

Towards 2011 output geographies: Exploring the need for, and challenges involved in, maintenance of the 2001 output geographies

Samantha Cockings, Andrew Harfoot,
Duncan Hornby
University of Southampton

This article describes and presents early results from the ESRC-funded Census 2011Geog project, which aims to develop and evaluate automated procedures to maintain (split, merge or re-design) the 2001 Census output geographies in order to create the 2011 output geographies for England and Wales. The article explores population change at the small area level between 2001 and 2005–06, and considers the extent to which the 2001 Census output geographies are likely to be appropriate for the release of 2011 Census data. It concludes that the vast majority of output geography areas are unlikely to have breached population thresholds by 2011, but that a small proportion of areas will require maintenance. The article finishes with a discussion of the key decisions that need to be made before the automated procedures can be implemented operationally.

Introduction

The 2001 Census output geographies for England and Wales were designed to be an optimised representation of the population distribution and socio-economic characteristics at that time. By the next Census in 2011 there will have been changes in the size and composition of population in most areas. While recognising this, the National Statistics Small Area Geography Consultation in 2007¹ revealed strong user demand for output geography stability. The challenges involved in creating 2011 output geographies that maintain both a high degree of stability and also reflect real-world population changes are non-trivial. This article introduces the ESRC-funded Census 2011Geog project, which aims to develop automated methods for maintaining (splitting, merging or re-designing) the 2001 output geographies in order to create output geographies for 2011. The article presents preliminary results from the first stage of this project, exploring small-area population change in England and Wales between 2001 and 2005–06, and considering the extent to which the 2001 output geographies are likely to be appropriate for use in 2011. It also reviews the key decisions that must be made before the maintenance procedures can be implemented. Note that this article is not specifically concerned with whether or not to re-align the boundaries of the output geographies to real-world features, although this is a relevant and related issue.

Background: the 2001 output geographies

The smallest zones for which Census data were released in 2001 in England and Wales were output areas (OAs). These OAs were created using a process of automated zone design following the collection and processing of the household-level 2001 Census data.^{2,3} Automated zone design involves two key methodological stages.⁴ First, a set of small

building blocks is created. Second, these building blocks are iteratively aggregated into larger zones, with the aim of optimising an objective function based on pre-specified design criteria. The building blocks employed for the 2001 Census OAs were postcode polygons. Geographic Information Systems (GIS) were used to create small, space-filling, polygons around the addresses of households enumerated by the Census. Adjacent address polygons belonging to the same postcode were then merged to create postcode polygons. The boundaries of wards and parishes were then intersected to create a set of 'ward-parts' and the postcode polygon boundaries were constrained to nest within these, as well as being made to coincide with road centre lines where possible.

These synthetic postcode polygons were then aggregated to create OAs. All OAs had to exceed specified minimum population (100) and household (40) thresholds in order to protect individuals from inadvertent disclosure in the aggregate data. Note that no maximum thresholds were specified for the OA creation process. The OAs within each ward-part were then iteratively re-combined, using multiple random restarts, in order to identify the set of OAs which best optimised a set of design criteria. The criteria were: homogeneity of population size across OAs (aiming for a target mean of 125 households); internal homogeneity of accommodation type and tenure within OAs; and compactness of shape. The OAs from all ward-parts were then merged to form a national set of OAs. These OAs subsequently became the building blocks for sets of larger 'neighbourhood' geographies, namely the Lower Layer and Middle Layer Super Output Areas (LSOAs and MSOAs respectively). These LSOAs and MSOAs are now well established geographies for the release of neighbourhood statistics at www.neighbourhood.statistics.gov.uk. Similar zone design criteria to those used to generate the OAs were employed in the creation of the LSOAs and MSOAs, including minimum population thresholds of 1,000 and 5,000 respectively. Note that the MSOAs were the only output geography layer to have a published upper threshold (4,000 households). Importantly, the boundaries of all of these output geographies were made freely available for non-commercial use.

Population change since 2001

While the output geographies were optimised for certain population and socio-economic characteristics in 2001, changes in the population size and distribution since then are likely to mean that in some areas the 2001 output geographies will no longer be appropriate for representing the population or for maintaining confidentiality. The key drivers of population change since 2001 have included migration, an ageing population, people marrying later and higher divorce and separation rates. These factors have led to a reduction in mean household size and a consequent rise in the number of residential properties required, together with a greater demand for smaller properties. Residential development has primarily comprised the building of new properties (mainly on either green-field or brown-field sites) and the sub-division of existing properties. A minority of areas since 2001 have experienced population decline; where this has happened, it has mainly been due to internal outward migration. All of the above changes will not only have led to changes in the population size and distribution within the output geographies since 2001 but also, potentially, to changes in the homogeneity of the socio-economic characteristics of the areas.

In planning for the 2011 Census it is important to estimate how much change there has been since 2001, and to what extent the output geographies will have breached population thresholds by 2011. It is also important to understand the nature of these changes, especially in terms of the types of breaches and their geographical distribution. This will enable the development and evaluation of methodologies that can take the 2001 output geographies and modify them, where appropriate, in order to create the 2011 output geographies, preferably using automated procedures.

The Census 2011Geog project

The Economic and Social Research Council (ESRC), via its Census Programme, is funding a collaborative research project (the 'Census 2011Geog' project, <http://census2011geog.census.ac.uk>) between the University of Southampton and the Office for National Statistics (ONS). The aim of the project is to create automated procedures for maintaining (that is, splitting, merging or re-designing) the 2001 output geographies in order to create the 2011 output geographies. It also aims to investigate the implications of using different building blocks, such as postcodes and street blocks, for these maintenance procedures. The project, which builds on previous experiments by ONS,⁵ will deliver prototype software that can be tested by ONS following the Census Rehearsal in 2009. This can then be adapted for operational use in the 2011 Census. It will also deliver an evidence base of the implications of using different building blocks and design criteria for the maintenance procedures. The first stage of the project has involved an exploration of the likely magnitude and geographical distribution of population change and consequent breaches in the output geographies. The results of this analysis form the basis of the findings presented here.

Use of ONS mid-year population estimates

Available ONS mid-year population estimates for 2001 to 2006 at local authority level and for 2001 to 2005 at OA, LSOA and MSOA levels were employed to investigate population change since 2001. Eighteen OAs were excluded from the OA-level analysis as they contained no postcodes and therefore did not receive any population via the postcode-best fit method employed to create the mid-year estimates at OA level. Note also that the MSOA counts include the Isles of Scilly pseudo-MSOA. The mid-year estimates of usual resident population include adjustments for births, deaths and migration.⁶ In addition, the 2001 mid-year estimates included specific adjustments/corrections for under-enumeration at Census. This under-enumeration arose for many reasons,⁷ but particularly as a result of problems experienced with the address register in certain areas.⁸ In assessing population change since the 2001 Census, it is therefore more appropriate to compare the 2005–06 mid-year estimates with the 2001 mid-year estimates, rather than with the actual 2001 Census counts, as this gives a more reliable estimate of population change. A further consideration when assessing population change over time is that some areas have high proportions of special populations that can be highly mobile, such as members of the armed forces. Changes in the geographical distribution of such populations can result in an apparent increase or decrease in an area's usually resident population since Census, even if the underlying non-special population is actually reasonably stable. This should be borne in mind when interpreting the results of the analyses presented here.

Population change at Local Authority level, 2001–06

Map 1 shows absolute population change between the 2001 and 2006 mid-year estimates for local authorities in England and Wales. The average population change across all local authorities in England between 2001 and 2006 was a 2.9 per cent increase, representing an average growth in population of 3,710 per local authority. By contrast, the levels of population change over the same time period in Welsh unitary authorities were lower, with an overall average increase of 1.8 per cent (2,523 people).

These averages hide considerable geographical variation between local authorities of different area types. In order to explore this further, local authorities were classified by their Department of Environment, Food and Rural Affairs (DEFRA) urban/rural category⁹ and then ranked within

Table 1

Local Authorities in England exhibiting highest population increases between mid-year estimates in 2001 and 2006, in each DEFRA urban/rural category

DEFRA Classification	Local Authority	Mid-2001 population	Mid-2006 population	Change (number)	Change (per cent)
1. Major Urban	Westminster	203,329	231,874	28,545	14.0
	Camden	202,567	227,453	24,886	12.3
	Kensington and Chelsea	162,199	178,021	15,822	9.8
	Manchester	422,915	451,984	29,069	6.9
	Tower Hamlets	201,090	212,804	11,714	5.8
2. Large Urban	Nottingham	268,939	286,378	17,439	6.5
	Bristol City of	390,049	410,487	20,438	5.2
	Portsmouth	188,043	196,379	8,336	4.4
	Wyre	105,800	110,371	4,571	4.3
3. Other Urban	Southampton	219,539	228,635	9,096	4.1
	Oxford	135,509	149,105	13,596	10.0
	Welwyn Hatfield	97,550	105,514	7,964	8.2
4. Significant Rural	Canterbury	135,381	146,181	10,800	8.0
	Exeter	111,180	119,606	8,426	7.6
	Cambridge	109,941	117,913	7,972	7.3
	South Derbyshire	81,738	89,779	8,041	9.8
	Colchester	156,016	170,846	14,830	9.5
	Ashford	103,024	111,177	8,153	7.9
5. Rural-50	Kettering	82,304	87,858	5,554	6.7
	Lancaster	134,049	143,033	8,984	6.7
	East Northamptonshire	76,835	83,954	7,119	9.3
	North Somerset	188,840	201,404	12,564	6.7
	Kerrier	92,634	98,008	5,374	5.8
6. Rural-80	Tonbridge and Malling	107,771	113,937	6,166	5.7
	Braintree	132,482	139,688	7,206	5.4
	South Northamptonshire	79,497	88,764	9,267	11.7
	Forest Heath	56,145	62,129	5,984	10.7
	Rutland	34,598	38,277	3,679	10.6
	North Kesteven	94,378	103,152	8,774	9.3
	Mid Bedfordshire	121,258	132,185	10,927	9.0

these categories according to their percentage increase or decrease in population between 2001 and 2006 (again using the mid-year estimates). **Table 1** and **Table 2** show the five English local authorities exhibiting the greatest population increases and decreases respectively, in each DEFRA urban/rural category. Note that for the most strongly rural categories ('significant rural', 'rural-50' and 'rural-80'), there were less than five local authorities experiencing a decrease in their population and so only those experiencing a decrease are shown. No similar urban/rural classification scheme was available for Welsh unitary authorities; instead population change for all Welsh unitary authorities is shown in **Table 3**.

It is clear from this analysis that a greater proportion of local authorities experienced population growth (90.4 per cent of English local authorities and 91.7 per cent of Welsh UAs) than population decline (9.6 per cent and 8.3 per cent for England and Wales respectively), and that the relative magnitude of the growth was greater than that of the decline. The levels of growth were reasonably similar across the various urban/rural categories: whilst the greatest increases were seen in the major urban areas (such as Westminster, Camden and Oxford), there was also significant growth in rural areas (for example, South Northamptonshire, Forest Heath and Rutland). By contrast, the levels of decline were less consistent: the greatest declines were seen in urban areas (such as Sefton, Middlesbrough and Rushmoor), whereas very few rural areas (other than Bridgnorth) declined substantially.

While the above analysis provides useful information about the general trends in population change since 2001, in planning for the 2011 Census it is more important to explore the extent to which the output geographies themselves have breached specified thresholds, as it is

stability of the output geographies (particularly at the OA and LSOA levels) which forms the basis of ONS's small area geography policy and users' preferred requirements (ONS, 2007). The mid-year estimates for 2001 to 2005 were therefore employed to investigate this in more detail.

Threshold breaches in the output geographies

The number of output geography areas breaching upper and lower thresholds by 2011 will be contingent upon the specific thresholds employed. ONS has not yet confirmed the thresholds to be employed in 2011, but it is likely that the levels will be similar to those used in 2001. Assuming this is the case, similar thresholds can be employed to explore the extent to which OAs, LSOAs and MSOAs had breached lower and upper thresholds by 2005. **Table 4** shows the population thresholds employed here, including our working definition of 'upper thresholds' which were not formally defined for Census purposes. It was not possible to explore household threshold breaches as inter-censal ONS mid-year estimates are not produced for households. Population thresholds were calculated by multiplying household thresholds by a factor of 2.5 (designed to approximate to mean household size).

Number of breaches

Table 5 shows the numbers and percentages of OAs, LSOAs and MSOAs in England and Wales breaching and/or within threshold, in either or both of 2001 and 2005. The columns show the number (or percentage) of areas below, within or above threshold in 2005, whereas the rows show the number (or percentage) below, within or above threshold in 2001. The cells in the matrix are therefore cross-tabulations of these various

Table 2

Local Authorities in England exhibiting highest population decreases between mid-year estimates in 2001 and 2006, in each DEFRA urban/rural category

DEFRA Classification	Local Authority	Population mid-2001	Population mid-2006	Change (number)	Change (per cent)
1. Major Urban	Sefton	282,884	277,421	-5,463	-1.9
	Sunderland	284,601	280,593	-4,008	-1.4
	Stockport	284,557	280,619	-3,938	-1.4
	Liverpool	441,858	436,072	-5,786	-1.3
	South Tyneside	152,793	151,020	-1,773	-1.2
2. Large Urban	Middlesbrough	141,233	138,434	-2,799	-2
	Bournemouth	163,560	161,169	-2,391	-1.5
	Reading	144,684	142,756	-1,928	-1.3
	Wirral	315,004	311,210	-3,794	-1.2
	Poole	138,368	136,869	-1,499	-1.1
3. Other Urban	Rushmoor	90,892	88,744	-2,148	-2.4
	Burnley	89,521	87,979	-1,542	-1.7
	Harlow	78,799	78,065	-734	-0.9
	Slough	120,577	119,516	-1,061	-0.9
	Stevenage	79,794	79,307	-487	-0.6
4. Significant Rural	Wycombe	162,050	161,326	-724	-0.4
5. Rural-50	Blyth Valley	81,334	81,204	-130	-0.2
6. Rural-80	Bridgnorth	52,458	51,808	-650	-1.2
	Isles of Scilly	2,140	2,126	-14	-0.7
	South Oxfordshire	128,307	128,124	-183	-0.1

combinations. The majority of OAs, LSOAs and MSOAs (98.9 per cent, 99.6 per cent and 99.8 per cent respectively) were within threshold in both 2001 and 2005. Even in areas with high population change, there were remarkably few OAs, LSOAs and MSOAs breaching the 2001 thresholds by 2005. Very few areas (<0.1 per cent of OAs; 0.1 per cent LSOAs; 0.1 per cent MSOAs) that were within threshold in 2001 had breached the lower threshold by 2005. Similarly low percentages (0.4 per cent of OAs; 0.2 per cent of LSOAs; zero MSOAs) were within threshold in 2001 but above threshold by 2005. 221 OAs (0.13 per cent) that were below threshold in 2005 were also below threshold in 2001.

Table 3

Population change between 2001 and 2006 mid-year estimates for Welsh Unitary Authorities

Unitary Authority	Population mid-2001	Population mid-2006	Change (number)	Change (per cent)
Powys	126,398	131,141	4,743	3.8
Monmouthshire	84,984	87,882	2,898	3.4
The Vale of Glamorgan	119,277	123,275	3,998	3.4
Denbighshire	93,070	96,089	3,019	3.2
Carmarthenshire	172,845	178,043	5,198	3.0
Bridgend	128,735	132,584	3,849	3.0
Ceredigion	75,083	77,160	2,077	2.8
Pembrokeshire	114,199	117,280	3,081	2.7
Cardiff	310,088	317,523	7,435	2.4
Neath Port Talbot	134,380	137,052	2,672	2.0
Wrexham	128,540	130,990	2,450	1.9
Newport	137,642	140,125	2,483	1.8
Swansea	223,463	227,079	3,616	1.6
Isle of Anglesey	67,806	68,884	1,078	1.6
Conwy	109,674	111,273	1,599	1.5
Gwynedd	116,844	118,250	1,406	1.2
Caerphilly	169,546	171,349	1,803	1.1
Flintshire	148,629	150,077	1,448	1.0
Rhondda, Cynon, Taff	231,910	233,936	2,026	0.9
Torfaen	90,912	91,022	110	0.1
Blaenau Gwent	70,000	69,341	-659	-0.9
Merthyr Tydfil	56,207	55,530	-677	-1.2

The only MSOA (in Cambridge) which was above threshold in 2005 was also above threshold in 2001. Only one OA had shifted from being below threshold in 2001 to above threshold in 2005.

Map 2 and Map 3 provide examples of how the magnitude and distribution of these breaches can vary geographically in specific areas. Map 2 and Map 3 show absolute population change between the 2001 and 2005 mid-year estimates for Camden and Liverpool respectively. OAs that have breached lower or upper thresholds have a semi-circle or circle symbol within their boundary; those breaching a threshold in 2001 have a left-handed semi circle; those breaching in 2005 have a right-handed semi-circle; breaches below the lower threshold are shown in grey; breaches above the upper threshold are in green. It is clear from these figures that, overall, Camden has experienced more population growth whereas Liverpool has experienced more decline. While the majority of OAs within both local authorities have not yet breached thresholds, both areas have a number of OAs that have breached either the lower or upper thresholds in 2001 and/or 2005. This illustrates the complexity of designing automated procedures that will maintain stability in the majority of OAs but split, merge or re-design only those that require change.

Table 4

Thresholds employed in the analysis

Geography	Lower thresholds		Upper thresholds	
	Household thresholds ²	Population thresholds ¹	Household thresholds ²	Population thresholds ¹
OA	40	100	250	625
LSOA	400	1,000	1,200	3,000
MSOA	2,000	5,000	6,000	15,000

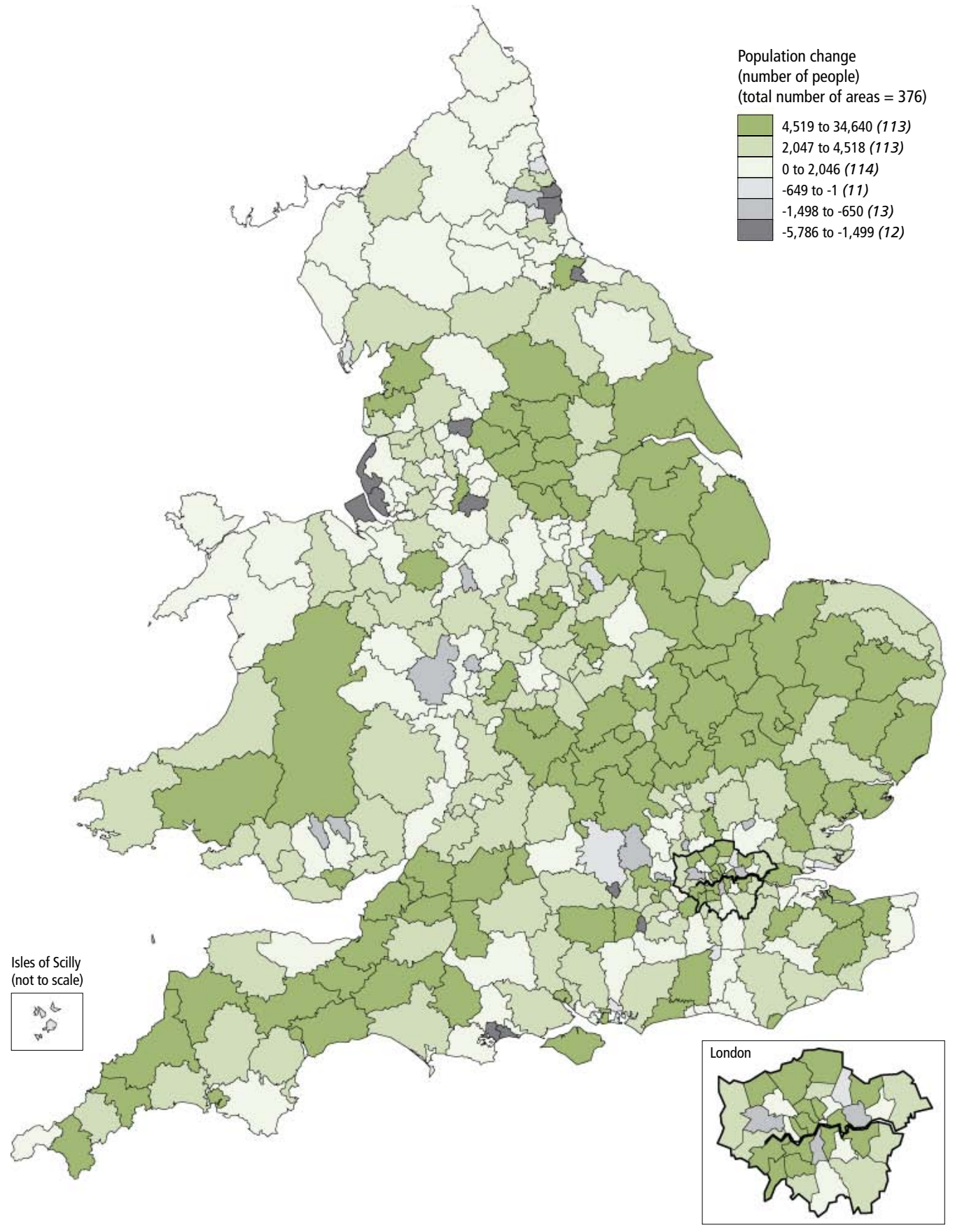
Notes:

- Population thresholds obtained by multiplying household thresholds by a factor of 2.5 (equating approximately to average household size).
- Household threshold values taken from Mitchell and Ralphs (2007), Table 1.1.
- No upper thresholds were published in 2001 for OAs or LSOAs. The values used here are the target mean employed for the automated zone design process multiplied by 2 (as in Ralphs and Mitchell, (2006)). MSOAs did have a published upper threshold of 4000 households, but here we use the value of 6000 households (as in Mitchell and Ralphs, (2007)) as this is consistent with the values used at the other levels.

Map 1

Absolute population change, by local or unitary authority, 2001–2006

England & Wales

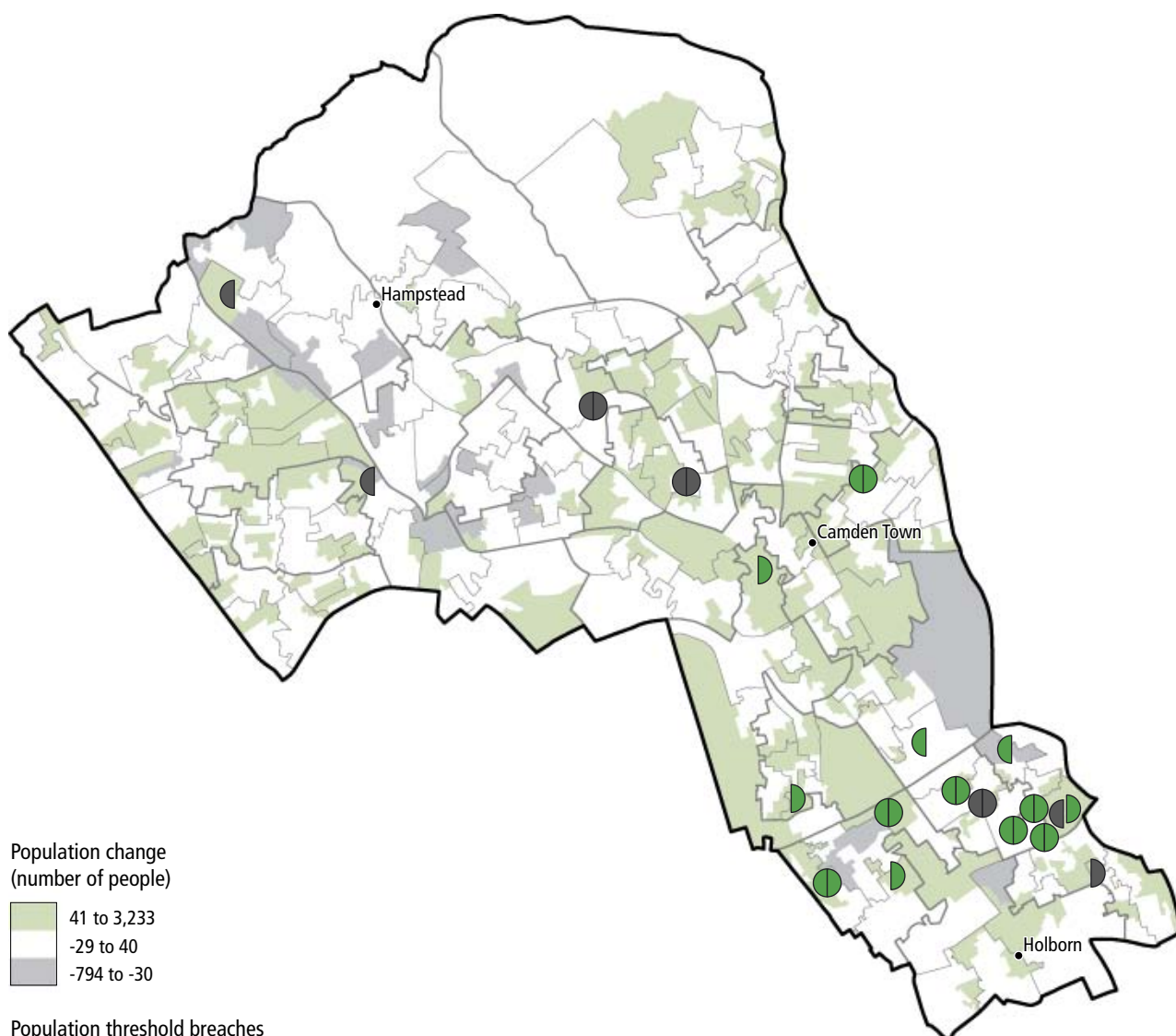


Source: Mid-year population estimates Office for National Statistics

Map 2

Absolute population change and threshold breaches for Output Areas, 2001–2005

Camden Local Authority



Population change
(number of people)

- 41 to 3,233
- 29 to 40
- 794 to -30

Population threshold breaches

- Above upper threshold for 2001
- Below lower threshold for 2001
- Above upper threshold for 2005
- Below lower threshold for 2005

Boundaries

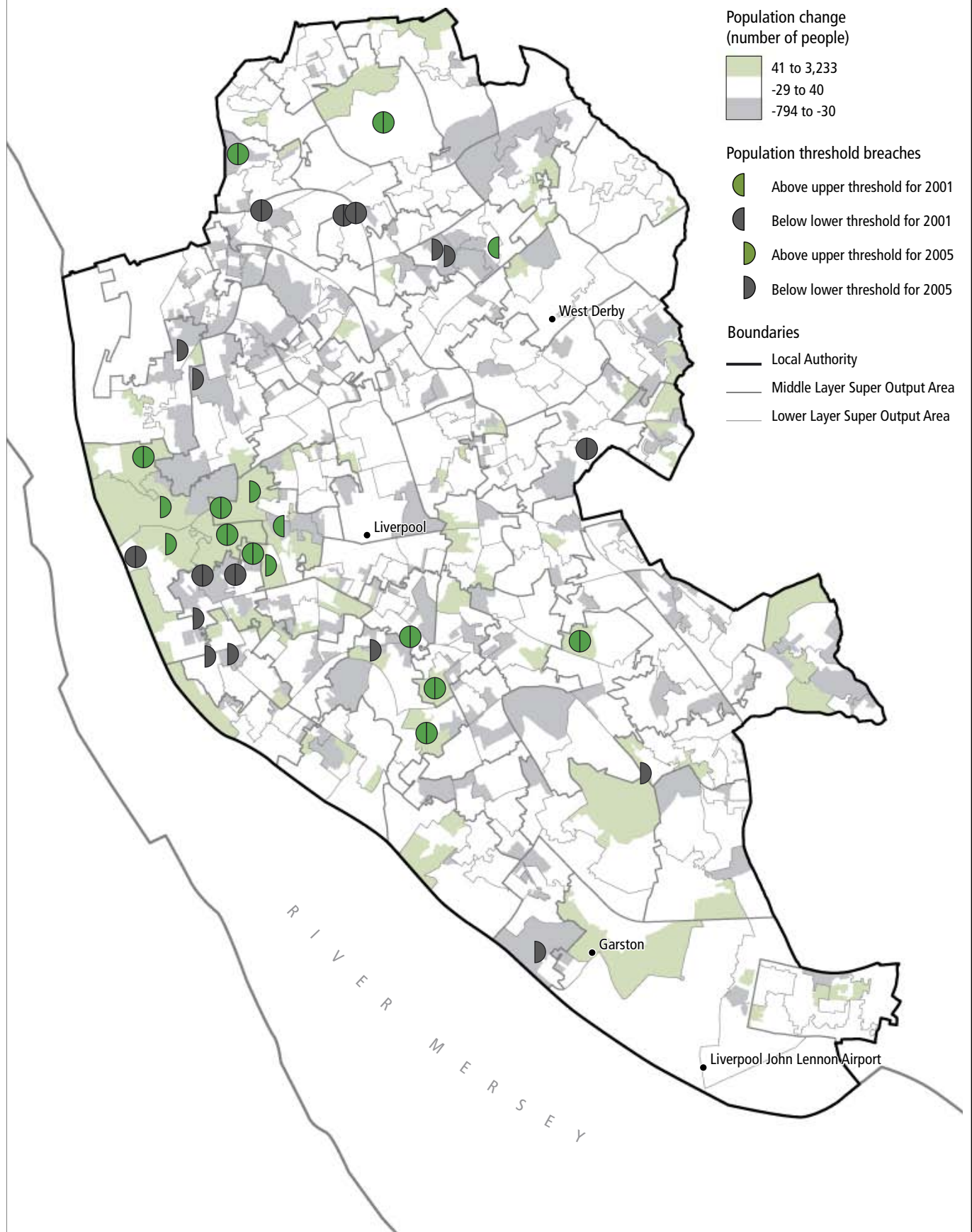
- Local Authority
- Middle Layer Super Output Area
- Lower Layer Super Output Area

Source: Mid-year population estimates Office for National Statistics

Map 3

Absolute population change and threshold breaches for Output Areas, 2001–2005

Liverpool Local Authority

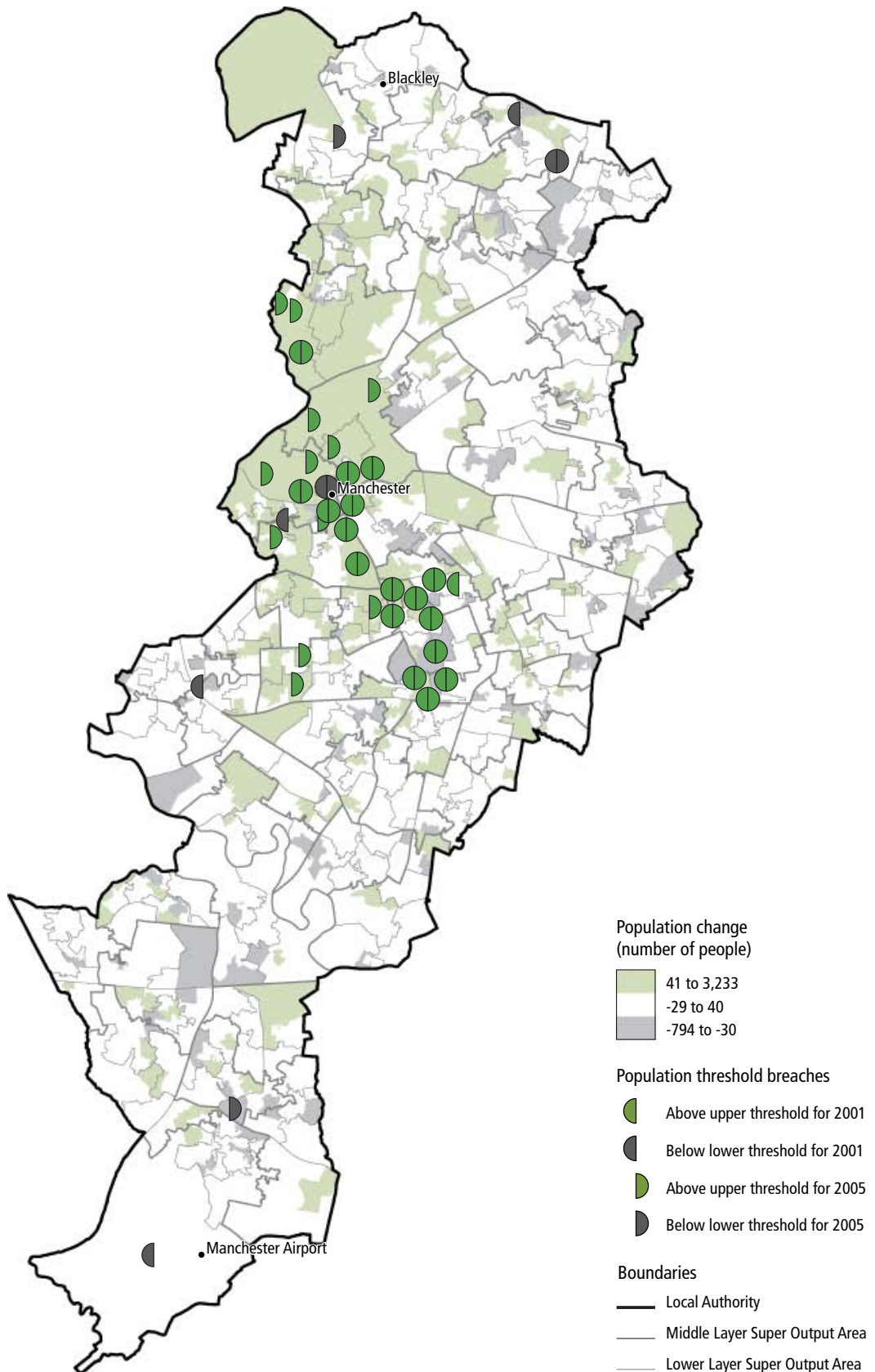


Source: Mid-year population estimates Office for National Statistics

Map 4

Absolute population change and threshold breaches for Output Areas, 2001–2005

Manchester Local Authority



Source: Mid-year population estimates Office for National Statistics

Table 5

Number of OAs, LSOAs and MSOAs breaching population thresholds in mid-year estimates for 2001 and 2005 in England and Wales

Output Areas (OA)

(a) Counts

	2005 Below	2005 Within	2005 Above	2001 Totals
2001 Below	221	228	1	450
2001 Within	147	173,553	682	174,382
2001 Above	0	78	506	584
2005 Totals	368	173,859	1,189	175,416

(b) Per cent (of total)

	2005 Below	2005 Within	2005 Above	2001 Totals
2001 Below	0.13	0.13	0.00	0.26
2001 Within	0.08	98.94	0.39	99.41
2001 Above	0.00	0.04	0.29	0.33
2005 Totals	0.21	99.11	0.68	100.00

Lower Level Super Output Areas (LSOA)

(c) Counts

	2005 Below	2005 Within	2005 Above	2001 Totals
2001 Below	6	8	0	14
2001 Within	34	34,242	58	34,334
2001 Above	0	3	27	30
2005 Totals	40	34,253	85	34,378

(d) Per cent (of total)

	2005 Below	2005 Within	2005 Above	2001 Totals
2001 Below	0.02	0.02	0.00	0.04
2001 Within	0.10	99.60	0.17	99.87
2001 Above	0.00	0.01	0.08	0.09
2005 Totals	0.12	99.63	0.25	100.00

Middle Level Super Output Areas (MSOA)

(e) Counts

	2005 Below	2005 Within	2005 Above	2001 Totals
2001 Below	3	4	0	7
2001 Within	8	7,178	0	7,186
2001 Above	0	0	1	1
2005 Totals	11	7,182	1	7,194

(f) Per cent (of total)

	2005 Below	2005 Within	2005 Above	2001 Totals
2001 Below	0.04	0.06	0.00	0.10
2001 Within	0.11	99.78	0.00	99.89
2001 Above	0.00	0.00	0.01	0.01
2005 Totals	0.15	99.84	0.01	100.00

Table 6

Threshold breaches for OAs, LSOAs and MSOAs over time, 2001–05 mid-year estimates, England and Wales

Census Output Areas (OA)

(a) Counts

Year	Zero people	Below threshold	Within threshold	Above threshold
2001	6	444	174,382	584
2002	8	307	174,465	636
2003	8	323	174,284	801
2004	4	339	174,055	1,018
2005	3	365	173,859	1,189

(b) Per cent (of total)

Year	Zero people	Below threshold	Within threshold	Above threshold
2001	<0.01	0.25	99.41	0.33
2002	<0.01	0.18	99.46	0.36
2003	<0.01	0.18	99.36	0.46
2004	<0.01	0.19	99.22	0.58
2005	<0.01	0.21	99.11	0.68

Lower Level Super Output Areas (LSOA)

(c) Counts

Year	Zero people	Below threshold	Within threshold	Above threshold
2001	0	14	34,334	30
2002	0	27	34,319	32
2003	0	25	34,307	46
2004	0	34	34,282	62
2005	0	40	34,253	85

(d) Per cent (of total)

Year	Zero people	Below threshold	Within threshold	Above threshold
2001	0.00	0.04	99.87	0.09
2002	0.00	0.08	99.83	0.09
2003	0.00	0.07	99.79	0.13
2004	0.00	0.10	99.72	0.18
2005	0.00	0.12	99.64	0.25

Middle Level Super Output Areas (MSOA)

(e) Counts

Year	Zero people	Below threshold	Within threshold	Above threshold
2001	0	7	7,186	1
2002	0	8	7,185	1
2003	0	10	7,183	1
2004	0	12	7,181	1
2005	0	11	7,182	1

(f) Per cent (of total)

Year	Zero people	Below threshold	Within threshold	Above threshold
2001	0.00	0.10	99.89	0.01
2002	0.00	0.11	99.88	0.01
2003	0.00	0.14	99.85	0.01
2004	0.00	0.17	99.82	0.01
2005	0.00	0.15	99.83	0.01

Change in number of breaches over time

Table 6 shows the numbers and percentages of OAs, LSOAs and MSOAs below, within and above threshold annually from 2001 to 2005 for England and Wales. **Table 7** explores the degree to which the output geography areas were approaching (within five or ten per cent of) the upper and lower thresholds by 2005 and the extent to which areas that had already breached by 2005 had exceeded the thresholds.

Table 6 shows that at the OA level between 2001 and 2005 there was a steady annual increase in the numbers above threshold whereas the number of OAs below threshold remained fairly stable over time. Table 7 reveals that the percentage of OAs less than five per cent above the lower threshold was only 0.1 per cent, and that this percentage had decreased very slightly year-on-year. Approximately 0.1 per cent of OAs were also within five per cent of the upper threshold but, by contrast, this percentage had grown slightly year-on-year. Overall, the number of OAs that had breached, or were near to breaching, the upper thresholds appeared to be steadily increasing annually, presumably reflecting general growth in population. Interestingly, where OAs had breached, they tended to have breached by at least ten per cent over the threshold, rather than having only just gone over threshold. The same is true of those breaching the lower threshold.

At the LSOA level, the trends of a small annual increase in the percentage of LSOAs breaching the lower threshold and a slightly larger increase in those breaching the upper threshold were similar to the trends observed amongst OAs. However, the degree to which non-breached LSOAs were approaching the thresholds, and the extent to which those already breaching had exceeded the thresholds, was different to the OAs. There was a smaller percentage of LSOAs near to the thresholds, with more being closer to the lower threshold than the upper threshold. Of those LSOAs that had already breached the lower threshold, most had done so by less than five per cent. Amongst those breaching the upper threshold, there appeared to be two distinct groups: those that had only just breached (by five per cent or less) and those that had breached more substantially (by greater than ten per cent) – with the second group being more numerous.

There was an inconsistent, but very slight, increase in the number of MSOAs breaching the lower threshold between 2001 and 2005. Of those breaching the lower threshold, most had only just dipped beneath it. Approximately two per cent of MSOAs were less than five per cent above the lower threshold in 2001: by 2005 this percentage was only 1.3 per cent, reflecting the general growth in MSOA populations. Encouragingly, despite this growth, only a tiny percentage (approximately 0.04 per cent) were within five per cent of the upper threshold by 2005. As noted previously, only one MSOA had breached

Table 7
Percentage of OAs, LSOAs and MSOAs near to thresholds, 2001–05 mid-year estimates, England and Wales
(a) Census Output Areas (OA) (n = 175,416)

Year	Below threshold			Around threshold			Above threshold		
	More than 10% below lower threshold	Between 5% to 10% below lower threshold	Less than 5% below lower threshold	Less than 5% above lower threshold	More than 5% above lower threshold and more than 5% below upper threshold	Less than 5% below upper threshold	Less than 5% above upper threshold	Between 5% to 10% above upper threshold	More than 10% above upper threshold
2001	0.12	0.05	0.08	0.09	99.24	0.07	0.05	0.05	0.23
2002	0.10	0.03	0.05	0.08	99.29	0.09	0.06	0.05	0.26
2003	0.11	0.04	0.04	0.08	99.16	0.11	0.07	0.06	0.32
2004	0.12	0.03	0.05	0.08	99.02	0.13	0.10	0.07	0.41
2005	0.12	0.03	0.05	0.08	98.87	0.16	0.10	0.08	0.49

(b) Lower Level Super Output Areas (LSOA) (n = 34,378)

Year	Below threshold			Around threshold			Above threshold		
	More than 10% below lower threshold	Between 5% to 10% below lower threshold	Less than 5% below lower threshold	Less than 5% above lower threshold	More than 5% above lower threshold and more than 5% below upper threshold	Less than 5% below upper threshold	Less than 5% above upper threshold	Between 5% to 10% above upper threshold	More than 10% above upper threshold
2001	0.00	0.00	0.04	0.45	99.39	0.03	0.02	0.02	0.05
2002	0.00	0.01	0.07	0.38	99.41	0.03	0.02	0.01	0.06
2003	0.01	0.01	0.05	0.39	99.35	0.05	0.03	0.03	0.07
2004	0.02	0.01	0.07	0.34	99.32	0.07	0.04	0.03	0.11
2005	0.02	0.02	0.08	0.32	99.23	0.09	0.06	0.04	0.15

(c) Middle Level Super Output Areas (MSOA) (n = 7,194)

Year	Below threshold			Around threshold			Above threshold		
	More than 10% below lower threshold	Between 5% to 10% below lower threshold	Less than 5% below lower threshold	Less than 5% above lower threshold	More than 5% above lower threshold and more than 5% below upper threshold	Less than 5% below upper threshold	Less than 5% above upper threshold	Between 5% to 10% above upper threshold	More than 10% above upper threshold
2001	0.01	0.01	0.07	1.96	97.93	0.00	0.01	0.00	0.00
2002	0.01	0.01	0.08	1.65	98.22	0.00	0.00	0.01	0.00
2003	0.01	0.00	0.13	1.49	98.35	0.01	0.00	0.01	0.00
2004	0.01	0.00	0.15	1.38	98.40	0.04	0.00	0.01	0.00
2005	0.01	0.01	0.13	1.26	98.53	0.04	0.00	0.00	0.01

the upper threshold by 2005, but this MSOA was already above threshold in 2001. It is worth noting though that it had moved further above threshold between 2001 and 2005.

It is likely that the differences in the patterns observed at OA-level compared to those at LSOA and MSOA levels are due to the scale and size of the geographical units, and due to differences in the 2001 mid-year population distributions (in terms of how close the 2001 mean OA, LSOA and MSOA populations were to the thresholds initially). Overall, these findings should be reassuring for ONS and users who are hoping that it is possible to retain stability in 2001, especially at the higher output geography levels.

Nested breaches

It is also important to understand what types of breaches are occurring in order to ensure that the maintenance procedures will be able to deal with them. For example, the procedures required to deal with instances where an LSOA has breached a threshold but its constituent OAs have not, would almost certainly be different to those required where an LSOA has breached as a result of a high proportion of its constituent OAs breaching.

Of the 85 LSOAs above threshold by 2005, 83 contained above-threshold OAs within them. The percentages of OAs breached within an LSOA ranged from 75 per cent (three out of four) to 13 per cent (one out of eight), indicating that in a minority of areas, the LSOA breaches were not just due to one OA going significantly above threshold, but rather due to breaches across a number of OAs. The one MSOA that had gone above threshold by 2005 contained breaches at both the LSOA level (three out of 29), and at the OA level (two out of four OAs within the three above-threshold LSOAs).

In terms of below threshold breaches, only eight of the 40 LSOAs that had gone below threshold by 2005 also contained under-threshold OAs (ranging from 60 per cent (three out of five) to 20 per cent (one out of five)). This suggests that the sub-threshold breaches have largely come about due to a general decrease in population across OAs within LSOAs rather than through significant decreases in specific OAs. Of the 11 MSOAs which were under threshold by 2005, only one contained any below-threshold areas within it (this one containing one out of four LSOAs breached but no OA level breaches).

Implications for maintenance of the 2001 output geographies

This analysis assumes that the ONS mid-year estimates provide an accurate picture of the rate and geographical distribution of population change since 2001. Any statistical or geographical bias in the mid-year estimates could significantly alter the levels and patterns observed. The number of breaches is of course dependent on the thresholds employed. It is possible that the factor (2.5) employed to calculate population thresholds from household thresholds over-estimates average household size and it has also been noted that average household size is decreasing over time. If this is the case, the number of threshold breaches reported here could under-estimate the scale of the problem. The magnitude of the breaches seen here is similar though to those previously reported by ONS,¹⁰ who employed different datasets and methodologies to explore potential output geography breaches by 2011.

It is also not clear whether population will continue to change at the same rate and in the same geographical areas. For example, it is possible that some of the areas that have already undergone significant growth since 2001 may now become more stable. Growth may shift to other geographical areas, leading to new breaches in those areas, but this will

be dependent on a number of factors such as trends in births, deaths, international and internal migration, economic prosperity and property development. Or growth may continue in the already breached areas, making the output geographies in these areas even more unsuitable.

There are also uncertainties surrounding the extent to which the socio-economic homogeneity of the output geographies will have deteriorated by 2011: this article was unable to evaluate this because accurate contemporary tenure and accommodation type data were not available at the small area level, and indeed are only collected by the decennial Census.

If the trends presented here are accurate and do continue and if similar population thresholds are employed in 2011, it is likely that the majority of output geography areas will remain within threshold by 2011. The fact that population change tends to be strongly geographically clustered does mean though that in a minority of areas the output geographies are likely to be unsuitable for the release of 2011 Census data. In these areas, maintenance procedures that split, merge or completely re-design the existing geographies will be needed. Further, as a result of known issues related to the address register database used in 2001 and due to the complex and dynamic nature of the population in some areas (such as variations in the number of armed forces being stationed in some areas, for example, Rushmoor), some areas appear to have 2001 output geographies that are unlikely to have been optimal for the representation of population, even in 2001. For example, **Map 4** shows absolute population change and threshold breaches between 2001 and 2005 for Manchester UA.

It is clear that a number of OAs had already breached the upper threshold by the time of the 2001 mid-year estimates, probably reflecting the fact that a large number of addresses were missing from the 2001 address register and were hence not used in the design of the output geographies. When corrections were made for these missing addresses, the size of the population in some areas will have increased, in some cases taking the OAs above threshold even by the time of the 2001 mid-year estimates. While these breaches arose for understandable, and often unavoidable reasons with respect to zone design, they do now present challenges for the maintenance of the geographies for 2011: should they be left as they are, maintained (that is split or merged), or completely re-designed?

Challenges involved in maintaining the 2001 output geographies

In 2007 ONS undertook a consultation on users' requirements for the 2011 Census small area output geographies. The consultation suggested that the majority of users would prefer to see the output geographies remaining stable rather than re-designing them completely for 2011. There were mixed views on the desirability of using postcodes as the building blocks for the 2011 geographies. Some users would prefer to see postcodes retained, while others would prefer the use of alternative building blocks such as street blocks. Some users argued for a better alignment of the output geography boundaries with real-world features. Some suggested that the OAs should better represent, and not split, 'neighbourhoods' or 'communities'. There was also some support for the resolution of known issues such as the address register problems experienced in Manchester and Westminster in 2001.

Population change since 2001, together with the requirements flagged by user consultation, present complex challenges for the design of the 2011 output geographies. Consequently, a series of important and potentially conflicting decisions need to be made concerning the processes, datasets and criteria employed to maintain 2001 output geographies:

- Perhaps foremost, is the decision as to whether geographic stability should be maintained at the OA, LSOA and MSOA levels where

- possible, or just at the LSOA and MSOA levels.
- For OAs which need to be split, what building blocks should be employed? The use of postcodes would allow datasets that are geo-referenced by postcode to be linked with census data. However, in some areas, the link between 2001 postcodes and OAs will have eroded during the inter-censal period making them less useful. Also, the use of postcodes for aggregation in the zone design process can be constraining in certain situations, for example, where postcodes are split across OAs or where vertical stacks (addresses with the same grid-reference but different postcodes) exist. Alternative building blocks, such as street blocks, have advantages in that they may appear to be more aligned to real-world features, but the ability to link via postcodes would be lost.
 - A closely related issue is whether the existing and/or maintained output geographies should be aligned or re-aligned to real-world features. The output geographies were always intended to be synthetic statistical boundaries, albeit constrained where possible to some geographical features, such as roads. Alignment to more geographical features would possibly make the boundaries appear more 'real' but attempting to do so would be arguably futile: there is no agreed set of 'real world' features to align to; such alignment would introduce conflicts with other zone design goals, such as maintaining inter-censal stability; and doing so would raise significant boundary copyright issues, thereby potentially impairing ONS' ability to freely distribute the boundary data as an integral part of the Census outputs.
 - Should the maintenance procedures ensure that the output geography boundaries nest within wards or local authority districts? This would be difficult given the regular changes in ward and LA boundaries and may lead to the need for a large number of changes to the output geography boundaries. It would also move away from ONS' stated policy of retaining stability of the output geographies over time.
 - Should the same design criteria and values as in 2001 be employed? For example, should the same thresholds, targets and shape constraints be used?

In addition, the process of maintaining the 2001 output geographies should ideally be automated to enable the systematic, objective and efficient creation of the geographies for all of England and Wales in a timely manner following the collection of census data in 2011.

Conclusions

This article has explored the magnitude and geographical distribution of population change since the 2001 Census in the context of maintenance of the 2001 Census output geographies. Using mid-year estimates, it concludes that virtually all output geography areas had not breached upper or lower thresholds by 2005, and are unlikely to do so by 2011. Nonetheless, because population change is usually strongly geographically clustered, in some areas there have already been significant breaches of population thresholds; the output geographies in these areas and others are therefore likely to need maintenance in order to be suitable for the release of 2011 Census data. The challenges involved in carrying out this maintenance are non-trivial and this article identifies some of the key decisions that need to be taken before the maintenance procedures can be developed and implemented. The ongoing ESRC-funded Census2011Geog project will develop prototype software for carrying out the automated maintenance procedures and will also evaluate the usefulness of different building blocks and maintenance methods. ONS will then need to evaluate these findings, make key policy decisions and then implement the procedures following the collection and collation of the 2011 Census data.

Key findings

- The vast majority of Output Areas, Lower Layer and Middle Layer Super Output Areas had not breached specified population thresholds by 2005, and seem unlikely to do so by the 2011 Census
- Population change is strongly geographically clustered so in a minority of areas the 2001 output geographies will not be appropriate for the release of 2011 Census data
- Automated maintenance procedures that split, merge or re-design the 2001 output geographies are being developed by the ESRC-funded Census2011Geog project in collaboration with ONS

Acknowledgements

This research is supported by Economic and Social Research Council Award RES-348-25-0019. It uses mid-year population estimate data at OA, LSOA, MSOA and local authority levels provided by ONS, the local authorities' urban/rural classification from DEFRA and 2001 Census digitised boundary data from UKBORDERS which is Crown copyright and reproduced with the permission of the Controller of HMSO. The authors are grateful to Professor David Martin, Andy Tait, Andy Bates, Donna Viney and three anonymous reviewers for useful comments on earlier drafts, and to Nick Richardson for help with production of the maps.

References

- 1 Office for National Statistics (2007) *National Statistics Small Area Geography Consultation 2007*. Available at: www.ons.gov.uk/about/consultations/closed-consultations/geography-policy-public-consultation/index.html
- 2 Martin D (1998) Optimizing census geography: the separation of collection and output geographies *International Journal of Geographical Information Science* 12, 673–685.
- 3 Martin D, Nolan A and Tranmer M (2001) The application of zone design methodology to the 2001 UK Census *Environment and Planning A* 33, 1949–1962.
- 4 Cockings S and Martin D (2005) Zone design for environment and health studies using pre-aggregated data *Social Science & Medicine* 60(12), 2729–2742.
- 5 Mitchell B and Ralphs M (2007) *Developing maintenance rules for the Neighbourhood Statistics Output Geographies*. Unpublished research report. Methodology Directorate, Office for National Statistics.
- 6 Office for National Statistics (2004a) *2004 Local Authority Studies: The Longitudinal Based Study Consequential Adjustment*. Available at: www.statistics.gov.uk/downloads/theme_population/LAStudy_LS_ConsequentialAdjustment.pdf
- 7 Office for National Statistics (2004b) *2001 Census Local Authority population Studies: Full report*. Available at: www.statistics.gov.uk/downloads/theme_population/LAStudy_FullReport.pdf
- 8 Office for National Statistics (2004c) *2001 Census: Manchester and Westminster Matching Studies Full Report*. Available at: www.statistics.gov.uk/downloads/theme_population/ManchesterandWestminster_FullReport.pdf
- 9 Department of Environment, Food and Rural Affairs (2005) *Defra Classification of Local Authority Districts and Unitary Authorities in England: A Technical Guide*. Available at: www.defra.gov.uk/evidence/statistics/rural/rural-definition.htm
- 10 Ralphs M and Mitchell B (2006) *Maintenance requirements for Super Output Area geographies: modelling changes from 2001–06*. Unpublished research report. Methodology Directorate, Office for National Statistics.