



ONS(ONC(SC))97/11

ONE NUMBER CENSUS STEERING COMMITTEE

Report on the Brent Census Coverage Survey (CCS) Pilot

1. The attached paper contains a report on the recent Brent Census Coverage Survey Pilot which sought to assess the practicalities of undertaking a postcode based coverage survey in 2001 as part of a One Number Census.
2. Further work is planned to:
 - a) consider optimum matching strategies and
 - b) investigate the potential of linking administrative records to the two current lists.
3. **The Steering Committee are asked to:**
 - a) **note the paper**
 - b) **provide any comments (at the forthcoming meeting or in writing by 10 December 1997) on the proposed plans for further research.**

**Lisa Buckner
Census Division
Office for National Statistics**

**Room 4200W
Segensworth Road
Titchfield
Fareham
Hampshire PO15 5RR**

November 1997

REPORT ON THE BRENT CENSUS COVERAGE SURVEY (CCS) PILOT

Faith Holland, Lisa Buckner, Ian Diamond

1. Introduction

1.1 In recent years the use of post enumeration surveys has become an important component of the Census process both in Britain and abroad. The Post Enumeration Survey (PES) generally carried out a few weeks after a Census has taken place in order to re-enumerate a sample of households. The discrepancy between the information collected from Census forms and the information obtained in the PES for individuals and households is used as the basis for estimating the level of underenumeration and the error that the Census has incurred.

1.2 Although previously, the PES in Britain had sought to evaluate both the coverage and the quality of the Census, the proposal for the 2001 Census is that the PES will focus solely on assessing the coverage. The quality of the Census will be investigated during Census Tests in 1997 and 1999.

1.3 The proposed design of the Census Coverage Survey (CCS) is such that it will be postcode based. For each county a random sample of enumeration districts will be selected. Within each of these, a fixed sample of postcodes will be chosen. The details of this procedure are addressed in paper ONS(ONC(SC))97/10. The proposal is to survey a large number of households using a short interviewer completion questionnaire.

1.4 The 1997 Census Test which took place on 15th June 1997 offered an ideal opportunity to assess the practicalities of undertaking a postcode based coverage survey.

2. The Brent Census Coverage Survey

2.1 The purpose of the Brent Census Coverage Survey pilot was:

- a) to assess the practicalities of undertaking a postcode based coverage survey;
- b) to ascertain the household and individual characteristics of non-responders in the coverage survey using associated 1997 Census Test data and 1991 Census data;
- c) to evaluate the ease with which the survey data could be matched to the Census Test data at an individual level.

2.1 The Brent CCS pilot took place during the period 21st-25th of July 1997 (Monday - Friday). Prior to the survey, four interviewers from the University of Southampton undertook a day of training at the Office for National Statistics in Titchfield. The interviewers were split into pairs:

- Pair A were assigned twelve postcodes from three enumeration districts (ED) in Brent A. These EDs were considered to be easier to enumerate and were graded B, B, and C on the hard to enumerate (HtE) scale (A being easiest and G being hardest to enumerate).
- Pair B were assigned eight postcodes from three enumeration districts in Brent B. These EDs were graded C, E and E.

Postcode Maps

2.2 Separate maps were provided for each individual postcode. In contrast to the Census Test where enumerators were supplied with a list of addresses, interviewers for the CCS pilot were only given an indication of properties which fell outside the allocated postcode area in order to maintain independence between the Census Test and the CCS pilot. This system of providing an indication of properties not to include in the enumeration was similar to the system employed in the 1991 Census. However, the construction of the postcode boundary was based on the address list used for the Census Test and so the two surveys were not wholly independent.

2.3 Another potential problem is how to deal with areas which are undergoing rapid development where the maps will be out of date. In these areas, which will usually be 'greenfield sites' of new housing, interviewers will encounter problems relating new streets onto the areas of 'greenfields' on the maps provided. In these areas it may be necessary to use specialist surveyors, or to try and make use of information collected by enumerators during the Census. However, it will probably be necessary to select the CCS areas before the Census is carried out due to the limited time available between Census and CCS. The effect of this strategy is to overestimate postcode size although this problem is of little consequence if there is consistency between the Census and CCS as to the postcode boundary.

Questionnaire

2.4 A copy of the interviewer completion questionnaire which was used in the survey can be seen in Annex A. This questionnaire consisted of three sections:

- 1) Household tenure
- 2) Individual demographic questions - name, sex, date of birth, marital status, ethnicity or ancestral origin, economic status
- 3) Household relationship matrix

2.5 Questions were asked in the same manner as they were worded in the 1997 Census Test. Care was taken to ensure that the ethnicity question was asked in areas which received this question in the Test and similarly for the question on ancestral origin. Show cards were provided to enable the interviewee to select the category of response which best described them. Interviewers were instructed to believe whatever the interviewee said, even if they believed it to be incorrect since this was most likely to be the answer which coincided to their answer on the Census Test.

Fieldwork

2.6 During the fieldwork week the pairs were expected to visit as many postcodes as possible in their given areas, and for each postcode carry out the following procedure:

- a) Using the postcode maps provided, locate and delimit the given postcode on the ground.
- b) Make a list of all addresses within the given postcode on the ***property listing sheet***.
- c) Identify and mark any addresses on the property listing sheet which are vacant or non-residential.
- d) Call at all remaining addresses and identify the households living within them.

- e) Make contact with each householder and complete the *coverage survey questionnaire* under their instruction. If the householder refuses to take part in the survey record this on the property listing sheet along with any explanation given for his/her refusal.
- f) Call back at non-contacted addresses/households as many times as possible at different times of the day during the week in order to obtain the highest possible coverage level.

Emphasis was placed on completion of postcodes rather than visiting all of the allocated postcodes in the given time.

3. Operational observations

Contact and response rates

3.1 Although some addresses were visited up to five times, contact was still not made. It was evident from neighbours that some households were away on holiday, however, this did not seem to be a major cause of non-contactability. In order to maximise response rate, it may be advisable to carry out the survey over one week (Thursday to Wednesday) and then after a break, e.g. of one or two weeks, have a 'mop-up' week where the households missed previously are contacted.

3.2 There was a significant number of households who were in when the interviewers called but did not answer the door. A few households were only willing to communicate with interviewers through entry 'phones or front doors making it almost impossible to carry out an interview with them.

3.3 Some residents felt short changed by society and were not open to discussion about the merits of the Census and refused to take part in the survey. Others, even after explanation could not understand why they needed to repeat the information that they had given in the Census Test.

3.4 In some parts of Brent B there were some incidences of language difficulty preventing residents from participating in the survey.

3.5 Response to particular questions

Tenure: this proved problematic as some respondents did not equate 'Buying with a loan' with having a mortgage and owner occupiers saw 'Owns outright' at the top of the list and gave it as their response without considering any of the other answer options.

Visitors: it was better to ask about visitors when identifying household residents in case they need to be classified as residents.

Name: some people were wary about giving this information and for some foreign names it was difficult to tell forenames and surnames apart.

Date of birth: this was the least well answered of all questions. Households where the members were unrelated, the date of birth of other residents was often unknown. Also, more importantly, many people were uncomfortable with this question as they felt that they were being asked to identify themselves too much over the doorstep.

Marital status: this question was order sensitive with respondents picking not the category which best described them but the category which described them nearest the top of the list.

Ancestry/ethnicity: the ancestry question was less well answered than the ethnicity question as it was felt that people were more used to answering ethnic group. On the doorstep it was difficult to tell into which 'other' category ('black other', 'mixed' or 'other') people's 'other' responses fell.

Economic status: Respondents who were self-employed, about to start a job and students working in the holiday were unsure of which occupational status category they fell into.

Household relationship: The major difficulty that was faced with the household relationship matrix was its format. The lines on the grid needed to be wider, the space to write in the 'other' category needed to be larger, and the rows needed to be alternately shaded to minimise 'ticking' error. It was also thought that it may be more appropriate to ask this question whilst the household members are being identified.

4. Results

Coverage

4.1 Within the time period, pair A visited ten of their postcodes and Pair B visited all eight. Contact with all identified households had been attempted up to five times at different times of day during the week.

Identification and listing of households

4.2 On the first day of the survey, interviewers located each of the postcodes on the ground and listed all properties found within the boundary, taking care to ascertain those which were non-residential, communal establishments or vacant.

4.3 In both Brent A and B there was little difference between the number of properties found by the interviewers and those located by the enumerators in the Census Test. However, there were significant differences in the number of households located, especially in Brent B, with the interviewers finding many fewer households than those found in the Census Test. The main reason for this appears to be the difficulty in determining the number of households resident in a multi-occupied building.

4.4 During the Census Test when enumerators were unable to contact householders a form was left in the letterbox. For multi-occupied buildings, which were particularly numerous in the enumeration districts surveyed in Brent B, when none of the householders in the building could be contacted, the enumerators estimated the number of resident households and left the required number of Census Test forms. The estimated number of households resident in the building was usually based on the number of doorbells, letterboxes, dustbins etc. The interviewers in the Brent CCS pilot were much more successfully at making contact with householders, especially in Brent B. From this they were able to ascertain the exact number of households living in most of the multi-occupied properties. For a number of multi-occupied properties the Census Test enumerators had overestimated the number of resident households which accounted for the discrepancy.

4.5 Problems were also encountered with property numbers. Some multi-occupied households were identified only by the position of their flat/bedsit in the house, this created difficulties when comparing properties/households located.

Contact and response rates

4.6 Questionnaires were completed for 196 households containing 479 individuals. Figure 1, shows that there was great variation between enumeration districts in the contact and response levels. Somewhat surprisingly, the higher contact and response levels were achieved in Brent B, an area which was thought to be harder to enumerate with the exception of area B3. This finding suggests that different criteria than those used previously may be more appropriate for classifying the ‘difficulty to enumerate’ index for a coverage survey area.

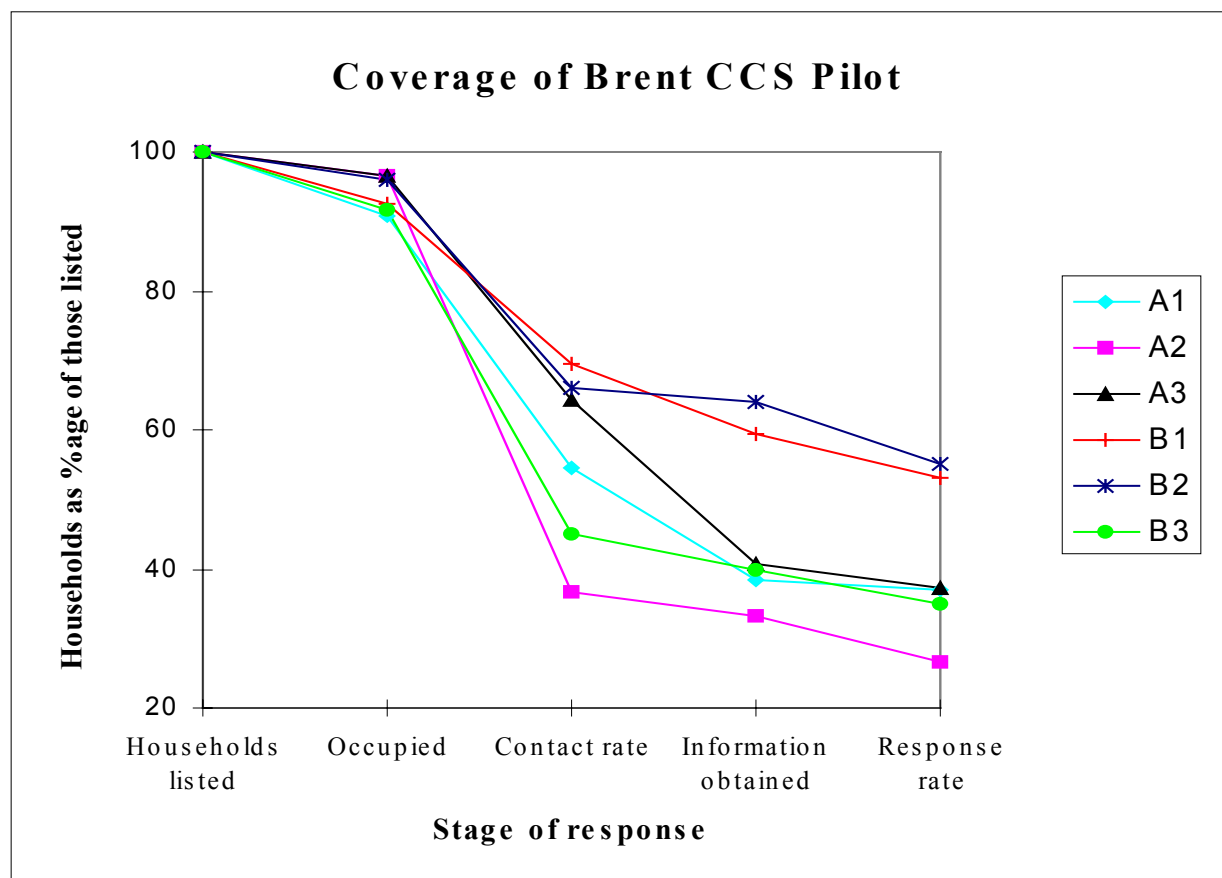


Figure 1. Contact and response levels by enumeration district

5. Comparison of respondents to the CCS, those responding to 1997 Census Test and the 1991 resident population

5.1 From the data obtained the variables of sex, date of birth, marital status, occupational status, household tenure and number of residents in the household were analysed. Ethnicity and ancestry responses were not examined as these questions were not considered to be comparable between enumeration districts. Results are presented in Annex B.

5.2 A comparison was made between the respondents of the coverage survey with those of the same areas in the Census Test to see if they contained similar respondents. Generally the

areas which had the highest response rate in the coverage survey had the lowest response rate in the Census Test, and *vice versa*. Data are presented in Table 1. Enumeration districts in which a low response rate in the Census Test was obtained were primarily those in Brent B which was piloting a post-back system as opposed to the usual collection. By comparing the profiles of the Census Test respondents with those of the coverage survey it may be possible to begin to identify survey non-respondents.

		<i>Census Test (provisional results based on enumerator response and data keyed in separately from the Census Test process) ^a</i>			Brent CCS Pilot		
ED	HtE	<i>Households listed</i>	<i>Contact rate (%) ^b</i>	<i>Response rate (%) ^b</i>	Households listed	Contact rate (%) ^b	Response rate (%) ^b
A1	B	129	98 (76)	70 (58)	130	71 (60)	48 (41)
A2	B	32	15 (47)	20 (67)	30	11 (38)	8 (28)
A3	C	61	57 (93)	35 (58)	59	38 (67)	22 (39)
B1	C	66	34 (52)	15 (23)	79	55 (75)	42 (58)
B2	E	64	30 (47)	26 (41)	100	66 (69)	55 (58)
B3	E	36	14 (39)	14 (39)	60	27 (49)	21 (38)

Table 1. Number of identified households and response rates for Brent CCS Pilot areas and corresponding Census Test areas.

^a Data presented for the surveyed postcodes only NOT complete EDs.

^b Contact and response rates are based on the number of listed households excluding those which were classed as non-residential or vacant.

5.3 The nearest data that there is to the resident population is the data from the 1991 Census obtained from SASPAC. However, in comparing the coverage survey with the 1991 Census data the following assumptions were made:

- the proportion of non-residential addresses and vacant addresses was the same in July 1997 as it was in the 1991 Census;
- the household and individual characteristic profiles of the enumeration districts was the same in July 1997 as it was in the 1991 Census.

5.4 In addition to these assumptions the enumeration district boundaries have changed since the 1991 Census therefore it was necessary to create dummy profiles of the 1997 enumeration districts by using a weighted average of the 1991 Census data.

5.5 Areas which obtained a high coverage survey response found:

- a low number of one person households, which was proportionately less than found by the Census Test and found in 1991 (in 1991 these areas had under 50% of households being single person households)
- low levels of full time employment and high unemployment (12%), but found proportionately less full time employment than was found in the Census Test (all of these areas had full time employment levels under 60% in the 1991 Census).

Conversely, in the areas which had a low coverage survey response rate there were:

- many single person households found, however again proportionately too few compared to 1991 which had over 50% of all households being single person households.
- low levels of full time employment, proportionately much lower than both the Census Test and the 1991 Census observed (1991 Census found full time employment among individuals in these areas above 60%).

6. Characteristics of respondents and non-respondents to CCS

6.1 The Census coverage survey found proportionately too few single person households and full time employees compared to both the 1997 Census Test Census and 1991 Census. Therefore it appears that full time employees were underenumerated in the CCS, this however is not surprising because in order to be enumerated residents had to have been at home during the week. Single person households also seem to have been this is possibly due to two reasons:

- Single person households, especially young single person households are less likely to be in.
- Single person households, especially if elderly and/or female may be reluctant to answer the door.

Therefore, in areas where there are low numbers of full time employees and single person households, although these two groups of people may be underenumerated the effect on the response level will not be significant. However in areas where there are many full time employees and single persons households the effect of their underenumeration on the response rate will be much greater.

6.2 Through a fairly detailed analysis of the individual and household profiles of the Census coverage survey, Census Test and 1991 Census respondents it has been possible to explain the spatial variations in the response levels of the coverage survey. It has been shown that the most difficult residents to enumerate in a coverage survey are those who live alone and/or are full time employed. Therefore in Brent A where there is a high proportion of persons displaying either one of both of these characteristics the coverage survey response level was much lower than in Brent B where there is a lower proportion of persons displaying these key characteristics.

7. Matching individual data

Estimating underenumeration

7.1 A standard method of estimating underenumeration from two independent surveys is the capture/recapture method. The method works by assuming that the proportion of persons included in the coverage survey who were also included in the Census Test is an estimate of the proportion of the total population who were enumerated in the Census. In order to calculate the proportion of the population who were enumerated by the Census the value of D, where $D = (B * C) / A$ [as independence means that $C / A = D / B$], has to be obtained as illustrated in Figure 2. From the equation it can be seen that in order to calculate D the number of individuals who were enumerated in both the Census Test and coverage survey

(A), the number of individuals who were enumerated only by the Census Test (B), and the number of individuals who were enumerated only by the coverage survey (C) need to be known.

		COVERAGE SURVEY	
		Enumerated	Not enumerated
CENSUS TEST	Enumerated	A	C
	Not enumerated	B	D

Figure 2. A simple capture/recapture model

Linkage of individual records

7.2 If every individual record from both the Census Test and the Census coverage survey were accurately recorded within a data file then the process of matching individual records and hence calculating A, B and C would be straightforward. However the 384 individual records of Census Test respondents and the 479 individual records of coverage survey respondents are far from being complete and accurate. The postcode and address number variables are considered to be recorded accurately and are complete for all individual records. The other variables recorded for each individual, day of birth, month of birth, year of birth, sex, marital status and occupational status, have many omissions and have been the subject of mis-reporting, mis-documenting and mis-inputting. Therefore the process of matching individual Census Test and coverage survey records is far from clear cut.

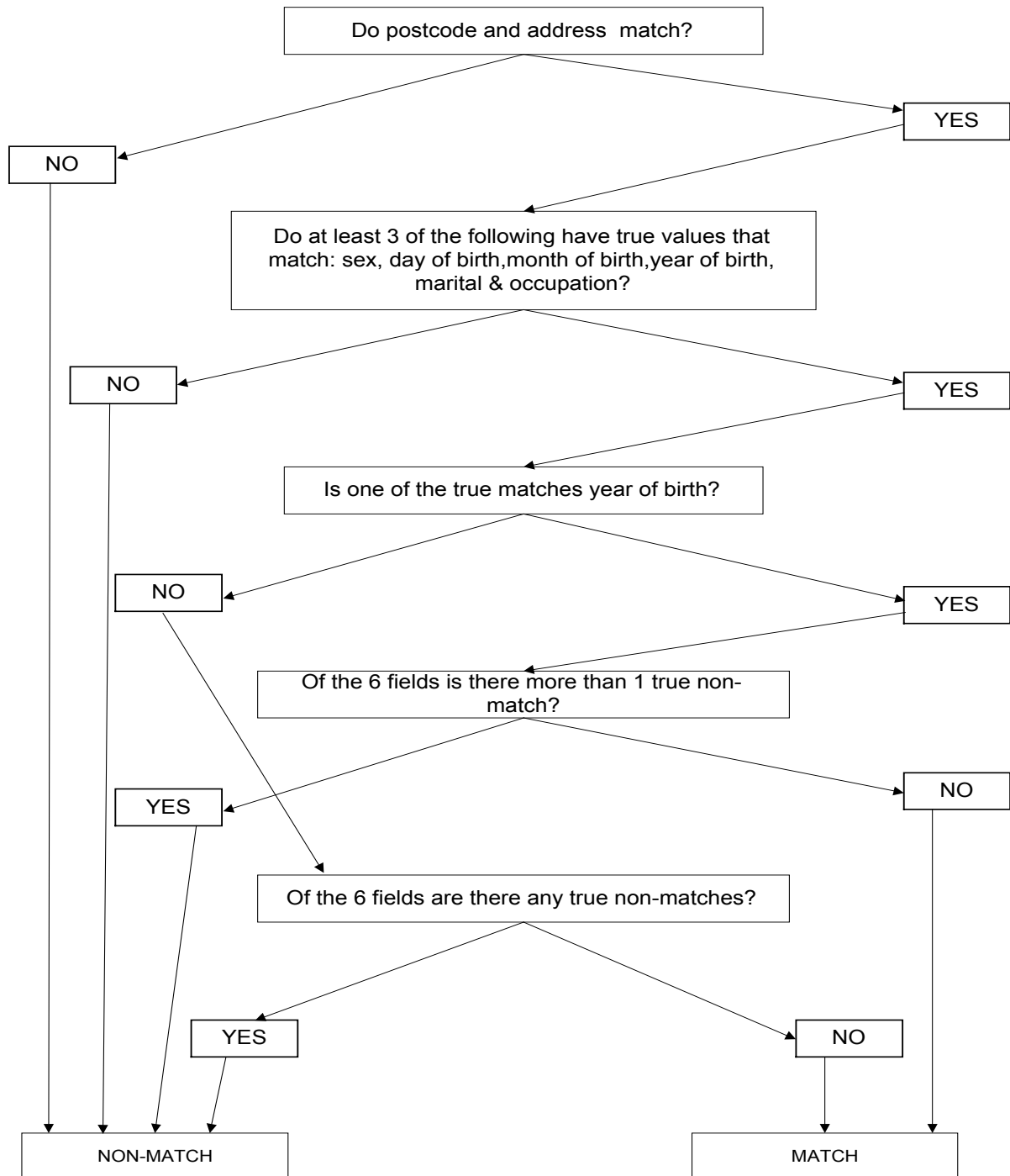
A matching strategy

7.3 In order to determine whether two records are a match, i.e. they refer to the same individual a strategy had to be formulated. In designing any strategy to determine the matching of data it is necessary to consider the relative importance of each variable in identifying an individual, e.g. is sex more important in classifying an individual than occupational status? In addition to this the relative accuracy with which each variable has been reported, recorded and inputted needs to be looked at, as well as the likelihood with which a variable contains missing values.

7.4 Having considered these points the matching strategy shown in Figure 3 was derived. It can be seen that in order for two records to be termed a match firstly their postcode and address number must be identical, these two variables were recorded the most accurately as they were the basis to identifying and labelling households in the survey. At this stage any records containing a unique postcode and address number combination were judged to be unmatchable records and were excluded from the subsequent stages of the matching process. The next stage of the matching procedure is to ascertain which of the other variables in a possible matching pair share the same true values (missing values are classified as untrue values). From Figure 3 it can be seen that year of birth plays a key role in determining whether records match. This is because year of birth has substantial discriminating powers as

it has the greatest range of values of any of the remaining variables. The final stage of the process is to sum up the number of matches achieved, the number of non-matches of Census Test records, and the number of non-matches of coverage survey records to find the values of A, B and C respectively.

Figure 3. Strategy for assessing whether cases match



8. Results of matching

8.1 Using the matching strategy described the following matches/non-matches were achieved.

		COVERAGE SURVEY	
		Enumerated	Not enumerated
CENSUS TEST	Enumerated	137	247
	Not enumerated	342	?

Figure 4. Matches and non-matches

As can be seen from Figure 4 there were many more non-matches than matches. The initial interpretation of this would be that the Census coverage survey mainly enumerated people who were missed by the Census Test. However although this may account for some of the non-matches, a proportion of these were due to the fact that there was a substantial amount of missing data within the individual records leading to a lack of matching evidence. Secondly some of the non-matches may be attributable to the high mobility of the respondents in these areas. For example, in one enumeration district five households were found who had not been living at their address at the time of the Census Test.

8.2 Therefore, the success of data matching depends not only on the formulation of a logical strategy, but more importantly on collecting and recording accurate and complete information about individuals and ensuring that the coverage survey is carried out soon after Census day to reduce discrepancies due to migration.

9. Conclusions and recommendations

9.1 The 1997 Brent pilot Census Coverage Survey has been a valuable insight both the feasibility of undertaking a postcode based Census Coverage Survey, and using the collected data, for record linkage with Census data. A series of recommendations have been drawn to improve the coverage of subsequent coverage surveys:

- In order to improve the matching of data the coverage survey needs to be carried out as soon as possible after the Census.
- It should be recognised that in areas where there is known to be both a high level of full time employment and single person households it will probably be more difficult to obtain a high response in a Census coverage survey.
- Ideally the coverage survey should run from Thursday to Wednesday to enumerate both residents who are at home during the week, and residents who are at home at weekends, (particularly important for enumerating full time employees). Also, to maximise coverage in the CCS, a 'mop-up' procedure should be instigated following the CCS to contact previously absent households.

- On the first morning of the survey whilst the properties are being listed a pre-warning postcard needs to be dropped through letter boxes explaining that an interviewer will be calling in the next few days. This should include a free 'phone number to call to answer any of their queries before the interviewer calls. This will hopefully encourage residents to answer the door to interviewers and may increase response levels by persuading people to participate in the survey through underlining the importance of the Census..
- Completeness of data is paramount for successful matching and this should be aimed for by interviewers. However, due to the reluctance of many people to give out their names and dates of birth it may be wise to only target collecting dates of birth.
- To maximise contact and response rates in the CCS it may be necessary to employ bilingual or multilingual interviewers. In the Brent CCS Pilot a bilingual interviewer carried out the survey in Brent B which was thought to have greatly enhanced the contact and response rate in this area.
- It is imperative that the interviewers are well motivated and flexible in their approach the survey. Another way of improving coverage may be to employ 'local' people as interviewers. Although there are issues of confidentiality associated with this the gain in confidence from the public may ultimately outweigh this.
- The interviewers for the pilot survey worked in teams and this proved very beneficial. This meant that it easier to deal with the problem of asking questions, writing answers and handling 'show cards' for questions with a number of possible answers. Also, working in teams meant that problems of interviewer safety were not encountered and team members could provide support for each other keeping motivation high.

9.2 In conclusion, a postcode based Census coverage survey will be practical in 2001, and if at least some of the recommendations outlined here are utilised then the survey will be one that also optimises coverage.

10. Further Research

- The Postcode boundary is created from the address list and therefore it is essential that this list is accurate as possible. Research is underway to investigate the accuracy of the address list as part of the evaluation of the 1997 Census Test.
- Collaborate with experts to improve matching techniques and consider optimum matching strategies.
- Investigate the potential of linking administrative records to the two current lists.

ANNEX A - Brent Coverage Survey Pilot Questionnaire

A1 - Property listing sheet

A2 - Household questions:

- address
- tenure

A3 - Questions on individual household members:

- name
- sex
- date of birth
- marital status
- ethnic group/ancestral origin
- economic status

A4 - Household relationship matrix and visitors

ANNEX B - Results from Brent Coverage Survey Pilot

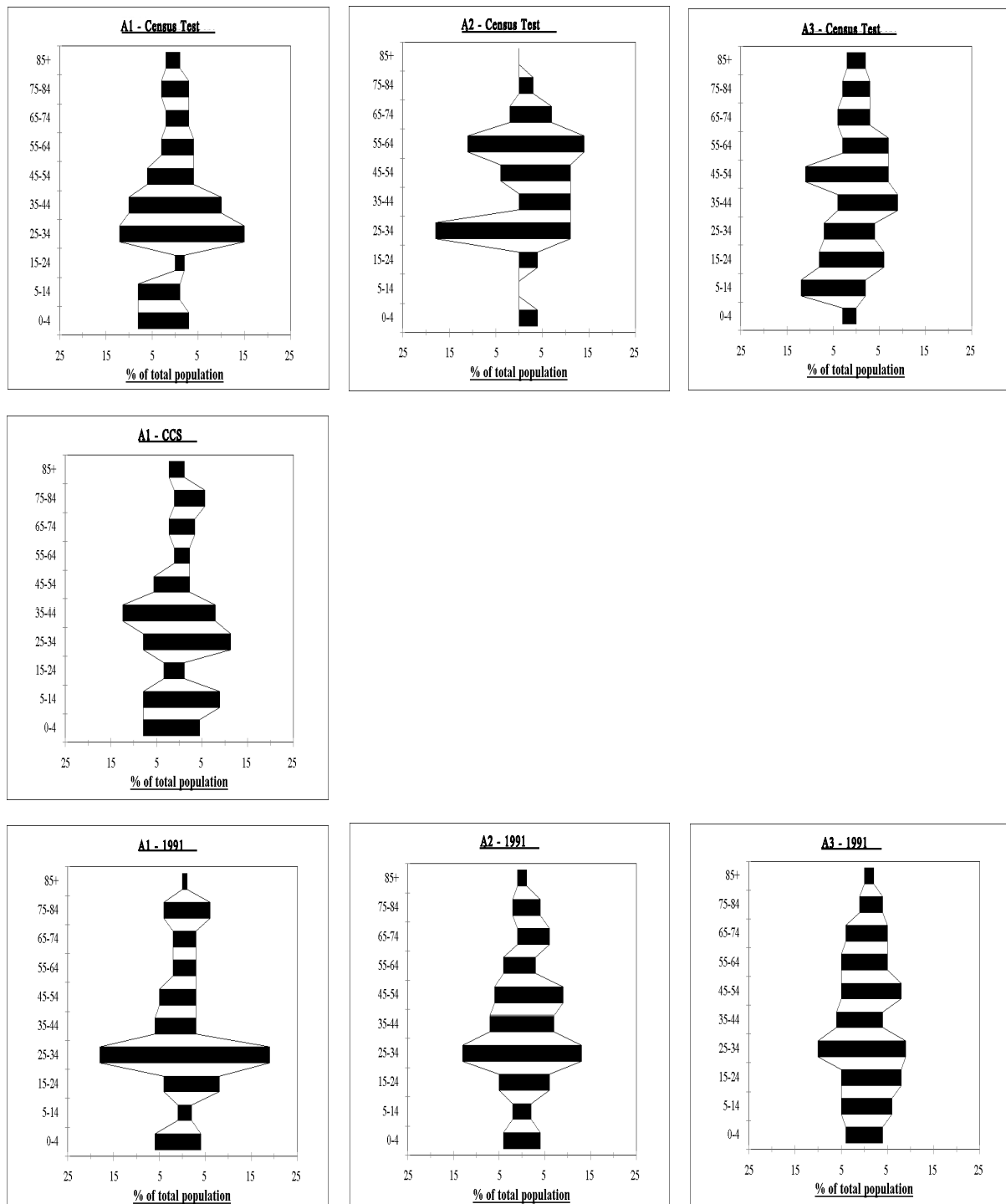


Figure B1. Population pyramids for enumeration districts in Brent A

Notes: n = number of cases (individuals) (**males - left, females - right**)

There is no graph for enumeration district A2 in the CCS due to the small number of respondents. Similarly for A3 where only a small number gave their date of birth.

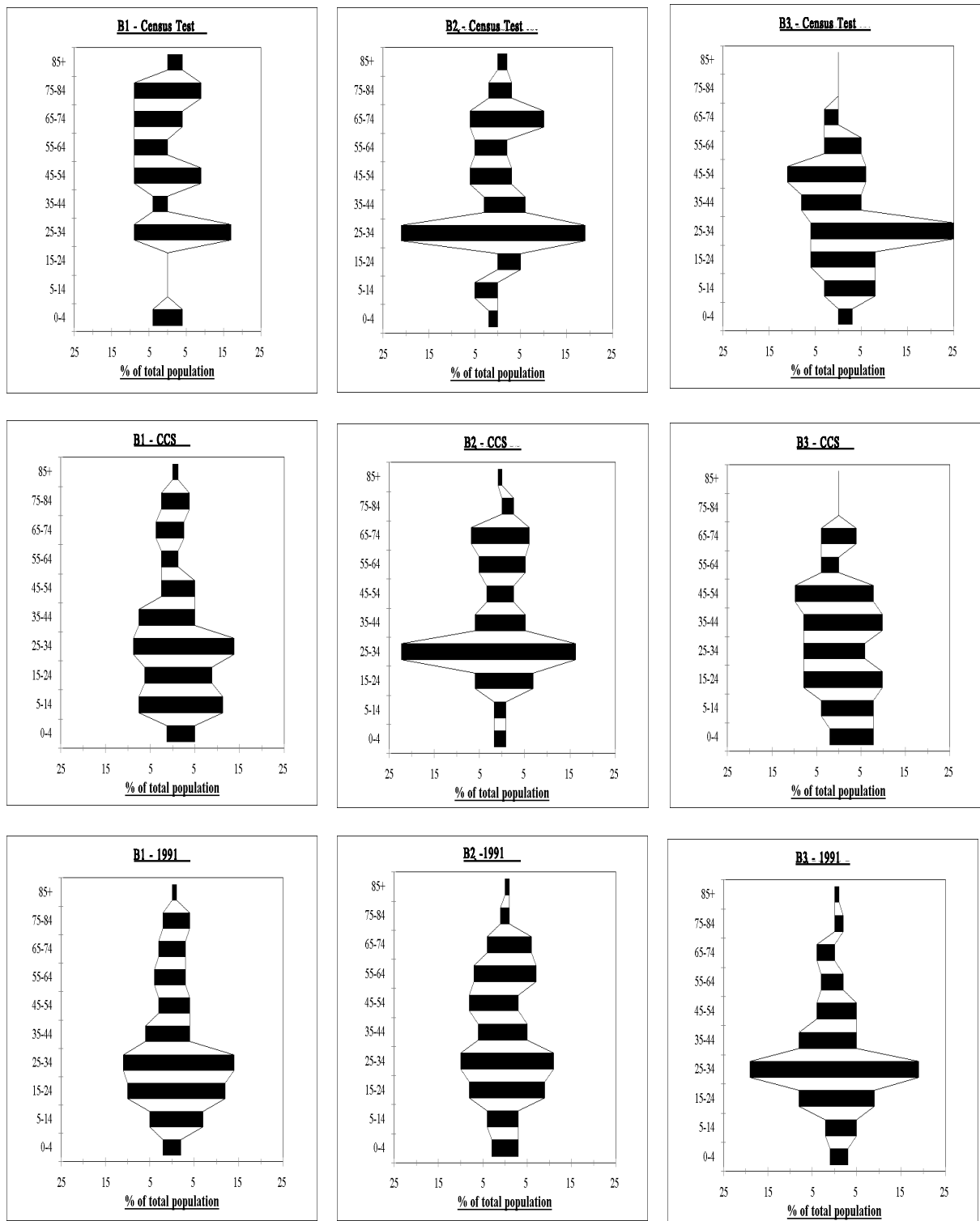


Figure B2. Population pyramids for enumeration districts in Brent B

Notes: n = number of cases (individuals) (males - left, females - right)

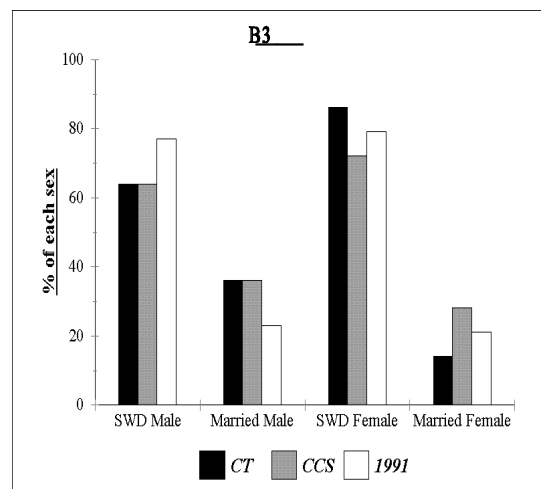
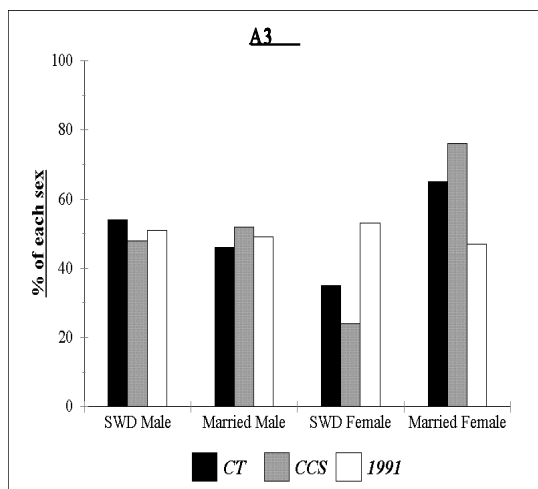
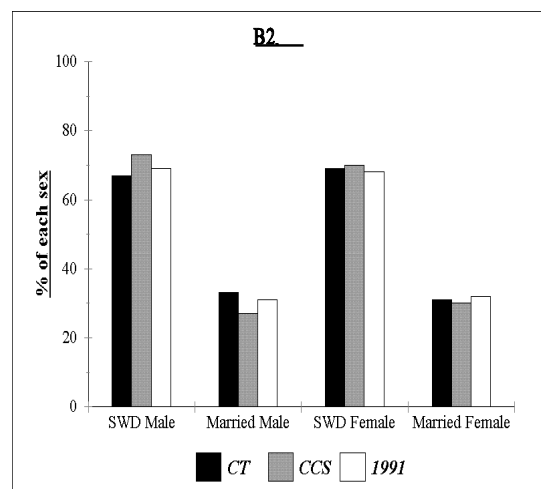
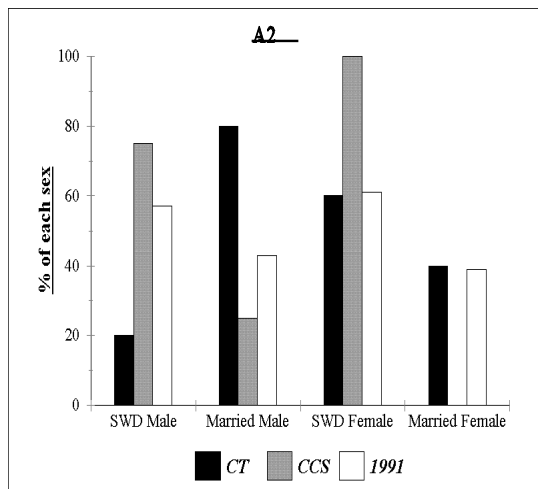
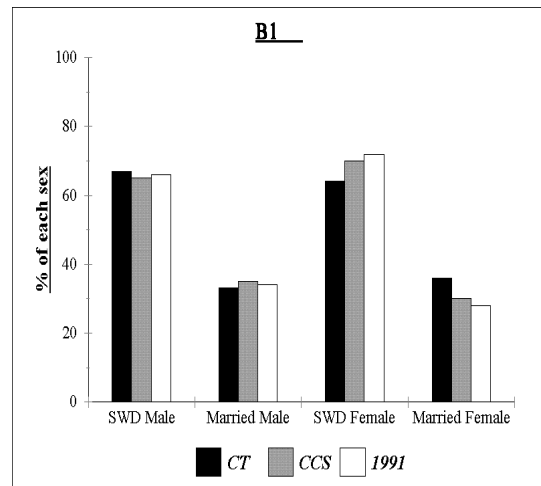
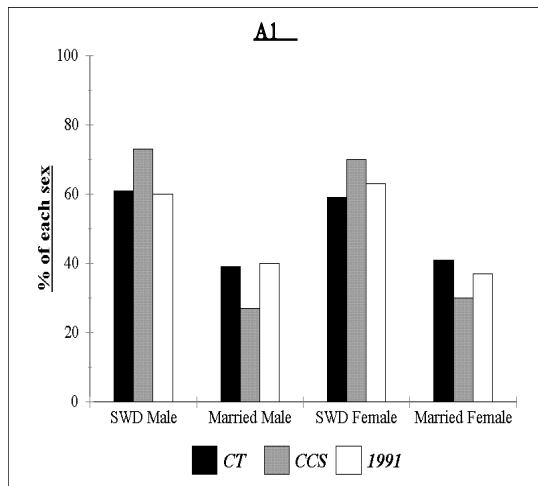


Figure B3. Marital status composition by enumeration district

Notes: n = number of cases (individuals)

SWD = single, separated and divorced, collectively referred to as single

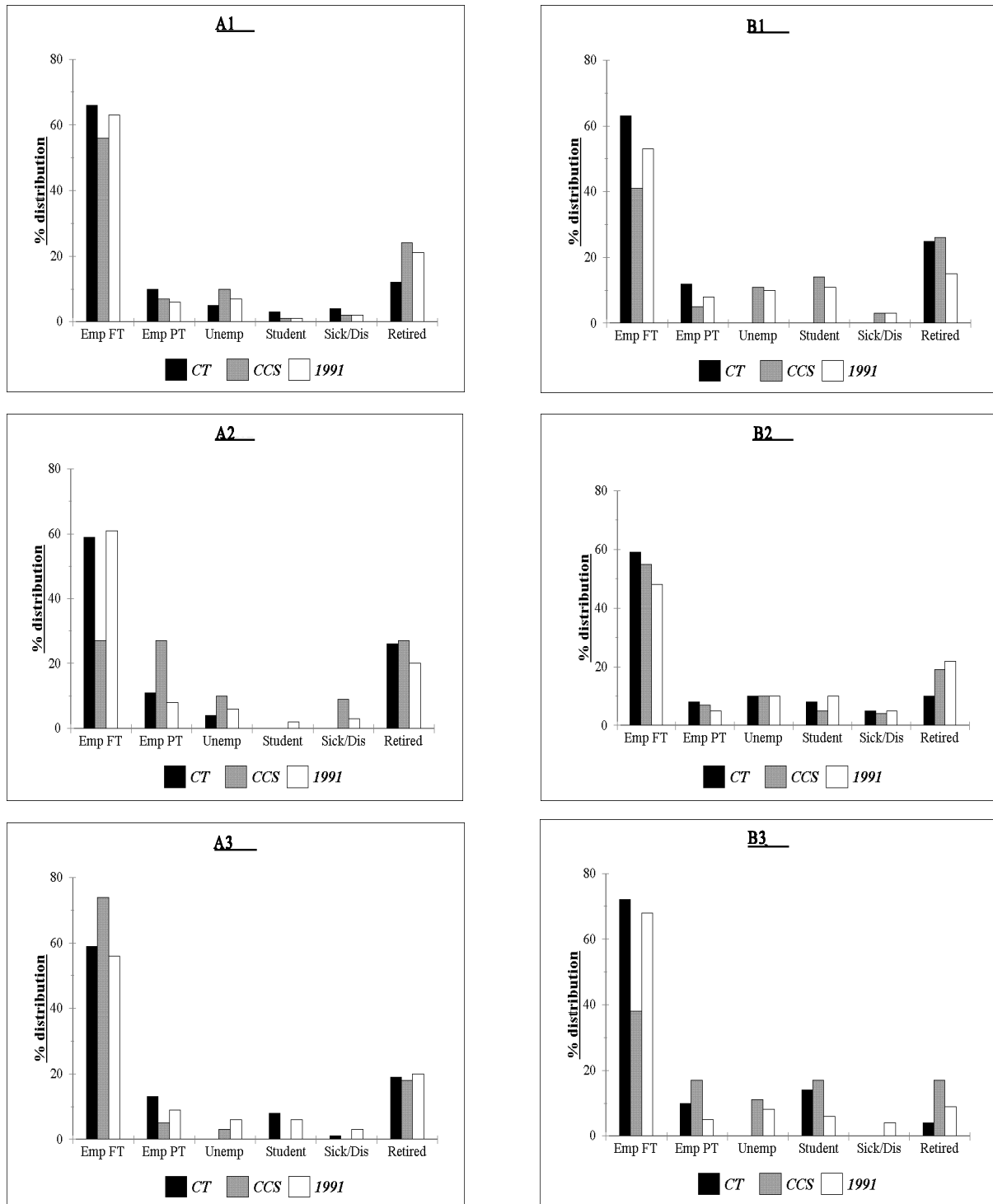


Figure B4. Occupational status by enumeration district

Note: n = number of cases (individuals)

Question on occupational status only asked about respondents aged 16 and over.

Respondents who answered 'Caring for children/house' were omitted for ease of comparison with 1991 Census data.

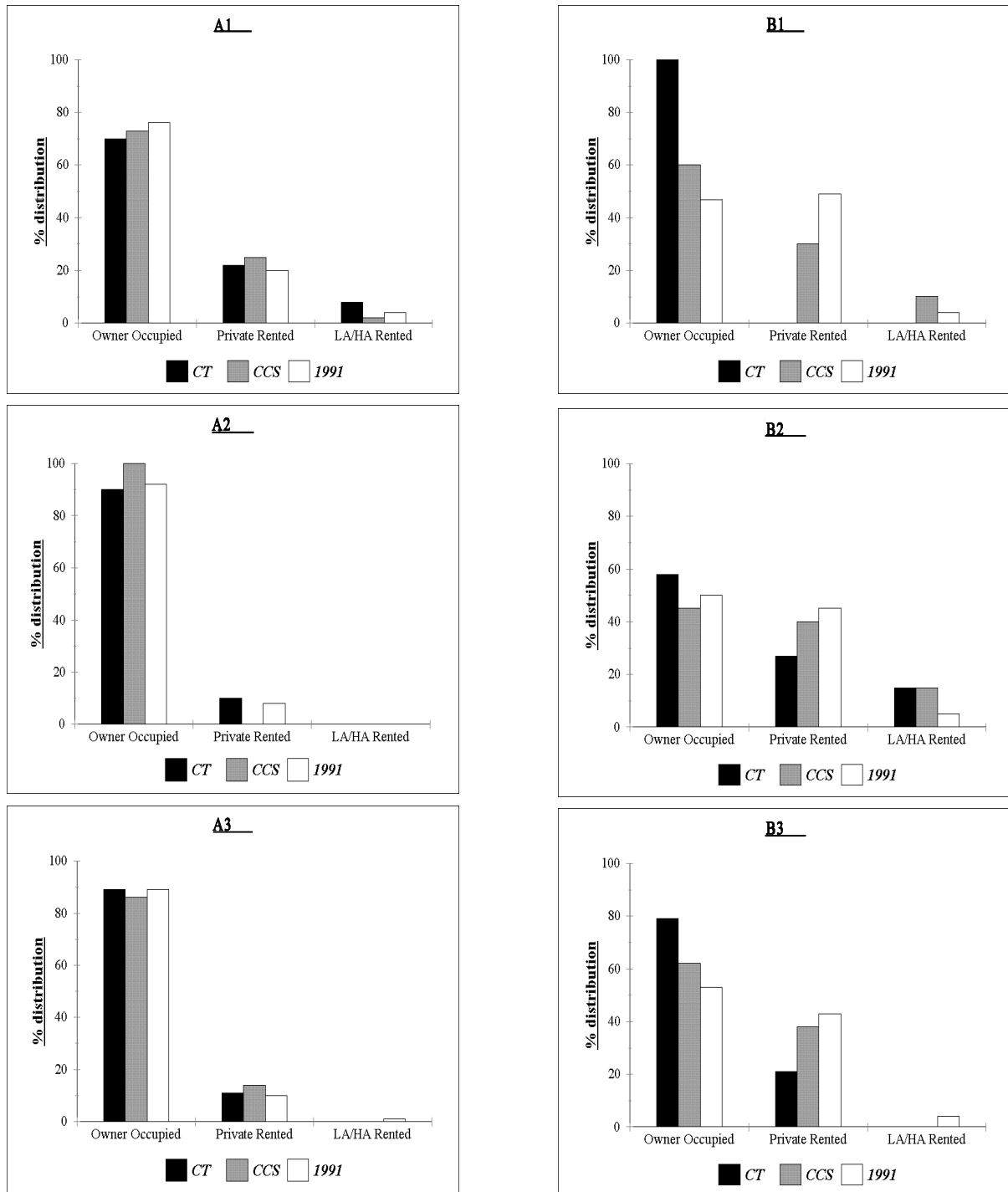


Figure B5. Tenure by enumeration district

Note: n = number of cases (individuals)

For ease of comparison with 1991 Census data those households who did not fall in to the categories: owner occupied and private rented or local authority/housing association rented were excluded from the analysis.

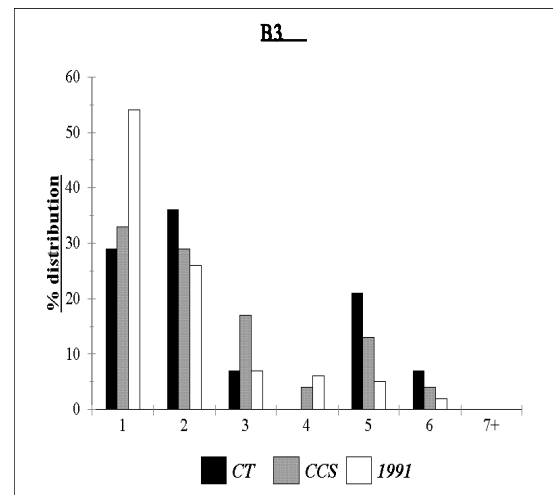
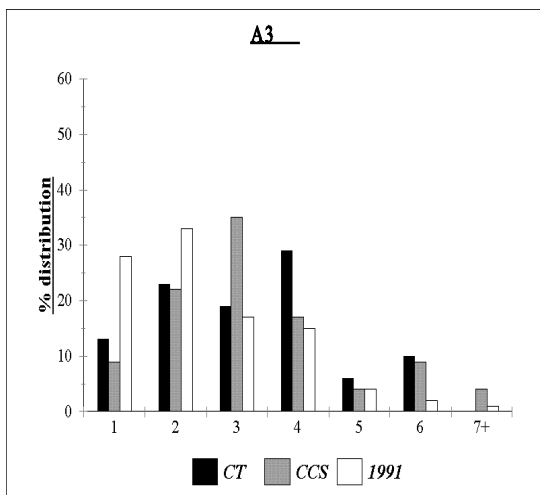
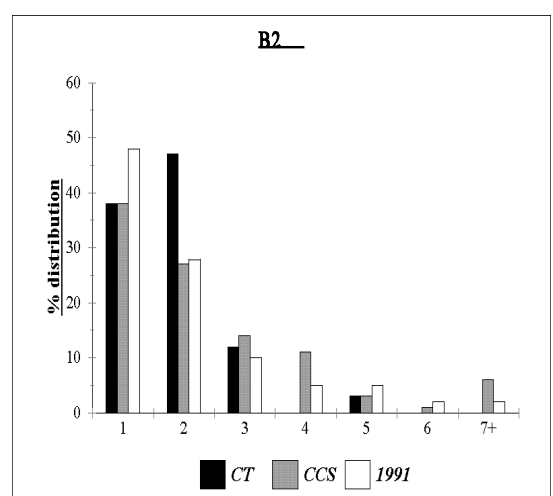
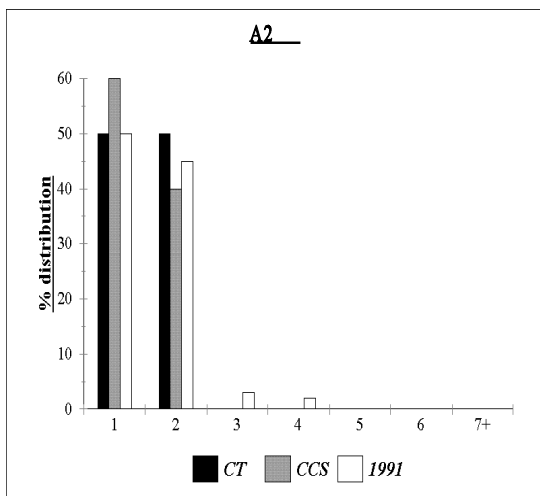
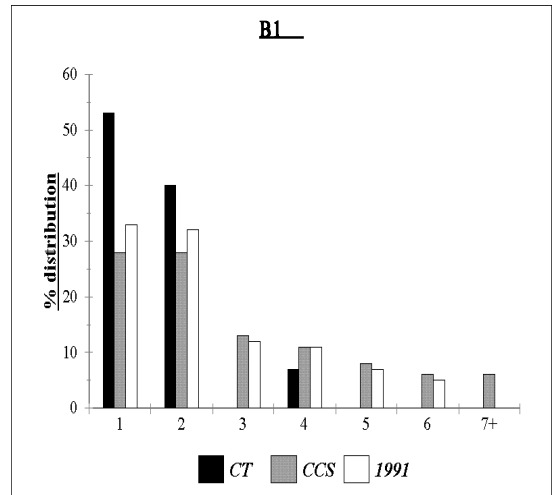
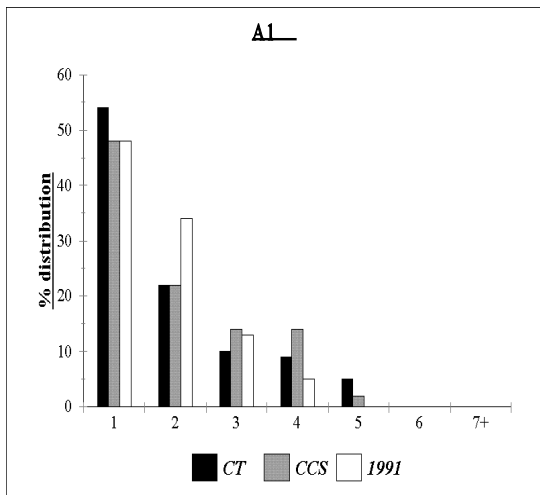


Figure B6. Number of persons in household by enumeration district

Note: n = number of cases (households)