

2001 Census: Manchester and Westminster Matching Studies Summary Report

Office for National Statistics

July 2004

2001 Census: Manchester and Westminster Matching Studies Summary Report

1. Summary

The Office for National Statistics (ONS) has now concluded a series of studies to improve population estimates in the areas that were hardest to count in the 2001 Census. This included address matching studies in two authorities, Manchester and Westminster, reported here. A parallel report [2001 Census: Local Authority Population Studies Progress Review](#) describes findings for other authorities.

Results from the 2001 Census: Manchester and Westminster Matching Studies show that the census failed to account for between 12,000 and 16,000 addresses in Manchester and between 11,600 and 14,600 addresses in Westminster.

ONS estimates the range of underenumeration in Manchester is 22,400 - 29,900 people, with a point estimate of 26,200. The estimate of the range of missed people in Westminster is 13,700 – 19,500 with a point estimate of 17,500.

Previous analysis has concluded that the One Number Census (ONC) worked well in most areas but that in exceptional circumstances it was not able to adjust sufficiently for underenumeration. The two areas where the combination of circumstances has led to most concern about the population estimates based on the 2001 Census are Manchester and Westminster. The results from detailed studies of these authorities arise from an extensive programme of research involving the two local authorities and a number of experts from within and outside ONS.

Some of the changes, especially in relation to enumeration failure, should have been taken into account in the original estimates published in September 2002 and this has been done now within these estimates. Most of the changes, however, rely on analysis that could not have been done within the time available or with the information that was known at that time. In relation to Manchester the findings largely confirm the interim results published on 4 November 2003, at the following link: http://www.statistics.gov.uk/downloads/theme_population/InterimReport_CensusMatching.pdf

These results will feed into revised mid-year population estimates (MYEs) to be published on 9 September 2004.

2. Introduction

The Census Manchester and Westminster Project was set up to investigate potential discrepancies between the administrative address lists of Manchester and Westminster City Councils, and the address list collated by the ONS for the 2001 Census. Quality Assurance of the ONC had shown that these two authorities are the ones with the largest differences. Manchester had the greatest difference between its census-based population estimates and the administrative sources used in the Quality Assurance process. Westminster had the largest population change between the 2000 MYE and the 2001 Census results. Further information about the context of this work is given in the accompanying paper [2001 Census: Local Authority Population Studies Progress Review](#).

The work of the project has been divided into several stages for each of the two Local Authorities (LAs). The first phase involved automatic matching of address lists from the Local Authority and ONS, and the second phase involved clerical matching. A third phase of work required a variety of approaches according to issues arising in each council area. This work included investigation of specific groups of addresses by reference to detailed council administrative address lists and matching by patterns of addresses in any location, termed 'extensive' matching.

In parallel with the matching work, assessment was made of the operation of the ONC in the exceptional circumstances found in Manchester and Westminster. This further informed the estimation process in these areas and provided a means of generalising to other areas, if necessary.

Once numbers of addresses had been estimated, estimates of the population living in them were produced, taking account of the numbers of people already imputed in the authority as part of the

ONC. Further information on the ONC is given at this link: <http://www.statistics.gov.uk/census2001/pdfs/oncguide.pdf> and a glossary of ONC terms and abbreviations is given at this link: <http://www.statistics.gov.uk/census2001/pdfs/oncinfopaper.pdf>

The outcome of this work is a set of census population ranges for Manchester and Westminster which will provide the basis for calculation of revised mid-year population estimates, due for publication on 9 September.

This report outlines the methods used to match the address datasets, the results of the matching, the revised ONC estimates and the calculation of census populations, and judgements surrounding the results and conclusions that can be drawn. Further details will be published in a full report on 9 September.

3. Matching

3.1 Methods

The matching process was planned in two steps: straightforward matching of consistent addresses was completed by automatic matching. More complex and inconsistent addresses were matched by a clerical team. Due to the large numbers of addresses involved, the clerical matching was undertaken on a sample of the unmatched addresses remaining after the automatic phase.

Automatic matching

Phase one of the Census Matching Project, automatic matching, was completed by Manchester Geomatics Limited (MGL) on behalf of ONS. This phase involved the automated matching of Manchester City Council's (MCC) and Westminster City Council's (WCC) address lists against the address lists of ONS for each area. The main city council address lists were council tax and electoral register lists. The ONS address lists were the enumerators address list (the list with which enumerators were sent out to enumerate the census in April 2001) and the list of additional addresses which enumerators found. All analyses requiring use of confidential census data were completed in a secure environment at ONS.

The addresses from all of these sources were matched to identify addresses that appeared in the local authority (LA) lists but not ONS's, and vice versa. Such addresses were termed 'unique' to the LA or ONS. Where there was some uncertainty about whether addresses matched, these addresses were classified as being in "suspense". This could arise if, for example, the LA list included 2 Acacia Avenue, but the census list recorded a 2a Acacia Avenue. It would not be clear whether these addresses represented the same property, whether 2a was a flat within 2 Acacia Avenue, or whether 2a was a property built between 2 and 4 Acacia Avenue. These types of addresses were put into "suspense" for clerical intervention. A small proportion of addresses from all sources were not included in the matching and classified as "unattempted" due to inadequate address details. Such addresses were included in later phases of the matching.

The automatic matching phase produced a Master Address File (MAF): an integrated list of all addresses, from all sources, remaining after the automatic matching had removed the duplicate records of identified matches.

Clerical matching concentrated on five key categories of addresses identified during phase one. These were:

LA unique addresses

ONS unique addresses

LA suspense addresses

ONS suspense addresses

Matched addresses for which no census form was received ('Matched No Form').

The final category represents addresses that appear in the LAs' address lists, and in the enumerators pre-issued address list, but for which no form was returned. These addresses contributed to the overall discrepancy between the census counts of dwellings and the LAs' address lists, as they would

not have been reflected in the published census results, since they were assumed to be either out of scope of the census (eg commercial or derelict) or not to exist.

Clerical matchers attempted to match a weighted, random sample of the unique and suspense records from phase one. This involved detailed searches for each address using a range of resources, including the full census address list, the Postcode Address File (PAF) and the enumerator's record book (ERB, a summary record of the enumeration process maintained by enumerators during the census and including information about visits and circumstances around each address on/added to the enumerator's list).

The clerical matching was designed to resolve discrepancies between the address lists. A resolved address was one that was matched to another address or one where a valid reason was identified for it not appearing on the address list. These reasons include the property being commercial, built since census day, derelict or demolished.

Extensive matching in Westminster

Due to the complexity of the housing stock and address information a further phase of matching was required for Westminster. For the initial clerical matching of addresses, the clerical staff only had access to the sample of addresses and the full Census database. At the end of this process, a substantial number of unique addresses remained on both the Westminster and ONS lists. It was clear that substantially more checks were needed to exclude further matches.

By looking at the full MAF, sorted by street order and then by property, it was possible to identify matches where, although there were differences in the exact addresses, there were the same number of flats (typically) in each source. For example, the MAF may contain basement, ground, first and second floor flats at 10 High Street as WCC Uniques, and flats 1-4 at 10 High Street as ONS Uniques. Although it was not possible to say exactly which individual flats match, both sources have 4 flats and could be matched as a group. This process was termed 'extensive' matching. By this process it was possible to resolve a further number of addresses.

Query work

Clerical matching was followed by a number of query resolution meetings between ONS and the LAs, to resolve further addresses. This work resolved addresses using the full council tax database, by identifying additional matches and address duplication between sources. Information on the council tax database was used to identify whether the property in question was known by another address, whether it was a new build or whether there were any other clues as to why a match could not be found for the address.

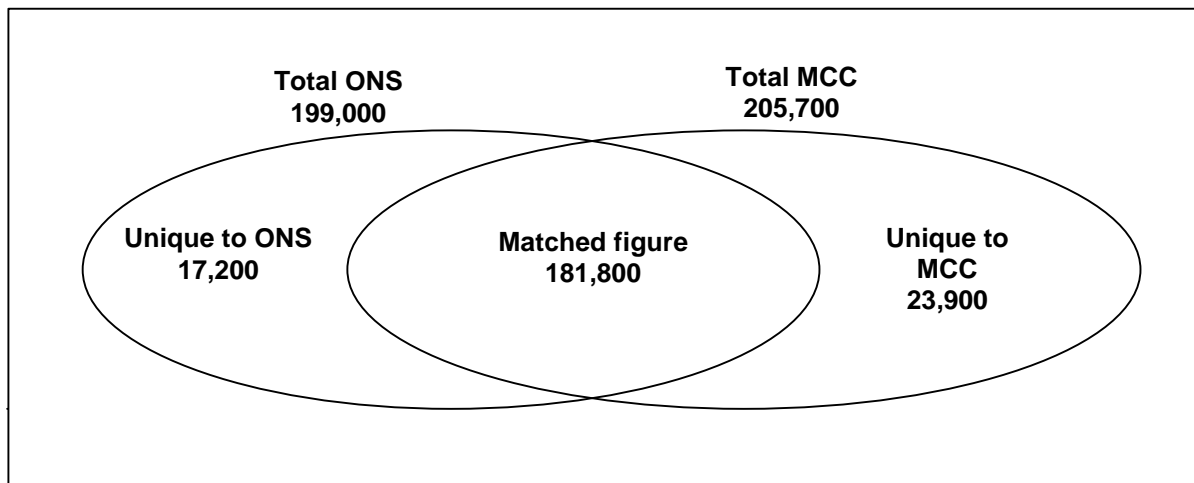
Validation fieldwork

The matching work was accompanied by several small fieldwork exercises. The purpose of this work was to confirm matches and investigate non-matches in order to validate the automatic and clerical work. Typically this involved experienced interviewers and enumerators checking addresses on the ground, using information collected at the time of the census, such as ERBs, and the MAF list. These exercises have largely confirmed the matching work, though in some cases the housing situation has changed in the three years since the census and renders comparison inconclusive. For example, if a block of flats has been demolished it cannot now be conclusively confirmed whether it was inhabited or not at the time of the census.

3.2 Results: Address matching in Manchester

Provisional results from the matching exercise for Manchester were published on 4 November 2003. The matching work estimated that the difference between the census and administrative lists is between 12,000 and 16,000 addresses. This range includes between 4,700 and 5,700 addresses which were on both census and MCC address lists for which no census form was received (Matched no Form).

The diagram below shows the results of the automatic matching stage of the project for Manchester. At the end of the phase 1 automatic matching 181,800 addresses had been matched, but significant numbers of addresses remained for investigation through clerical matching.



Note: numbers may not add exactly due to rounding.

The outcome of the clerical work on a sample of the remaining unmatched addresses resulted in an estimated number of additional addresses of between 12,000 and 16,000 in Manchester. Some of these addresses were vacant and this was taken into account in estimating the population resident in these addresses.

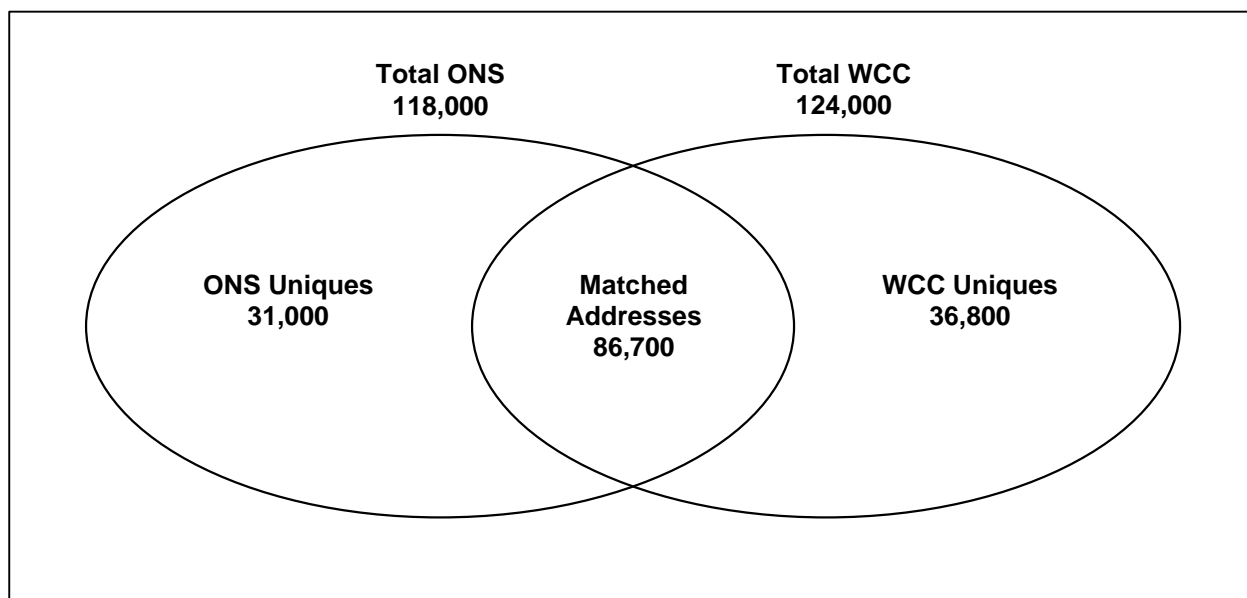
There were a number of addresses that appeared on both census and MCC address lists but for which no census form was received. Some of these addresses were discounted as missed addresses as the enumerator had recorded a reason for not collecting a form, for example that the property was derelict or had been demolished. According to council tax records the addresses were simply vacant. This gave rise to a range of missed addresses. The range is between 4,700 taking the enumerators definition of a derelict property and 5,700 taking council tax records.

Further information about missed addresses will be given in the full report.

3.3 Results: Address matching in Westminster

The matching estimated that the difference between the census and administrative lists is between 11,600 and 14,600 addresses. This range includes the 5,700 addresses, which were on both census and WCC address lists, for which no census form was received (Matched no Form).

The diagram below shows the results of the automatic matching stage of the project for Westminster. At the end of the phase 1 automatic matching 86,700 addresses had been matched, but significant numbers of addresses remained for investigation through clerical matching.



Note: numbers may not add exactly due to rounding.

The outcome of the clerical work on a sample of the remaining unmatched addresses resulted in an estimated number of additional addresses of between 11,600 and 14,600 in Westminster. Some of these addresses were vacant and this was taken into account in estimating the population resident in these addresses. It also took account of the 1,690 households that had been imputed into random postcodes by the ONC. In accordance with the agreed methodology this approach was adopted in any case where there was not an appropriate specific address in which to impute a missing household.

Further information about missed addresses will be given in the full report.

4. Converting from addresses to people

The matching work estimated the extent of discrepancies between the address lists created by the 2001 Census and those held by the City Councils. Once the number of addresses missed by the census was established, it was necessary to estimate the impact of this on the 2001 Census population estimate, taking account of the numbers of people already 'imputed' in the authority as part of the ONC.

Since missed addresses do not appear on the census database it is very difficult to identify their characteristics. No single administrative source, or combination of sources that is available gives enough information to construct a reliable picture of the households at an address. Different methods were developed to account for specific local circumstances and a judgement made by ONS on the point estimate appropriate for each authority.

4.1 Manchester

In Manchester in order to estimate the characteristics of the missed addresses the characteristics of their neighbours were used as proxies. The nearest neighbour to each MCC unique address was identified through a clerical process. The results were split into several categories:

Next door neighbour: A census household was found within 2 house numbers of the missed MCC address, in the same road. Where households are found on either side, the lower numbered address was chosen.

Near neighbour: The closest address to the missed MCC address on the same road.

No neighbour found: No neighbour was found in the same street or block of flats.

If there were a significant number of addresses for which no neighbour was found then a weighted average household size was used. Up to four neighbours were identified for each missed address in case the closest (next door) neighbour was not a good match for the missed address (for example, a bedsit identified as a neighbour for a communal establishment).

Once the neighbours of the missed MCC addresses were identified their characteristics were drawn from the Census database. This provided an average household size for the missed addresses. The people in those households provided an estimate of the age and sex characteristics of those missed.

The estimation process was divided into two stages, firstly an estimate of the number of people missed by the census, secondly an estimate of the age and sex of these people. The estimate of population was derived as follows:

Population missed by census = Missed Addresses * Ave HH size of neighbours

In order to ensure that the estimates resulting from this methodology are the best possible, plausibility was checked at every stage. For example the average household size of neighbours of MCC unique addresses was checked against an estimate of household size at the following geographical levels: postcode, enumeration district (ED) and Manchester as a whole. The estimates for postcode and ED were weighted according to the number of discrepancies identified in each area. In addition a separate exercise was carried out to test the method using samples of census addresses. This confirmed the robustness of the near neighbour methodology in Manchester. Further details will be published in the full report on 9 September.

The near neighbour method was used to calculate an average household size for missed addresses in Manchester. The provisional estimate for Manchester published in November 2003 (http://www.statistics.gov.uk/downloads/theme_population/pe_manchester.pdf) used a household size of 1.75 that included vacant properties in the calculation. A further review of this work to produce a final revised estimate for Manchester has highlighted the need for a further adjustment, for example to avoid the double counting of vacant and second homes.

The revised average household size for missed addresses in Manchester is 1.87 persons.

4.2 Westminster

In contrast to Manchester, when the near neighbour methodology was applied to the Westminster unmatched sample it produced an age-sex distribution which was implausible and for which there was no clearly identifiable explanation. A separate validation exercise showed that this method was not robust in Westminster reflecting the greater heterogeneity of small areas within the authority. As a result the method could not be confidently used for Westminster.

An alternative approach was adopted which first deducted, from the number of missed addresses, an estimate of the number of addresses with no residents (vacants and second homes) and then multiplied the missed addresses with residents by an average household size. The number of empty households was estimated as the difference between the 2001 Census count of vacant and residential households and the much higher council tax count of these households. The average household size was calculated as the average of the ward level average household size from the

2001 Census, weighted by the distribution of the ward level estimates of the number of WCC uniques and Matched no Form addresses from the matching work. The population missed was calculated as :

Population missed by census = Missed Addresses with residents* weighted Ave HH size for wards

The age-sex distribution of the missed population was taken as the published 2001 Census age-sex distribution in Westminster.

The Westminster calculation of average household size for missed addresses uses a census derived weighted average of 1.98 for occupied households.

5. One Number Census (ONC) review and re-estimation

This section describes the work undertaken to examine the ONC process in the light of the findings from the matching projects for Manchester and Westminster. This sought to identify factors that may have led to insufficient compensation for undercount in the original estimates. Further details of the ONC methodology can be found at:<http://www.statistics.gov.uk/census2001/pdfs/oncguide.pdf>

The analytical framework for assessing risk is described in [2001 Census: Local Authority Population Studies - Progress Review](#) that accompanies this report.

Manchester

The conduct of the Census

There is evidence of very poor enumeration in some parts of Manchester, although the exact reasons for this are not clear. The address lists used by enumerators did not capture the substantial redevelopment in the area, particularly in the years leading up to 2001, and enumerators may have missed addresses that did not appear on their lists. There are some wards where this problem appears to be particularly severe, stretching the ability of the ONC to make a robust adjustment for undercount.

Census Coverage Survey (CCS) sample balance and the conduct of the CCS

There is evidence that the CCS sample in Manchester was not balanced with respect to updated (2001 Census-based) hard to count scores (it was, however, well designed with respect to the 1991 scores which were the only information available before the census itself). The CCS did not include a balanced proportion of areas that had a high proportion of 'non-responding' households. This indicates that, by chance, the CCS tended to miss areas in Manchester where there was extremely poor census response. Moreover, the CCS fieldwork was incomplete in a number of postcodes that were some of the 'harder' postcodes selected.

Additionally, analysis, based on matching information from three sources (the census, the CCS and the Manchester Council Tax data) showed that the levels of dependence in Manchester are significantly higher than that used in the ONC process in that area. Further details of this analysis will be published in the full report.

The methodology for re-estimating the ONC population for Manchester relates to the unique combination of circumstances found there: in 3 wards, on account of poor enumeration, it was not possible for the ONC to make a robust adjustment; evidence has been identified of sample imbalance in the CCS; and detailed use of the council tax list has identified additional dependence between the census and CCS.

The results of this work produced an ONC estimated range of 22,000 to 23,000¹ additional people in Manchester.

¹ A different range had been included incorrectly in the version posted on the web on the morning of 8 July 2004.

Westminster

The conduct of the Census

There is evidence, both from the processed data and from the matching studies, that there were pockets of extremely high non-response in 5 wards in Westminster.

Furthermore, in Westminster there were 1,400 unprocessed forms. These were completed census forms that for various reasons were not processed. Investigation into their distribution found them to occur outside of the CCS sample. Therefore, the CCS had no information to be able to adjust for these.

CCS sample balance

The CCS sample in Westminster was balanced across the area in terms of the hard to count index and areas that were considered to be difficult. However, it did not include a balanced proportion of areas that had a high proportion of 'non-responding' households. This indicates that by chance, the CCS tended to miss areas in Westminster where there was extremely poor census response (60% or lower).

ONC Estimation issues

For the original ONC estimation Westminster was part of an Estimation Area that contained Camden, Kensington & Chelsea, the City of London and Westminster. The patterns and extent of undercount in Westminster were very different from the other LAs within the Estimation Area. Therefore Westminster was re-estimated separately.

Westminster (and the Estimation Area to which it originally belonged) had a large number of outliers identified within the ONC estimation process. At Estimation Area level, there were 31 outliers, the largest number nationally. This may not be surprising, as these areas have a large undercount and therefore there is a greater chance that observations would be higher than the pre-defined cut off points. However, this may provide an indication that the outlier strategy works less well in such areas of high underenumeration. The outlier strategy is described in:

<http://www.statistics.gov.uk/census2001/pdfs/sc0003a.pdf#page=18>

Additionally, our analysis, based on matching information from these sources (the census, the CCS and the Westminster council tax data) showed that the levels of dependence in Westminster were significantly higher than that used in the ONC process in that area. Further details of this analysis will be published in the full report.

The methodology for re-estimating the ONC population for Westminster related to the unique combination of circumstances found there: in 5 wards on account of poor enumeration it was not possible for the ONC to make a robust adjustment; evidence has been identified that there were unprocessed forms not previously taken into account in the census estimate; evidence that Westminster was very different from other authorities forming its original ONC Estimation Area; presence of large numbers of outliers and additional levels of dependence (dependency is described here: <http://www.statistics.gov.uk/census2001/pdfs/dependency.pdf>).

The results of this work produced an ONC estimated range of 16,700 to 18,300² additional people in Westminster.

6. Reaching a statistical judgement on a census estimate in Manchester and Westminster: assumptions and uncertainties.

For both Manchester and Westminster, the matching approach provides a range of missed addresses which is then converted to a range of people in missing addresses to calculate the adjustment required to the census 2001 population. The starting point is that the census point estimate should be somewhere within that range. ONS's judgement about where within the range the best census point estimates lies uses additional qualitative information about the uncertainties and assumptions around the matching and near neighbour/average household size process. The estimates from the ONC

² A different range had been included incorrectly in the version posted on the web on the morning of 8 July 2004.

approach for which we have qualitative information about uncertainties and assumptions also inform this decision.

The matching process was conducted to minimise residual uncertainty over the numbers of remaining unmatched addresses. However uncertainty remains, most notably from matching error. There are two main types of error that may have occurred in the matching work:

- Measurable error - this could include clerical matching error, or error introduced at the automated matching stage. This is most likely to be "false-negative" matching error (failure to match two addresses that in reality should be matched). This would lead to overestimation of the number of missed addresses. This type of error is identifiable through QA and was corrected when found.
- Unmeasurable error - this arises where two addresses look completely different and could not be matched clerically, or even in some cases in the field, due to the changes in the period since census day. This is also most likely to be false-negative error and would lead to over estimation of the number of missed addresses.

In addition the method for converting addresses to people is approximate and rests upon assumptions about the numbers of vacant addresses or second or other homes containing people who are not usually resident according to the census definitions, as well as the distribution of the characteristics of households and people missed. These assumptions could be in error in either direction.

Sampling errors, resulting from selecting only a sample of unmatched addresses from computer matching for follow-up, have been calculated and accounted for in the range. Extensive matching has sought to resolve cases where it is necessary to examine all the addresses in a small area to make a match.

6.1 Census Estimates

Results of the matching study for Manchester indicate a range of additional addresses of between 12,000 and 16,000. On the assumption of an average household size of 1.87 this results in an addition to the population of between 22,400 and 29,900. Whilst unresolvable uncertainty remains, ONS judges the point estimate to be the mid point of the matching range: 26,200. This is based on the strength of the near neighbour methodology for Manchester together with concern that the ONC re-estimation depended on an understanding of the patterns of underenumeration. The Manchester matching exercise did not conclusively identify why the census had performed so poorly in finding and enumerating addresses.

Results of the matching study for Westminster indicate a range of additional addresses of between 11,600 and 14,600. It is estimated that between 6,900 and 9,800 of these addresses included residents. On the assumption of an average household size of 1.98, the matching approach implies an addition to the census population of between 13,700 and 19,500. Considerable uncertainty remains and for Westminster the near neighbour methodology proved unsuitable. Other approaches were developed which relied on the ONC re-estimation. Moreover, the matching studies indicated strongly that there were particular areas within Westminster where the census estimation was poor. These were identifiable, and therefore the adjustments made within the ONC framework were judged by ONS to be reasonably robust. We judge on the basis of this evidence that the best point estimate within this range is the mid point of the ONC range: 17,500.

6.2 Links with the Mid Year Estimates (MYEs)

The 2001 MYEs will use the adjustments published above. The population estimates will also take into account the likelihood that some of the additional people identified from these studies are the same people as were identified from the Longitudinal Study adjustment made in September 2003. They will also take into account other minor amendments and corrections arising from the usual quality assurance procedures for the MYEs.

6.3 Census 2011

Through the 2001 Census: Manchester and Westminster Matching Studies and Local Authority Population Study, ONS has gained a further insight into specific problems in different areas of the country and characteristics that can be difficult to enumerate. This knowledge will be taken forward and fed into the planning and development of the 2011 Census. Further details will be included in the full report.