POPULATION AND MIGRATION STATISTICS COMMITTEE (SCOTLAND)

Comparisons between Census Test and DNA Address Lists

1. At the last PAMS meeting, PAMS members asked to see a paper produced by Sharon Meighan, comparing the number of dwellings recorded in each postcode covered in the 2006 Census Test, with those from the Definitive National Address list for Scotland (DNA-S), and the Royal Mail's Postal Address Fife (PAF). Sharon has now left GROS, but her paper is provided for information.

GROS: Household estimates and projections branch
November 2007

Comparisons between Census Test and DNA Address Lists

Contents

2. Summary

National address lists are used by the Scottish Government (SG) and GROS as a sample frame for surveys, an address list for the Census, and for producing statistics on the housing stock.

This paper contains a summary of analysis of address lists from the Definitive National Address list for Scotland (DNA-S), and the address lists used in the 2006 Scottish Census Test. The Census Test covered around 50,000 households across several local authority areas. The address lists used for the Census Test were based on the Royal Mail's Postal Address File (PAF) from October 2005, and had been checked against the actual number of dwellings on the ground, by address checkers in November/December 2005, and then by Enumerators during the Census Test in April 2006.

Also included in this paper is a comparison between the original PAF list, and the final list used for the Census Test, after the changes had been made by the address checkers and enumerators. This analysis provides some information about the usefulness of the DNA-S and PAF address lists, for GROS and SG purposes.

Overall, the quality of both PAF and DNA-S address lists appeared to be good, with no major differences in quality, which is re-assuring. No address list is likely to be perfect, and in each area a number of differences were identified between the different address lists.

The biggest problem, from the perspective of GROS and the Scottish Government, is that for the Census and household surveys it is very helpful to be able to separately identify residential and non-residential properties. However, the PAF 'small user file' also includes some small businesses, and the DNA-S address lists can contain non-residential properties (parks, car parks etc). The DNA-S project consists of a single address list for each local authority area known as the Corporate Address Gazetteer (CAG). GROS asked councils to provide a download from their CAG, and some removed non-residential properties but others included these addresses. We have been told that a residential flag will be added to addresses in the DNA-S address lists later this year, which would certainly make it more useful for users such as GROS and the SG.

3. Aim

To investigate the quality of the new Definitive National Address list for Scotland (DNA-S) by comparing it with Census Test address list and the original Royal Mail Postal Address File (PAF) from October 2005, in the areas chosen for the 2006 Census Test in Scotland (covering approximately 50,000 addresses).

4. Background

National address lists are used by the SG and GROS as a sample frame for surveys, an address list for the Census, and for providing information on the housing stock. In the past, the SG and GROS have relied on the Royal Mail's Postal Address File, PAF, for providing an address list for the Census and household surveys. In addition to this, a new national address list is being developed - the Definitive National Address list for Scotland (DNA-S). This will be officially launched in October 2007. The DNA-S project consists of a single address list for each local authority area known as Corporate Address Gazetteer (CAG), brought together to form a single 'definitive' address list for the whole of Scotland. This paper describes additional analysis, carried out by GROS, which consisted of comparing the DNA-S addresses lists from a few areas with an address list which had been checked against actual dwellings on the ground.

A Census Test was carried out in six areas of Scotland on April 23rd 2006, containing a total of around 50,000 households. This provided an excellent opportunity to investigate the quality of different address lists in these areas. The address lists used for the Census Test were based on PAF (frozen as at October 2005). This was frozen at an earlier date than would normally be chosen for a Census, so that the address list could be checked against dwellings on the ground by address checkers in November/December 2005. Further checks were carried out by enumerators in April 2006. The address checkers checked the PAF-based list against every address on the ground by eye, and amended the list with new addresses and removed non-residential, derelict, demolished properties, and addresses outside the enumeration district. Enumerators then took the address checkers' amended address list and found some further inaccuracies which had been missed by the address checkers, as well as changes that had happened in between. This provides a base address list, which could be used for comparisons with the DNA-S.

The Census Test included areas expected to be difficult to enumerate:

- Part of North Glasgow with a high number of asylum seekers;
- Part of South Glasgow with high ethnic diversity;
- An area of West Dunbartonshire with poor housing stock, deprivation and large numbers of young males – one of the hardest groups to enumerate;
- Breadalbane/Lochaber (comprising parts of Highland, Stirling, Perth & Kinross and Argyll & Bute council areas). This area was selected because of the large number of holiday homes and the presence of a number of gypsy/traveller sites.

More information on the Census Test, including maps of the areas covered, is available from the GROS website at http://www.gro-scotland.gov.uk/census/censushm2011/2006-census-test/index.html.

This paper includes an analysis of the address lists from North Glasgow, South Glasgow, West Dunbartonshire and Highland. Argyll and Bute council provided GROS with a copy of their CAG, but there were problems accessing the postcode information, so it was not possible to include analysis of this area. The areas of Perth & Kinross and Stirling included in the Census Test were considered too small to be adequately covered on their own.

5. Sources of the address lists

GROS asked the councils to provide an extract from their Corporate Address Gazetteer (CAGs), but did not specify what information was required. The information provided was very different in format and content. Glasgow City Council were able to provide just residential dwellings for the Census Test area but Highland and West Dunbartonshire Councils included non-residential addresses, including possible residential areas such as parks, car parks, land opposite etc. There was no flag to remove these non-residential addresses, so they had to be removed from the data by hand so that they could be compared to the residential areas in the Census Test. As this was done by hand, there is a possibility that some non-residential properties were missed, and so were included in this exercise.

The Census Test did not cover the entire local authority areas, so the relevant addresses were extracted from the CAG on the basis of the postcode, to allow a comparison between the two sources.

Not all the address lists were provided at the same date, and the timing of when a new address is added may differ between different council areas, and the Royal Mail. Most of the CAGs were provided from approximately the time that the Census Test was carried out, but some were later than this. In contrast, the PAF comparisons were based on the original ‘frozen’ PAF, from several months before the Census Test. The different dates can have some impact on the analysis included here, as later address lists may be more likely to reflect any new builds and demolitions during the intervening period.

The CAGs were provided 1-2 years before the DNA-S project went ‘live’ so it is likely that improvements will have been made to the CAGs since this analysis was carried out.

It is important to remember that neither PAF nor CAGs are maintained for the purpose of conducting a Census or household surveys – PAF is maintained for the purpose of delivering mail, and is a file of delivery points, and CAGs are maintained for local authorities’ benefits. Therefore, the analysis contained here is not intended as a criticism of their primary purpose – it is intended to help GROS and the SG consider their own needs for address lists. For example, the fact that address lists may include non-residential properties may be a problem for GROS/SE, as we are mainly interested in residential properties for carrying out surveys and the Census. However, this does not indicate that the address lists are not fit for Royal Mail or Local Authorities’ purposes.

6. Method of Comparing Address Sources

The address lists used in the comparisons contained just a list of addresses and postcodes. GROS staff removed any CAG addresses that were clearly non-residential, by hand.

Both the Census Test and CAG address lists were aggregated using SAS to record the number of dwellings in each postcode. Then these two datasets were combined into a single database, with the postcode and the number of dwellings recorded in each address list.

The difference between the number of dwellings per postcode in each address list was then calculated by subtracting the number of dwellings in the CAG from the number in the Census Test.

The new dataset was uploaded into ArcGIS (mapping software). Maps were produced to visually compare the difference between the two address lists to see if there were particular areas where there are more dwellings in one list than the other.

For each area, a graph was also produced which plotted the number of dwellings in each postcode, in the Census Test versus CAG, for each of the Census Test Areas. This provides a summary of how similar the address lists are, and major problems such as areas completely missed by one or the other or both, or areas with large differences. Summary tables were also produced for each Census Test Area showing the number and percentage of postcodes in each area with the same number of dwellings in each list, and those in one list and not the other.

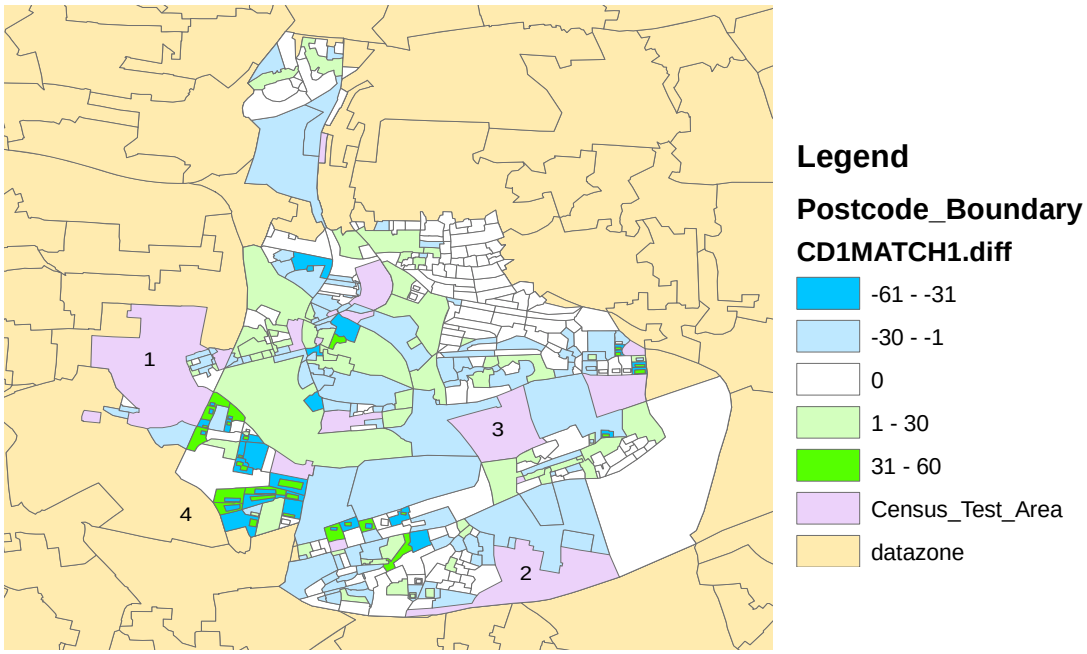
7. Results – CAG vs Census Test

Glasgow North

7.1.1 A map showing the actual difference between the Census Test address list and the CAG in each postcode can be seen in Figure 6.1.

- The **green** areas show postcode areas where there are more dwellings in the Census Test than in the CAG .
- The **blue** areas show the opposite effect where there are more dwellings per postcode in the CAG than in the Census Test.
- The **white** areas represent a complete match in the number of dwellings in the postcode boundary in both address lists.
- The **pink** areas represent areas where there are no residential dwellings recorded in either address list.

Figure 6.1: Glasgow North: Number of dwellings recorded in Census Test address list, minus the number of dwellings recorded in CAG



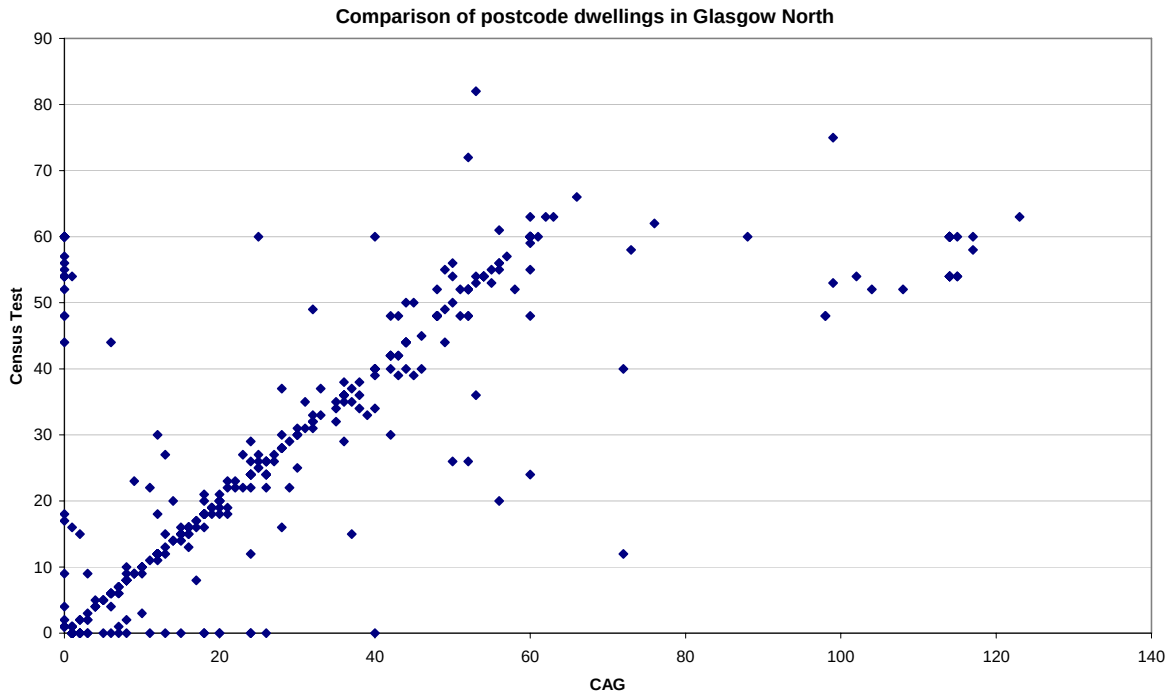
7.1.2 The pink areas shown in Figure 6.1 illustrate areas where there are no residential dwellings in the postcode boundary. Upon further investigation into the numbered areas in Figure 6.1, it was observed that the large pink area labelled 1 to the left includes a school and Gowlairs Park; the area labelled 2 to the bottom has a large market area and the area labelled 3 in the middle has railway tracks and a driving test centre.

7.1.3 Also shown in Figure 6.1 is the quite large area of dark green and dark blue colours in the left of the map. This is labelled 4 to the left of the area. Upon further investigation of these postcodes it was noted that the Census Test Address lists had a postcode boundary within another postcode boundary – these are known as 'island postcodes'. Therefore the number of dwellings within the CAG postcode boundary

matched the sum of dwellings within the two postcode boundaries. This was similar to the rest of the dark areas in Glasgow North, where a dark green area is a neighbouring postcode of the dark blue area so dwellings were not missed in the CAG but were simply recorded as having a different postcode in each source. This means that the two address lists are actually far more similar than it first appears. A result such as this could have been caused by postcode changes between the times that the two address lists were produced. To check whether this was the case, GROS checked the date at which each postcode in the Glasgow North Census Test area was introduced. It was found that none of the postcodes had changed since 2002, so postcodes changes cannot explain the differences – it is just that the PAF and CAG have dwellings allocated to different, adjacent, postcodes.

7.1.4 The overall differences can be seen more clearly in Figure 6.2.

Figure 6.2: Number of dwellings in the Census Test compared to those in the CAG



7.1.5 Figure 6.2 shows the number of dwellings per postcode in the Census Test plotted against the number of dwellings per postcode recorded in the CAG. There is obviously a clear overall correlation between the number of dwellings recorded in both address lists but there are also a number of difference. There are a few postcodes where there are double the number of dwellings in the CAG compared to the Census Test and postcodes where the Census Test have recorded around 50 or 60 dwellings but the CAG have none. From Figure 6.1 and Figure 6.2, there appears to be a lot more dwellings within the CAG address list than in the Census Test. These differences and others are summarised in the following Table 6.1.

Table 6.1: Summary Table showing the differences observed between the CAG and Census Test address lists for Glasgow North

	No. postcodes	% of postcodes
>30 more dwellings recorded in the CAG than the Census Test	33	7.7%
1-30 more dwellings recorded in the CAG than the Census Test	119	27.6%
The same number of dwellings per postcode in both address lists	190	44.1%
1-30 more dwellings recorded in the Census Test than the CAG	61	14.2%
>30 more dwellings recorded in the Census Test than the CAG	28	6.5%
Total number of Postcode Boundaries	431	

Footnote: Percentages may not sum to 100 due to rounding

7.1.6 Table 6.1 shows that less than half of the postcode areas have exactly the same number of dwellings recorded within both address lists and there are more dwellings recorded in the CAG address list than recorded in the Census Test. Glasgow City Council omitted non-residential addresses from their original CAG so they will not contribute to the extra addresses. One possible reason for the difference would be if some multi-occupancy addresses (where a single property is divided into several addresses) were included just once in the Census Test address list, but several times in the CAG. However, as noted above, the actual address lists are more similar than they appear, because there were a number of cases where the same addresses were included in both address lists, and it was just the postcode that was different.

Table 6.2: Total number of residential dwellings recorded in CAG and Census Test

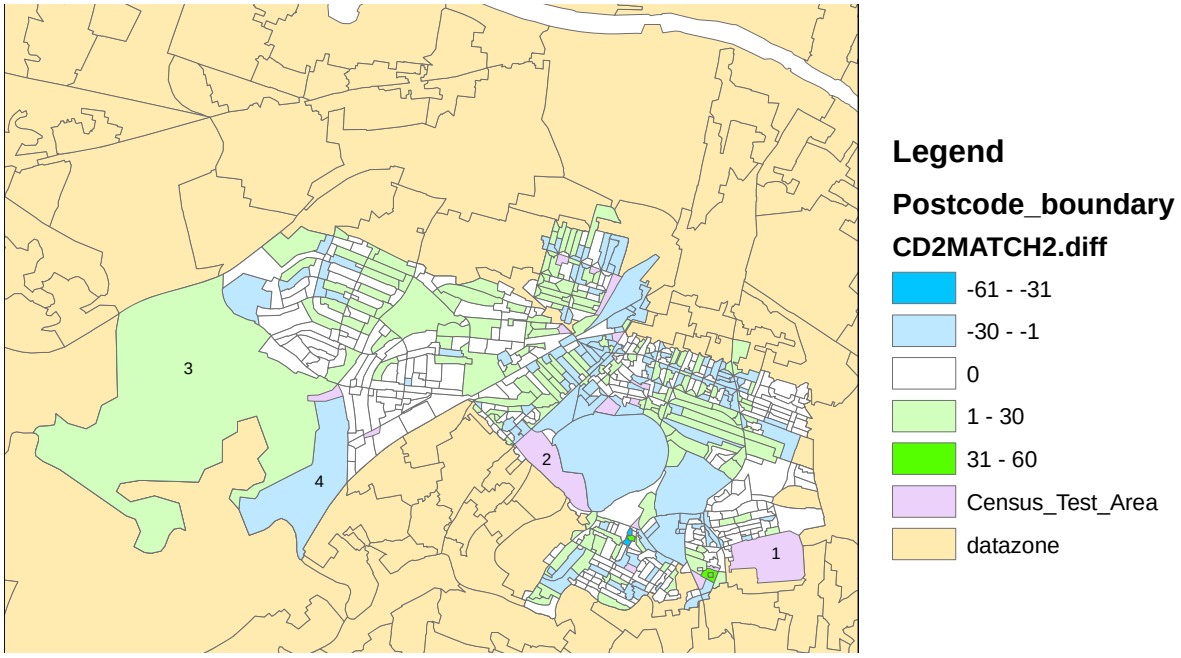
	Census Test	Corporate Address Gazetteer
Total number of residential dwellings recorded	12,094	12,635

7.1.7 Table 6.2 shows that there are approximately 600 more addresses present in the CAG compared to the Census Test address list for Glasgow North.

Glasgow South

7.1.8 Figure 6.3 shows the map of the Census Test area in Glasgow South. This map provides the same information as that described in Section 6.1.

Figure 6.3: Glasgow South: Number of dwellings recorded in Census Test address list, minus the number of dwellings recorded in CAG

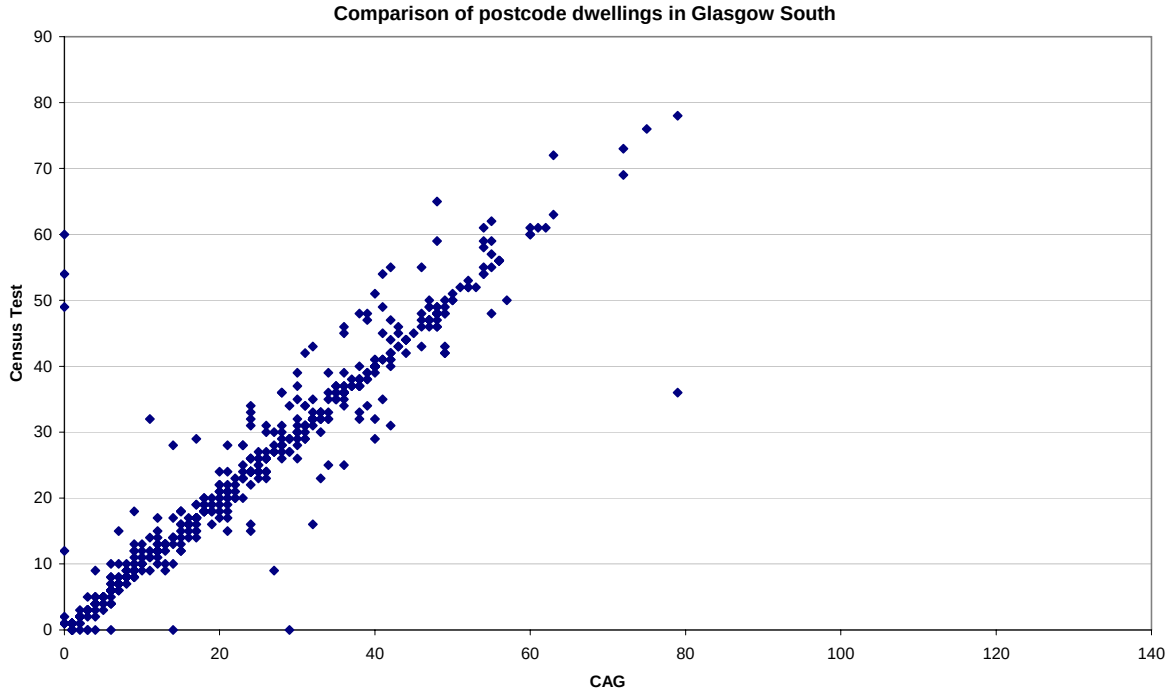


7.1.9 Figure 6.3 illustrates that there are very few postcode areas with large differences between the Census Test address lists and the CAG. The pink area labelled 1 represents Hampden Park football ground and the other large pink area (labelled 2) represents a bowling club. This means that there are no residential dwellings in these areas and so they do not appear on either of the address lists.

7.1.10 The large green postcode area labelled 3, shows that there are more dwellings within the Census Test address list compared to the CAG. However, with further investigation it appears that this postcode boundary includes a country park and so only a few dwellings are residential. In this postcode boundary, the CAG recorded 2 residential dwellings and the Census Test address list recorded 3 residential dwellings. Therefore the difference is only 1 dwelling in the postcode.

7.1.11 The postcode area labelled 4 is very similar to the above point except that there are more dwellings observed in the CAG address list than in the Census Test. Upon further investigation of this area, it was found that the CAG list has recorded 6 residential dwellings and the Census Test recorded 4. Therefore, there is only a difference of 2 residential dwellings in the CAG address list. This may simply be a misclassification of dwellings in one postcode rather than the other.

Figure 6.4: Number of dwellings in the Census Test compared to those in the CAG



7.1.12 Figure 6.4 shows the scatter plot of the number of dwellings per postcode in the CAG against those in the Census Test. It can be seen that there is a clear correlation between the CAG and Census Test address lists for Glasgow South showing that the lists are quite closely matched. There are, however, a few postcodes which differ greatly in the number of dwellings in each of the address lists. For example, one clear outlier in the data would be where the CAG has recorded 79 dwellings in a postcode (G42 9PR) but the Census Test has only recorded 36. Further investigation shows that 49 flats belonging to Hanover Close were recorded under G42 9PR in the CAG address list but recorded under G42 9AX in the Census Test address list. This will also account for a postcode that has 49 dwellings in the Census Test but none in the CAG address list.

7.1.13 There are also two postcodes (G42 9ER and G42 9ES) above this point in Figure 6.4, which have 60 and 54 dwellings in each respectively. Further investigation into these postcodes shows a number of flats called Battlefield Court, 15 Cathkinview Place. These flats do not appear to exist in the CAG address list.

Table 6.3: Summary Table showing the differences observed between the two address lists for Glasgow South

	Actual Number of Postcode Boundaries	Percentage of Total Postcode Boundaries
>30 more dwellings recorded in the CAG than the Census Test	1	0.2%
1-30 more dwellings recorded in the CAG than the Census Test	145	21.6%
The same number of dwellings per postcode in both address lists	359	53.5%
1-30 more dwellings recorded in the Census Test than the CAG	163	24.3%
>30 more dwellings recorded in the Census Test than the CAG	3	0.4%
Total number of Postcode Boundaries	671	

7.1.14 Table 6.3 shows that over 50 per cent of the postcodes showed a complete match between the number of dwellings recorded in the Census Test and CAG address lists for Glasgow South. Also, there were very few large differences between the number of dwellings per postcode recorded in the two address lists.

Table 6.4: Total number of residential dwellings recorded in CAG and Census Test

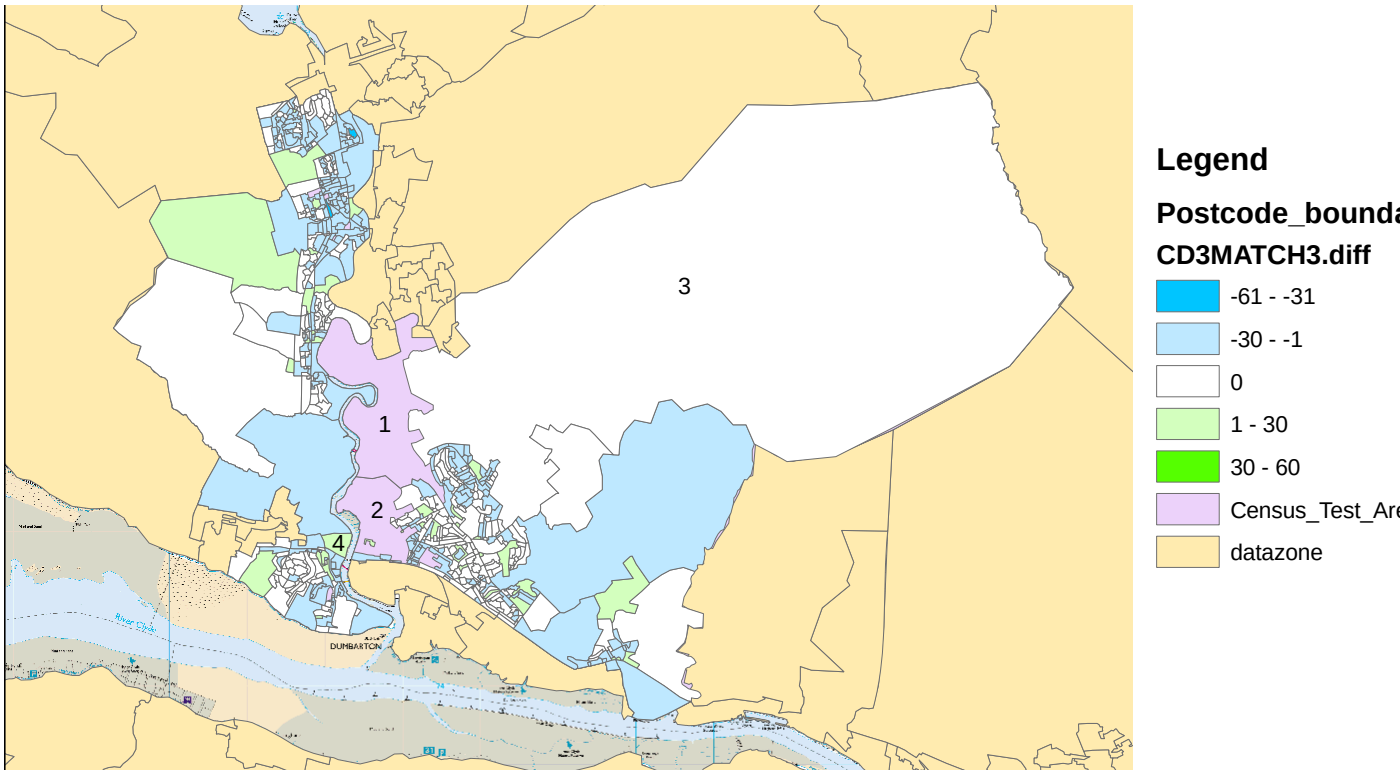
	Census Test	Corporate Address Gazetteer
Total number of residential dwellings recorded	15,703	15,424

7.1.15 Table 6.4 shows that there are approximately 300 more addresses present in the Census Test address list compared to the CAG for Glasgow South.

West Dunbartonshire

7.1.16 GROS did not specify what should be included in the CAGs provided by the councils, or what format the information should be provided in. The West Dunbartonshire CAG needed a lot of data cleaning before the comparison could be performed with the Census Test address list. This is because the Council provided information on non-residential sites as well residential dwellings and included plots of land, playing fields, car parks etc as potential residential sites. These areas needed to be removed by hand by GROS staff so that they were not included in the comparison exercise. Figure 6.5 shows the map of the Census Test area in West Dunbartonshire.

Figure 6.5: West Dunbartonshire: Number of dwellings recorded in Census Test address list, minus the number of dwellings recorded in CAG

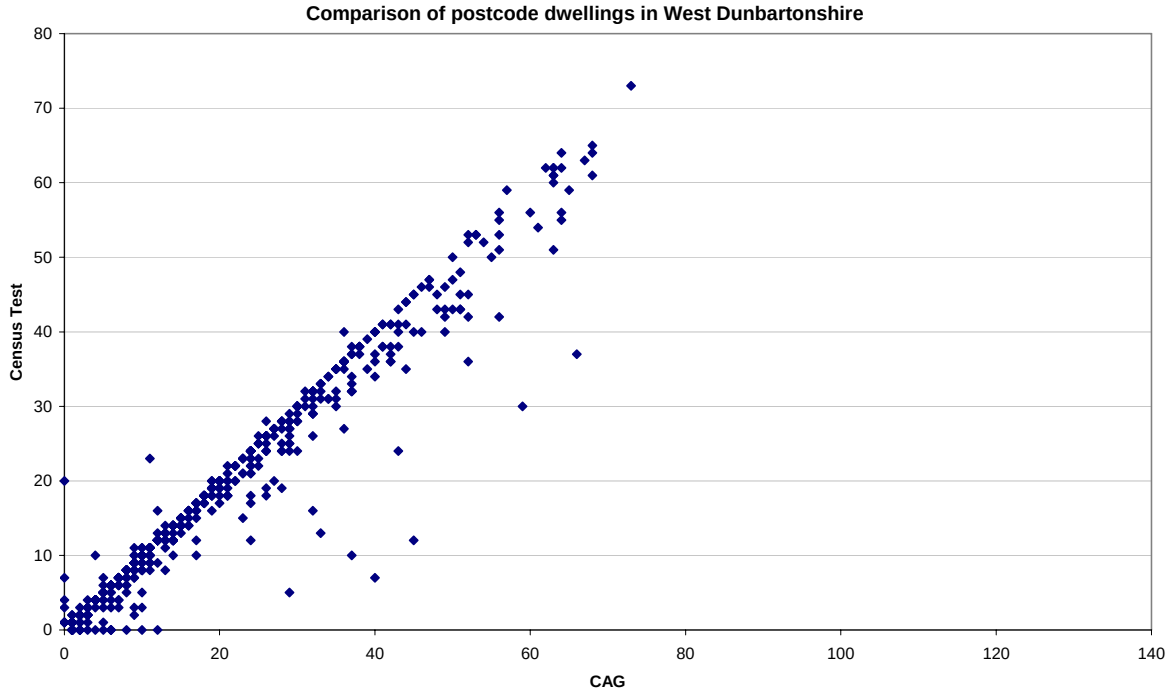


7.1.17 In Figure 6.5, the two large pink areas labelled 1 and 2 represent postcode boundaries where there are no residential dwellings recorded in either of the two address lists, CAG and Census Test. Upon further investigation of these areas, it was found that large industrial estates occupy the two postcode areas. The large postcode area labelled 3, represents a complete match of the number of dwellings in the postcode in both address lists. The area has only 7 residential dwellings as the area consists of the Kilpatrick Hills.

7.1.18 The pale green area labelled 4, represents a postcode area where there are 20 more residential dwellings within the Census Test address list compared to the CAG. With further investigation, it was observed that the postcode (G82 4BW) contained 20 residential dwellings. These were all recorded in the Census Test address list as Dennystoun Forge but were fully omitted from the CAG. A neighbouring postcode area to this one shows over 16 residential dwellings more in the CAG compared to the Census Test. So it initially appeared that there was a misclassification of dwellings but this does not appear to be the problem. There is a Gypsy/Travellers site at Dennystoun Forge, which may explain the difference.

7.1.19 The dark blue area represents almost 30 dwellings difference in postcode G82 4AB for West Bridgend (66 dwellings recorded in CAG and 37 recorded in Census Test). There is a caravan park with 20 dwellings recorded in this postcode in the CAG, which may explain some of this difference.

Figure 6.6: Number of dwellings in the Census Test compared to those in the CAG



7.1.20 Figure 6.6 illustrates that the CAG and Census Test address lists are highly correlated with the number of dwellings per postcode in the CAG almost mirroring the number of dwellings per postcode in the Census Test. There are very few postcodes where there are more dwellings recorded in the Census Test compared to the CAG; however, there are a lot more vice versa. The reason for there being more dwellings in the CAG compared to the Census Test may be partly due to the manual cleaning of the CAG address list.

Table 6.5: Summary Table showing the differences observed between the two address lists for West Dunbartonshire

	Actual Number of Postcode Boundaries	Percentage of Total Postcode Boundaries
>30 more dwellings recorded in the CAG than the Census Test	2	0.3%
1-30 more dwellings recorded in the CAG than the Census Test	266	42.5%
The same number of dwellings per postcode in both address lists	323	51.6%
1-30 more dwellings recorded in the Census Test than the CAG	35	5.6%
>30 more dwellings recorded in the Census Test than the CAG	0	0.0%
Total number of Postcode Boundaries	626	

7.1.21 Table 6.5 shows that over 50 per cent of the postcode boundaries completely matched in the number of residential dwellings in the CAG and Census Test address lists for West Dunbartonshire. It can also be seen in this table that there are a lot more dwellings recorded in the CAG than in the Census Test (approximately 43%)

but as previously mentioned, this could be due to many non-residential dwellings not being removed from the original CAG address list.

Table 6.6: Total number of residential dwellings recorded in CAG and Census Test

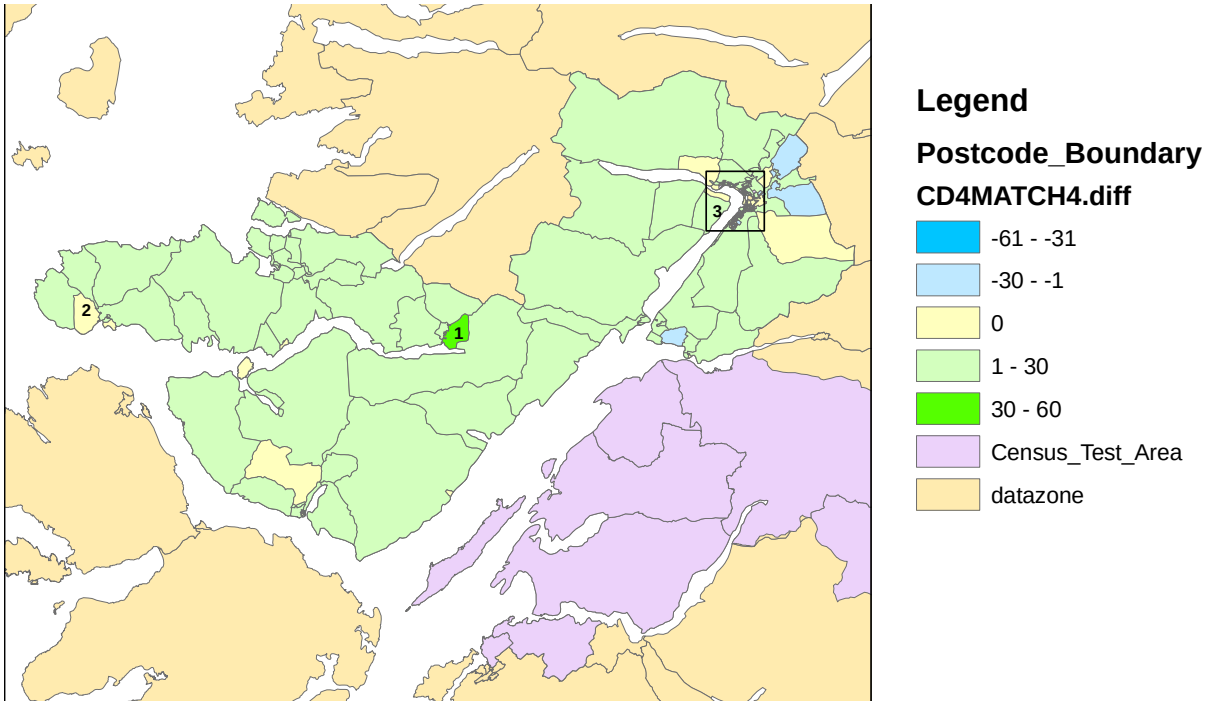
	Census Test	Corporate Address Gazetteer
Total number of residential dwellings recorded	11,429	12,360

7.1.22 Table 6.6 shows that there are approximately a thousand more addresses present in the CAG compared to the Census Test address list for West Dunbartonshire.

Highland

7.1.23 GROS did not specify what should be included in the CAGs provided by the councils, or what format the information should be provided in. The Highland CAG also required a lot of data cleaning in order to attain consistency with the Census Test address list. Non-residential areas needed to be identified and removed from the data so that they could be compared to the residential areas in the Census Test. This was a manual task and with some addresses, it was not clear whether or not they were residential. The map of the Census Test area in Highland can be seen in Figure 6.7.

Figure 6.7: Highland: Number of dwellings recorded in Census Test address list, minus the number of dwellings recorded in CAG



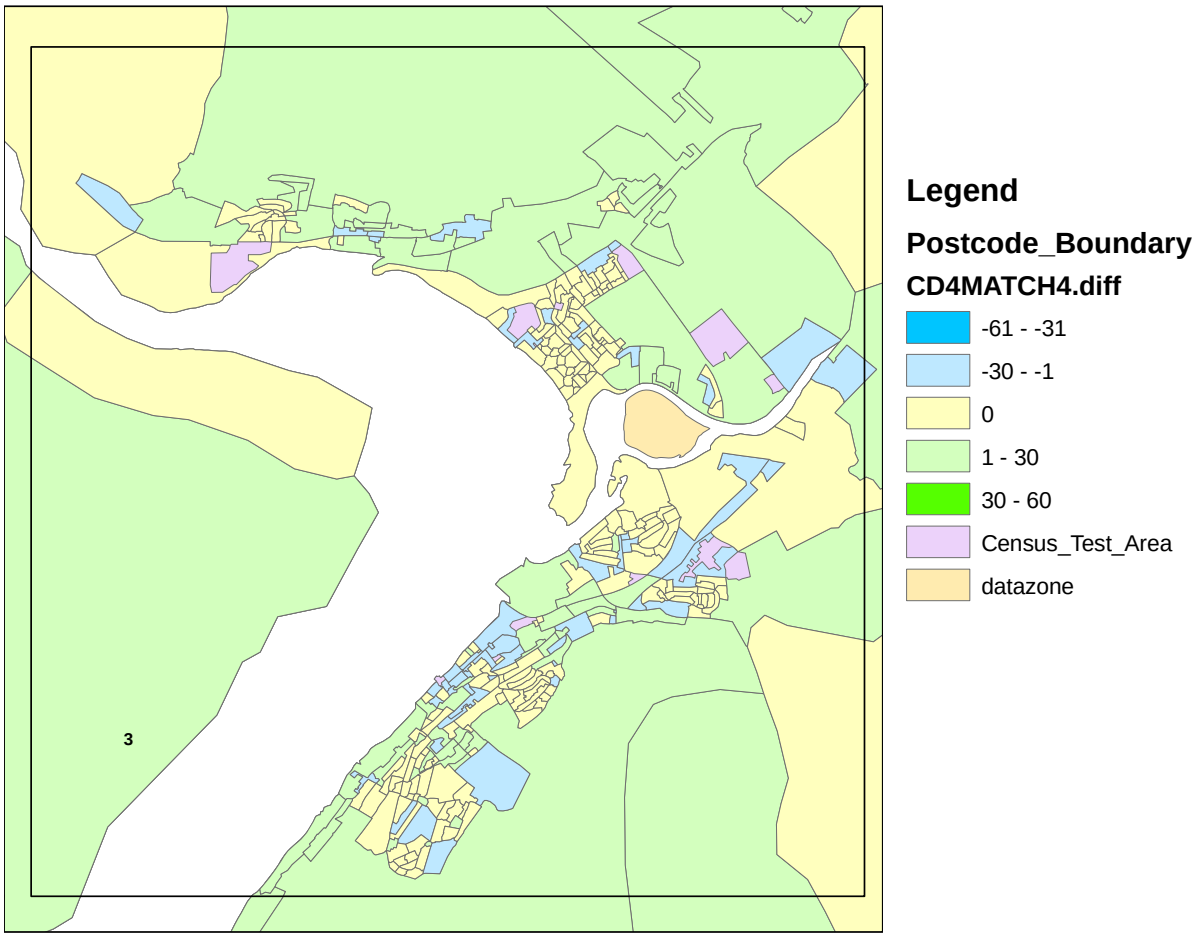
7.1.24 Figure 6.7 illustrates that there are more dwellings per postcode in the Census Test address list than the CAG overall, in the rural area of Highland. This is

very different to all the results noted previously where there were more dwellings recorded in the CAG. This may be due to the area being rural as there can be duplication of addresses where names of houses or official addresses or both may be have been recorded. Many of these postcodes contain very few dwellings, e.g., under ten, so in many cases the overall difference is very small.

7.1.25 The colour of the areas representing the completely matched postcodes had to be changed to pale yellow because the background area of water is white. The dark green area labelled 1 represents more than 30 residential dwellings in the postcode (PH36 4HZ) recorded in the Census Test compared to the CAG. This postcode has 91 residential dwellings recorded in the Census Test and 57 recorded in the CAG. This partly may be due to some additional holiday cottages (e.g., log cabins) being included in the Census Test. The other postcode labelled 2 represents an area where all the residential dwellings fully matched between both sources of addresses.

7.1.26 The area highlighted in the box and labelled 3 has been magnified and can be seen in Figure 6.8. This demonstrates a different trend where the differences are more even in the urban areas.

Figure 6.8: Highland: Number of dwellings recorded in Census Test address list, minus the number of dwellings recorded in CAG: Area 3 magnified



7.1.27 Figure 6.8 shows the small area in Figure 6.7 labelled 3 in more detail. It can be seen that there are a lot of postcode areas which have fully matched, and areas where there are more residential dwellings in the CAG than in Census Test. There are also some areas where there are no residential dwellings. Upon further investigation, it was found that industrial estates, a school, and Highland Communities NHS Trust occupy these areas.

7.1.28 The pale orange area represents Eilean Bhealaidh, an island where there are no residents and therefore not included in the comparison exercise.

7.1.29 Figure 6.9 illustrates the differences between the number of residential dwellings per postcode for all the postcodes in the Census Test address list and CAG.

Figure 6.9: Number of dwellings in the Census Test compared to those in the CAG

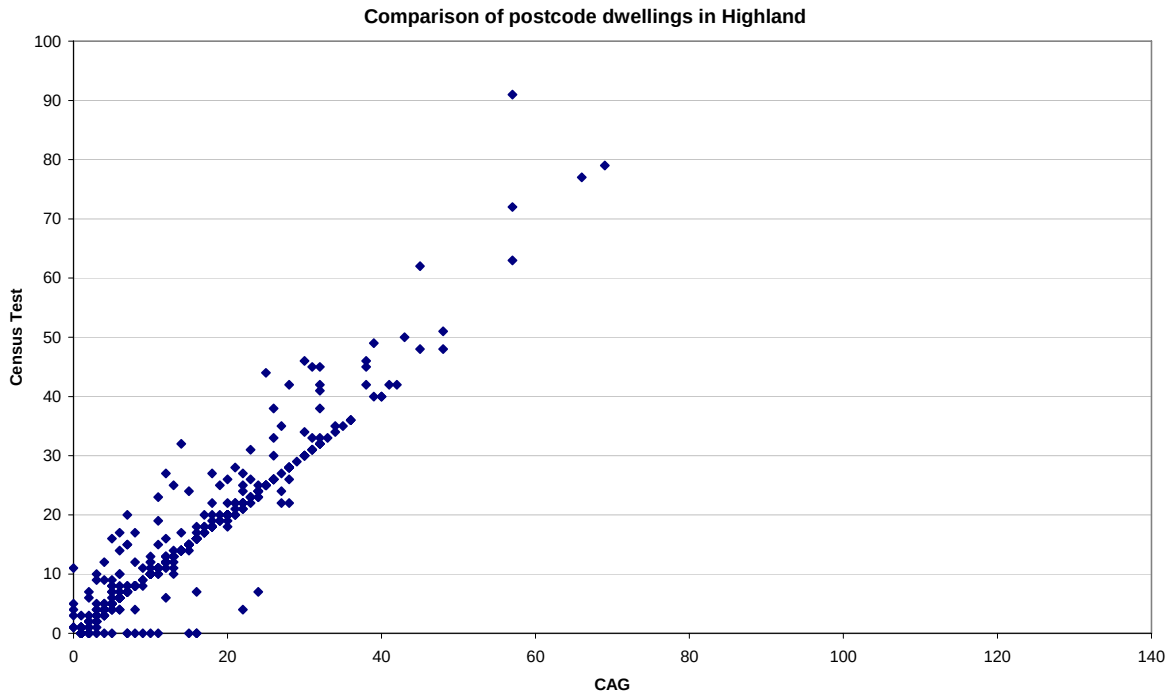


Figure 6.9 shows that there is a correlation between the number of residential dwellings per postcode in each of the address lists. It shows that in many cases, there are more dwellings per postcode in the Census Test compared to the CAG and that there are very few postcodes with a large number of dwellings per postcode. This is in contrast to the other CAGs included in this report, where there tended to be more dwellings included in the CAG than the Census Test.

7.1.30 There appears to be an outlier in this data for the postcode PH36 4HZ as there are over 90 dwellings recorded in the Census Test and fewer than 60 recorded in the CAG. This was mentioned previously in paragraph 6.4.3.

7.1.31 There are a few postcodes where a number of dwellings were recorded in the CAG but not in the Census Test. These may be non-residential areas, but it is not possible to be certain of this. For example, postcode PH33 6PW has 16 dwellings recorded in the CAG but none in the Census Test – though 16 dwellings have been

added to more recent versions of PAF, which suggests that the discrepancy in this case is due to the different dates used.

7.1.32 Table 6.7 shows the summary of matched and unmatched postcodes between the two address lists.

Table 6.7: Summary Table showing the differences observed between the two address lists for West Dunbartonshire

	Actual Number of Postcode Boundaries	Percentage of Total Postcode Boundaries
>30 more dwellings recorded in the CAG than the Census Test	0	0.0%
1-30 more dwellings recorded in the CAG than the Census Test	80	18.8%
The same number of dwellings per postcode in both address lists	235	55.2%
1-30 more dwellings recorded in the Census Test than the CAG	110	25.8%
>30 more dwellings recorded in the Census Test than the CAG	1	0.2%
Total number of Postcode Boundaries	426	

7.1.33 Table 6.7 shows that there are very few postcodes with large differences between the Census Test address list and the CAG in Highland. Again over half of the postcode areas fully matched in the number of residential dwellings per postcode and there were slightly more dwellings recorded in the Census Test address list compared to the CAG – mainly in the rural areas, as shown in the maps in figures 6.7 and 6.8.

Table 6.8: Total number of residential dwellings recorded in CAG and Census Test

	Census Test	Corporate Address Gazetteer
Total number of residential dwellings recorded	6,440	6,076

7.1.34 Table 6.8 shows that there are approximately 400 more addresses present in the Census Test address list compared to the CAG for Highland.

8. Results – Census Test vs PAF

No address list is ever likely to be perfect, and it is inevitable that there will be some changes since the address list was ‘frozen’. When new dwellings are recorded in an address list, the information tends to be accurate, but there can be timing differences when information about a property is added to different address lists. It is also possible that historic errors occur in address lists, e.g., where duplicate entries exist if a property has more than one name, or is recorded under two different postcodes. This is why there was a need for the address checkers and enumerators to carry out their address checking work in order to produce a comprehensive address list.

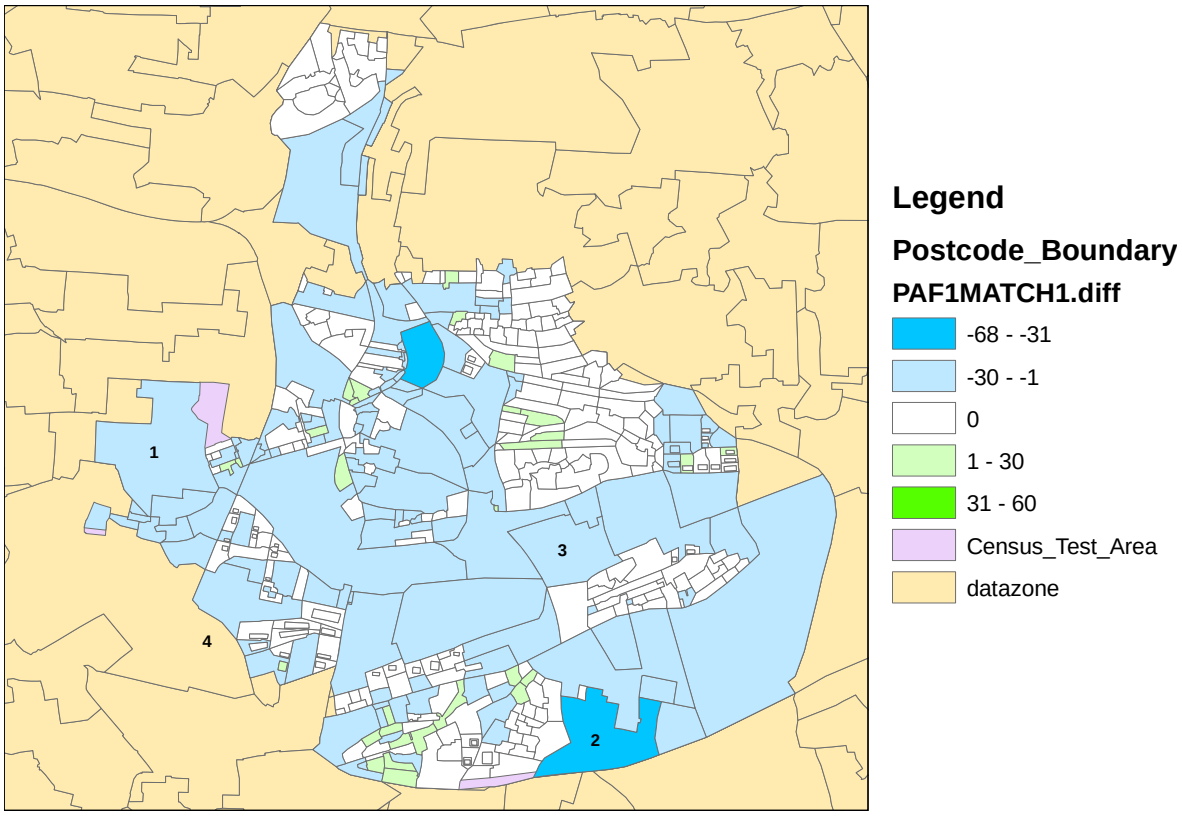
In order to provide some context to the analysis of the quality of the CAG address lists, we also carried out a comparison of the final Census Test address list with the original Royal Mail Postal Address File (PAF), before the Address checkers and Enumerators amended it. This provides a useful comparison, because it indicates the comprehensiveness of the original PAF file.

As the CAG address list is based on a different source to the Census Test, it is possible for an address to appear in both the CAG and Census Test but be given a different postcode in each file, which gives the impression that there are large numbers of discrepancies (as was seen in the Glasgow North CAG/Census Test comparison - Section 6.1). Because the Census Test address list was based on PAF, we would expect dwellings in each address list to have the same postcode, so we would not expect to find the same problem in this analysis.

Glasgow North

8.1.1 Figure 7.1 shows the map of that Glasgow North Census Test area, with the difference between the Census Test addresses and the original PAF. Colours and scale of maps are the same as those used in Section 6.

Figure 7.1: Glasgow North: Number of dwellings recorded in Census Test address list, minus the number of dwellings recorded in PAF



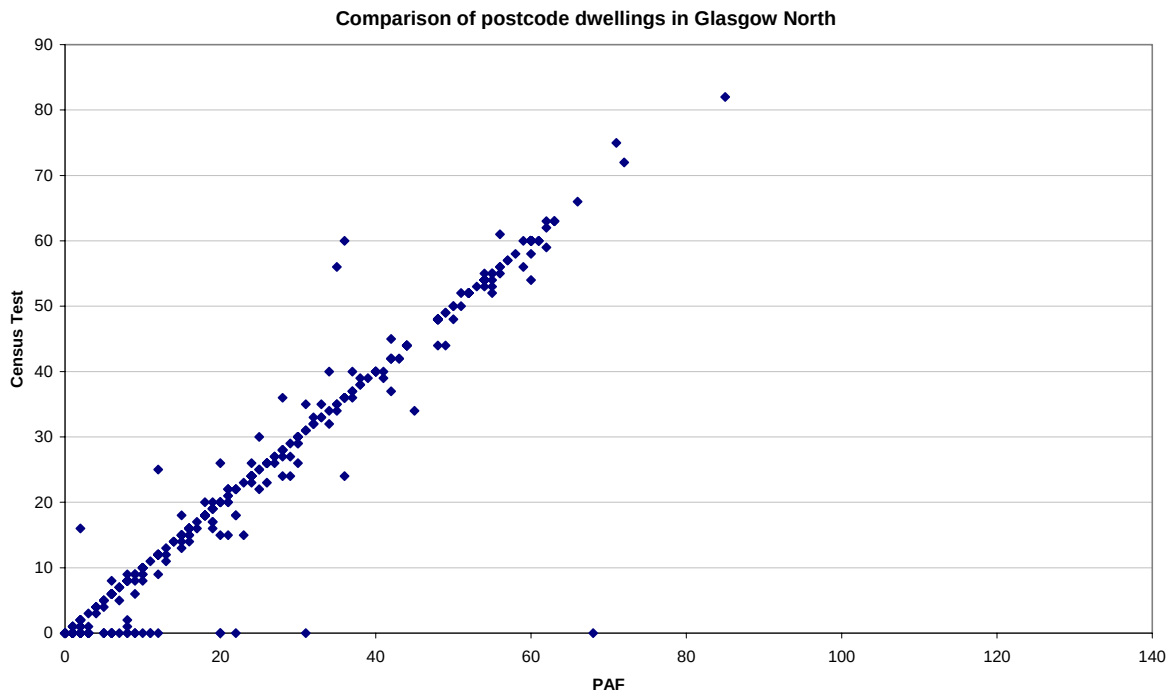
8.1.2 Figure 7.1 illustrates that there were a lot more dwellings recorded in the PAF than in the Census Test, as would be expected. The reason for this is that the non-residential properties were included in the original PAF.

8.1.3 The postcode areas labelled above are the same as those explained in Section 6.1 so explanations will now be provided for the differences. The postcode areas labelled 1, 2 and 3 in the CAG analysis (Figure 6.1) were pink because there are no residential properties in these areas. However in Figure 7.1 they are blue, indicating some non-residential properties included in the PAF.

8.1.4 The area labelled 4 now shows a very different picture to that in Figure 6.1 as most of the areas are fully matched with the exception of a few misclassified flats. Upon further investigation, these blue areas only account for a difference of one flat which may have been missed by the address checkers and enumerators.

8.1.5 Figure 7.2 will show the extent of the difference between these two address sources.

Figure 7.2: Number of dwellings in the Census Test compared to those in the PAF



8.1.6 Figure 7.2 shows the number of dwellings per postcode in the Census Test plotted against the number of dwellings per postcode recorded in the PAF. It can be seen that these two sources have a strong correlation with a few non-residential properties included in the PAF but which were removed from the address list in the Census Test.

8.1.7 There are a couple of postcodes (G21 2JF and G21 2LQ) with large numbers of dwellings recorded in the Census Test but few recorded in the PAF. These may have been multi-occupancy addresses. Upon further investigation, it appears that the number of dwellings recorded in the Census Test for the first postcode is correct so perhaps a new block of flats was not recorded in the PAF in October but were

added by the Address checkers on the ground after this date. The other postcode boundary appears to have some flats where some are numbered and get their own mail and some only have one address point. For example, the enumerators added 21 individual flats to the Census Test address list but PAF only had the address recorded once.

Table 7.1: Summary Table showing the differences observed between the PAF and Census Test address lists for Glasgow North

	Actual Number of Postcode Boundaries	Percentage of Total Postcode Boundaries
>30 more dwellings recorded in the PAF than the Census Test	2	0.4%
1-30 more dwellings recorded in the PAF than the Census Test	127	27.7%
The same number of dwellings per postcode in both address lists	300	65.5%
1-30 more dwellings recorded in the Census Test than the PAF	29	6.3%
>30 more dwellings recorded in the Census Test than the PAF	0	0.0%
Total number of Postcode Boundaries	458	

Footnote: Percentages may not sum to 100 due to rounding

8.1.8 Table 7.1 shows that over 65% of postcode areas fully matched in the number of residential dwellings per postcode and that there are more dwellings recorded in the PAF compared to the Census Test address list. These properties are mainly non-residential properties recorded in the PAF and subsequently removed by the Address checkers and Enumerators. There are very few postcodes with large differences between the Census Test address list and the PAF in Glasgow North.

Table 7.2: Total number of residential dwellings recorded in PAF and Census Test

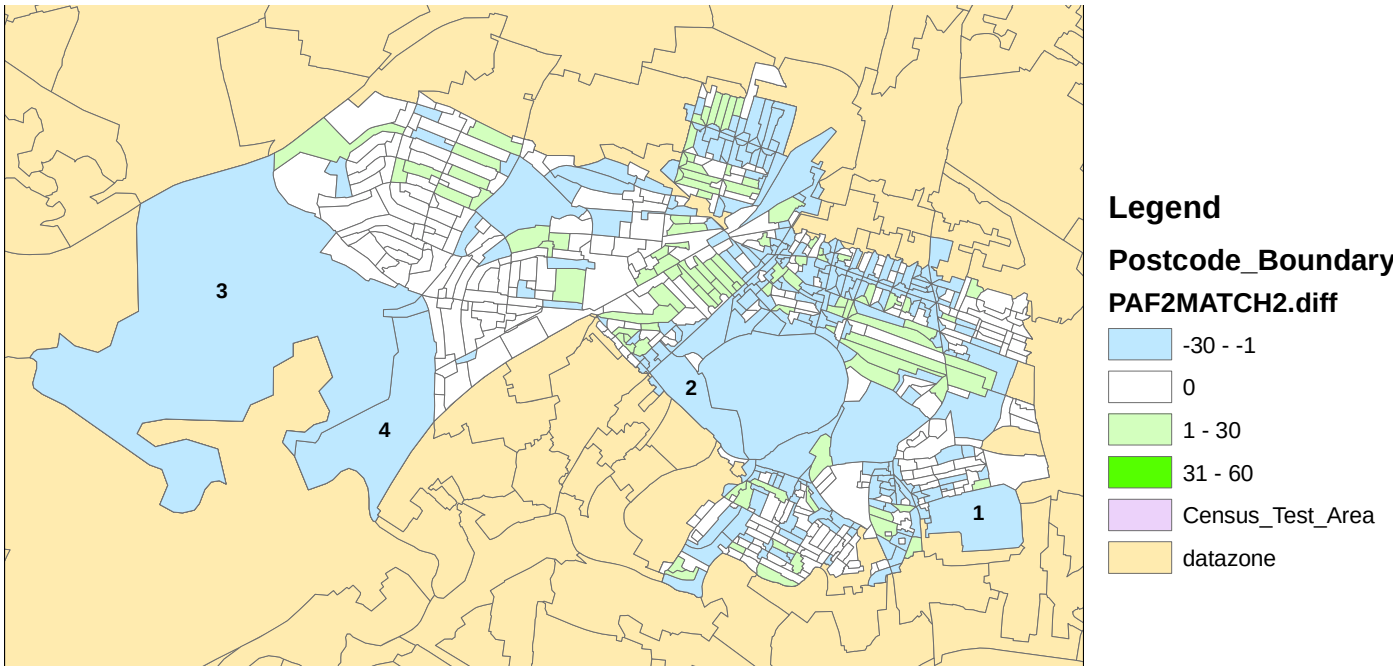
	Census Test	Postal Address File
Total number of residential dwellings recorded	12,094	12,486

8.1.9 Table 7.2 shows that there are approximately 400 more addresses present in the PAF compared to the Census Test address list for Glasgow North.

Glasgow South

8.1.10 Figure 7.3 is a map showing the difference between the number of dwellings per postcode recorded in the original PAF and the final Census Test address lists.

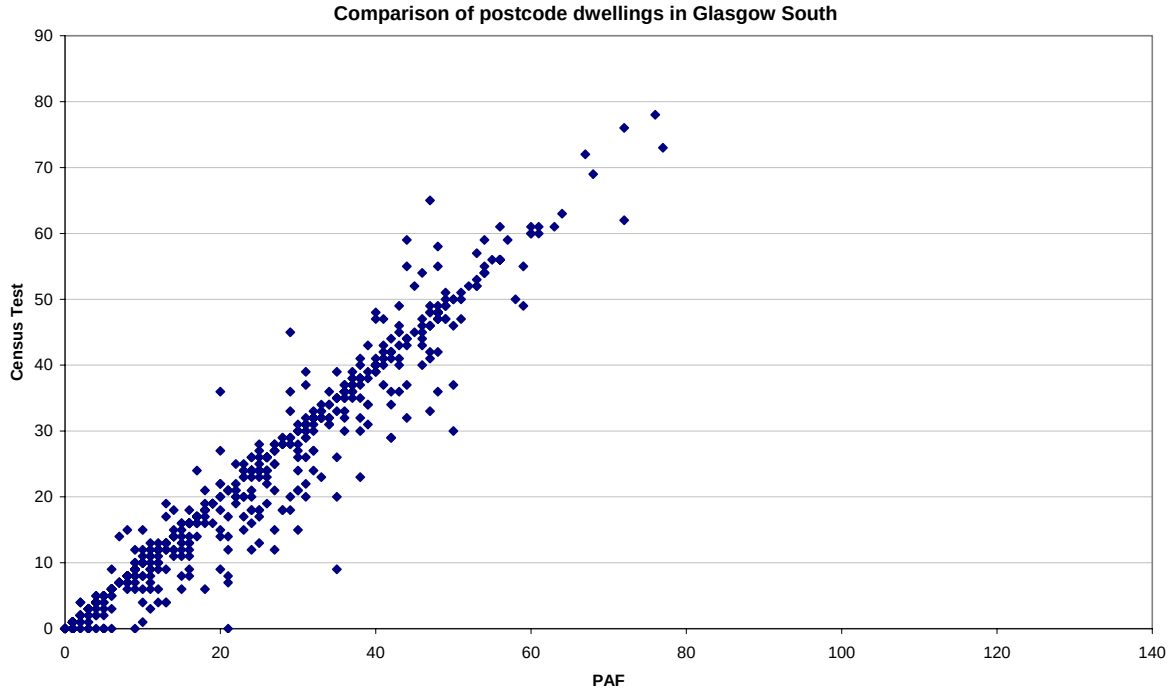
Figure 7.3: Glasgow South: Number of dwellings recorded in Census Test address list, minus the number of dwellings recorded in PAF



8.1.11 Figure 7.3 illustrates that there are more dwellings per postcode recorded in the PAF compared to the Census Test address list. Again, where the postcode boundaries labelled 1 and 2 were pink areas in Figure 6.3, they are now pale blue. This represents 1-30 more dwellings recorded in the PAF than recorded in the Census Test but as these were found to be purely non-residential areas, which confirms that PAF has included non-residential properties in these postcode areas.

8.1.12 The area labelled 3 was a pale green area in Figure 6.3 which suggested that there were more dwellings recorded in the Census Test compared to the CAG, however, in Figure 7.3 there are more dwellings recorded in the PAF. Finally in the area labelled 4, there were more dwellings recorded in the CAG compared to the Census Test in Figure 6.3, but there are now more dwellings in the PAF compared to the Census Test. This suggests that perhaps properties were removed from the Census Test after the PAF which should not have been removed, or the CAG has included non-residential properties. Upon further investigation, it was found that there was a difference of 2 between the CAG and the Census Test, whereas there is a difference of 6 between the PAF and the Census Test.

Figure 7.4: Number of dwellings in the Census Test compared to the PAF



8.1.13 It can be seen in Figure 7.4 that there is a clear association between the number of dwellings per postcode recorded in the PAF and the number of dwellings per postcode recorded in the Census Test. It can also be determined that overall, there are more dwellings recorded in the PAF because there are more data points lying to the right of the diagonal line. However, there are also a number of postcodes where the enumerators and address checkers identified dwellings which were not recorded in PAF (where the number of dwellings recorded in the Census Test was higher).

Table 7.3: Summary Table showing the differences observed between the PAF and Census Test address lists for Glasgow South

	Actual Number of Postcode Boundaries	Percentage of Total Postcode Boundaries
>30 more dwellings recorded in the PAF than the Census Test	0	0.0%
1-30 more dwellings recorded in the PAF than the Census Test	252	36.6%
The same number of dwellings per postcode in both address lists	338	49.1%
1-30 more dwellings recorded in the Census Test than the PAF	98	14.2%
>30 more dwellings recorded in the Census Test than the PAF	0	0.0%
Total number of Postcode Boundaries	688	

Footnote: Percentages may not sum to 100 due to rounding

8.1.14 Table 7.3 shows that in almost 50% of postcode areas there was a complete match between the number of dwellings recorded in the PAF and the Census Test

address lists. It also shows that over 35% of postcode boundaries had more dwellings recorded in the PAF. These are mainly due to the inclusion of non-residential properties. In 14% of postcodes, the enumerators and address checkers identified additional dwellings that were not included in PAF.

Table 7.4: Total number of residential dwellings recorded in PAF and Census Test

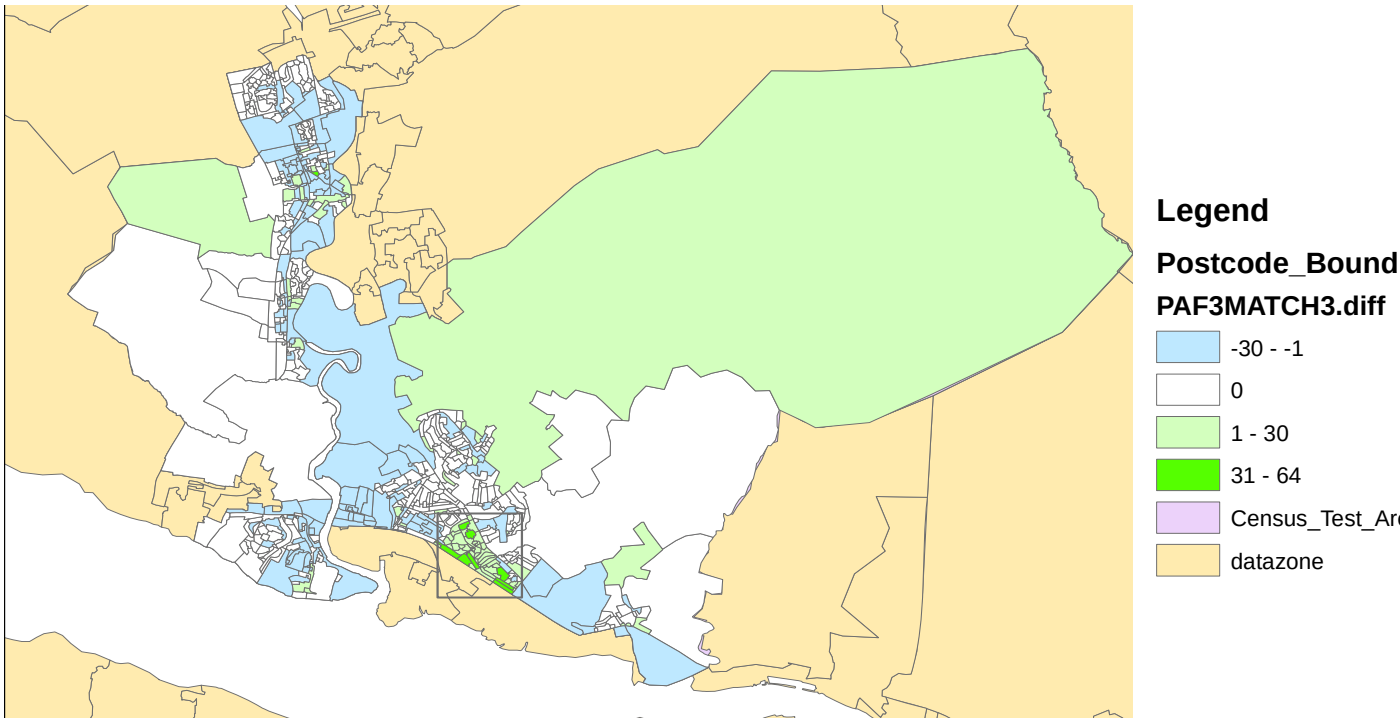
	Census Test	Postal Address File
Total number of residential dwellings recorded	15,703	16,433

8.1.15 Table 7.4 shows that overall, there are approximately 700 more addresses present in the PAF compared to the Census Test address list for Glasgow South.

West Dunbartonshire

8.1.16 It has been discovered that two files, covering parts of West Dunbartonshire, were missing from the original October 2006 PAF provided for this work – therefore, there will be several postcode areas where there will be no recorded dwellings for the PAF. These files have been requested and will be included if possible. This affects the area of Figure 7.5 below, which is highlighted by the box.

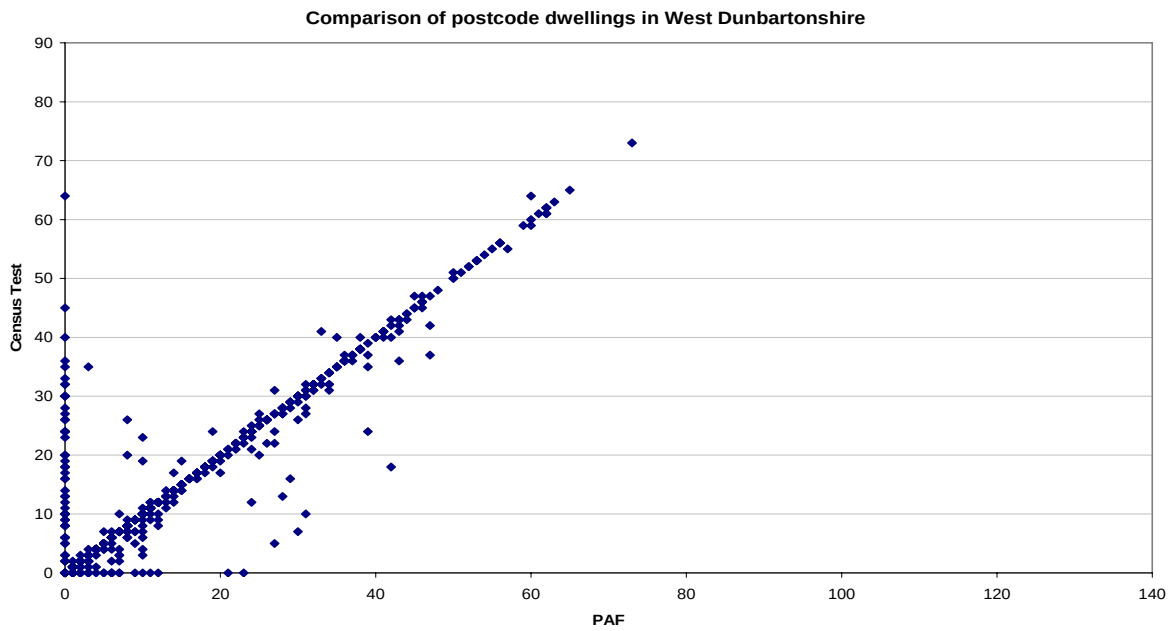
Figure 7.5: West Dunbartonshire: Number of dwellings recorded in Census Test address list, minus the number of dwellings recorded in PAF



8.1.17 Figure 7.5 illustrates that there were a lot of postcode areas where more properties were recorded in the PAF compared to the Census Test. This is especially true in the middle band which was found to be completely non-residential in Figure 6.5. This indicates that there were more non-residential properties included in the PAF and removed by the Address checkers and Enumerators on the Census Test.

8.1.18 The area highlighted with the surrounding box shows the postcode areas where the missing files from the PAF are affected. There are up to 64 dwellings per postcode missing from the PAF but which had been recorded in the Census Test. This will therefore affect the scatter plot in Figure 7.6 below.

Figure 7.6: Number of dwellings in the Census Test compared to the PAF



8.1.19 Figure 7.6 illustrates a clear correlation between the number of dwellings recorded per postcode in the two address sources. However, there are a large number of postcodes which have dwellings recorded in the Census Test but not in the PAF. This is due to the missing files in PAF as explained previously. There are some postcode boundaries where more dwellings were recorded in the PAF than in the Census Test but again these are probably due to non-residential properties being included in the PAF.

Table 7.5: Summary Table showing the differences observed between the PAF and Census Test address lists for West Dunbartonshire

	Actual Number of Postcode Boundaries	Percentage of Total Postcode Boundaries
>30 more dwellings recorded in the PAF than the Census Test	0	0.0%
1-30 more dwellings recorded in the PAF than the Census Test	136	21.1%
The same number of dwellings per postcode in both address lists	422	65.5%
1-30 more dwellings recorded in the Census Test than the PAF	77	12.0%
>30 more dwellings recorded in the Census Test than the PAF	9	1.4%
Total number of Postcode Boundaries	644	

Footnote: Percentages may not sum to 100 due to rounding

8.1.20 Even excluding the files missing from the PAF, there are still over 65% of postcode boundaries with the same number of dwellings in both sources, and over 21% have more dwellings recorded per postcode in the PAF compared to the Census Test.

Table 7.6: Total number of residential dwellings recorded in PAF and Census Test

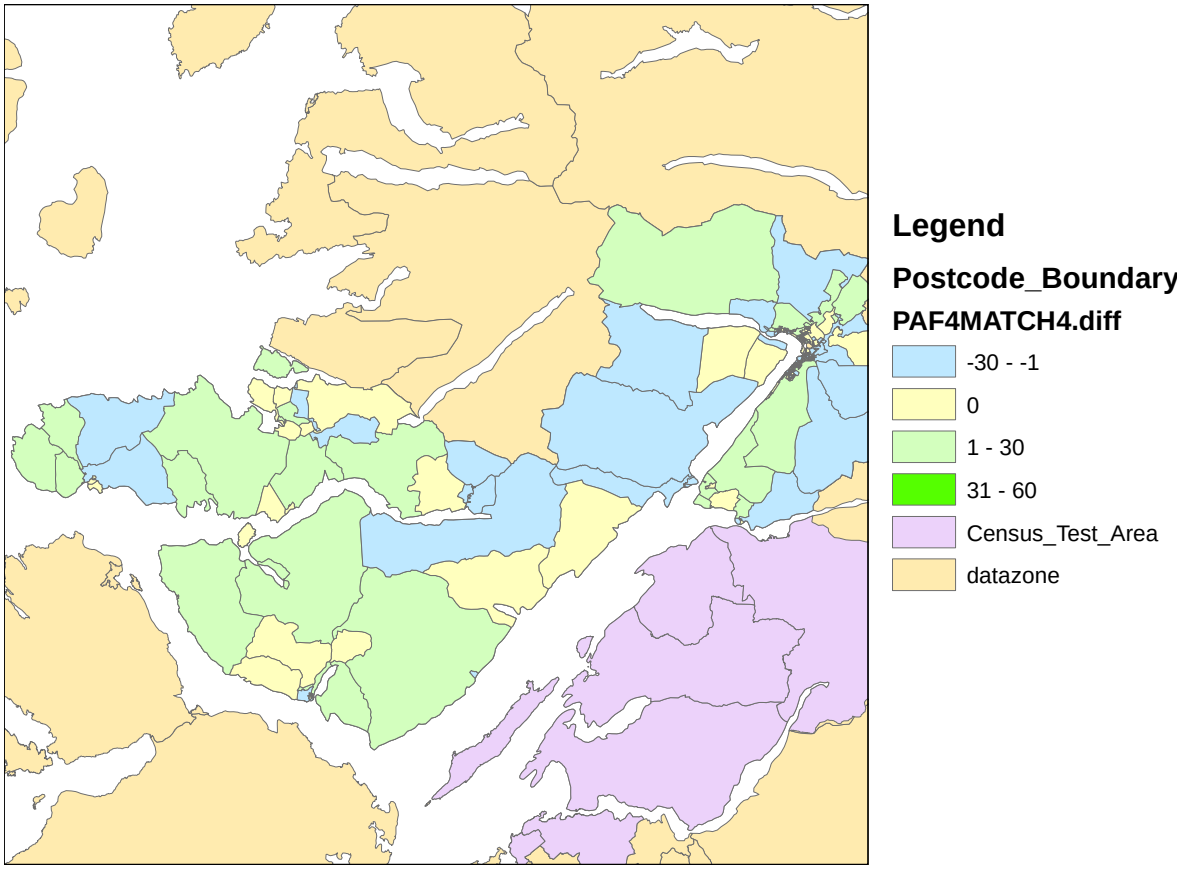
	Census Test	Postal Address File
Total number of residential dwellings recorded	11,429	10,839

8.1.21 Table 7.6 shows that there are approximately 600 more addresses present in the Census Test address list compared to the PAF for West Dunbartonshire.

Highland

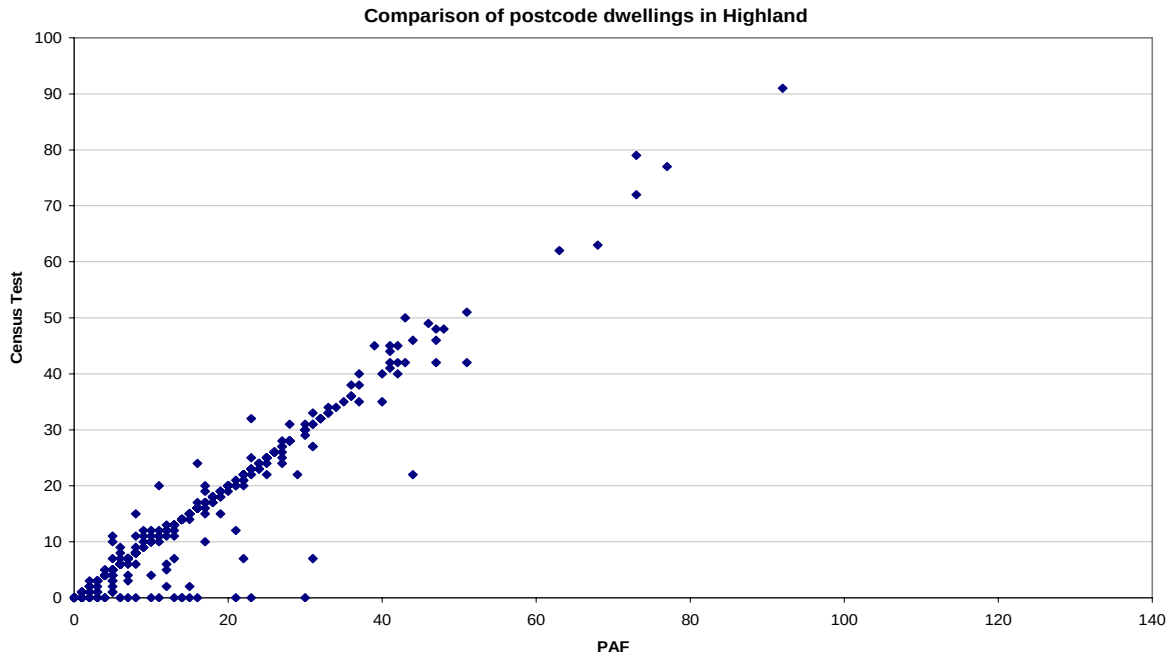
8.1.22 Figure 7.7 is the final map illustrating the comparisons between the number of dwellings per postcode recorded in the PAF and those recorded in the Census Test.

Figure 7.7: Highland: Number of dwellings recorded in Census Test address list, minus the number of dwellings recorded in PAF



8.1.23 It can be seen that in Figure 7.7, there are more colours incorporated in the map when compared to Figure 6.7. In Figure 6.7, the majority of the postcode boundaries were pale green, which meant that there were more dwellings recorded in the Census Test compared to the CAG. In contrast, more areas are pale yellow (an exact match) and even pale blue (more dwellings in PAF) in Figure 7.7. This means that the Census Test address checkers either found no changes to make, or removed some dwellings from the PAF for these areas. There are however some pale green areas in Figure 7.7, which means that some properties were added to the PAF address list. These same areas in Figure 6.7 also had more dwellings recorded in the Census Test compared to the CAG.

Figure 7.8: Number of dwellings in the Census Test compared to the PAF



8.1.24 Figure 7.8 shows that there are more dwellings recorded in the PAF compared to the Census Test, overall.

Table 7.7: Summary Table showing the differences observed between the PAF and Census Test address lists for Highland

	Actual Number of Postcode Boundaries	Percentage of Total Postcode Boundaries
>30 more dwellings recorded in the PAF than the Census Test	0	0.0%
1-30 more dwellings recorded in the PAF than the Census Test	100	22.6%
The same number of dwellings per postcode in both address lists	299	67.5%
1-30 more dwellings recorded in the Census Test than the PAF	44	9.9%
>30 more dwellings recorded in the Census Test than the PAF	0	0.0%
Total number of Postcode Boundaries	443	

Footnote: Percentages may not sum to 100 due to rounding

8.1.25 Table 7.7 shows that in over 67% of postcode areas in Highland, the number of dwellings completely matched. Over 22% of postcode areas recorded more properties in the PAF than the Census Test address list (probably non-residential properties), and in 10% of postcodes the address checkers and enumerators identified properties which had not been included in PAF.

Table 7.8: Total number of residential dwellings recorded in PAF and Census Test

	Census Test	Postal Address File
Total number of residential dwellings recorded	6,440	6,784

8.1.26 Table 7.8 shows that there are approximately 350 more addresses present in the PAF compared to the Census Test address list for Highland.

9. Potential for future work

There would be potential to take this work further – some suggestions are listed below. However, GROS does not currently have the resources to do this, so it would be necessary to identify the need and appropriate resources.

GROS did not have access to a complete CAG with postcodes for Argyll and Bute Council, so were unable to carry out the comparison to the Census Test. This information is now available.

Two files were missing from the West Dunbartonshire PAF, so the results provided above in Section 7.3 are partial and incomplete.

Due to time limits, it was not possible to conduct the same comparison work using the Assessor's Portal data. This would provide another view of the quality of the address lists. Unlike PAF and CAGs, the Assessors Portal should only contain residential properties (once the data have been filtered to remove garages and storage facilities), which may make it more useful for household surveys and the Census.

There may be scope for future work incorporating additional information, to see whether factors such as level of deprivation, urban/rural indicator and dwelling type have an impact on the quality of address lists.

10. Conclusions

No address list will ever be perfect, and there will always be changes since the date at which the address list was 'frozen'. For both the CAGs and PAF, there was a clear positive correlation between the number of dwellings recorded in the address list, and those identified on the ground during the Census Test. Overall, the quality of both address lists appeared to be very good but not perfect – additional properties were identified by the address checkers and enumerators carrying out the Census Test, which had not been included on the CAG or PAF. Some of these differences will be the result of the different dates that the address lists referred to (e.g., the October 2005 PAF was used, but the Census Test was not carried out until April 2006), and non-residential properties being included. While there were a few discrepancies between the two address lists, but the overall quality of the CAGs and PAF address lists appeared to be similar, which is re-assuring.

Both the PAF and CAG address lists contain some non-residential properties. This is not a problem for the Royal Mail or Local Authorities, but it would greatly enhance their usefulness for GROS/SE purposes if it was possible to identify non-residential properties by some form of flag. PAF does have an indicator to show whether a postcode is residential, non-residential, or mixed – but this is just at the postcode level, not individual properties. We have been told that a residential flag will be added to each address in the DNA-S address lists later this year.

For an exercise such as the Census, which has to cover every dwelling in Scotland, an address list provides an excellent starting point, and there may be scope to carry out an administrative exercise to 'clean-up' the address list used. But there will still be a need for address checkers and/or enumerators to check for any additional addresses, on the ground. Where Census forms are hand-delivered by enumerators, the enumerators can check the address lists and make any necessary changes.