
Report on the Comparison of Populations of Non-Standard Geographies Constructed from Different Small Area Geographies

Nick Wright & Donata Wasiuk

1. Summary.....	2
2. Introduction	3
3. Method	5
4. Results - National Parks:.....	6
5. Results - Community Health Partnerships:	7
6. Results - Westminster Parliamentary Constituencies:	9
7. Results - Scottish Parliamentary Constituencies:	12
8. Results - Multi-Member Wards:	15
9. Summary & Conclusions:	18

1. Summary

- 1.1. Five different geographies (National Parks, Community Health Partnerships, Westminster Parliamentary Constituencies, Scottish Parliamentary Constituencies and Multi-Member Wards) were constructed from three types of small area geography - Postcode Areas, Census Output Areas and Data Zones, using the 2001 boundaries for each.
- 1.2. Using the 2001 Census populations of the small area geographies, we estimated populations for the 5 large geographies. For each geography, this resulted in three different population estimates – one based on Postcodes, one based on Output Areas and one based on Data Zones.
- 1.3. The populations based on Postcode estimates were assumed to be the most accurate, as Postcodes are the smallest area for which populations are available and so fit most precisely into larger geographies. The accuracy of Output Area and Data Zone based population estimates was then compared against Postcode Based estimates.
- 1.4. Output Area based population estimates are very close to Postcode based population estimates for all geographies.
- 1.5. Data Zones produce very close estimates for Community Health Partnerships, Westminster Parliamentary Constituencies and Scottish Parliamentary Constituencies. They can also be used to produce close estimates for National Parks and MMWs – with an average (absolute) difference of 3% between Data Zone populations and Postcode populations for Multi-member Wards.

2. Introduction

- 2.1. The purpose of this report is to detail the work carried out in August 2007, on analysing how the populations of five different non-standard geographies varied if they were constructed out of Postcodes, Output Areas or Data Zones. There is an increasing demand for non-standard geography populations and this work was carried out in order to establish whether they could be built up from Data Zones or Output Areas (as Data Zones populations are updated every year, it would be very useful if they could be used to construct different geographies).
- 2.2. This work was carried out in conjunction with Scottish Neighbourhood Statistics (SNS), Scottish Executive Geographic Information Service (SEGIS) and the Geography and Population & Migration Statistics branches of the General Register Office for Scotland (GROS).
- 2.3. The purpose of the work was to assess how well three small area geographies – Postcodes, Output Areas and Data Zones could be matched to 5 different non-standard geographies, through seeing how the population estimate changed depending on which small area was used.
- 2.4. Population estimates were based on 2001 Census population estimates for Postcodes, Output Areas & Data Zones.
- 2.5. Non-standard geographies that were not based on Postcodes, Output Areas or Data Zones were chosen – and each of the three ‘small area’ geographies was then mapped to each of the non-standard geographies. The non-standard geographies investigated were:
- National Parks – There are 2 National Parks in Scotland – The Cairngorms National Park (created in 2003) and The Loch Lomond and The Trossachs National Park (created in 2002). Their boundaries were set by wide-spread public consultation with interested groups.
 - Community Health Partnerships (CHPs) – There are 40 Community Health Partnerships in Scotland, for the most part, these have been based on the 32 Local Authorities however 3 Local Authorities (Fife, Glasgow and Highland) were split into multiple CHPs.
 - Westminster Parliamentary Constituencies (WPCs) – Since 2005 there have been 59 Westminster Parliamentary Constituencies. Their boundaries are set by the Boundary Commission for Scotland and are based on Electoral Wards.
 - Scottish Parliamentary Constituencies (SPCs) – There are 73 Scottish Parliamentary Constituencies – they were set in 1990 by the Boundary Commission for Scotland.

- Multi-Member Wards (MMWs) – There are 353 Multi-Member Wards that were created in 2007 by the Local Government Boundary Commission as a replacement for the Electoral Wards. Their boundaries are loosely based on the old Electoral Ward boundaries.

2.6. Three small area units were used to construct the non-standard geographies listed above, all using their 2001 boundaries (as Postcode and Output Area populations are only available in the 2001 Census).

- Postcodes – In 2001, there were 139,045 Postcodes – these are the smallest geographical area for which population data is available (they only have the numbers, no age / sex breakdown). These were assumed to provide the best estimate for the non-standard geographies.
- Output Areas – Output Areas are the standard geographies for reporting the 2001 Census results and contain an age / sex breakdown of the population. There are 42,604 Output Areas, constructed from Postcodes.
- Data Zones – There are 6,505 Data Zones, Data Zones are the standard Small Area geography used by the Scottish Government and were constructed from Output Areas in 2001 – again, these have an age / sex breakdown of the population.

2.7. Although each 2001 Postcode fitted perfectly into a 2001 Output Area, which in turn fitted perfectly into 2001 Data Zones, this had no bearing on how each was assigned to non-standard geographies, as the population weighted centroid of each small area unit was used to assign them to non-standard geographies, without reference to the small area hierarchy of Postcodes within Output Areas within Data Zones.

2.8. Population weighted centroids are a standard GIS procedure for assigning the population of a small geography to a large geography if the small geography does not wholly fit within the boundaries of the large geography – or lies across the border of two large geographies. The population weighted centroid is essentially the point in the area where population density is the same all around the point. For many cases, this will be somewhere with the highest population density within the area. The small geography is assigned to whichever larger geography the population weighted centroid lies in.

2.9. As Postcodes are the smallest geographic area, in practice, the Postcode construction of non-standard geographies was the closest to their actual boundaries. Therefore, the Postcode based population estimates were considered closest to the actual populations of the non-standard geographies (however, most indicators, including age and sex breakdowns are only available at Output Area and / or Data Zone level).

3. Method

- 3.1. For each non-standard geography, three look-up tables (showing how each type of small area unit could be used to construct the larger non-standard geographies) were provided by the Geography branch of GROS. Following SNS policy, the population weighted centroid of each small area unit (Postcodes, Output Areas & Data Zones) was used to assign it to the non-standard geographies.
- 3.2. Each Postcode, Output Area and Data Zone could only be assigned to one area in each non-standard geography set. Each of the three types of small area unit was assigned to the non-standard geographies independently of how other types of small area units were assigned.
- 3.3. For example, Postcode 'X' could be contained within Data Zone 'G' and assigned to Multi-Member Ward 'A'. However, this assignment would be based entirely on where the population weighted centroid of Postcode 'X' fell, regardless of whether Data Zone 'G' was assigned to Multi-Member Ward 'A' or not.
- 3.4. Using the 2001 populations of each Postcode, Output Area and Data Zone (from the 2001 Census Household Populations) and the look-up of each small area unit to each non-standard geography, populations for each construction of each non-standard geography were produced. For Data Zones and Output Areas, age and sex breakdowns of the population were also available. Postcodes are too small to provide this level of detail at, so just the total population of each Postcode is available.
- 3.5. For example, for each Community Health Partnership, the populations for the geography were constructed from Postcodes, Output Areas and Data Zones. Therefore, for each non-standard geography, three different population estimates were produced.
- 3.6. Once the population estimates were completed for each non-standard geography, they were compared against each other, using the estimates based on Postcodes – assumed to be the closest estimate to the actual population.
- 3.7. The results for each non-standard geography are listed fully in separate Excel workbooks (as reproducing the entire set would take up a large amount of space). The summarised results are presented below.

4. Results - National Parks:

4.1. The table below shows compares different population estimates for National Parks.

Table 1 – National Park Population Estimates

National Park	Postcode based population	Output Area based population	Difference – OA / Postcode (%)	Data Zone based population	Difference – DZ / Postcode (%)
The Cairngorms National Park	15,759	15,864	0.67	15,461	-1.89
The Loch Lomond and The Trossachs National Park	15,261	15,365	0.68	14,685	-3.77

4.2. Overall, the differences in estimated populations for National Parks are relatively small (reflecting their relatively remote locations).

5. Results - Community Health Partnerships:

5.1. CHPs are generally based on local authority boundaries which are a standard geography (which match Data Zone boundaries and, by extension, Output Area and Postcode boundaries). For 26 of the 40 CHPs, the populations are the same, whether they are constructed from Postcode, Output Areas or Data Zones. The other 14 arise from CHPs which have been constructed from 'split' Local Authorities – or where the CHP boundary does not exactly match the local authority boundary.

5.2. Table 2 below shows the CHPs where the population estimates do not match. In the case where a local authority has been divided up, aggregating the CHPs back to local authority level resolves the differences in population estimates (for example, the three Highland CHPs in the table above). In cases where the problem is a slight mismatch between local authority borders and CHP borders, adding the two local authorities together (and so removing the border) resolves the error (for example, Aberdeen City and Aberdeenshire CHPs above).

5.3. Chart 1 below details the frequency of proportional differences between Data Zone based estimates vs. Postcode based estimates and Output Area based estimates vs. Postcode based estimates.

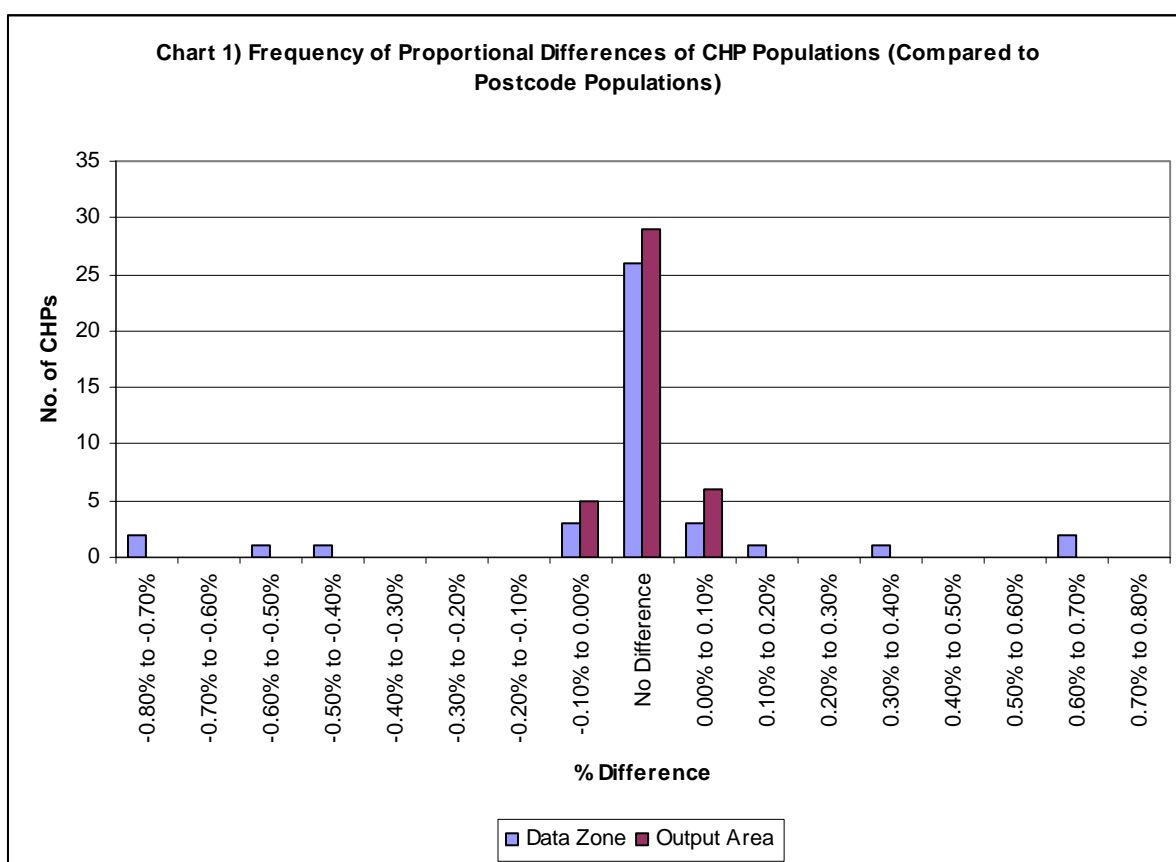


Table 2 – CHP Population Estimates

CHP	Postcode based population	Output Area based population	Difference – OA / Postcode (%)	Data Zone based population	Difference – DZ / Postcode (%)
Aberdeen City CHP	212,098	212,125	0.01	212,125	0.01
Aberdeenshire CHP	226,898	226,871	-0.01	226,871	-0.01
East Glasgow CHCP	127,321	127,323	0.00	127,556	0.19
North Glasgow CHCP	100,538	100,521	-0.02	99,821	-0.71
West Glasgow CHCP	133,478	133,493	0.01	133,960	0.36
South East Glasgow CHCP	101,939	101,863	-0.08	101,176	-0.75
South West Glasgow CHCP	114,593	114,669	0.07	115,356	0.67
West Dunbartonshire CHP	93,395	93,378	-0.02	93,378	-0.02
Argyll & Bute CHP	91,289	91,306	0.02	91,306	0.02
Mid Highland CHP	86,653	86,653	0.00	87,215	0.65
North Highland CHP	38,669	38,669	0.00	38,462	-0.54
South East Highland CHP	83,592	83,592	0.00	83,237	-0.43
West Lothian CHCP	158,720	158,714	0.00	158,714	0.00
Edinburgh North CHP	448,618	448,624	0.00	448,624	0.00

Average population of a Community Health Partnership area = 126,550

6. Results - Westminster Parliamentary Constituencies:

- 6.1. As the WPC boundaries were drawn up by the Boundary Commission for Scotland, they have no relation to any standard geography, so their borders can cut across Postcode, Output Areas and Data Zones.
- 6.2. When population estimates based on Output Areas are compared to population estimates based on Postcodes, 9 are the same, 24 WPC have higher Output Area based population estimates and 26 have higher Postcode based population estimates.
- 6.3. When comparing Output Area derived estimates to Postcode derived estimates, the largest positive difference comes to 425 extra people in the Paisley & Renfrewshire North WPC (+0.49%) and the largest negative difference comes to 425 fewer people in Paisley & Renfrewshire South WPC (-0.50%). On average, the differences between Output Area and Postcode derived populations are small – around 0.12 per cent per WPC.
- 6.4. When population estimates based on Data Zones are compared to population estimates based on Postcodes, 8 are the same, 25 WPC have higher Data Zone based population estimates and 26 have higher Postcode based population estimates.
- 6.5. When comparing Data Zone derived estimates to Postcode derived estimates, the largest positive difference comes to 3,004 extra people in the Glasgow North West WPC (+3.7%) and the largest negative difference comes to 2,443 fewer people in Glasgow North WPC (-3.5%).
- 6.6. On average, the differences between Data Zone and Postcode derived populations are around minus 0.53 per cent per WPC. Assuming a cut-off of plus or minus 2.5 per cent for Data Zone vs. Postcode populations, only 2 WPCs fall outside the cut-off points.
- 6.7. Charts 2a below details the frequency of proportional differences between Data Zone based estimates vs. Postcode based estimates and Chart 2b details Output Area based estimates vs. Postcode based estimates.
- 6.8. The quartile absolute differences and quartile percentage differences of the estimates based on Data Zones or Output Areas against estimates based on Postcodes for WPCs are shown in Table 3 below.

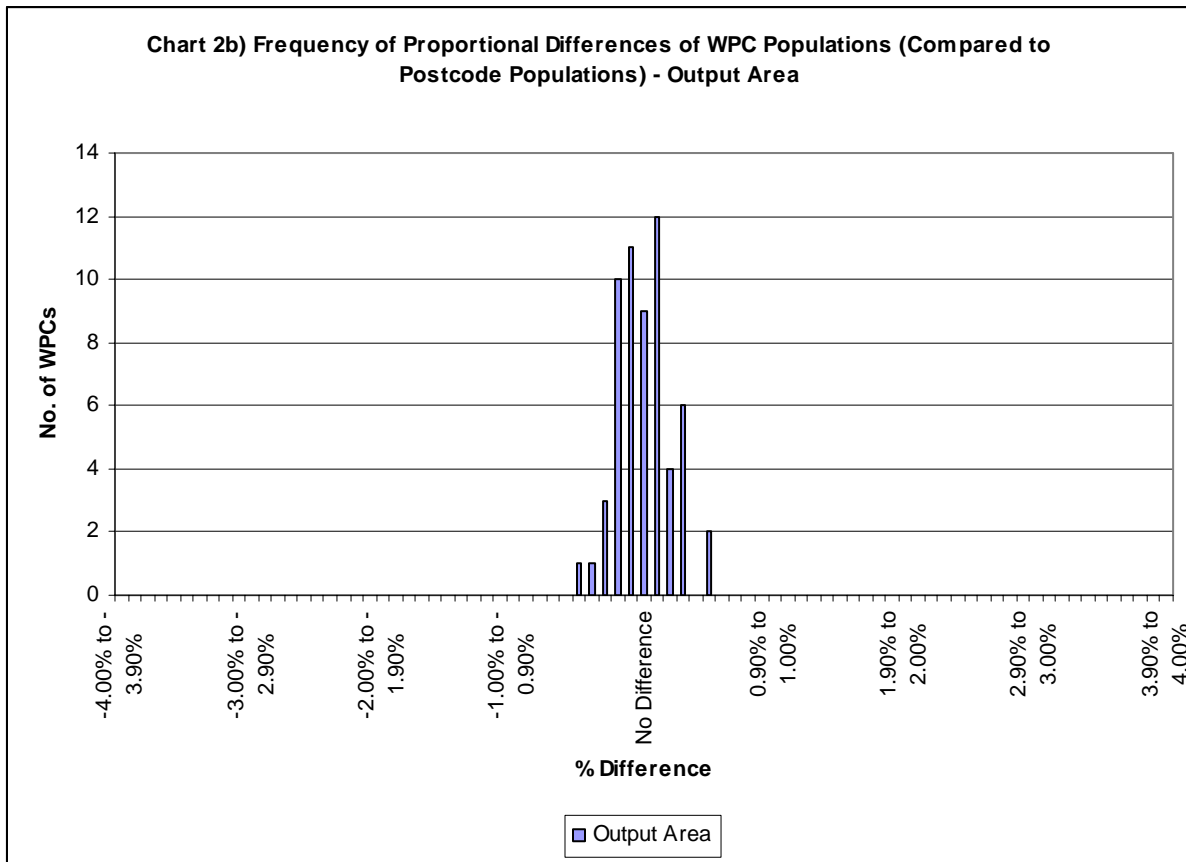
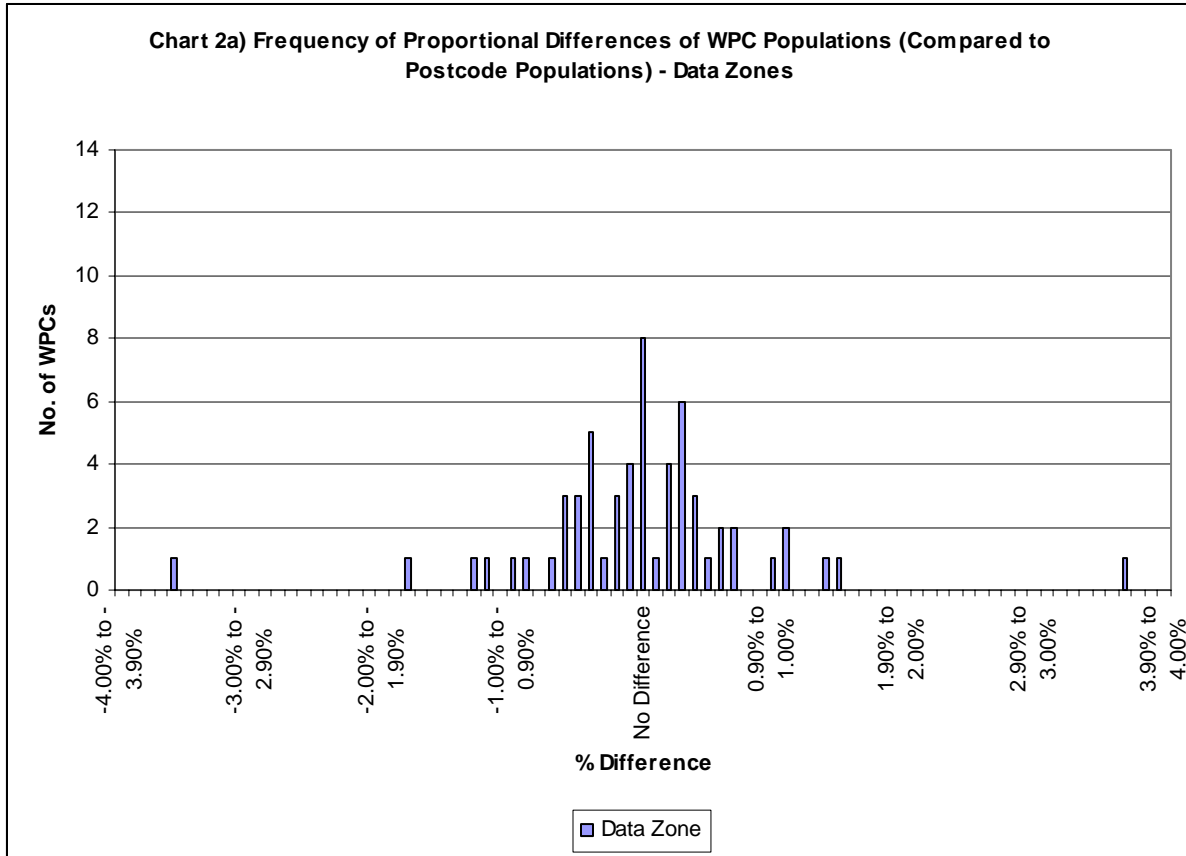


Table 3 – WPC Population Estimates

WPC Summary	Output Area vs. Postcode (Abs.)	Data Zone vs. Postcode (Abs.)	Output Area vs. Postcode (%)	Data Zone vs. Postcode (%)
Min	-425	-2,443	-0.50	-3.53
Q1	-87.5	-351	-0.10	-0.37
Q2 (Median)	0	0	0.00	0.00
Q3	55	245	0.06	0.28
Max	425	3,004	0.49	3.70

Average population of a Westminster Parliamentary Constituency area = 85,980

7. Results - Scottish Parliamentary Constituencies:

- 7.1. Similarly to WPCs, SPC boundaries were drawn up by the Boundary Commission for Scotland and, as a result, there are several differences in the population estimates, depending on which small area unit was used to estimate them.
- 7.2. When population estimates based on Output Areas are compared to population estimates based on Postcodes, 6 are the same, 34 SPCs have higher Output Area based population estimates and 33 have higher Postcode based population estimates.
- 7.3. When comparing Output Area derived estimates to Postcode derived estimates, the largest positive difference comes to 457 extra people in the Glasgow Kelvin SPC (+0.7%) and the largest negative difference comes to 458 fewer people in the Glasgow Anniesland SPC (-0.7%). On average, the differences between Output Area and Postcode derived populations are small – around 0.13 per cent per SPC.
- 7.4. When population estimates based on Data Zones are compared to population estimates based on Post Codes, 5 are the same, 35 SPCs have higher Data Zone based population estimates and 33 have higher Postcode based population estimates.
- 7.5. When comparing Data Zone derived estimates to Postcode derived estimates, the largest positive difference comes to 1,036 extra people in the Dumfries SPC (+1.3%) and the largest negative difference comes to 1,036 fewer people in the Galloway & Upper Nithsdale SPC (-1.6%).
- 7.6. On average, the differences between Data Zone and Postcode derived populations are around 0.51 per cent per SPC. Assuming a cut-off of plus or minus 2.5 per cent for Data Zone vs. Postcode populations, no SPCs fall out of the cut-off points.
- 7.7. Charts 3a below details the frequency of proportional differences between Data Zone based estimates vs. Postcode based estimates and Chart 3b details Output Area based estimates vs. Postcode based estimates.
- 7.8. The quartile absolute differences and quartile percentage differences of the estimates based on Data Zones or Output Areas against estimates based on Postcodes for SPCs are shown in Table 4 below.

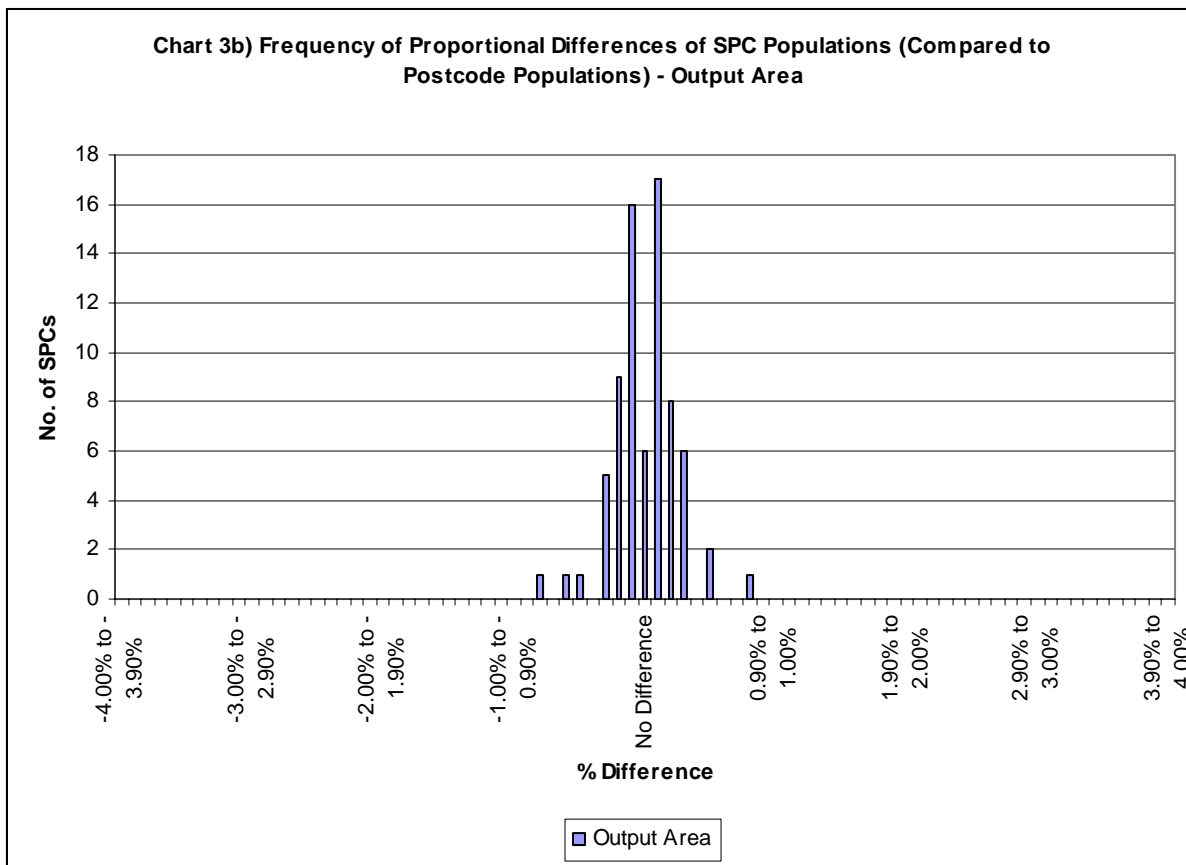
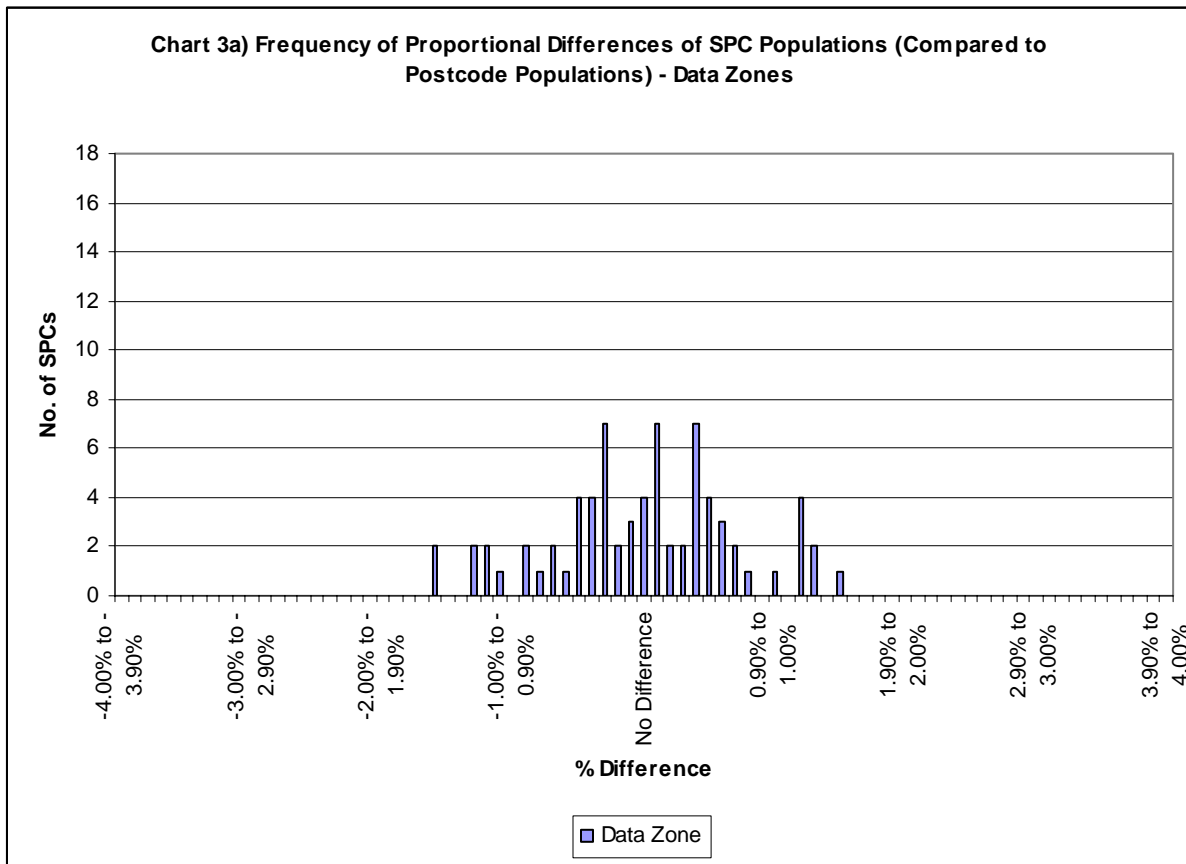


Table 4 – SPC Population Estimates

SPC Summary	Output Area vs. Postcode (Abs.)	Data Zone vs. Postcode (Abs.)	Output Area vs. Postcode (%)	Data Zone vs. Postcode (%)
Min	-458	-1,036	-0.71	-1.58
Q1	-45	-232	-0.06	-0.34
Q2 (Median)	0	0	0.00	0.00
Q3	58	287	0.09	0.38
Max	457	1,036	0.70	1.49

Average population of a Scottish Parliamentary Constituency area = 69,340

8. Results - Multi-Member Wards:

- 8.1. As a relatively small geography, with non-standard boundaries drawn up by the Local Authority Boundary Commission, matches to MMWs are likely to be quite difficult.
- 8.2. When population estimates based on Output Areas are compared to population estimates based on Postcodes, 11 are the same, 174 MMWs have higher Output Area based population estimates and 168 have higher Postcode based population estimates.
- 8.3. When comparing Output Area derived estimates to Postcode derived estimates, the largest positive difference comes to 371 extra people in the Lenzie and Kirkintilloch South MMW (+2.6%) and the largest negative difference comes to 304 fewer people in the Partick West MMW (-2.6%). On average, the differences between Output Area and Postcode derived populations are around 0.50 per cent per MMW.
- 8.4. When population estimates based on Data Zones are compared to population estimates based on Postcodes, 9 are the same, 174 MMWs have higher Data Zone based population estimates and 170 have higher Postcode based population estimates.
- 8.1. When comparing Data Zone derived estimates to Postcode derived estimates, the largest positive difference comes to 1,486 extra people in the Dunfermline North MMW (+12%) and the largest negative difference comes to 1,505 fewer people in the Rutherglen Central & North MMW (-11%).
- 8.2. On average, the differences between Data Zone and Postcode derived populations are around 3.07 per cent per MMW. Assuming a cut-off of plus or minus 2.5 per cent for Data Zone vs. Postcode populations, 150 MMWs fall out of the cut-off points.
- 8.3. Charts 4a below details the frequency of proportional differences between Data Zone based estimates vs. Postcode based estimates and Chart 4b details Output Area based estimates vs. Postcode based estimates.
- 8.4. The Min, 10th, 25th, 50th, 75th, 90th & Max percentile absolute differences and quartile percentage differences of the estimates based on Data Zones or Output Areas against estimates based on Postcodes for MMWs are shown in Table 5 below.

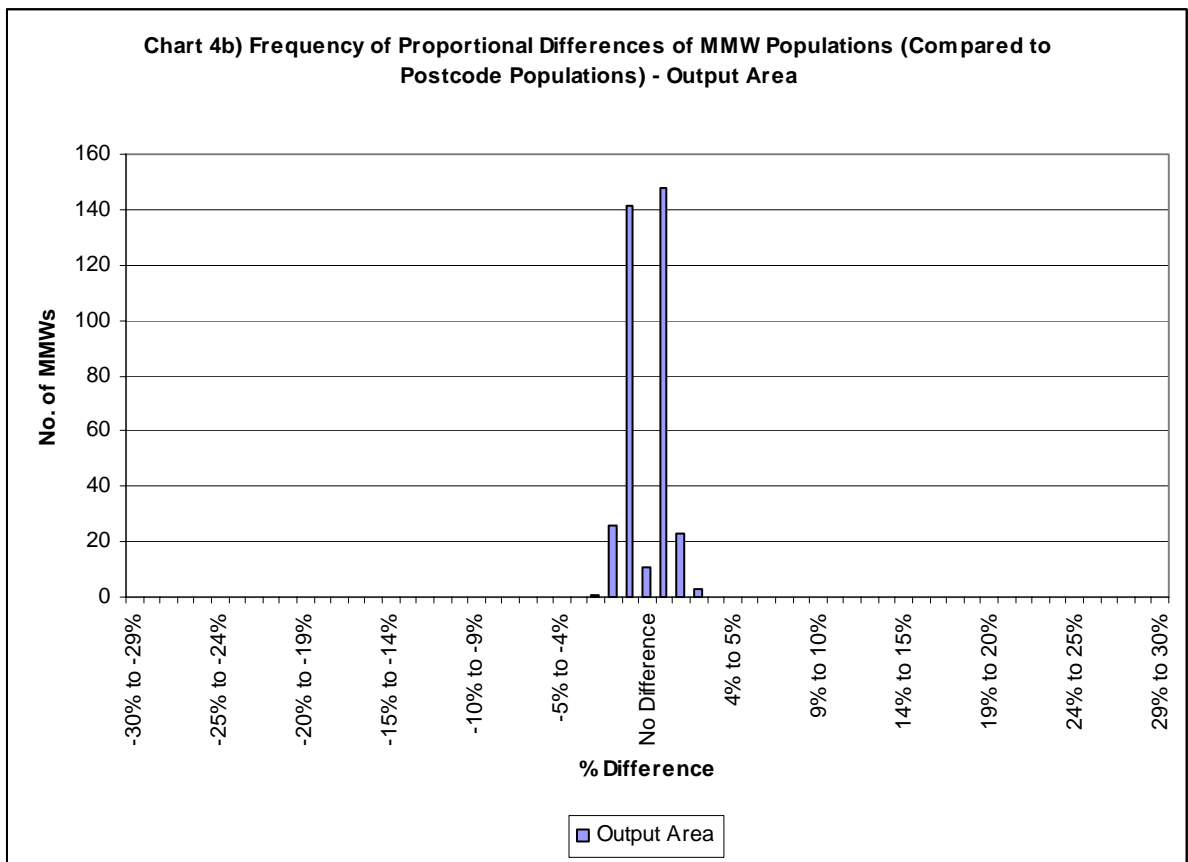
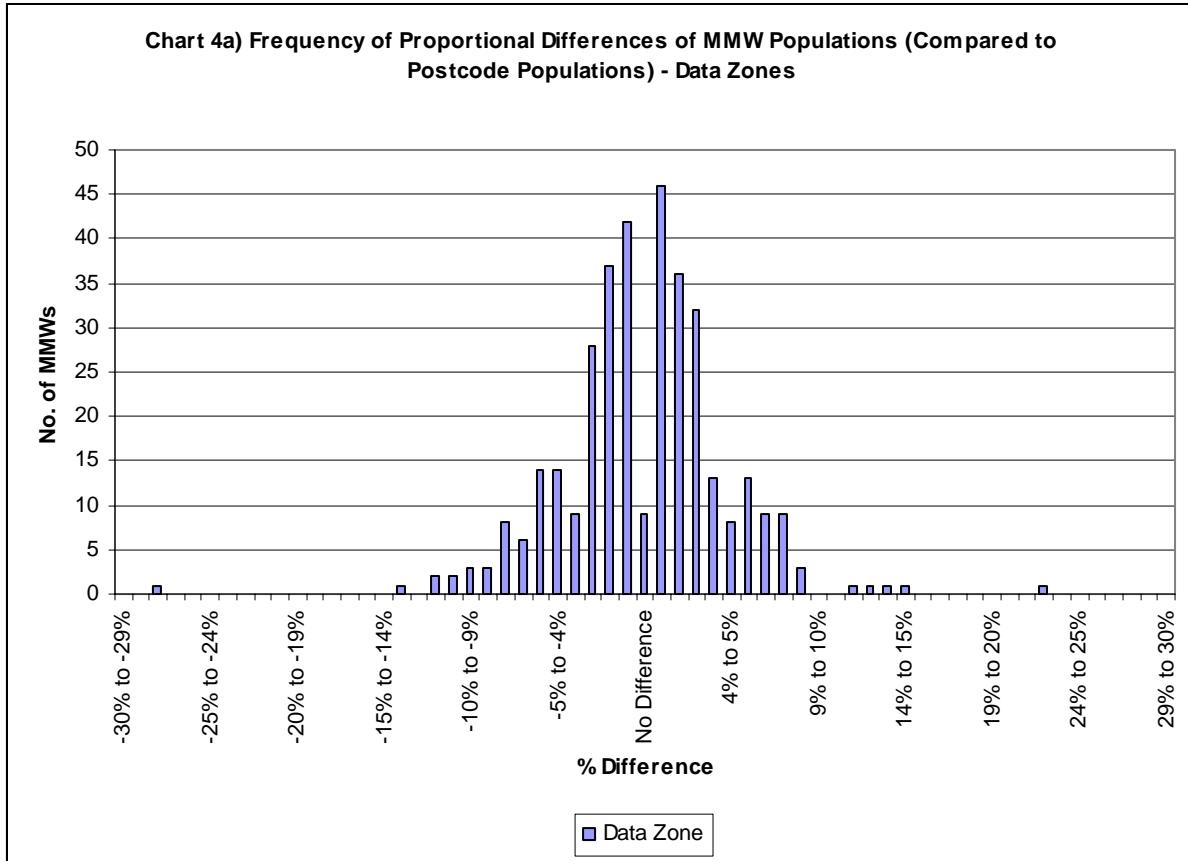


Table 5 – MMW Population Estimates

MMW Summary	Output Area vs. Postcode (Abs.)	Data Zone vs. Postcode (Abs.)	Output Area vs. Postcode (%)	Data Zone vs. Postcode (%)
Min	-304	-1,505	-2.59	-27.04
10%	-109	-597	-0.90	-5.26
25% (Q1)	-46	-287	-0.33	-2.12
50% (Median)	0	0	0.00	0.00
75% (Q3)	43	285	0.31	2.11
90%	101	624	0.77	5.39
Max	371	1,486	2.62	22.31

Average population of a Multi-member Ward area = 14,340

9. Summary & Conclusions:

9.1. The average proportional difference of the estimates based on Data Zones or Output Areas against estimates based on Postcodes are shown in the table below.

Table 6 – Summary Table

		Output Area based estimates		Data Zone based estimates	
Non-standard Geography	Average Population	No. of areas with different populations to Postcode based estimates	Av. difference to population estimated by postcodes (%)	No. of areas with different populations to Postcode based estimates	Av. difference to population estimated by postcodes (%)
National Parks (2)	15,510	2 / 2 (100%)	0.67	2 / 2 (100%)	2.83
CHPs (40)	126,550	11 / 40 (27.5%)	0.06	14 / 40 (35%)	0.11
WPCs (59)	85,980	50 / 59 (84.7%)	0.12	51 / 59 (86.4%)	0.53
SPCs (73)	69,340	67 / 73 (91.8%)	0.13	68 / 73 (93.2%)	0.51
MMWs (353)	14,340	342 / 353 (96.9%)	0.50	344 / 353 (97.5%)	3.07

9.2. Assuming that Postcode based estimates are the closest available estimates to the true population of non-standard geographies, then Output Areas (only available for the 2001 Census) provide estimates that are very close to Postcode based estimates for all geographies.

9.3. Data Zones (available from 2001 – current) can be used to produce very close estimates for CHPs, WPCs and SPCs. They can also be used to produce close estimates for National Parks and MMWs (averaging 3% difference in population estimates).