Office for National Statistics

An Overview of Best-fitting: Building 2011 Census Estimates from Output Areas

Edition: **October 2012**

Editor: **ONS Geography**

Office for National Statistics

October 2012

A National Statistics Publication

National Statistics are produced to high professional standards set out in the Code of Practice for Official Statistics. They are produced free from political influence.

About Us

Office for National Statistics

The Office for National Statistics (ONS) is the executive office of the UK Statistics Authority, a non-ministerial department which reports directly to Parliament. ONS is the UK government's single largest statistical producer. It compiles information about the UK's society and economy, and provides the evidence-base for policy and decision-making, the allocation of resources, and public accountability. The Director-General of ONS reports directly to the National Statistician who is the Authority's Chief Executive and the Head of the Government Statistical Service.

Government Statistical Service

The Government Statistical Service (GSS) is a network of professional statisticians and their staff operating both within the ONS and across more than 30 other government departments and agencies.

Contacts

This publication

For information about the content of this publication, contact ONS Geography Customer Services

Tel: 01329 444971
Email: ons.geography@ons.gsi.gov.uk

Other customer enquiries

ONS Customer Contact Centre
Tel: 0845 601 3034
International: +44 (0)845 601 3034
Minicom: 01633 815044
Email: info@ons.gsi.gov.uk
Post: Room 1.101, Government Buildings,
Cardiff Road, Newport, South Wales NP10 8XG
www.ons.gov.uk

Media enquiries

Tel: 0845 604 1858
Email: media.relations@ons.gsi.gov.uk

Copyright and Reproduction

Contents

1. What is best-fitting?

Best-fitting is the method used to produce estimates for any output geography, by aggregating whole statistical building blocks, even where these may be nested within a higher geography. It is the method used to produce all national and 2011 Census statistics, so that estimates produced are consistent, comparable and non-disclosive.

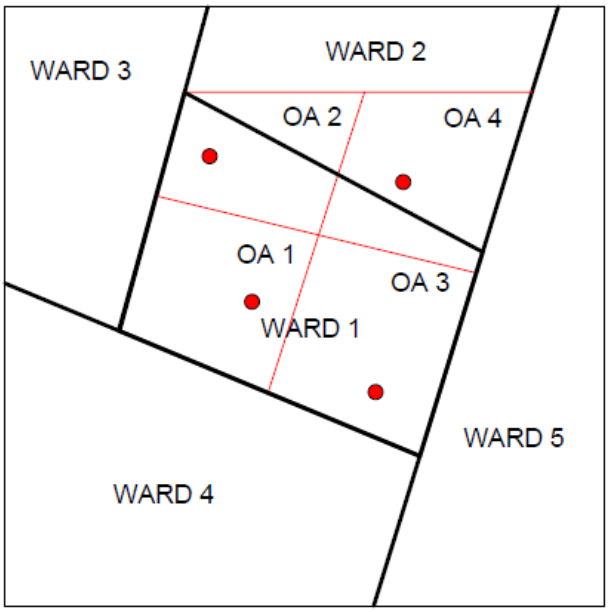
2. How have best-fit estimates been produced?

For the 2011 Census ONS has simplified the process of producing best-fit of Output Area (OA) to higher geographies. This has been done in a way that allows anyone to produce lookups in the same way, without needing to access 2011 Census source data.

Each geographic instance of the 2011 OA, Lower Layer and Middle Layer Super Output Areas (LSOA and MSOA) geographies has a population weighted centroid. This centroid, described below, is a single summary point that reflects the spatial distribution of the 2011 Census population in each instance of those geographies. **Figure 1** shows how an OA is allocated to a higher geography if the OA population weighted centroid falls within the boundary of the higher geography.

Generating estimates using these OA best-fit allocations provides an approach that is flexible enough to allow users to build their own best-fit OA allocations. This is done by plotting these frozen OA population weighted centroids to the boundaries of any new or changed geographies.

This approach to building estimates for any geography minimises the risk of disclosure by preventing slivers containing very small populations.



OA-Ward Lookup	
OA	**WARD**
OA 1	WARD 1
OA 2	WARD 1
OA 3	WARD 1
OA 4	WARD 2

Figure 1: 2011 Best fit of Output Areas to Higher Geographies. OA 4 is allocated to WARD 2 because OA 4's population weighted centroid falls within the WARD 2 boundary

3. What data was used to produce best-fit estimates?

Best-fit estimates are produced using the 2011 OAs and their population weighted centroids (PWC).

A PWC represents how the population is spatially distributed within the OA, summarised to a single point on the ground.

PWCs were calculated using a median centroid algorithm, which is less influenced by outliers, i.e. a household location that is located a long way from the rest of the households in the OA - than the mean centroid algorithm that was used to produce a set of PWCs for the 2001 Census[1].

The Median Center *(sic)* function in ArcGIS 10.0 was run against the coordinates and the populations of each household in each OA, LSOA and MSOA.

Where the calculated centroid fell outside the boundary of the area for which it was being calculated or within two metres of the area boundary, it was moved to the nearest location at least two metres inside the area boundary.

More information about the Median Center algorithm is available here.

4. So why was best-fit used this time?

The Geography Policy for National Statistics sets out best practice for producing national and official statistics by geography, to ensure outputs are accurate, consistent and comparable in their use of geography. One element of the policy prescribes that statistical estimates should be produced for the core statistical geographies of OAs and SOAs and that estimates for all geographies should be built from aggregations of OAs or SOAs. For this reason OAs and SOAs are referred to as building block geographies.

When OAs were created in 2001, they aligned with the administrative ward, parish and community boundaries current at the time. Many of the administrative boundaries have changed significantly since the 2001 Census, so they no longer align with the OAs and SOAs. In addition, a number of the OAs and SOAs have been redesigned as part of the 2011 Census OA maintenance, which means they no longer align with unchanged ward and parish boundaries. No decision has yet been made on whether any future maintenance will be applied, but any decision will be made in the context of the Beyond 2011 programme, which is looking at the future of the England and Wales Census.

The fact that OAs and administrative boundaries no longer align means it is not possible to aggregate OAs and SOAs to exactly fit these changed administrative boundaries. This poses a number of problems for producing statistics for an area that does not now align with OAs.

An exact-fit estimate could be produced for an area that does not align with an OA by plotting the grid reference of each household address directly into the boundary of the output geography. It would then be possible to compare this estimate with an estimate for the overlapping but unaligned OA and identifying the difference.

Figure 2 shows that the difference between the two geographies produces a 'sliver' containing two households. By taking one estimate from another it is possible to expose very small numbers of households or persons. This is called 'disclosure by differencing', and this problem applies to all exact-fit estimates produced using the same data source for different overlapping geographies.

[1] The 2001 population weighted centroids were not used for the production of statistical estimates or geographic lookups as part of the 2001 Census

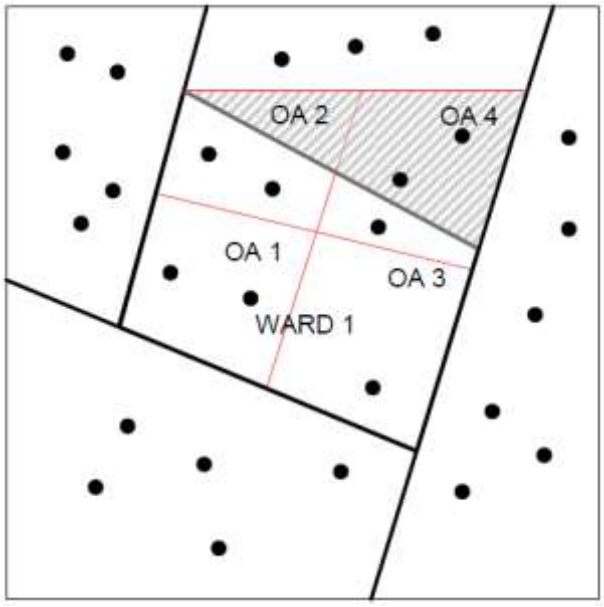


Figure 2: In 2011, the difference between WARD 1 and the OAs means a disclosive sliver is produced

In 2001, the risk of disclosure was addressed by adjusting cells in tables that had very small values and which could be disclosive. This small cell adjustment proved unpopular with users and was not favoured for use as part of the 2011 Census.

It is possible to provide estimates for different overlapping geographies from the same data by always aggregating OA estimates for each geography, using a best-fit methodology to determine the OAs that are allocated to each geography. This is the approach prescribed by the Geography Policy for National Statistics and the solution adopted for 2011 Census outputs. It means that estimates for any geography are always aggregations of whole OAs' estimates.

Best-fit has also been adopted for the 2011 Census because the PWCs that are used to create the allocations can be provided as part of the 2011 Census geography outputs. Doing this allows users to produce their own best-fit estimates in line with the Geography Policy for National Statistics. Having access to the PWC makes best-fitting a far more flexible methodology than alternative methods because it allows statistical users to produce their own estimates to any geography during the intercensal period.

5. What impact does best-fitting have on statistical estimates?

ONS has previously conducted research into the impact of adopting a best-fit policy which was used to inform the Geography Policy for National Statistics[2].

Overall, the research found that the adoption of best-fit for producing estimates did not have a significant impact, and was an adequate methodology to be used for 2011 Census outputs.

Analysis of best-fitting 2001 OAs to 2009 electoral wards found that 94.4 per cent of wards had best-fit estimates that were within 10 per cent tolerance of exact-fit estimates to the same ward.

[2] http://www.ons.gov.uk/ons/guide-method/census/2011/the-2011-census/producing-and-delivering-data/output-geography/best-fit-policy/exploring-the-performance-of-best-fitting-to-produce-ons-data-for-non-standard-geographical-areas.pdf

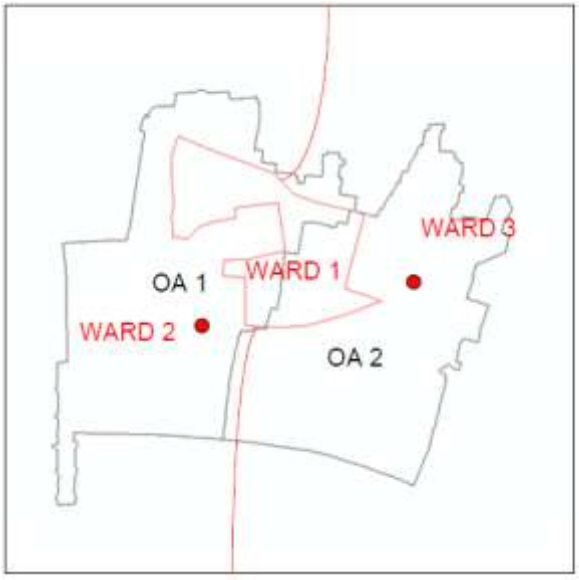
6. What about instances of geographies that are smaller than an OA?

In 2001 some wards and parishes contained less than 100 people or 40 households so therefore fell below the population threshold required for the Census Area Statistics (CAS) outputs. Each of these wards and parishes with small population were merged with a neighbouring ward or parish until the minimum population threshold was reached.

These merged wards and parishes were added to the unmerged wards and parishes to become two separate geographies, known as the 2001 CAS parishes and the 2001 CAS wards, adding to an already complex geographic hierarchy.

As estimates were produced using the best-fit approach for 2011, it is not possible to identify geographies with exact-fit populations below the minimum population threshold for OAs of 100 population and 40 households.

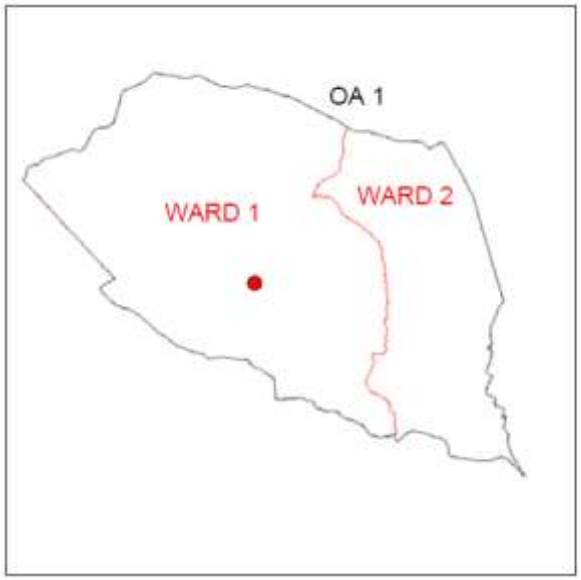
It was however possible to identify those geographies that were smaller than an OA and which did not contain an OA population weighted centroid. **Figure 3** shows an area where a ward does not contain an OA population weighted centroid. Without applying some additional adjustment to the data, this ward will not have any OA estimates attached to it and the ward would be returned as having no population or households. For this reason, ONS made the decision to apply a manual adjustment to the data to allocate population weighted centroids to geographies without them.



Ward-OA Lookup	
WARD	**OA**
WARD 1	no allocation
WARD 2	OA 1
WARD 3	OA 2

Figure 3: Best fit showing a ward (WARD 1) with a small population. As it contains no OA population weighted centroids, it would not be included in the best fit lookup without a manual adjustment being applied

In the majority of cases, such as **figure 4**, geographies with a small population are contained within an OA that has its PWC attached to another geography within the OA. Users accessing the OA level statistics will therefore be seeing the estimates for both geographies.



Ward-OA Lookup	
WARD	**OA**
WARD 1	OA 1
WARD 2	OA 1*

* nearest population weighted centroid

Figure 4: Best fit showing a ward (WARD 2) with a small population contained within a single OA. As the ward is within a single OA, it is allocated the OA centroid nearest to its boundary

ONS decided not to merge geographies together as they did for the CAS tables in 2001 and also that they would allocate every instance of a geography to an OA, even without a PWC. ONS identified instances where an area contained no OA PWC, and ensured that at least one PWC is allocated based on the nearest OA PWC for an OA that overlaps the higher geography. This means that a single OA can be allocated to more than one instance of a geography. This also means that each instance of a geography is allocated to an OA, even one which has a very small population and doesn't contain an OA PWC.

Users of these estimates for wards and parishes with very small populations should therefore be aware that the estimates produced for wards and parishes cannot be aggregated together to produce totals for higher geographical areas, as this would be double counting the same OA's estimates.

ONS is aware of the risk of double counting, but an indicator within the lookups will clearly set out where areas with very small populations have been allocated to the same OA. Users who want to access estimates at a higher geographic level will be able to do this directly from the census tables (where estimates for this geography are available) or by best-fitting OA PWCs themselves to a higher geography, rather than by aggregating wards or parishes.

The problems created by geographies with a small population are limited to administrative wards, parishes and communities, and built-up areas (BUA). These instances are limited in the context of the overall geographies for which statistics are published and are limited to 0.2 per cent (18) of wards, 10 per cent (1141) of parishes and 10 per cent (626) of BUAs.

7. Are there any instances where best-fitting cannot be applied?

There are some geographies for which best-fitting directly from OA is not appropriate. ONS has identified three of these for which a more appropriate solution was provided.

a) Local authority districts

Local authority districts (LAD) are the most common output geography for which statistics are produced. Due to the large amount of change at this geographic level 2011 OAs have been realigned to LAD boundaries as at 31 December 2011. As the primary administrative geography used as a basis for funding allocation, OAs and SOAs were designed not to cross LAD boundaries as it would mean

two different LADs being responsible for the same OA's population. This change only affects two LADs in Wales as demonstrated for Merthyr Tydfil in **figure 5**. This ensures that exact-fit estimates can still be produced for LADs by aggregating OAs that constrain to LAD boundaries, consistent with ONS's 2001 Census approach. The exact-fit of data to changed LAD boundaries will only be done for Census and as part of any future Census maintenance to the OA /SOA hierarchy. ONS will not modify OAs to fit any LADs as they change.

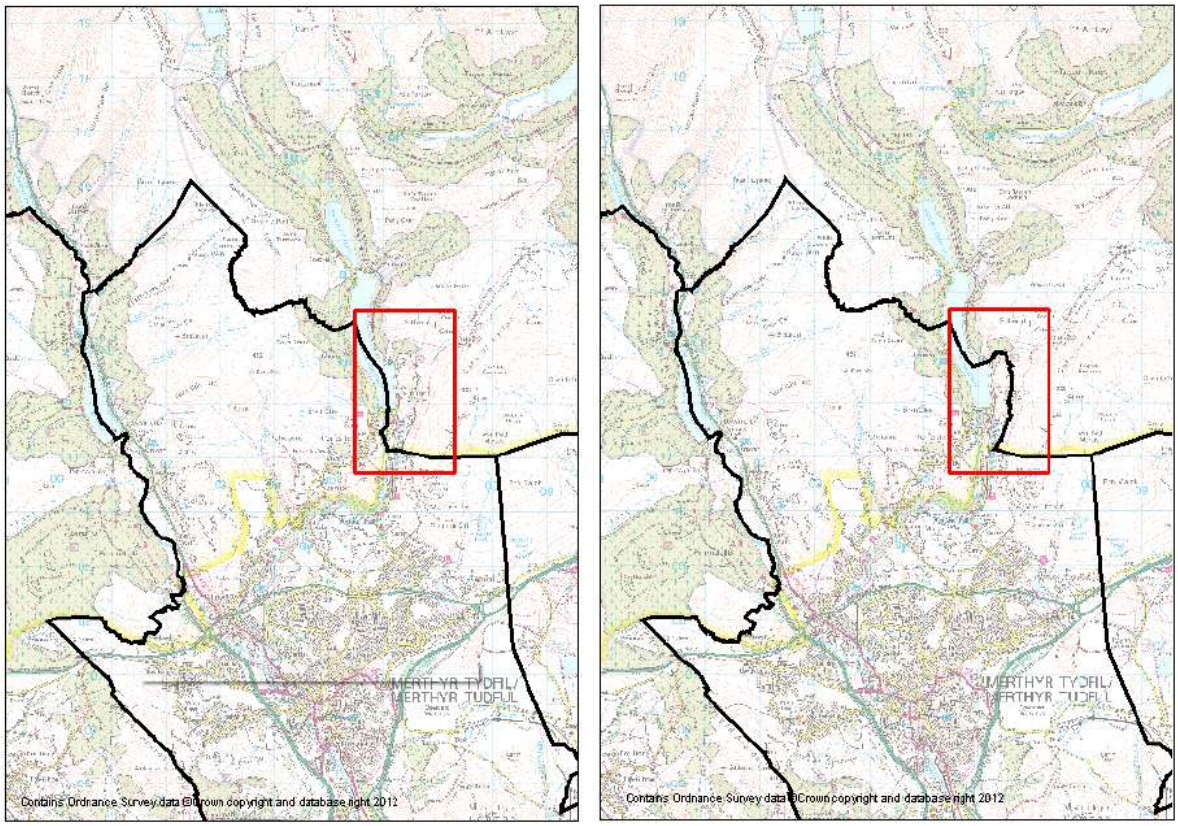


Figure 5: Changes in the local authority boundary for Merthyr Tydfil between 2003 and 2011

b) National parks

National parks are a unique geography that do not align with any other geographic boundaries in England and Wales. They tend to be sparsely populated and the small pockets of population that do live within a national park are often in an OA that straddles the national park boundary. The OA PWC often falls outside the national park as shown in **figure 6**. This means the OA's estimates would not be included in the estimates for the national park.

Other research has concluded that best-fit estimates for national parks can have large differences to exact-fit estimates and are therefore not fit for purpose.

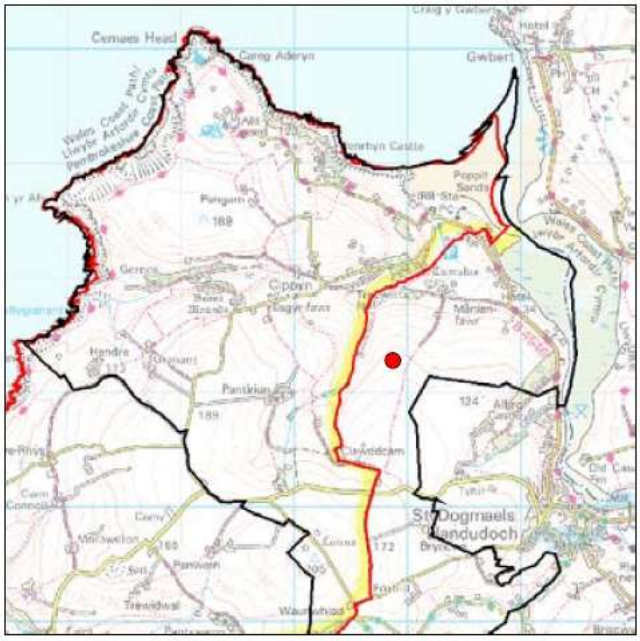


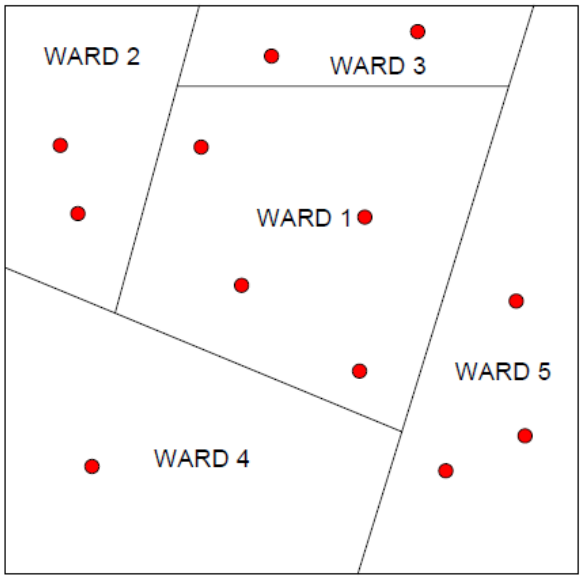
Figure 6: The villages within the national park boundary (red/yellow line) are excluded from any best fit counts for the national park as the population weighted centroid falls outside the park boundary

For this reason, ONS decided that exact-fit estimates will be produced for national parks from the 2011 Census. This geography takes its estimates directly from the census data rather than from aggregations of OAs, LSOAs and MSOAs.

c) 2011 Census merged wards

In 2001 Standard Table (ST) wards were used to produce multivariate statistics at a higher population threshold than OA. For ST wards, higher thresholds of 1,000 persons and 400 households were applied. Wards that fell below this threshold were merged with adjacent wards until they reached the threshold. ST wards are roughly equivalent to the 2011 Detailed Theme tables.

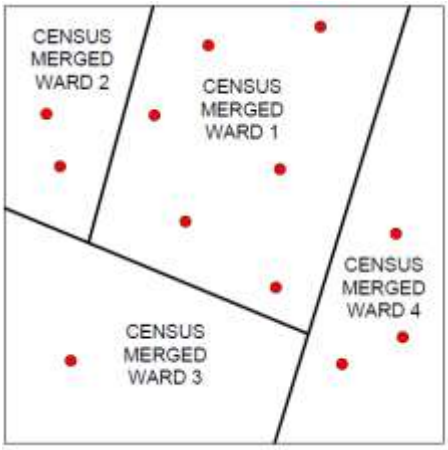
For the purpose of providing multivariate statistics in 2011, wards below the population threshold of 1000 persons or 400 households were merged with an adjacent ward until the population of the merged wards was above threshold (see **figures 7 and 8**). This effectively forms a separate geography to the 2011 wards (called 2011 Census merged wards). In line with the policy, estimates for the 2011 Census merged wards are best-fitted from OA.



Ward Counts	
WARD	**Count**
WARD 1	800
WARD 2	1200
WARD 3	1500
WARD 4	1300
WARD 5	1100

Figure 7: Ward estimates best fitted from the OA population weighted centroid. As WARD 1 contains less than 1000 population, it will need to be merged to provide enough population to take it above the lower threshold

The solution is to identify the nearest OA PWC to the disclosive ward and merge the ward containing the OA PWC with the ward with the disclosive population. This creates a new geography (roughly comparable with ST wards in 2001) for which detailed characteristics tables can be produced without the risk of disclosure, by best-fitting OAs to the instances of 2011 Census merged wards. As this is a new and separate geography all the areas are coded differently even where they represent the same boundary for the 2011 wards.



Census Merged Ward Counts		
WARD	**Census Merged Ward**	**Count**
WARDS 1 and 3	CENSUS MERGED WARD 1	2300
WARD 2	CENSUS MERGED WARD 2	1200
WARD 4	CENSUS MERGED WARD 3	1300
WARD 5	CENSUS MERGED WARD 4	1100

Figure 8: WARD 1 has been merged with WARD 3 to increase the population count to a non-disclosive level. This has been done as WARD 3 contains the nearest OA PWC to WARD 1

8. What best-fit tools and products are available?

As part of the 2011 Census geographic outputs, lookups are available to support the use of best-fit. These lookups include an allocation of OA to higher geographies to allow the aggregation of census estimates up to higher geographies[3].

[3] Not all BUAs are assigned an OA, as many are functional sites that contain little or no population. Details of what is included are available in the metadata for this product.

ONS Geography has also published a complete set of OA population weighted centroids to allow users to produce their own aggregations to higher geographies. Using these population weighted centroids allows users to accommodate new, changed or their own geographies but still be able to produce National Statistics in line with the Geography Policy for National Statistics.

In addition, ONS Geography publishes the National Statistics Postcode Lookup (NSPL). The NSPL is a product that allocates live and terminated postcodes to each OA, based on the geometric centroid of the postcode and to a range of higher geographies best-fitted from the postcode's OA. This allows users to aggregate postcode referenced data in line with the Geography Policy for National Statistics.

These products allow users of statistical data to generate estimates that are consistent and comparable. ONS will however, continue to investigate opportunities for new tools to improve the usability of statistics in line with geographic policy. Should you have recommendations or suggestions, please direct them to ONS Geography Customer Services at ons.geography@ons.gsi.gov.uk.

Appendix I

Production of 2001 Exact-fit Estimates

2001 Census estimates for output geographies were produced on an **exact-fit** basis. Exact-fit describes the method of allocating estimates to a geography by plotting the grid reference of each census household address, and the populations in them, into the boundary of the output geography and aggregating the population counts up to the geography level. A grid reference is a point on the ground that will always either fit within or outside the boundary of the geography. This is why this method of allocating estimates to geographies is known as exact-fit, as estimate and the households it is comprised of fit exactly to its output geography.

OAs and SOAs are the core statistical geographies from which building blocks of statistics are built for all higher geographies. They were created from 2001 Census data and are therefore known as 2001 OAs and SOAs.

2001 OAs and SOAs were designed to contain consistent numbers of census population, and were intended to be stable geographies that would allow reporting of statistics across time on a consistent geographical base. This is why they are known as building block geographies. They had a set minimum population size, so that statistics released for them would not be disclosive and had a target population size to make them as comparable as possible.

The 2001 OA and SOA geographies fitted exactly within the administrative geography boundaries, like LADs, wards and parishes, that were current at the time the OAs and SOAs were created. Because the OA and administrative geographies aligned, exact-fit estimates were created by ONS plotting the grid reference of each household address directly into the boundary of the output higher geography. **Figure 9** shows how all households fall into Ward 1 and OA 1, OA2, OA3 and OA4 that are exactly contained within Ward 1.

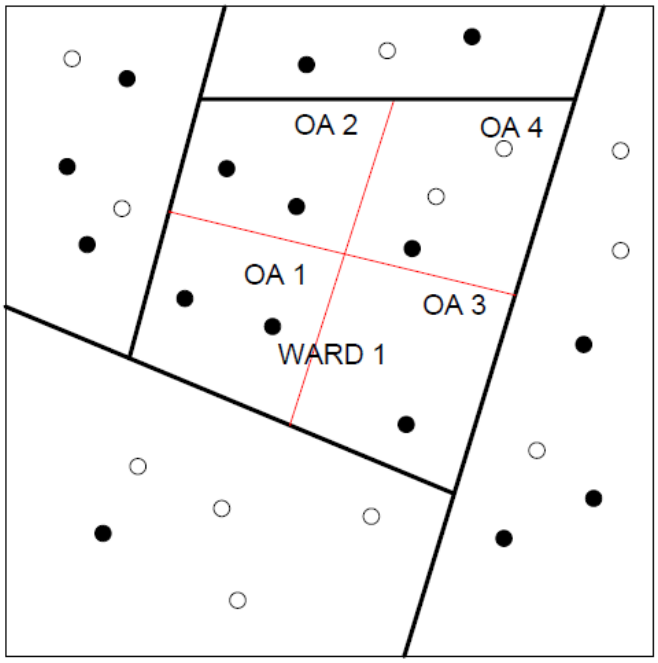


Figure 9: In 2001, households fit exactly into WARD 1. WARD 1 has 8 households

Therefore OAs are the lowest level at which non-disclosive, exact-fit statistical estimates were produced for 2001 Census outputs. However, there is also a requirement to produce statistics for a range of other, higher-level administrative, electoral, and health geographies, as well as other current and historical geographies. This paper explores the options that ONS considered for producing statistical estimates and lookups, along with detailing the solution adopted for 2011 Census.

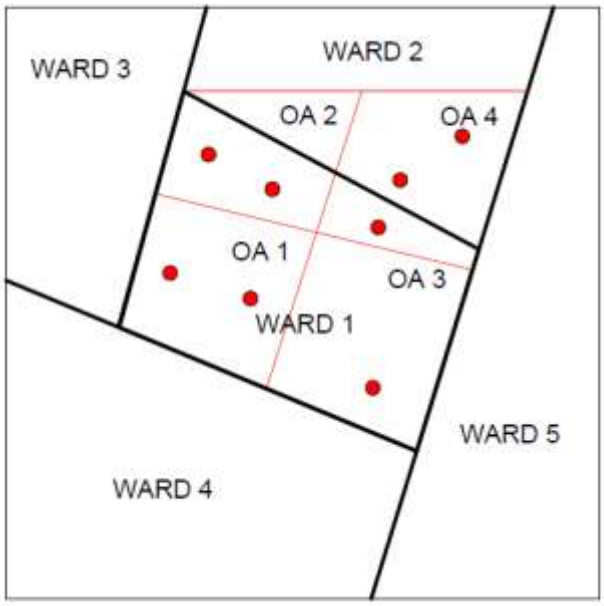
Appendix II

2001 Best-fitting to Higher Geographies

After 2001, lookups were produced that best-fitted OAs to higher geographies as they changed, using the 2001 Census data population. In the majority of cases, the OAs were entirely contained within each higher geography boundary. No additional methodology was needed as these OAs and all their associated estimates were included within the estimates for the larger higher geography.

In some cases, however the OA did not sit entirely within the higher geography, and was split between two instances of the higher geography. In these cases a decision had to be made as to which higher geography the OA should be allocated, i.e. best-fitted.

In 2001 an OA that was split across two instances of the higher geography was mapped to the higher geography in which most of the OA population fell. **Figure 10** illustrates how these allocations were made.



OA-Ward Lookup	
OA	**WARD**
OA 1	WARD 1
OA 2	WARD 1
OA 3	WARD 1
OA 4	WARD 2

Figure 10: 2001 Best fit of Output Areas to Higher Geographies. OA4 is allocated to WARD 2 because it contains more of the OA population

ONS used this methodology to provide annual lookups from OA to other geographies, as these higher geographies frequently changed their boundaries between 2001 and 2011. This allowed users to aggregate estimates produced for an OA, LSOA or MSOA to produce estimates for any higher geography for which ONS provided an OA best-fit lookup[4]. However, an issue with this methodology is that in order to reallocate OAs to higher geographies you need access to the confidential 2001 Census source data, to which only ONS has access.

[4] An Upper Layer SOA (USOA) dataset exists for Wales, but not for England