

Interim Report on Census Matching Project for Manchester

Introduction

The population census results were significantly different from the previously available population estimates, in several local authorities. The estimates of external migration flows between censuses were regarded as the weakest part of the whole population estimates system, over the full ten year period. In most cases, this was because of the ability to measure departures, but in some cases too large a share of incoming migrants had been allocated, particularly to the inner London boroughs. The two particular local authorities that were selected for more intensive study were the City of Manchester, which was the local authority which least fitted the mix of evaluation criteria, and Westminster. Westminster had a poor response rate, and the largest difference between the census and its previously estimated population. The method of evaluation that was chosen was matching, because it was the most demanding means of comparing alternative estimates, and when fully completed, the studies will give a large amount of information which will allow simpler methods of validation to be adopted. While the Manchester study has proceeded with growing reservations about the census count, that has not proven the case yet with Westminster.

The census is a count of persons, and is defined to enumerate people where they usually live. The fundamental enumeration unit for the census is the dwelling, and the households that they contain. Counting dwellings alone is quite inadequate as a measure of the population, as the mix of persons within households is not constant, and it is evolving as social groupings change. The quality of the initial address list available for any census is usually inadequate, and in the 2001 census, as with earlier censuses, a large part of the task of enumerators is to identify unlisted dwellings. These may be additional households, operating in distinct ways, in a single dwelling, with few facilities in common. In both Westminster and Manchester, the enumerators had to add many households that were not included in the initial address list. These two local authorities have unusual features about their population, which makes it difficult to generalise about person household relationships. As a consequence, the concluding stages of the two matching studies will involve some fieldwork enabling on the ground comparisons of dwellings, that has not yet been possible. This work will be completed before the studies are finalised.

The first goal of the Census Matching Project was to investigate potential discrepancies between the administrative address lists of Manchester and Westminster City Councils, and the address list collected by the Office for National Statistics (ONS) for the 2001 Census. The work has been divided into two stages for each of the Local Authorities. The first phase involves automatic matching of address lists from the Local Authority and ONS, and the second phase involves further clerical matching. Both phases of this very demanding and key stage is now complete for Manchester, and the results of this have been used to revise the population estimates for Manchester, released today. These estimates may be modified when the study is fully completed, but any revisions that are required as a consequence of this work will be included the next time the population estimates for England and Wales are revised, in August 2004.

This report outlines the method used to match all the Manchester address datasets, the results of the matching, issues surrounding the results and conclusions that can be drawn from the project at this stage.

Method

The automatic matching phase of the Census Matching Project has been undertaken by Manchester Geomatics Limited (MGL) on behalf of ONS. This phase involved the automated matching of Manchester City Council's (MCC) address lists with the address lists of ONS. The Manchester City Council's address lists included the following:

- Council Tax
- Electoral Register
- Registered Social Landlords
- Communal Establishments
- Council Housing

The ONS address lists were as follows:

- Enumerators address list - the list with which enumerators were sent into the field
- Additional addresses - addresses that enumerators collected in the field, in addition to their pre-prepared lists

No information other than addresses was used in the matching.

The addresses from all of these sources were matched by MGL to identify addresses that appeared in Manchester City Council's lists but not ONS's, and vice versa. Such addresses were termed 'unique' to MCC or ONS. A small proportion of addresses from both MCC and ONS were not included in the matching and classified as "unattempted" due to inadequate address details.

Phase two of the matching project, carried out by ONS to ensure that confidentiality of Census information was protected, concentrated on five key categories of addresses identified during phase one. These were as follows:

- MCC unique addresses
- ONS unique addresses
- MCC suspense addresses: addresses where there was uncertainty about any match
- ONS suspense addresses
- Matched addresses for which no Census form was received.

The final category represents addresses which appear in MCC's address list, and in the enumerators pre-planned address list, but for which no Census form was returned. These addresses contributed to the overall discrepancy between the Census results and MCC's address lists, as they would not have been reflected in the published Census results.

Clerical matchers attempted to match the unique and suspense records from phase one. This involved detailed searches for each address utilising a range of sources.

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

Clerical matching is very resource intensive so samples of each of the five categories of addresses, mentioned above, were matched. The samples were representative and sufficient addresses were selected to ensure a 95% confidence interval of less than 2%.

Results

The results of the matching project for Manchester were calculated by taking the unresolved proportion of each sample of addresses and applying this to the population from which the sample was drawn. This calculation was made for each of the five key categories of addresses outlined in the method section above. This enabled an estimate to be made of the number of addresses that MCC has on its lists which do not appear on the Census and vice versa.

MCC were able to assist in remedying any spurious results using their detailed Council Tax and Electoral Register records. This included investigating high numbers of derelict or demolished properties recorded by the Census, and the status of MCC addresses which appeared only on the Electoral Register. It was assumed that many of these were built after Census day and therefore should not contribute to the total number of MCC addresses not appearing on the Census. This assumption was made because the Electoral Register used in the exercise was taken 18 months after Census day but this needed verification. By working together with MCC and their detailed records it was possible to identify which addresses were built after Census day. Census enumerators had occasionally identified a property as derelict, although it still appeared on MCC's Council Tax register. This situation arose because of differing definitions over what was a derelict property. Differences in interpretation of this issue in particular necessitated an estimate of a range for missed addresses.

The estimated range of addresses that appear on MCC's address lists but not on the Census is as follows:

12,000 - 16,000: Mid-point of the range - 14,000

The estimate of addresses that appear on the Census address list, but not MCC's address lists is:

2,000

The mid-point of the range has been adopted as the final estimate of addresses missed by the Census. This decision was taken as the effect of the issues creating the range, for example derelict and demolished properties, appeared to be evenly balanced.

The best estimate of addresses in Manchester missed by the Census according to the Census Matching Project is 14,000.

Turning an estimate of the number of households into a population estimate

The Census Matching Project used households neighbouring missed addresses in order to estimate the size and characteristics of the missed population. For each address identified as having been missed by the Census, neighbouring households were identified from the Census. These neighbouring households were used to construct an average household size for the missed addresses. By multiplying the average household size by the estimated number of missed addresses an estimate was produced of the missed population. The characteristics of the neighbouring households were then applied to this estimate. Several neighbours were identified for each of the missed addresses, and ranked by their proximity to the address in question. These were used to quality assure the average household size and characteristics.

Once the estimate of population missed, by sex and single year of age, is derived this is then 'aged on' by 9 weeks to reflect the time difference between Census day (29 April 2001) and mid-year (30 June 2001).

The final stage in deriving 2001 population estimates is to take account of any overlap between Census Matching and the revisions to 2001 population estimates published on 26 September 2003¹.

The main component of the 26 September revision was to add men who were not measured as usually resident by the Census. For men aged 25 to 34, the revisions published on 26 September have been retained, these added more men than those identified as missing by the Census Matching. For men aged 35 to 49, the addition due to Census Matching is used, this adds more men than were identified as missing by the 26th September revision. (The 26 September revision was made for males age 25 to 49 only.)

The effect of the Census Matching exercise has also been carried forward into the 2002 estimates. In large part this is effectively just changing the 2001 base used in the 2002 calculations and therefore the same additions that have been made to 2001 estimates can be added to 2002 estimates as well, after ageing on a further year. However there is a complicating factor - the adjustment for unattributed population change. Since the matching exercise adds in significant numbers to the 2001 population estimate the amount of unattributable change for Manchester is substantially reduced. Thus, there is a consequent reduction to the size of the adjustment for unattributable population change required. The calculations for this adjustment for Manchester have been reworked to reflect the new base and the results are reflected in the 2002 estimate.

¹ For more information see:

http://www.statistics.gov.uk/about/methodology_by_theme/revisions_to_population_estimates/default.asp

Issues:

The Census Address Matching study for Manchester is an evidence based project which has provided a best estimate of the number of addresses missed in Manchester by the Census. There are several issues surrounding this estimate which could have an impact. These are dealt with below.

The Matching project will have been subject to matching error. The extent of this error is very difficult to measure. There are two types of matching error, false negative (or missed matches) and false positive matches (or erroneously made matches). These errors could result in two different ways, clerical mistakes or underlying error. Clerical mistakes can be measured by quality assuring the work of the matchers. This suggests that false negative errors caused by clerical mistakes run at around 1%, and false positive errors are very rare. The underlying matching error is far more difficult to measure. This relates to addresses that would never be matched clerically but represent the same entity in reality. An example would be 12a Acacia Avenue being the same dwelling in reality as Top Flat, Kings Court, Didsbury. Resolving these sort of issues would normally require some form of fieldwork, which has not been possible for this project to date.

The results shown in the previous section are based on samples. They are therefore subject to sample error. Sufficient records were sampled to ensure that the standard error (SE) for each sample was ≤ 1 . This has resulted in the confidence interval for each sample group being below 2%. Whilst the sampling error has been kept to the minimum level practicable for this project, it should be noted that the cumulative effect of sampling error on each of the five samples could have a positive or negative effect on the result.

It is difficult to judge whether a unique MCC record is actually a unique address. Addresses from different sources can be recorded completely differently, and can be impossible to resolve only having recourse to the address lists. The assumption has been made that all MCC unique addresses are indeed unique.

The matching has identified around 14,000 missed MCC addresses and around 2,300 unique ONS addresses. It would be very difficult to identify the extent which the ONS unique addresses offset the number of MCC uniques. It would seem implausible to suggest that there is no crossover between the two unique address sets, however it would require fieldwork to properly assess the effect of this.

The One Number Census attempted to estimate and impute for the number of un-enumerated households, NOT addresses. ONC imputation added approximately 14,000 households to Manchester. It would seem plausible that a proportion of this adjustment should offset the estimate of 14,000 missed addresses. The issue here is that ONC imputation filled addresses which were captured by the enumeration process. The likelihood is that the extent to which the ONC imputation offsets the difference in addresses is small, however further work would be needed to establish the full effect, which will be taken into account in the population estimates published in September 2004.

An adjustment has already been made to the ONC estimates for dependency between the Census and the CCS. This was a measure of household dependence and therefore was an attempt to correct for wholly missed households. Again, this adjustment was made into existing Census addresses, instead of increasing the address list. It is likely that a proportion of this adjustment would offset the total number of MCC uniques.

There will be at least one more stage of validation before the results have the precision that ONS believes can be finally achieved. The results now available show that a sizeable adjustment should be made to the City of Manchester population estimates, and that any risk of overestimation would be offset by the most probable larger bias from making no adjustment until September 2004, by when the window of opportunity with respect to the annual local authority funding allocations for 2003 would have passed. This study would have benefited from a preliminary stage of analysis, whereby an up to date alternative population estimate for the local authority prepared independent of the existing methodology might have helped identify at risk population groups or areas for the matching study to focus on.