STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

COMMISSION OF THE EUROPEAN COMMUNITIES

CONFERENCE OF EUROPEAN STATISTICIANS

EUROSTAT

Joint ECE/Eurostat Work Session on Statistical Data Confidentiality
(Thessaloniki, Greece, 8-10 March 1999)

Working Paper No. 26
English only

Topic (i): new applications of disclosure control methods

RECORD SWAPPING - A POSSIBLE DISCLOSURE CONTROL APPROACH FOR THE 2001 UK CENSUS

Submitted by the Office for National Statistics, United Kingdom[1]

Contributed paper

Abstract

The UK Census Offices make a commitment that all data collected from the Census will be used for statistical purposes only and that no data will be released that identifies individuals. In previous Censuses the confidentiality of the data has been protected by a variety of techniques that adjust tabular output. In 1991 these techniques led to inconsistencies between the output for small areas, where protection was implemented, and larger areas, which did not require protection. There is a demand from users of Census data for consistency between outputs and for more flexible outputs to meet a variety of user needs. To meet these requirements whilst protecting the confidentiality of individual data, the UK Census Offices are considering adjusting the database prior to output. This paper discusses preliminary results of one of these 'pre-tabulation' techniques, namely record swapping. It shows how the database is altered by the method and illustrates that the effect on the messages coming from the database is generally small.

I. INTRODUCTION

1. The 2001 UK Census will take place in April 2001 and will be conducted by the Office for National Statistics (ONS) for England and Wales, the General Register Office (Scotland) (GRO(S)) and the Northern Ireland Statistical and Research Agency (NISRA). It will be the 20th decennial census to take place, the first being conducted in 1801 for Great Britain.

2. Although Censuses in the UK and most other countries are compulsory, it is important that a Census is seen to be, and is, confidential, in order to gain widespread public co-operation. In the UK this has traditionally been achieved firstly by giving assurances about confidentiality on Census forms and in publicity and secondly by protecting personal information using legislation in the 1920 Census Act and 1991 Census Confidentiality Act. These two measures are relatively easy to implement, however it is much more difficult to protect against inadvertent disclosure (whether real or perceived) through the authorised production of apparently anonymous statistical information. The rest of this paper concentrates on this concern.

3. In 2001, in contrast to previous Censuses, the UK Census Offices are considering protecting the database before any results are given. These methods are known as pre-tabulation techniques. This paper reports on why pre-tabulation techniques may now be more appropriate than the post-tabulation techniques which have

[1] Prepared by Maggie Boyd and Paul Vickers.

been used in previous Censuses. One possible pre-tabulation technique, record swapping, is discussed in some detail and the results of some preliminary analysis showing the effect that record swapping has on the structure of the database are given.

II. PRE-TABULATION AND POST TABULATION TECHNIQUES

4. Since 1971 GB Census data has been protected against disclosure using the post-tabulation techniques of *Barnardisatio*n and *thresholding*. *Barnardisation* is a procedure that randomly modifies individual cell counts by plus or minus one, while *thresholding* limits the minimum size of population for which tables are published. Advantages of such methods lie in their transparency - it is clear that some protection has been applied and *thresholding* is generally easily understood and easy to explain. The disadvantages, however, of post-tabulation methods are substantial: production of amended tables can be cumbersome and the data that can be released is limited in order to prevent 'differencing' (disclosure by subtracting close, but non-identical areas, from one another to identify a smaller area). Other problems are also apparent. For example, users experienced problems with the outputs from the 1991 GB Census when the totals of tables for small geographic areas were summed, they did not equal the values in the tables for the corresponding larger geographic areas.

5. User expectations of Census outputs have increased since 1991 and this has resulted in plans to produce flexible and customised output in 2001. Post-tabulation methods are time consuming and are not well-suited to the sort of output that is now demanded. In addition, customised outputs will increase the risk of 'differencing' and post-tabulation techniques are unlikely to provide sufficient protection. Finally, powerful computing technology has made it possible to develop sophisticated matching software and the risk of disclosure by matching Census micro-data with data collected from other sources has increased (Winkler 1998).

6. It is for these reasons that we are considering the use of a pre-tabulation method in 2001. Pre-tabulation techniques include record swapping, data switching and over-imputation. Such methods modify the database before data is aggregated. The aim is to adjust the Census data sufficiently so that a level of uncertainty becomes attached to every record. This attempts to ensure that if an intruder identifies an individual, they will not know whether it is the information given by the individual or whether it has been adjusted in some way.

7. The major advantage of such methods is that once an adjusted database has been created it can be aggregated in any way. This means that the degree of flexibility for outputs is maximised and the output produced will be consistent: tables produced for small areas will add up to tables produced for larger areas. Additionally, each table produced will not need to be examined for disclosure risk.

8. There are, however, disadvantages to such methods. Users may not like the fact that the data is actually altered by the disclosure control technique, and the methods can be difficult to explain to the public. The level of pre-tabulation disclosure control applied to the 2001 data will be a trade off between the need to protect the data by altering individual records and the need to preserve the statistical structure of the data (de Waal and Willenborg 1994). Additionally it is debatable whether or not pre-tabulation methods protect individuals with rare characteristics. For these reasons it is likely that a post-tabulation method such as thresholding will be used alongside a pre-tabulation method (if pre-tabulation is acceptable) in the 2001 Census.

III. RECORD SWAPPING METHODOLOGY

9. Record swapping involves swapping a complete household record with a similar record in the same geographic area. This is achieved by swapping the geographical references of two records. The method has also been studied for US Censuses (see for example, Griffin et al 1989 and Steel 1998).The UK is divided into Local Authority Districts (LADs). The size of each LAD varies but generally they contain around 50,000 households. All swapping in this research has been conducted within one LAD.

10. Record swapping maintains the internal consistency of household records and ensures that at the level of geography at which the swapping occurs, the structure of the data will be perfectly preserved. Generally record swapping will make all areas smaller than the LAD become more like the LAD as a whole. Hence areas that are more homogenous than the LAD will become less homogenous as a result of record swapping, and likewise, areas that are more heterogeneous than the LAD will become less heterogeneous.

11. We investigated record swapping using 1991 UK Census data from an LAD containing approximately 80,000 households. The results presented here consider the effect of swapping among the 73,509 households that contain four or less people. Investigations into record swapping among larger households are underway, and will be presented at the conference.

12. The LAD was split into 17 Processing Units (PUs), each of which consists of 25 Enumeration Districts (EDs). The PUs contain between 3,500 and 5,000 households and can be considered analogous to Wards, for which information was not available in the dataset used. Five PUs and three EDs were analysed in detail in order to see the effect of record swapping at small geographic levels. Local Authorities in Britain are often very varied in terms of the types of population that they contain, and analysis of a variety of PUs and EDs allowed us to consider the effects of record swapping in different types of area.

13. It is important to be able to let users know the possible effects that record swapping may have on the data. This analysis looks at how different levels of record swapping alter the structure of the data. This will inform Census Offices of how much record swapping can be achieved before the integrity of the data is harmed. If a pre-tabulation method is used in the 2001 Census the precise level of pre-tabulation will not be released. It is likely that we would say that between 0 and X% of the data was affected. X would not exceed a level that would harm the data. This analysis looked at five different proportions of records being swapped, 1%, 3%, 5%, 10% and 20%.

14. The swapping process consisted of three main stages:

i) A subset of records were selected for swapping using Systematic Random Sampling (SRS)

ii) Records were paired within the subset using the following criteria:
- Number of people in the household
- Sex of each individual in the household
- Age band of each individual in the household (0-15, 16-29, 30-44, 45-64, 65+)

iii) Paired records were swapped with each other. If no exact match was found the record was not swapped.

15. Because all swapping occurred among the X% of records selected out from the main data, generally slightly less than X% of records were actually swapped as there were always some records for which no exact match was found. The level of swapping that actually occurred is shown in table 1 below.

Table 1: Records selected for swapping

% Swap	Total no. households	No. households selected for swapping	% households selected for swapping	Actual no. of households swapped	Actual % of households swapped
1%	73509	718	0.977%	634	0.862%
3%	73509	2162	2.941%	2024	2.753%
5%	73509	3648	4.963%	3463	4.711%
10%	73509	7356	10.007%	7132	9.702%
20%	73509	14962	20.354%	14748	20.063%

16. Because records are matched on certain criteria the process does not alter the population or sex distribution of the data at any geographic level. The number of households in each area remains the same as well. The affect of swapping on the age structure was minimised by matching records on a grouped age variable (5 groups). Matching on exact age of each individual proved to be too much of a constraint on finding a match.

IV. EVALUATION CRITERIA FOR RECORD SWAPPING

17. The purpose of this analysis was to see how record swapping affects the internal structure of the data. We analysed how the marginal distributions and joint distributions between the key variables changed as a result of record swapping. The following key variables were considered:

- Sex (male, female)
- Age (0-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90+)
- Marital Status (single, married, remarried, divorced, widowed)
- Ethnic Group (white, non-white)
- Country of Birth (UK/ Ireland, other)
- Primary Activity Last Week (employed, unemployed, economically inactive, not applicable/ missing)
- Tenure (owner occupied, rent council, rent private)

18. Cramers V Statistic (C) was used to assess how record swapping altered the distributions of these variables and the relationships between them at the PU and ED levels. C is a measure of association based upon the chi-square statistic. It allows us to consider how close the post swapping distributions are to the original distributions. The closer C is to +1, the higher the level of association.

19. We also compared a number of typical Census output tables that were created from the original data and the post swapping data to illustrate the effect of swapping.

V. ANALYSIS

20. Analysis was conducted on five out of 17 PUs and three out of 204 EDs. The discussion in this section will draw on results from all five PUs but will focus on two PUs that are substantially different from each other. PU1 has the highest proportion of ethnic minorities (12%), the highest proportion of council housing (households renting their house from the public sector) (38%) and the highest unemployment rate (18%). In contrast PU16 has a small ethnic minority population (1%), lower levels of council housing (19%) and a low unemployment rate (8%) compared to the LAD average. The EDs were selected to illustrate the effect of swapping on an average ED and two 'extreme' EDs based upon their unemployment rates. Therefore ED1 has an unemployment rate close to the LAD average of 10%, ED2 has a very low rate (2%) and ED3 has a high rate (21%).

VI. MARGINAL DISTRIBUTIONS

21. A comparison of the distribution of each of the key variables before and after swapping in PUs 1 and 16 is shown in Annex 1. It shows that swapping, even at the 20% level, has a small effect on the structure of the data in each PU. As we would expect, the difference between the distributions pre and post swapping increases as the proportion of records swapped increases. However, even at the 20% level, in the 5 PUs analysed, the percentage difference caused by swapping is less than 1 percent in the majority of cases. Where the difference exceeds 1 percent the numbers are generally small. The exception is 'ethnic group' and 'country of birth' in PU1 at the 20% level. Tables 2 and 3 below look at this more closely and we can see that up to the

10% level of swap the difference between pre and post swapping is small, but once swapping hits the 20% level the difference between distributions increases substantially.

Table 2: Affect of different levels of swap on the distribution of Ethnic Group and Country of Birth for PU1 (frequencies)

PU1		**Before Swapping**	1% Swap	% Difference	3% Swap	% Difference	5% Swap	% Difference	10% Swap	% Difference	20% Swap	% Difference
Ethnic Group	Missing	**98**	98	0.00	98	0.00	96	-2.04	97	-1.02	96	-2.04
	White	**5743**	5740	-0.05	5752	0.16	5761	0.31	5749	0.10	5782	0.68
	Other	**775**	778	0.39	766	-1.16	759	-2.06	770	-0.65	738	-4.77
Country of Birth	Missing	**62**	63	1.61	61	-1.61	63	1.61	62	0.00	62	0.00
	UK/Ireland	**5754**	5749	-0.09	5755	0.02	5760	0.10	5759	0.09	5779	0.43
	Other	**800**	804	0.50	800	0.00	793	-0.88	795	-0.63	775	-3.13

Table 3: Affect of different levels of swap on the distribution of Ethnic Group and Country of Birth for PU1 (%)

PU1		**Before Swapping %**	1% Swap %	Difference	3% Swap %	Difference	5% Swap %	Difference	10% Swap %	Difference	20% Swap %	Difference
Ethnic Group	Missing	**1.48**	1.48	0.00	1.48	0.00	1.45	-0.03	1.47	-0.02	1.45	-0.03
	White	**86.80**	86.76	-0.05	86.94	0.14	87.08	0.27	86.90	0.09	87.39	0.59
	Other	**11.71**	11.76	0.05	11.58	-0.14	11.47	-0.24	11.64	-0.08	11.15	-0.56
Country of Birth	Missing	**0.94**	0.95	0.02	0.92	-0.02	0.95	0.02	0.94	0.00	0.94	0.00
	UK/Ireland	**86.97**	86.90	-0.08	86.99	0.02	87.06	0.09	87.05	0.08	87.35	0.38
	Other	**12.09**	12.15	0.06	12.09	0.00	11.99	-0.11	12.02	-0.08	11.71	-0.38

22. Swapping does not seem to adversely affect the variables ‘sex’, ‘age’, ‘marital status’, ‘primary activity last week’ and ‘tenure’. ‘Sex’ is clearly not affected because it was one of the control variables. The reason that age differs slightly (despite being controlled for during the swapping process) is because the groups used for the analysis are not identical to those used during swapping.

23. Annex 2 shows the distribution of key variables at the ED level. The results show that swapping has a small impact on the unemployment rate in each of the three EDs considered. In the ED with an average unemployment rate, swapping at the 20% level causes the unemployment rate to fall to 8%, but this only reflects a drop of three in the number of unemployed. In ED2 (the low unemployment ED) swapping caused the unemployment rate to rise slightly towards the LAD average and in ED3 (the high unemployment ED) the rate rises slightly at the 1% level swap and falls towards the LAD average at the 20% swap level.

24. In order to further investigate how the swapping process altered the structure of the marginal variables at the PU level of geography, Cramers V was used to assess the association between variables before and after swapping (see table 4). The association between each variable before and after swapping is high in all PUs and C is greater than 0.75 in every case. Not surprisingly the value of C falls as the level of swapping increases. C also tends to be lower for the variables ‘ethnic group’ and ‘country of birth’ than for any of the other variables that we looked at.

25. Results obtained at the ED level are shown in table 5 below. The results are similar to those for PUs in EDs 2 and 3 and C is greater than 0.78 in all cases. However ED1 shows considerably lower levels of association, especially when the variables ‘ethnic group’, ‘tenure’ and ‘country of birth’ are analysed. At the 20% level ‘country of birth’ gives a value of C of 0.532, however the numbers involved are small (see table 7 in annex 2).

Table 4: Cramers V (C) - association between marginal variables before and after swapping for PU1 and PU16

	PU1					PU16				
Variables	1% Swap	3% Swap	5% Swap	10% Swap	20% Swap	1% Swap	3% Swap	5% Swap	10% Swap	20% Swap
Age2	0.998	0.982	0.974	0.949	0.907	0.996	0.984	0.973	0.955	0.89
Sex	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Marital Status	0.995	0.982	0.974	0.942	0.884	0.994	0.986	0.976	0.947	0.888
Primary Activity Last Week	0.996	0.986	0.976	0.951	0.897	0.997	0.990	0.981	0.959	0.896
Ethnic Group	0.993	0.976	0.953	0.939	0.825	1.000	0.943	0.967	0.889	0.828
Country of Birth	0.993	0.974	0.955	0.924	0.822	0.996	0.963	0.959	0.894	0.795
Tenure	0.992	0.977	0.959	0.929	0.872	0.993	0.981	0.966	0.926	0.836

Table 5: Cramers V (C) - association between marginal variables before and after swapping for 3 Eds

	ED1			ED2			ED3		
Variables	1% Swap	10% Swap	20% Swap	1% Swap	10% Swap	20% Swap	1% Swap	10% Swap	20% Swap
Age	0.998	0.946	0.900	0.997	0.954	0.886	1.000	0.964	0.923
Sex	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Marital Status	0.987	0.950	0.876	0.981	0.949	0.889	0.994	0.938	0.866
Primary Activity Last Week	1.000	0.947	0.884	0.997	0.924	0.822	0.998	0.953	0.893
Ethnic Group	1.000	1.000	0.773	1.000	0.885	0.945	1.000	1.000	0.813
Country of Birth	0.952	0.832	0.532	1.000	0.847	0.869	1.000	0.985	0.785
Tenure	0.975	0.823	0.750	1.000	0.843	0.825	0.992	0.886	0.875

VII. JOINT DISTRIBUTIONS

26. When the association between each of the key variables is compared pre swapping and post swapping we find that the relationships between variables are generally well maintained during the swapping process (see annex 3). In the 5 PUs analysed the largest difference between the value of C before swapping and after swapping is 0.021 (PU16 for 'ethnic group' and 'country of birth'). There are three other cases where the difference is greater than 0.01. Each of these also involved the 'ethnic group' variable, perhaps suggesting that the relationship between other variables and 'ethnic group' are most altered by the swapping process. The differences are not always greatest at the 20% swap level, and there are some cases where the largest differences occur at the 1% level. This reflects our decision to choose independent samples so that the 1% sample was not a subset of the 20% sample.

27. These results are encouraging and perhaps not surprising as you would expect the relationships between variables to be preserved because whole households are being swapped.

28. At the ED level the relationships between variables appears to be less well maintained than at the PU level (see annex 4). Comparing the association (C) between key variables pre and post swapping results in larger differences in the value of C than at the PU level. The largest effect in the EDs analysed was at the 3% level for the variables 'ethnic group' and 'country of birth' where the difference in the value of C pre and post swapping was 0.200. As at the PU level it is not true that the largest differences always occur at the 20% swap level.

[2] These are not 1 as the C statistic is calculated on 10 age groups and not the 5 age groups used in swapping.

VIII. CENSUS TABLES

29. A final piece of analysis involved setting up a mock Census table of marital status by age and sex and comparing the results using pre swapping data and post swapping data. An example for an ED with an average unemployment rate (ED1) is shown in annex 5. This enables us to see that the shape of the table is largely unaltered by the swapping process. Comparing the original table to the table using the 20% swap data we find that some marginal totals have been changed by up to eight people. The marginal totals in the table based upon the 1% swap data are changed by up to one person.

IX. CONCLUSIONS AND FURTHER WORK

30. This paper has considered the need for a pre-tabulation disclosure control method in the light of two opposing demands: the need to assure respondents that the information they supply will not be disclosed versus the demands of users for more flexible outputs.

31. The results are encouraging. They show that quite high levels of record swapping can be achieved without substantially harming the integrity of the data.

32. We recognise, though, that record swapping will not protect households that are unique at the Local Authority District (or higher) level. Therefore, regardless of whether or not pre-tabulation is applied there will still be a requirement for some form of post-tabulation protection such as thresholding.

33. This analysis has only focused on households containing up to four persons and has only been done using the data from one LAD. Over the next month or so we will complete the analysis for larger households and hope to conduct it on at least one more LAD. These will be presented at the conference

Acknowledgements

34. We are indebted to Chris Skinner for his support and advice during the course of this work. We are also grateful to Sarah White and Carol Gregory for developing the record swapping system and to Peter Stokes for helping to produce SAS tabulations.

References

Griffin, R.A.; Navarro, A.; Flores-Baez, L. (1989): Disclosure Avoidance for the 1990 Census. Proceedings of the Section on Survey Research Methods, American Statistical Association, 516-521.

Steel, Philip and Zayatz, Laura (1998): Disclosure Limitation for the 2000 Census of Housing and Population. U.S. Bureau of the Census, Washington D.C. Paper presented at Statistical Data Protection ’98 Conference, Lisbon, March 1998.

U.S. Office of Management and Budget (1994): Report on Statistical Disclosure Limitation Methodology. Washington D.C.

de Waal, A.G. and Willenborg, L.C.R.J. (1994): Principles of Statistical Disclosure Control. Department of Statistical Methods, Statistics Netherlands.

Winkler, William E. (1998): Re-identification Methods for Evaluating the Confidentiality of Analytically Valid Microdata. Paper presented at Statistical Data Protection ’98 Conference, Lisbon, March 1998.

Annex 1

Table 6: Distributions of key variables before and after swapping

		PU1					PU16				
		Before Swap	1% Swap		20% Swap		**Before Swap**	1% Swap		20% Swap	
			After Swap	% Difference	After Swap	% Difference		After Swap	% Difference	After Swap	% Difference
Total Persons		**6616**	6616	0.00	6616	0.00	**10575**	10575	0.00	10575	0.00
Sex	Missing	**27**	27	0.00	27	0.00	**35**	35	0.00	35	0.00
	Male	**3360**	3360	0.00	3360	0.00	**5120**	5120	0.00	5120	0.00
	Female	**3229**	3229	0.00	3229	0.00	**5420**	5420	0.00	5420	0.00
Age	Missing	**60**	60	0.00	60	0.00	**31**	31	0.00	31	0.00
	0 to 9	**636**	636	0.00	635	-0.16	**1219**	1219	0.00	1215	-0.33
	10 to 19	**475**	475	0.00	469	-1.26	**985**	987	0.20	978	-0.71
	20 to 29	**1803**	1803	0.00	1810	0.39	**1979**	1977	-0.10	1990	0.56
	30 to 39	**816**	814	-0.25	815	-0.12	**1411**	1411	0.00	1414	0.21
	40 to 49	**627**	629	0.32	619	-1.28	**1430**	1430	0.00	1424	-0.42
	50 to 59	**681**	680	-0.15	684	0.44	**1188**	1190	0.17	1189	0.08
	60 to 69	**708**	709	0.14	715	0.99	**1073**	1072	-0.09	1067	-0.56
	70 to 79	**560**	559	-0.18	559	-0.18	**899**	896	-0.33	914	1.67
	80 to 89	**229**	230	0.44	229	0.00	**328**	330	0.61	322	-1.83
	90+	**21**	21	0.00	21	0.00	**32**	32	0.00	31	-3.13
Marital Status	Missing	**37**	37	0.00	36	-2.70	**30**	30	0.00	27	-10.00
	Single	**3312**	3315	0.09	3306	-0.18	**3767**	3766	-0.03	3764	-0.08
	Married	**1705**	1703	-0.12	1715	0.59	**4659**	4661	0.04	4662	0.06
	Remarried	**380**	381	0.26	381	0.26	**675**	670	-0.74	687	1.78
	Divorced	**615**	613	-0.33	608	-1.14	**642**	643	0.16	640	-0.31
	Widowed	**567**	567	0.00	570	0.53	**802**	805	0.37	795	-0.87
Ethnic Group	Missing	**98**	98	0.00	96	-2.04	**89**	89	0.00	92	3.37
	White	**5743**	5740	-0.05	5782	0.68	**10348**	10348	0.00	10344	-0.04
	Other	**775**	778	0.39	738	-4.77	**138**	138	0.00	139	0.72
Country of Birth	Missing	**62**	63	1.61	62	0.00	**32**	31	-3.13	35	9.38
	UK/Ireland	**5754**	5749	-0.09	5779	0.43	**10266**	10267	0.01	10264	-0.02
	Other	**800**	804	0.50	775	-3.13	**277**	277	0.00	276	-0.36
Primary	Employed	**2551**	2552	0.04	2568	0.67	**5089**	5083	-0.12	5085	-0.08
	Unemployed	**579**	576	-0.52	572	-1.21	**463**	464	0.22	463	0.00
	Economically inactive	**2614**	2616	0.08	2605	-0.34	**3232**	3237	0.15	3234	0.06
	No code required / missing	**872**	872	0.00	871	-0.11	**1791**	1791	0.00	1793	0.11
	Unemployment rate (%)[3]	**18.5**	18.41	-0.09	18.22	-0.28	**8.4**	8.37	-0.03	8.35	-0.05
Tenure	Missing	**49**	49	0.00	47	-4.08	**35**	35	0.00	34	-2.86
	Own	**1224**	1226	0.16	1230	0.49	**3351**	3349	-0.06	3353	0.06
	Rent LA	**1338**	1336	-0.15	1335	-0.22	**849**	852	0.35	848	-0.12
	Rent Privately	**892**	892	0.00	891	-0.11	**317**	316	-0.32	317	0.00

[2] Percentage difference for employment rate is calculated as employment rate after swapping minus employment rate before swapping

Annex 2

Table 7: Distributions of key variables before and after swapping in three different types of ED

		Average unemployment ED					Low Unemployment ED				
		Before Swap	1% Swap		20% Swap		**Before Swap**	1% Swap		20% Swap	
			After Swap	% Difference	After Swap	% Difference		After Swap	% Difference	After Swap	% Difference
Total Persons		**443**	443	0.00	443	0.00	**449**	449	0.00	449	0.00
Sex	Missing	**1**	1	0.00	1	0.00	**0**	0	0.00	0	0.00
	Male	**211**	211	0.00	211	0.00	**207**	207	0.00	207	0.00
	Female	**231**	231	0.00	231	0.00	**242**	242	0.00	243	0.41
Age	Missing	**1**	1	0.00	1	0.00	**0**	0	0.00	0	0.00
	0 to 9	**41**	41	0.00	44	7.32	**37**	37	0.00	36	-2.70
	10 to 19	**48**	48	0.00	44	-8.33	**34**	34	0.00	35	2.94
	20 to 29	**51**	51	0.00	52	1.96	**32**	32	0.00	32	0.00
	30 to 39	**60**	59	-1.67	67	11.67	**43**	43	0.00	41	-4.65
	40 to 49	**68**	69	1.47	57	-16.18	**56**	56	0.00	60	7.14
	50 to 59	**52**	52	0.00	56	7.69	**61**	61	0.00	58	-4.92
	60 to 69	**51**	51	0.00	55	7.84	**74**	74	0.00	73	-1.35
	70 to 79	**46**	46	0.00	44	-4.35	**70**	70	0.00	72	2.86
	80 to 89	**24**	24	0.00	22	-8.33	**40**	40	0.00	39	-2.50
	90+	**1**	1	0.00	1	0.00	**2**	2	0.00	3	50.00
Marital Status	Missing	**5**	5	0.00	5	0.00	**0**	0	0.00	0	0.00
	Single	**147**	147	0.00	142	-3.40	**116**	117	0.86	119	2.59
	Married	**217**	217	0.00	216	-0.46	**213**	210	-1.41	213	0.00
	Remarried	**18**	19	5.56	25	38.89	**37**	39	5.41	35	-5.41
	Divorced	**22**	21	-4.55	19	-13.64	**16**	15	-6.25	16	0.00
	Widowed	**34**	34	0.00	36	5.88	**67**	68	1.49	66	-1.49
Ethnic Group	Missing	**8**	8	0.00	8	0.00	**0**	0	0.00	2	200
	White	**432**	432	0.00	430	-0.46	**430**	430	0.00	428	-0.47
	Other	**3**	3	0.00	5	66.67	**19**	19	0.00	19	0.00
Country of Birth	Missing	**1**	1	0.00	1	0.00	**0**	0	0.00	0	0.00
	UK/Ireland	**431**	432	-0.23	427	-0.93	**416**	416	0.00	416	0.00
	Other	**11**	10	-9.09	15	36.36	**33**	33	0.00	33	0.00
Primary	Employed	**190**	190	0.00	190	0.00	**176**	174	-1.14	176	0.00
	Unemployed	**20**	20	0.00	17	-15.00	**4**	4	0.00	5	25.00
	Economically inactive	**164**	164	0.00	166	1.22	**214**	216	0.93	213	-0.47
	No code required / missing	**69**	69	0.00	70	1.45	**55**	55	0.00	55	0.00
	Unemployment rate (%)[4]	**9.52**	9.52	0.00	8.21	-1.31	**2.22**	2.25	0.02	2.76	0.54
Tenure	Missing	**0**	0	0.00	0	0.00	**0**	0	0.00	0	0.00
	Own	**406**	406	0.00	400	-1.48	**344**	344	0.00	340	-1.16
	Rent LA	**5**	5	0.00	12	140.00	**83**	83	0.00	86	3.61
	Rent Privately	**32**	32	0.00	31	-3.13	**22**	22	0.00	23	4.55

[3] Percentage difference for employment rate is calculated as employment rate after swapping minus employment rate before swapping

Annex 2 continued

Table 7 continued: Distributions of key variables before and after swapping in three different types of ED

		High unemployment ED				
		Before Swap	1% Swap		20% Swap	
			After Swap	% Difference	After Swap	% Difference
Total Persons		**386**	386	0.00	**386**	0.00
Sex	Missing	**0**	0	0.00	**0**	0.00
	Male	**165**	165	0.00	**165**	0.00
	Female	**221**	221	0.00	**221**	0.00
Age	Missing	**1**	1	0.00	**1**	0.00
	0 to 9	**35**	35	0.00	**35**	0.00
	10 to 19	**16**	16	0.00	**16**	0.00
	20 to 29	**44**	44	0.00	**44**	0.00
	30 to 39	**37**	37	0.00	**38**	2.70
	40 to 49	**45**	45	0.00	**40**	-11.11
	50 to 59	**45**	45	0.00	**48**	6.67
	60 to 69	**68**	68	0.00	**72**	5.88
	70 to 79	**66**	66	0.00	**66**	0.00
	80 to 89	**28**	28	0.00	**25**	-10.71
	90+	**1**	1	0.00	**1**	0.00
Marital Status	Missing	**0**	0	0.00	**0**	0.00
	Single	**136**	136	0.00	**141**	3.68
	Married	**102**	102	0.00	**93**	-8.82
	Remarried	**35**	35	0.00	**46**	31.43
	Divorced	**44**	43	-2.27	**35**	-20.45
	Widowed	**69**	70	1.45	**71**	2.90
Ethnic Group	Missing	**2**	2	0.00	**2**	0.00
	White	**350**	350	0.00	**355**	1.43
	Other	**34**	34	0.00	**29**	-14.71
Country of Birth	Missing	**0**	0	0.00	**0**	0.00
	UK/Ireland	**351**	351	0.00	**356**	1.42
	Other	**35**	35	0.00	**30**	-14.29
Primary	Employed	**110**	109	-0.91	**106**	-3.64
	Unemployed	**30**	30	0.00	**27**	-10.00
	Economically inactive	**201**	202	0.50	**208**	3.48
	No code required / missing	**45**	45	0.00	**45**	0.00
	Unemployment rate (%)[5]	**21.43**	21.58	0.15	**20.30**	-1.13
Tenure	Missing	**0**	0	0.00	**0**	0.00
	Own	**68**	68	0.00	**74**	8.82
	Rent LA	**284**	283	-0.35	**277**	-2.46
	Rent Privately	**34**	35	2.94	**35**	2.94

[4] Percentage difference for employment rate is calculated as employment rate after swapping minus employment rate before swapping

Annex 3

Table 8: Cramers V (C) - association between variables before and after swapping for PU1 and PU16

		PU1						PU16					
Variables		**Before Swap**	1% Swap	3% Swap	5% Swap	10% Swap	20% Swap	**Before Swap**	1% Swap	3% Swap	5% Swap	10% Swap	20% Swap
Age	Sex	**0.136**	0.135	0.134	0.136	0.137	0.136	**0.073**	0.073	0.072	0.073	0.073	0.073
Age	Marital Status	**0.424**	0.424	0.425	0.425	0.424	0.425	**0.472**	0.473	0.472	0.472	0.472	0.473
Age	Activity	**0.590**	0.590	0.590	0.590	0.592	0.591	**0.638**	0.638	0.639	0.637	0.638	0.638
Age	Ethnic Group	**0.137**	0.139	0.132	0.134	0.134	0.130	**0.068**	0.068	0.065	0.070	0.068	0.058
Age	Country of Birth	**0.154**	0.153	0.154	0.153	0.146	0.149	**0.077**	0.076	0.077	0.077	0.076	0.077
Age	Tenure	**0.256**	0.256	0.255	0.259	0.257	0.258	**0.143**	0.143	0.143	0.142	0.141	0.147
Sex	Marital Status	**0.205**	0.202	0.206	0.203	0.203	0.205	**0.182**	0.181	0.181	0.182	0.182	0.180
Sex	Activity	**0.218**	0.215	0.220	0.214	0.218	0.217	**0.237**	0.237	0.238	0.237	0.239	0.236
Sex	Ethnic Group	**-0.012**	-0.013	-0.011	-0.011	-0.015	-0.018	**-0.009**	-0.009	-0.013	-0.008	-0.008	-0.003
Sex	Country of Birth	**-0.009**	-0.008	-0.008	-0.009	-0.012	-0.008	**0.013**	0.012	0.010	0.014	0.010	0.010
Sex	Tenure	**0.099**	0.099	0.097	0.097	0.096	0.102	**0.054**	0.053	0.054	0.055	0.055	0.055
Marital Status	Activity	**0.264**	0.264	0.265	0.264	0.265	0.265	**0.398**	0.398	0.399	0.398	0.398	0.397
Marital Status	Ethnic Group	**0.129**	0.130	0.127	0.125	0.132	0.128	**0.037**	0.037	0.039	0.037	0.037	0.034
Marital Status	Country of Birth	**0.190**	0.191	0.187	0.189	0.185	0.185	**0.048**	0.047	0.046	0.048	0.047	0.051
Marital Status	Tenure	**0.205**	0.204	0.204	0.206	0.203	0.202	**0.143**	0.144	0.143	0.143	0.146	0.142
Activity	Ethnic Group	**0.099**	0.100	0.094	0.099	0.096	0.088	**0.047**	0.047	0.047	0.047	0.048	0.043
Activity	Country of Birth	**0.102**	0.101	0.102	0.101	0.101	0.101	**0.059**	0.059	0.059	0.059	0.060	0.058
Activity	Tenure	**0.205**	0.206	0.204	0.202	0.208	0.205	**0.143**	0.143	0.142	0.145	0.144	0.146
Ethnic Group	Country of Birth	**0.513**	0.514	0.513	0.514	0.515	0.512	**0.315**	0.315	0.325	0.311	0.336	0.309
Ethnic Group	Tenure	**0.104**	0.103	0.103	0.096	0.096	0.091	**0.025**	0.025	0.027	0.025	0.020	0.024
Country of Birth	Tenure	**0.109**	0.107	0.109	0.103	0.103	0.103	**0.029**	0.029	0.031	0.030	0.024	0.025

Annex 4

Table 9: Cramers V (C) - association between variables before and after swapping for EDs

		ED1						ED2					
Variables		**Before Swap**	1% Swap	3% Swap	5% Swap	10% Swap	20% Swap	**Before Swap**	1% Swap	3% Swap	5% Swap	10% Swap	20% Swap
Age	Sex	**0.119**	0.115	0.118	0.112	0.111	0.117	**0.104**	0.108	0.106	0.109	0.121	0.104
Age	Marital Status	**0.487**	0.487	0.480	0.483	0.488	0.490	**0.509**	0.507	0.502	0.510	0.508	0.506
Age	Activity	**0.646**	0.646	0.646	0.646	0.644	0.646	**0.663**	0.663	0.664	0.667	0.665	0.668
Age	Ethnic Group	**0.109**	0.109	0.136	0.111	0.111	0.169	**0.226**	0.226	0.224	0.225	0.209	0.227
Age	Country of Birth	**0.179**	0.185	0.208	0.186	0.192	0.155	**0.264**	0.264	0.260	0.254	0.245	0.249
Age	Tenure	**0.148**	0.157	0.150	0.158	0.156	0.184	**0.457**	0.457	0.436	0.450	0.469	0.452
Sex	Marital Status	**0.210**	0.211	0.207	0.216	0.213	0.222	**0.217**	0.220	0.222	0.215	0.221	0.213
Sex	Activity	**0.238**	0.238	0.228	0.235	0.240	0.242	**0.248**	0.247	0.248	0.240	0.250	0.235
Sex	Ethnic Group	**-0.032**	-0.032	-0.005	-0.032	-0.032	-0.027	**-0.072**	-0.072	-0.072	-0.072	-0.078	-0.072
Sex	Country of Birth	**0.037**	0.023	0.043	0.054	0.032	0.029	**0.004**	0.004	0.004	0.021	0.006	0.044
Sex	Tenure	**0.030**	0.027	0.015	0.041	0.022	0.042	**0.119**	0.119	0.116	0.126	0.132	0.110
Marital Status	Activity	**0.400**	0.400	0.398	0.399	0.400	0.395	**0.419**	0.416	0.409	0.421	0.420	0.414
Marital Status	Ethnic Group	**0.040**	0.040	0.058	0.040	0.043	0.082	**0.124**	0.125	0.123	0.125	0.137	0.121
Marital Status	Country of Birth	**0.105**	0.087	0.083	0.110	0.123	0.216	**0.073**	0.079	0.077	0.082	0.087	0.071
Marital Status	Tenure	**0.190**	0.189	0.185	0.194	0.253	0.162	**0.346**	0.343	0.335	0.331	0.325	0.344
Activity	Ethnic Group	**0.072**	0.072	0.029	0.073	0.072	0.094	**0.054**	0.055	0.054	0.053	0.145	0.055
Activity	Country of Birth	**0.080**	0.079	0.091	0.079	0.079	0.096	**0.104**	0.106	0.095	0.092	0.167	0.085
Activity	Tenure	**0.114**	0.122	0.076	0.105	0.126	0.090	**0.316**	0.314	0.307	0.291	0.302	0.338
Ethnic Group	Country of Birth	**0.358**	0.378	0.558	0.378	0.325	0.469	**0.577**	0.577	0.577	0.577	0.604	0.577
Ethnic Group	Tenure	**0.085**	0.087	0.051	0.074	0.087	0.058	**0.100**	0.100	0.099	0.103	0.226	0.103
Country of Birth	Tenure	**0.071**	0.081	0.063	0.065	0.065	0.046	**0.112**	0.112	0.111	0.120	0.225	0.115

		ED3					
Variables		**Before Swap**	1% Swap	3% Swap	5% Swap	10% Swap	20% Swap
Age	Sex	**0.232**	0.232	0.229	0.245	0.212	0.231
Age	Marital Status	**0.442**	0.447	0.440	0.454	0.439	0.432
Age	Activity	**0.636**	0.635	0.640	0.635	0.641	0.629
Age	Ethnic Group	**0.254**	0.254	0.255	0.266	0.257	0.251
Age	Country of Birth	**0.133**	0.133	0.133	0.141	0.141	0.150
Age	Tenure	**0.253**	0.251	0.259	0.253	0.268	0.256
Sex	Marital Status	**0.300**	0.293	0.301	0.300	0.284	0.306
Sex	Activity	**0.367**	0.310	0.311	0.297	0.348	0.302
Sex	Ethnic Group	**0.047**	0.047	0.047	0.039	0.047	0.048
Sex	Country of Birth	**0.054**	0.054	0.054	0.036	0.061	0.036
Sex	Tenure	**0.161**	0.168	0.161	0.163	0.148	0.168
Marital Status	Activity	**0.341**	0.342	0.345	0.345	0.336	0.342
Marital Status	Ethnic Group	**0.145**	0.145	0.147	0.141	0.144	0.143
Marital Status	Country of Birth	**0.231**	0.230	0.229	0.228	0.214	0.206
Marital Status	Tenure	**0.208**	0.204	0.209	0.213	0.201	0.154
Activity	Ethnic Group	**0.237**	0.237	0.237	.0243	0.238	0.185
Activity	Country of Birth	**0.112**	0.110	0.113	0.100	0.109	0.119
Activity	Tenure	**0.265**	0.265	0.246	0.261	0.264	0.221
Ethnic Group	Country of Birth	**0.548**	0.548	0.548	0.526	0.538	0.515
Ethnic Group	Tenure	**0.097**	0.093	0.105	0.102	0.090	0.174
Country of Birth	Tenure	**0.109**	0.106	0.115	0.103	0.120	0.111

Annex 5

Table 10: Marital Status by Age and Sex using before swapping data for ED1

Age	Marital Status														
	Male							Female							Total
	Total	Missing	Single	Married	Re-married	Divorced	Widowed	Total	Missing	Single	Married	Re-married	Divorced	Widowed	
Missing	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1
0 to 9	20	0	20	0	0	0	0	21	0	21	0	0	0	0	41
10 to 19	26	0	26	0	0	0	0	22	1	21	0	0	0	0	48
20 to 29	22	0	16	5	0	1	0	29	0	19	10	0	0	0	51
30 to 39	28	0	3	21	1	3	0	32	0	2	23	2	3	0	60
40 to 49	39	0	5	28	3	3	0	29	0	1	22	2	4	0	68
50 to 59	23	0	2	18	3	0	0	29	1	0	21	2	1	4	52
60 to 69	23	0	1	21	0	1	0	28	0	3	17	1	2	5	51
70 to 79	18	1	1	11	3	1	1	27	0	2	11	1	2	11	45
80 to 89	11	1	2	5	0	0	3	13	1	1	3	0	0	8	24
90+	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1
Total	211	2	77	109	10	9	4	231	3	70	109	8	12	29	442

Table 11: Marital Status by Age and Sex using 1% swap data for ED1

Age	Marital Status														
	Male							Female							Total
	Total	Missing	Single	Married	Re-married	Divorced	Widowed	Total	Missing	Single	Married	Re-married	Divorced	Widowed	
Missing	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1
0 to 9	20	0	20	0	0	0	0	21	0	21	0	0	0	0	41
10 to 19	26	0	26	0	0	0	0	22	1	21	0	0	0	0	48
20 to 29	22	0	16	5	0	1	0	29	0	19	10	0	0	0	51
30 to 39	28	0	3	21	1	3	0	32	0	2	24	2	4	0	60
40 to 49	39	0	5	28	3	3	0	29	0	1	22	2	4	0	68
50 to 59	23	0	2	17	4	0	0	29	1	0	21	2	1	4	52
60 to 69	23	0	1	21	0	1	0	28	0	3	17	1	2	5	51
70 to 79	18	1	1	11	3	1	1	27	0	2	11	1	2	11	45
80 to 89	11	1	2	5	0	0	3	13	1	1	3	0	0	8	24
90+	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1
Total	211	2	77	108	11	9	4	231	3	70	108	8	13	29	442

Table 12: Marital Status by Age and Sex using 20% swap data for ED1

Age	Marital Status														
	Male							Female							Total
	Total	Missing	Single	Married	Re-married	Divorced	Widowed	Total	Missing	Single	Married	Re-married	Divorced	Widowed	
Missing	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1
0 to 9	20	0	20	0	0	0	0	24	0	24	0	0	0	0	44
10 to 19	25	0	25	0	0	0	0	19	1	18	0	0	0	0	44
20 to 29	23	0	18	4	0	1	0	29	0	17	12	0	0	0	52
30 to 39	35	0	4	25	2	4	0	32	0	2	25	3	2	0	67
40 to 49	31	0	2	22	5	2	0	26	0	1	18	3	3	1	57
50 to 59	25	0	2	19	4	0	0	31	1	0	22	2	1	5	56
60 to 69	24	0	1	21	1	1	0	31	0	3	19	1	2	6	55
70 to 79	17	1	0	11	3	1	1	26	0	1	12	1	2	10	43
80 to 89	10	1	2	4	0	0	3	12	1	1	2	0	0	8	22
90+	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1
Total	211	2	75	106	15	9	4	231	3	67	110	10	10	31	442