Incorrect flows within 2011 Census Origin and Destination statistics

1. Summary of the issue

A number of public workplace and second residence origin-destination tables were released on 25th July 2014 (at both Local Authority and MSOA level). Since the release a number of queries have come into Census Customer Services questioning unusual flows in the data, and upon investigation it has been found that some anomalies exist within the workplace data. These anomalies include flows of people between Local Authorities where such flows would not be expected, for example because of very large distances between them.

This report explains the issue in more detail, why it occurred, the extent of the issue, the impact upon the data, and the reasons behind ONS's decision not to correct the data.

2. What is the cause of the errors?

In the vast majority of cases, the workplace addresses from the 2011 Census questionnaires were captured correctly, and the suite of tables included in the origin-destination product contains a high degree of statistical utility.

However, it is reasonable to expect that the coders would have difficulty in coding some workplace addresses. For example, this can occur given that a lot of respondents will only have provided partial or incorrect information. In these instances, some workplace addresses were coded to a different address of the same, or a similar, name. Some of the cases also appear to have been caused issues in some of the data processing stages, for example imputation.

A lot of the errors appear to have happened in clusters, where people who actually work at one specific workplace address have been incorrectly coded to another place. This may have been the result of manual coding, as mentioned above, or could have been caused by errors in the coding system or address register. More information on how the coding of addresses was carried out is provided in Annex A of this paper.

3. What is the extent of the issue?

ONS believes that this issue occurred at a very early stage of data capture and processing, therefore it is difficult to assess an overall total for the numbers of people or the areas affected. The table included in Annex B of this document includes a list of potential errors identified and the numbers affected. However, the flows and numbers provided here are approximations and this is not necessarily a fully comprehensive list.

4. What is the impact on the data?

The impact on the data is that a very small proportion of the flows reported in the origin-destination statistics involving workplace are incorrect. Annex B lists the potential errors spotted by ONS; this is not a comprehensive list as there may be others which are as yet unidentified. Any additional errors are also likely to have a similarly small impact.

As well as the workplace origin-destination tables (SWS), some of the second residence origin-destination tables (SRS) also show flows to place of work. Of the 20 tables that were released on 25th July 2014, 9 of the tables would be affected by errors in the workplace address data - 6 SWS tables and 3 SRS tables.

For the remaining priority origin-destination tables that are due to be released shortly as either safeguarded or secure data, 6 of the 14 safeguarded tables are affected (all 6 SWS tables), and 6 of the 10 secure tables are affected (both SWS tables).

For the non-priority origin-destination tables that are scheduled for release towards the end of the year, all SWS tables and any SRS tables that show flows to place of work will be affected by the errors.

In addition to the origin-destination data, there are a number of other census products that have used information on workplace address which would also be impacted:

Workday population - 22 tables that redistribute the usually resident population to their place of work, while residents who are not in work remain in their area of usual residence. 13 of these tables were released in October 2013, accompanied by a short story, an infographic, interactive maps, and population pyramids. The remaining 9 tables were released in May 2014 and at the same time, 10 of the 13 tables released in October were corrected due to a production error found affecting 8 output areas.

Workplace population - 32 tables that redistribute the usually resident working population to their place of work. These tables were released in May 2014 and accompanied by a short story and a set of interactive maps.

Distance travelled to work - 20 tables that show the distance travelled to work for usual residents in employment the week before the census. These tables were released in March 2014 and accompanied by a short story.

Armed forces - 1 of the Armed Forces tables provides information on members of the armed forces by workplace address by sex by age. This table was released in May 2014.

Microdata variables - 5 of the microdata variables are derived using information on workplace, 4 are used in the secure file and 3 are used in the safeguarded file.

Eurostat - 6 of the data cubes and 1 of the quality cubes provided to Eurostat in March 2014 use information on workplace, as well as metadata and related notes.

Longitudinal Study - the census extract provided to the LS team includes data on workplace and distance travelled to work.

5. How is this issue being resolved?

ONS have taken the decision to leave the workplace data as they are and not attempt to make any revisions.

ONS considered correction of the data and all affected outputs but it recognised that any work carried out would not result in a fully correct set of outputs; because only the errors that have been identified would be fixed, there would likely be more errors that have not been identified and would remain in the data. It would therefore be difficult to provide a message to users about the nature of the corrections, as it would not be possible to confirm that all errors in the data have been corrected.

ONS also considered the suppression of the incorrect flows in the origin-destination data. Although incorrect flows would no longer be visible to users, users would be able to compare the origin-destination data with previously published workplace statistics to calculate the suppressed flows. Again, it would also only be possible to suppress the incorrect flows that have been identified.

Lessons learnt from this issue will be taken forward for the 2021 Census. ONS apologises for any inconvenience caused by this situation.

Annex A - How workplace address was coded

The following table shows how the address information provided by respondents was coded:

Information provided on questionnaire	**If matched against Address Register**	**If no match found (1)**	**If no match found (2)**	**If no match found (3)**
Postcode and address	Output postcode and Unique Address Identifier to WKPLUAI. (NB: postal town from enumeration address used if postal town not provided. Expert coders only).	Add organisation name to workplace address and attempt an address match.	If address match fails, attempt a postcode match.	Output captured postcode and captured address.
Postcode only	Output postcode.	Add organisation name and attempt a match.	Output captured postcode.	N/A
Address only	Output matched address postcode and Unique Address Identifier. (NB: postal town from enumeration address used if postal town not provided. Expert coders only).	Add organisation name to workplace address and attempt a match.	Output captured address.	N/A

Any postcodes that remained unmatched after coding was completed were sent to ONS Geography for resolution. Where a postcode was found by Geography, this was loaded into the database prior to Post Coverage Item Imputation.

Annex B – List of potentially incorrect flows

This list has been compiled by looking at the top 10 inflows and outflows for every local authority using the workplace data on NOMIS to identify any flows that appear unusual (down to a threshold of 100 usual residents). For a sample of the unusual flows, the original images of the census questionnaire were checked, to see how the respondent had recorded their workplace address.

Enumeration address	Workplace address	Number of usual residents	Proportion of outflow (%)	Proportion of inflow (%)
Sheffield	Bury	1193	2.56	4.43
Swindon	Plymouth	480	1.94	1.85
Sheffield	South Lanarkshire	478	1.03	1.38
Nottingham	Luton	461	1.20	1.34
Barnsley	Shropshire	387	1.02	1.32
South Gloucestershire	Manchester	354	0.66	0.20
Wolverhampton	Midlothian	332	0.82	2.99
South Gloucestershire	Rugby	316	0.59	1.80
Newport	Isle of Wight	283	1.32	13.42
Bristol	Calderdale	262	0.48	0.97
Newcastle upon Tyne	Ealing	260	0.66	0.41
Plymouth	Fife	241	1.13	1.79
Newcastle upon Tyne	Newcastle under Lyme	219	0.56	1.02
Sheffield	Lincoln	211	0.45	0.82
North Lincolnshire	North Lanarkshire	200	1.25	0.54
St. Helens	Swansea	193	0.54	0.69
Bradford	Torridge	192	0.34	5.36
South Tyneside	Aberdeen City	191	0.67	0.37
Manchester	Bath and NE Somerset	181	0.25	0.62
Flintshire	South Lakeland	178	0.69	1.83
Cornwall,Isles of Scilly	Wiltshire	175	0.88	0.43
South Ayrshire	Outside UK	156	1.23	0.30
Havering	East Dorset	143	0.24	1.06
Corby	North Lanarkshire	136	1.52	0.37
Isle of Wight	Gloucester	131	2.73	0.50
Conwy	Liverpool	119	0.96	0.13
Denbighshire	Manchester	115	1.00	0.06
Dudley	Midlothian	115	0.20	1.04
Luton	Teignbridge	105	0.31	0.86
North Tyneside	Newcastle under Lyme	100	0.23	0.47