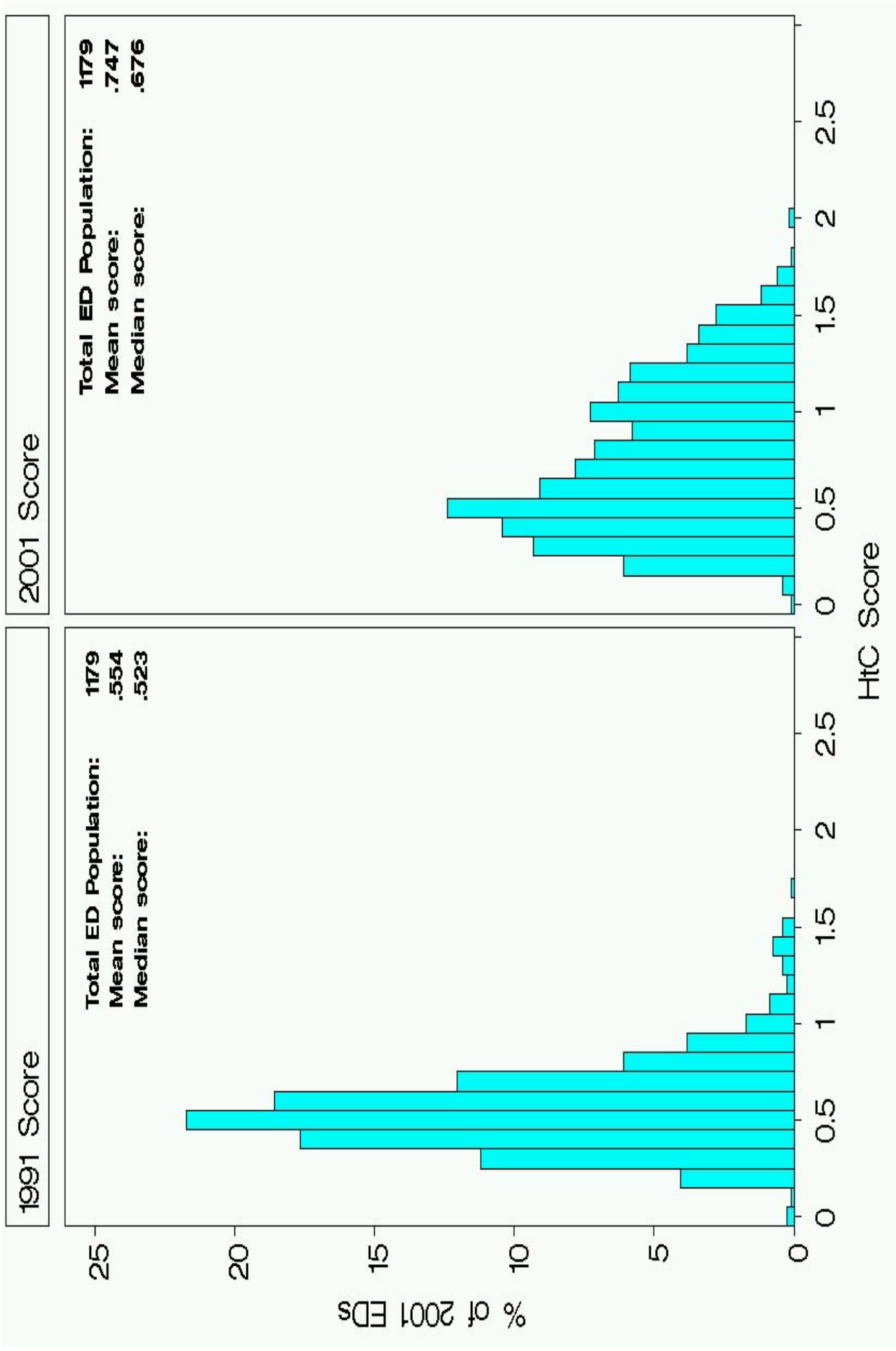


Figure 3.3.2 on page 23 shows how the HtC score distribution has changed from the 1991 Census to the 2001 Census for Southwark. As with the EA graph, this shows that Southwark is harder to count - EDs scoring above 0.43 would fall into HtC stratum 3, and the 2001 score shows a higher proportion of EDs above this level.

Figure 3.3.3 on page 24 shows how individual EDs have changed from the 1991 score to the 2001 score for the Local Authority District. This confirms the conclusions from the previous graphs, indicating that Southwark is now harder to count, although it was already fairly hard to count in 1991.

This is not an area for concern as the sample size was large in the hardest to count strata. Therefore the change in the area will be reflected in the ONC estimated undercount.

e) Household and person imputation results

Detailed information on imputation can be found at www.statistics.gov.uk/census2001/pdfs/sc9908.pdf

The ONC imputation process imputed people as part of wholly imputed households and also into counted households. There was no explicit restriction on how many people were imputed into these two categories, the only restriction was on the total number of people imputed into a Local Authority. Since the imputation of households (rather than people) was carried out as the first step, any people remaining to be imputed were then placed into counted households. Therefore, the patterns observed in the CCS may be very different to those created by the imputation methodology. There are a number of factors that will contribute to any extreme differences, mainly relating to households and people being estimated separately. This means that any under or over-estimation of either will affect the proportions of households or people that are imputed.

For areas where the imputed proportion is significantly higher than that measured by the CCS there are three interpretations:

- 1) There was an underestimate of total households, therefore more imputed people went into counted households.
- 2) There was a high level of dependence between Census and CCS for people within counted households, and so the CCS pattern shows an underestimate. More information on dependency can be found at www.statistics.gov.uk/census2001/pdfs/dependency_paper.pdf

- 3) There was an overestimate of total people, therefore the remainder were placed into counted households.

Of these three scenarios, generally the first two are the most likely to have occurred. However, a firm conclusion could only be reached if supported by other external evidence, of dependence, under-estimation of households or over-estimation of people. Conversely, for areas where the imputed proportion is significantly lower than that measured by the CCS, there are two interpretations:

- 1) There was an overestimate of total households, therefore too many imputed people went into wholly imputed households.
- 2) There was a high level of dependence for missed households, which results in an underestimate of total households and total people. Because the under-estimate of people will be within wholly missed households, the CCS pattern will show an overestimate.

In general, of these two scenarios, the second is the most likely to have occurred. However, a firm conclusion could only be reached if supported by other external evidence, of dependence or of over-estimation of households.

Within this Estimation Area, the imputation process imputed 15.8 per cent of the imputed persons into counted households. The CCS measured this proportion to be 15.3 per cent. This was the 27th largest difference in this direction across all 101 Estimation Areas.

At LA level, the imputation process imputed 19.7 per cent of the imputed persons into counted households. The CCS measured this proportion to be 15.8 per cent. This was the equal 92nd largest difference in this direction across all 376 Local Authority Areas.

These figures (where the imputation process has imputed higher proportions than the CCS measured) could indicate a potential underestimate of total households or a high level of dependency between Census and CCS, as described above. However, there is not a large difference between the imputation process and the CCS with Southwark's figures (ranked 87th of 376) so there is unlikely to be a problem here.

Figure 3.3.2
Distribution of 1991 and 2001 Hard to Count scores for Southwark

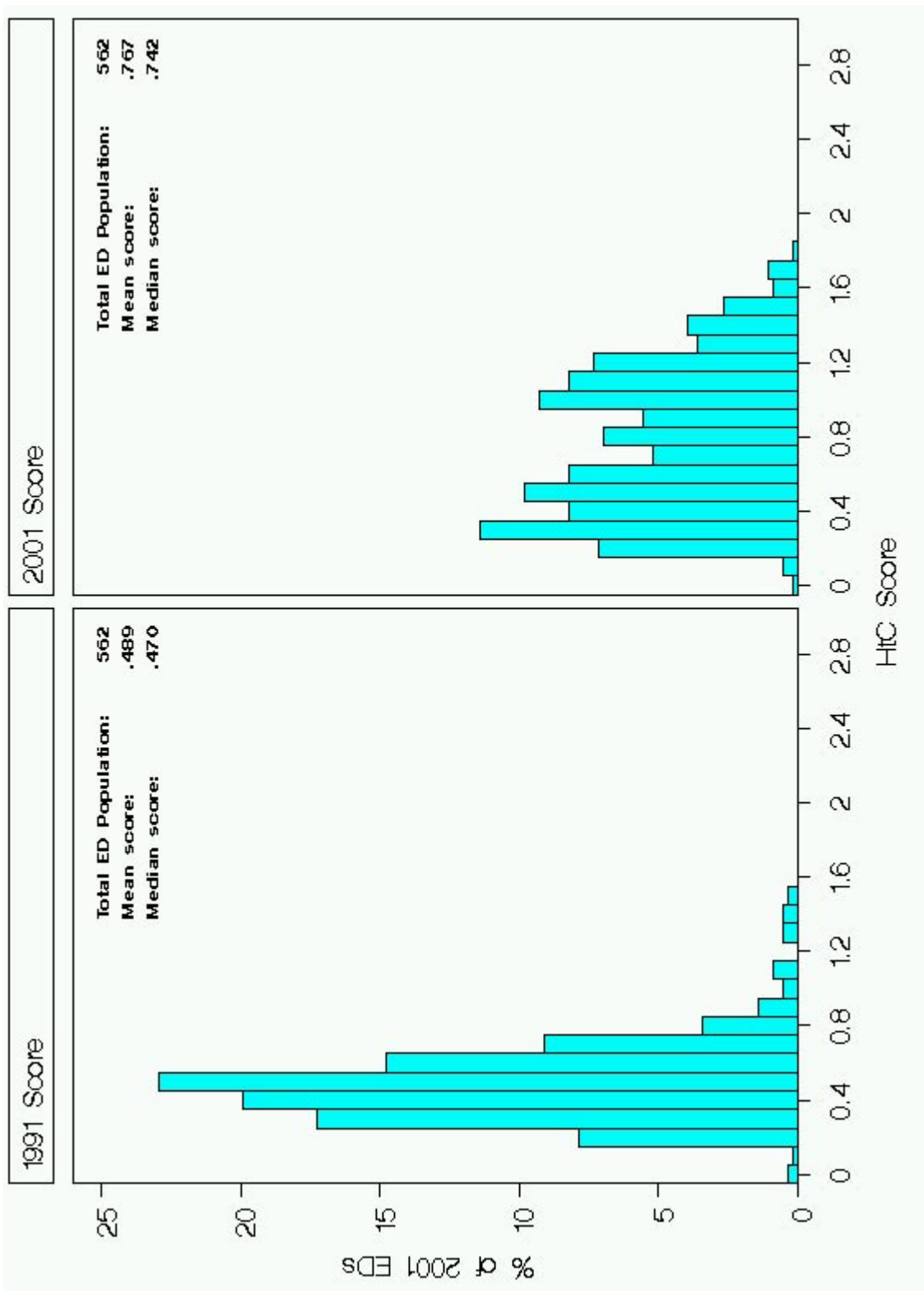
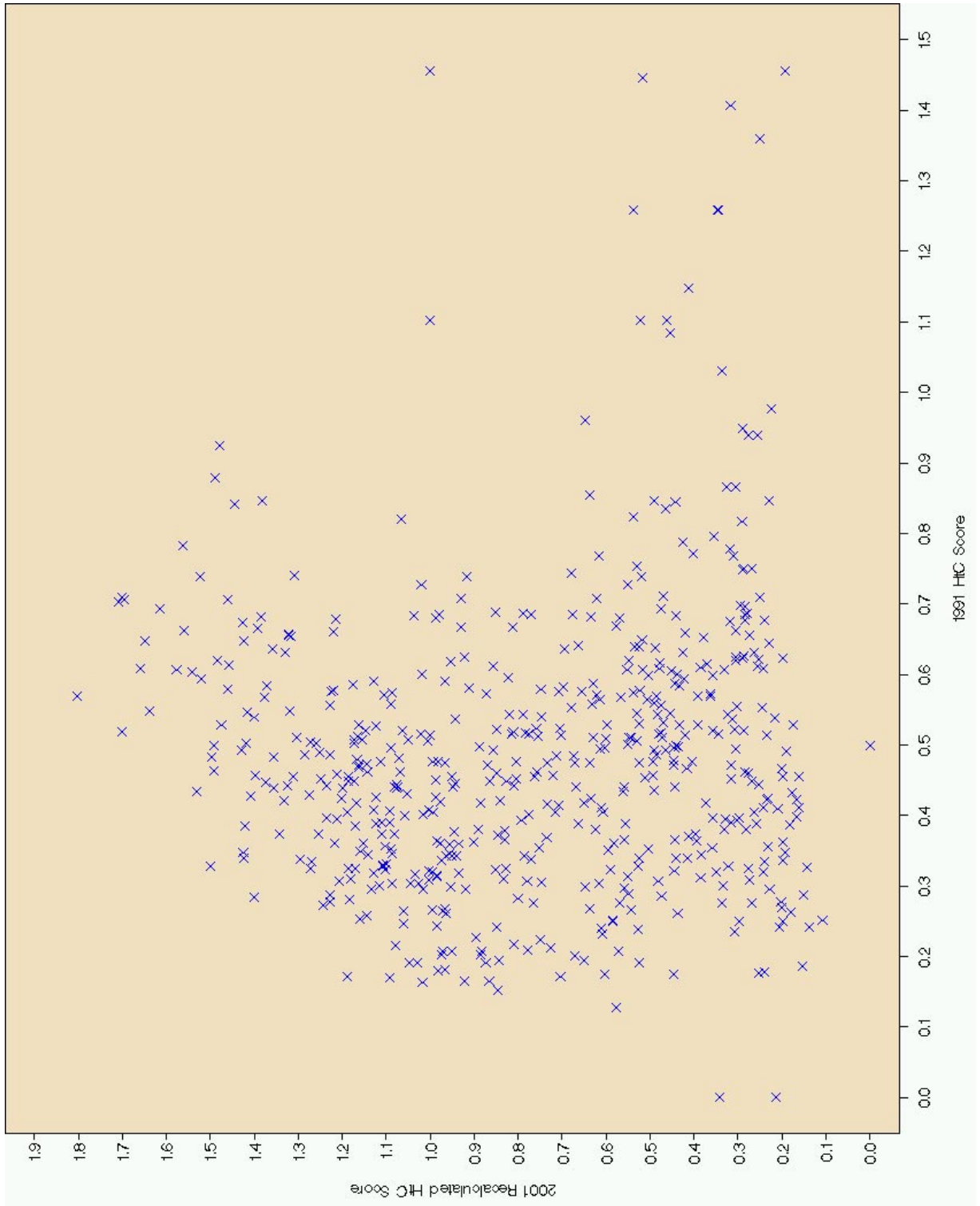


Figure 3.3.3
Scatter plot of 1991 and 2001 Hard to Count scores for each Enumeration District in Southwark



4 Population definitions

The 2001 Census was conducted on a usual residence base, that is, people were asked to fill in details on a form at their place of usual residence.

It is likely that areas which have high numbers of mobile people where usual address is not easy to define will be difficult to count. People with second homes, students who live at different addresses during term time and holidays fit this category, as do many members of the armed forces who are often moved from base to base. Quantifying the numbers of people that are wrongly missed off forms for this reason is very difficult, due to the lack of evidence that is available - the 2001 Census was conducted solely on a usual residence base, with little information on visitors collected, therefore comparisons with figures collected on a different base cannot be made.

Students

A quality assurance of student estimates was carried out as part of the ONC process. More information can be found in the following links:

www.statistics.gov.uk/census2001/pdfs/students.pdf

www.statistics.gov.uk/census2001/pdfs/onc_key_findings.pdf (under section 4.2)

The number of students that live in Southwark during vacation time but were living away in term-time was 1,465. This is ranked 346th among all LAs, so is not particularly high.

An analysis on student numbers based on comparisons with census figures and council tax records has been conducted. This analysis highlights areas that have a large difference between these two figures.

This analysis shows that Southwark (shown below) is ranked 6th highest of the 376 LAs in terms of the difference between student houses in council tax records and student properties counted in the Census. The second column shows households which are occupied solely by students and therefore gaining an exemption from council tax. There will also be many households in Southwark which contain students living with non-students that will not appear in these figures.

Table 4.1

Comparison of students in the Census and Council Tax records

Southwark	Total Exemption Class N (CTB1 Student Exemption)	2001 Census - All Student Properties	Difference
	2,014	866	1,148

5 Processing

As the data was processed, checks were carried out to ensure that the delivery of the data for the Estimation Area data was complete. The data was checked to ensure that the Census Household form numbers were in sequence with no unexpected gaps. The identities of the missed forms were then compared to other data and information (Enumerator Record Books (ERB), enumerator-completed completed summary forms and the geography database). If Household forms had been missed for a valid reason (ie non residential, demolished, derelict or 'Late returns' (ie returned after the processing cut off date) these were excluded from the analysis. If, however, the absence could not be explained the details of the individual records were recorded and summary reports were produced for each Estimation Area (EA), recording the number of forms missed.

The main reasons for forms not being processed were:

- Forms were not received from the field or received too late.
- Forms lost in the system during system crashes.
- Some forms misrecognised and therefore delivered as part of a different EA/LA

For the Lambeth and Southwark EA as a whole, there were 3,390 unprocessed forms, 1,392 of which were late forms. In Southwark LA there were 1,518 unprocessed forms distributed over a number of EDs. This is ranked 1st in terms of the number of unprocessed forms in each LA. Analysis was carried out to identify where the unprocessed forms were in relation to the CCS, to see whether the ONC could have made an appropriate adjustment. The evidence suggests that it is unlikely that the CCS would have picked up this source of error. There were 5 EDs with 25 or more forms in a block missing. This is a very high number of unprocessed forms and gives cause for concern.

Figure 5.1 on page 26 shows the number of CCS postcodes for Southwark, by ED, and the location of the CCS postcodes. This map shows that many of the EDs with a high number of unprocessed forms did not contain CCS postcodes, and therefore an adjustment may need to be made for unprocessed forms.

6 Other - communal establishments

This section contains information on communal establishments (such as residential homes, student halls).

The 2001 Census shows that Southwark has 5,185 persons living within 129 Communal Establishments. There is no evidence of problems with enumeration of Communal Establishments in Southwark.

7 Other - administrative sources

The ONC Quality Assurance (QA) process involved comparisons with various administrative data sources, as detailed in the individual QA information packs for each Local Authority.

Further analysis of administrative sources was completed as part of the LA Studies program. The administrative sources used, both by the ONC QA processes and the LA Studies, are detailed below.

Council Tax - Council Tax dwelling counts for 2001 were compared with the Census results for each Local Authority. More detailed comparisons of Council Tax counts of occupied and vacant dwellings with the Census results were also completed.

Electoral Roll - The revised 2001 Mid -Year Estimates (MYEs) for people aged 18 years and over were compared with the Electoral Roll data.

Patient Register - The revised 2001 MYEs have been compared with the NHS Patient Register data (adjusted and unadjusted) for 2001.

Pensions - The revised 2001 MYEs for people aged 65 years and over have been compared with the year 2000 Department for Work and Pensions (DWP) Pensions data.

Child Benefit - The revised 2001 MYEs for people aged 0 – 14 were compared with the 2001 Child Benefits data.

Schools Census - The revised 2001 MYEs for people aged 5 – 14 years inclusive were compared with the 2001 Schools Census data.

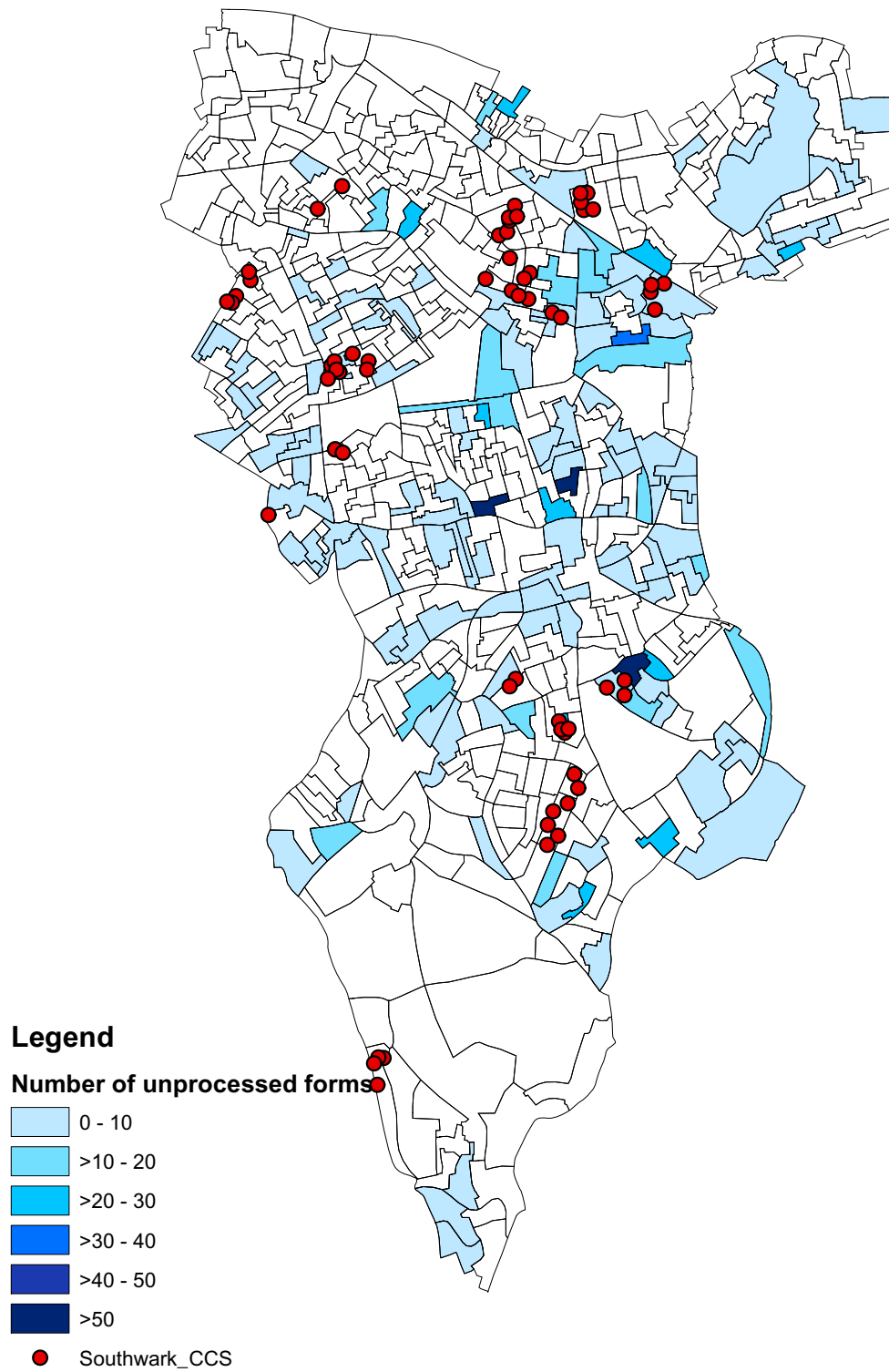
Table 7.1 compares the 2001 MYEs and Census counts, as appropriate, with each of the administrative data sources and the previous (2000) MYE. This table also gives Southwark's ranking amongst local authorities in England and Wales, for each comparison, with LAs ranked first having the largest difference between the two counts.

Table 7.1
Comparisons of Revised 2001 MYE with administrative sources

Source compared to 2001 MYE	Ranking (/376)	Source Population	Revised 2001 MYE	Source as % of MYE	National Average of source as % of MYE
2000 MYE	364	238,709	251,100	95.1%	101.3%
2001 Electoral Roll data (18+ Only)	365	171,122	196,100	87.3%	98.5%
2000 Pensions data (65+ Only)	359	24,199	25,400	95.3%	98.5%
2001 School Census Data (5-14 Only)	68	31,345	29,600	105.9%	100.6%
2001 Child Benefit Data (0-14 Only)	60	47,729	47,000	101.6%	99.2%
2001 Health Register Data	9	291,945	251,100	116.3%	105.0%
2001 Adjusted Patient Record Data	13	283,969	251,100	113.1%	104.4%
Source compared to Census Data	Ranking (/376)	Source Households	Census Households	Source as % of Census	National Average
2003-04 Council Tax Data	3	115,217	107,663	107.0%	101.2%

NB: The Mid Year Estimates in this table have been rounded.

Figure 5.1
Map to show the number of unprocessed forms for Southwark, by ED, with CCS postcodes



The differences between Southwark's 2001 MYE and the administrative sources shown in the table above are higher than the national average for all of the sources except for 2000 MYE, 2001 Electoral Roll and 2001 Pensions data. The difference is particularly large in comparisons with Health Register data, Adjusted Patient Record data and Council Tax data. Southwark's administrative data is at the extreme of the rankings for many of the sources, and therefore gives some cause for concern.

8 Conclusions and recommendations for adjustment

There were two issues that indicated that Southwark's population estimate had a high risk of underestimation:

- a) Analysis undertaken as part of the Local Authority Population studies has established that there is evidence to suggest that there are issues concerning enumeration and unprocessed forms. Investigations in Southwark have identified 1,518 unprocessed forms, ranking first in both absolute form numbers and as a per cent of total households. An assessment has been made of the number of forms that would not have been accounted for by the CCS, for which a revision to the population estimate is required.
- b) There is also evidence to suggest that insufficient Dummy Forms were created by Enumerators in the wards with the largest Council Tax difference (Friary, Consort and Basset), and the response rate implied by the Council Tax was significantly below the response rate calculated as part of the ONC process. Since there was no CCS sample in these wards, and there was insufficient sample in similar areas, the ONC process will not have made a robust adjustment for the LA and therefore a revision to the population estimate is required. Evidence from the ERB analysis found a high number of derelict and demolished properties in the Liddle ward. As this may explain the difference between the Council Tax and Census counts, an adjustment is not required in this ward.

9 Adjustment and methods used

Methods of adjustment were developed to address the issues that arose from the LA studies. The specification and rationale that underpin these methods are fully documented at www.statistics.gov.uk/downloads/theme_population/LAStudy_AdjustmentMethodology.pdf

The unprocessed forms methodology assessed how well the CCS catered for the unprocessed forms. By establishing how many of the unprocessed forms were in CCS postcodes, the expected number that would be found by the CCS has been calculated and an assessment made about whether the ONC will have resulted in an adjustment that corrects the census data for the unprocessed forms. The average household size of occupied households was calculated by dividing the overall population of Southwark by the number of occupied households. This gave a figure of 2.31, and multiplying this by the number of unprocessed/missing forms (1,518), gives the unprocessed forms adjustment of 3,507. This overwrites the previous adjustment made to the LA for around 90 unprocessed forms in September 2003. There are three 2001 wards, Friary, Consort and Basset, considered to be beyond the scope of the ONC to make a robust adjustment, and they have been removed from the data and considered as a separate stratum. The population of this stratum is then re-estimated based on the Council Tax count of addresses and the household size from the existing Census Ward level results. This results in an estimate for Friary, Consort and Basset wards of 9,183, 7,285 and 6,571 persons respectively, where 5,939, 4,925 and 4,480 were counted in the Census.

10 Summary of adjustments

Tables 10.1 and 10.2 show how the application of these processes altered the estimates for Southwark for each level of the Hard to Count index. Note that the population totals will not match the original census results published in 2002, as they do not include Communal Establishment residents or alterations that were made to the data following the ONC process that affect the numbers of residents (such as geographical re-coding). The following tables therefore detail the difference between the original and the recalculated ONC estimates. The Census counts published in 2002 will not be revised. These figures will be used in the calculation of the MYEs.

Table 10.1
Original ONC population Estimates by HtC group

	Census count	ONC estimate	Response Rate
Hard to count 1	n/a	n/a	n/a
Hard to count 2	n/a	n/a	n/a
Hard to count 3	182,601	239,513	76.2
Overall	182,601	239,513	76.2

Table 10.1 shows that there was no HtC 2 stratum, because in the original ONC processing it was collapsed with the HtC 3 stratum to stabilise the estimation by increasing the effective sample size.

Table 10.2
Census counts by HtC group with Friary, Consort and Barsest removed from the ONC process

	Census count	ONC estimate	Response Rate
Hard to count 1	n/a	n/a	n/a
Hard to count 2	n/a	n/a	n/a
Hard to count 3	167,257	219,519	76.2
Overall	167,257	219,519	76.2

Table 10.3 shows how the reworked ONC estimates and the ward specific estimates are brought together to provide an overall adjustment to the census results for Southwark.

Table 10.3
Components of revised population estimate and adjustment

	Persons
Original ONC estimate	239,513
ONC estimate excluding 3 wards	219,519
Adjusted Ward Estimate - Friary	9,183
Adjusted Ward Estimate - Consort	7,285
Adjusted Ward Estimate - Barsest	6,571
Unprocessed Form Adjustment	3,507
Revised ONC Population estimate	246,065
Adjustment	6,552

The size of the revision for Southwark is 6,552. This is 2.7 per cent of the Census population.

11 Quality assurance

A quality assurance of the adjustments made for each LA receiving a revision was undertaken. This QA procedure was formed of the following elements:

- Administrative data comparisons - the differences between administrative data sources and the new estimates were calculated to see whether the comparisons were plausible for adjusted areas.
- Demographic analysis - the plausibility of the new age-sex profiles was investigated.
- General results QA - checks were made on whether the adjustment had caused the LA to change from a growing area to a declining area, and whether patterns of population change looked plausible. Pattern of response rates were also analysed and compared with other similar areas.
- With regards to quality assuring the methods used for adjustment - external experts were consulted on the methodology for making the adjustments.