WP. 19
ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

Joint UNECE/Eurostat work session on statistical data confidentiality
(Geneva, Switzerland, 9-11 November 2005)

Topic (iii) Confidentiality aspects of statistical information taking into account register-based data

ASSESSMENT OF STATISTICAL DISCLOSURE CONTROL METHODS FOR THE 2001 UK CENSUS

Invited Paper

Submitted by the Office for National Statistics, University of Southampton, United Kingdom, Hebrew University, Israel[1]

[1] Prepared by Natalie Shlomo.

Assessment of Statistical Disclosure Control Methods for the 2001 UK Census

Natalie Shlomo*

* Southampton Statistical Sciences Research Institute, University of Southampton, Department of Statistics, Hebrew University, Office for National Statistics

Abstract: We define the disclosure risk scenarios that led to the statistical disclosure control (SDC) methods for the 2001 UK Census. We examine the SDC methods that were implemented based on a disclosure risk-data utility framework and assess whether the methods managed the disclosure risk while maintaining the utility and quality of the outputs. We conclude with final remarks and goals for forming strategies for future Censuses.

1. Introduction

Beginning with the 2001 UK Census, the Office for National Statistics (ONS) re-examined its statistical disclosure control (SDC) policies and methods for protecting standard Census tabular outputs. The initial SDC method that was planned for the 2001 Census was random record swapping on the microdata prior to tabulating the data (defined in Section 2) and higher population thresholds for released tables. This method was shown to give about the same level of protection as the method that was used for the 1991 UK Census based on a post-tabular variation of record swapping that was applied to the tables. However, prior to releasing the 2001 Census tables, it was decided that an additional disclosure control method of small cell rounding would also be applied to the tabular outputs. This was due to the following reasons:

- 100% of the questionnaire was coded compared to only 10% in the 1991 Census;
- Increasing IT technologies and the wealth of available public data, including the Neighborhood Statistics Service (NeSS) website which provides detailed small area social and economic statistics from both administrative and census sources, raised the level of disclosure risk compared to 1991;
- Pre-tabular record swapping leaves the perception that no SDC method is applied at all to the tables, thus raising concerns about the impact on future response rates for ONS Censuses and surveys.

Scotland, however, did not include the small cell rounding in their SDC strategies and this led to differential SDC methods across the Statistical Offices of the UK.

In this paper, we examine the SDC methods that were applied to the 2001 UK Census tabular outputs based on a disclosure risk – data utility framework (Duncan, et. al. (2001)). The purpose is to assess whether the methods managed the disclosure risk against the risk of re-identification while maintaining the utility and quality of the Census outputs. Section 2 describes the SDC methods and Section 3 the data used for the analysis. Section 4 presents the disclosure risk and data utility quantitative measures with results as well as an R-U confidentiality map. Section 5 concludes with a discussion of the analysis and goals for forming strategies for future Censuses.

The views expressed are those of the author and do not necessarily reflect the views of the University of Southampton and the ONS.

2. SDC Methods Used in the 2001 UK Census

The SDC method implemented on the 2001 Census tables was a combination of a pre-tabular method of random record swapping and a post-tabular method of small cell rounding.

2.1 Random Record Swapping

The most common pre-tabular method of SDC for Census outputs is record swapping. As defined in Willenborg and de Waal (2001), each record i is partitioned into three sub-vectors: x_i, y_i and z_i. Controlling for x_i, a household is selected for swapping having the same sub-vector x_j. In this case, the distributions of the pairs of values (x_i, y_i) and (x_i, z_i) are preserved after swapping. If X is chosen so that Y and Z are conditionally independent given X then swapping will not affect the joint distribution of X, Y and Z. For example, let Z define geographical variables, X the household characteristics (household size, age-sex composition of the household, ethnic background, etc.), and Y all other variables. The above method will swap households across geographical areas Z while ensuring that swapped households have the same characteristics on X. This protects against disclosure risk by perturbing the relationship between y_i and z_i in the record. Note that this method distorts the joint distribution of Y and Z though marginal distributions are maintained at a higher geographical level. In addition, because of the conditional independence, we obtain less inconsistencies and edit failures as a result of swapping records. This method also gives slight protection for the disclosure risk resulting in differencing two tables which are nested and non-coterminous because of the uncertainty introduced in the data.

For the 2001 UK Census, the random record swapping of households was carried out within a large geographical area defined by the local authority (LA). A random sample within strata defined by control variables was selected using a fixed swapping rate *f*. The control variables that were used were: hard-to-count index, household size, sex and broad age distribution of the household (0-25, 25-44, 45 and over). For each household selected, a paired household is found and all geographical variables are swapped. Note that this has the same effect as swapping all other variables and leaving geography fixed.

For this analysis we carried out random record swapping as implemented for the 2001 UK Census at the following swapping rates: 1%, 10%, and 20%. In addition, we carried out some modifications of the random record swapping in order to compare the disclosure risk – data utility across the different methods:

- As carried out in the 2001 UK Census, records were swapped on imputed records as well as non-imputed records. Imputed records arise from two sources: records that have missing items and whole records that were imputed for correcting the coverage of the Census based on the Census Coverage Survey. Since imputed records are a priori protected records, there is no need to perturb them and therefore we carried out the random record swapping only on the non-imputed records.

- Based on the tables used in the analysis (see Section 3) we identified and flagged all the small cells of the tables. We implemented a targeted record swapping by pairing and swapping households that matched not only on the control variables but also on the flagged variable. If, however, a household that was selected for swapping did not have a match on the control variables from among the flagged households, a match was found outside the flagged households.

Note that on average, about 0.15% of the households selected for record swapping were not swapped because no matching household was found for them. In general, those records would have to be swapped outside the large geographical area (LA) but this was not carried out in this analysis. Table 1 presents advantages and disadvantages of the record swapping.

Advantages	Disadvantages
Consistent totals for all tables	Leaves a high proportion of risky (unique) records unperturbed
Preserves marginal distributions at higher aggregated levels	Errors (bias) in data and in particular joint distributions distorted
Some protection against disclosure by differencing two non-coterminous tables	Effects of perturbation hidden and can't be measures or accounted for in statistical analysis, i.e. a number in a table is not the true value
Less inconsistencies and edit failures when swapping geographies	Method is not transparent to users and appears as if no SDC method is used
Targeted swapping lowers disclosure risk	Targeted swapping causes more distortion in the distribution of the table

Table 1 : Advantages and Disadvantages of Record Swapping as a Pre-Tabular SDC method for Census Tabular Outputs

2.2 Small Cell Rounding

In comparison to pre-tabular record swapping where effects are hidden, the post-tabular rounding procedures are transparent to users and the stochastic forms of rounding can be taken into account when carrying out statistical analysis. For the 2001 Census tables (not including Scotland) small cells were rounded. The method used was an unbiased random rounding. Let x be a small cell and let $Floor(x)$ be the largest multiple k of the base b such that $bk < x$ for an entry x. In addition, define $res(x) = x - Floor(x)$. For an unbiased rounding procedure, x is rounded up to $(Floor(x) + b)$ with probability $\frac{res(x)}{b}$ and rounded down to $Floor(x)$ with probability $(1 - \frac{res(x)}{b})$. If x is already a multiple of b, it remains unchanged. The expected value of the rounded entry is the original entry. Each small cell is rounded independently in the table, i.e. a random uniform number u between 0 and 1 is generated for each cell. If $u < \frac{res(x)}{b}$ then the entry is rounded up, otherwise it is rounded down. As mentioned, the expectation of the rounding is zero and no bias should remain in the table. However, the realisation of this stochastic process on a finite number of cells in a table may lead to overall bias since the sum of the

perturbations (i.e., the difference between the original and rounded cell) going down may not equal the sum of the perturbations going up.

When only small cells are rounded, the margins of the tables are obtained by aggregating the rounded and non-rounded cells, and therefore tables with the same population base will have different totals. The confidence interval for the expected differences from the true totals as a result of the small cell rounding procedures depends on the number of small cells that are adjusted in the table. Figure 1 presents the confidence intervals for the expected differences from true totals when rounding small cells to base 3.

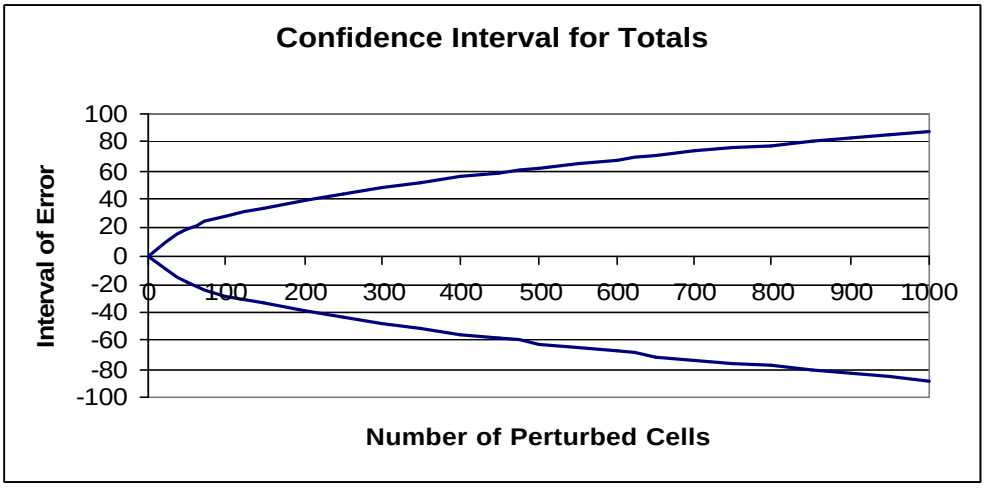

Figure 1: Confidence Intervals for Random Rounding to Base 3

In addition, we also carried out modifications of the random rounding procedure for this analysis in order to compare the disclosure risk – data utility across the different methods:

- Since different totals are obtained for tables with the same population base, we carried out a semi-controlled small cell rounding where the overall total of the table is preserved. This method can also preserve some of the marginal totals in the tables as well (Shlomo and Young (2005)).
- A full (semi-controlled) random rounding was also carried out. This is implemented as described above for the small cells after first converting the entries x to residuals of the rounding base $res(x)$.

Table 2 presents advantages and disadvantages of the rounding procedures.

Advantages	Disadvantages
Full protection for the high-risk (unique) cells	Inconsistent totals between tables when margins aggregated from rounded and non-rounded cells
Full rounding protects against disclosure by differencing two non-coterminous tables.	Small cell rounding gives little protection against disclosure by differencing so only one set of geographies and other variables disseminated
Small cell rounding has less information loss	Full rounding has margins rounded separately and tables aren't additive
Methods clear and transparent to users	Stochastic methods of rounding are easier to unpick and tables may need to be audited prior to release
Stochastic methods can be accounted for in statistical analysis	

Table 2 : Advantages and Disadvantages of Rounding as a Post-Tabular SDC method for Census Tabular Outputs

Note that fully controlled rounding which preserves the marginal totals of the tables as developed within the Tau-Argus framework (Salazar, et al (2004)) is not at the moment a viable option for the size and scope of Census tables and therefore will not be examined further in this analysis.

3. Data Used

To carry out the disclosure risk – data utility analysis on 2001 Census data, we obtained unperturbed microdata from different Estimation Areas of the UK. In this report, we will show results for one Estimation Area: SJ - (Southampton, Eastleigh, Test Valley) 437,744 persons, 182,337 households, 1,487 Output Areas (OA). For this Estimation Area (EA), we defined five standard census tables (the number of categories of the variable are in parenthesis):

(1) Religion(9) * Age-Sex(6) * OA
(2) Travel to Work(12) * Age-Sex(12) * OA
(3) Country of Birth (17) * Sex (2) * OA
(4) Economic Activity (9) * Sex (2) * Long-Term Illness (2) * OA
(5) Health status (5) * Age-Sex (14) * OA

The microdata was perturbed according to the record swapping scenarios (random, random without imputed records and targeted) and then tabulated and rounded according to the rounding procedures: small cell rounding (SCA), semi-controlled small cell rounding (CSCA) and semi-controlled full random rounding (CRND).

4. Disclosure Risk – Data Utility Analysis

In this Section we assess the methods used in the 2001 UK Census based on a disclosure risk-data utility framework. This will examine whether an optimal balance was found between managing the disclosure risk and maximizing the utility of the data for the standard 2001 Census tables.

4.1 Disclosure Risk

The disclosure risk in population-based Census tables arises from small cells or small cells obtained from differencing tables. The record swapping will not inhibit small cells from appearing in the tables and therefore we need a quantitative disclosure risk measure which reflects whether the ones or twos in the table are true values or perturbed values.

The quantitative measure of disclosure risk for assessing record swapping is the proportion of records in the small cells that have not been perturbed. The perturbation comes from two sources: the record swapping procedure and imputation. Imputed records can be viewed as protected records and therefore there is no need to apply SDC methods on those records nor include them in the quantitative risk measures.

Let R_i represent the record i, I the indicator function having a value 1 if true and 0 if false, C_1 the set of cells with a value of 1, C_2 the set of cells with a value of 2, $|C_1 \cup C_2|$ the number of small cells with a value of 1 or 2. The disclosure risk measure is: $DR = \dfrac{\sum_{i \in C_1 \cup C_2} I(R_i \quad perturbed \quad or \quad imputed)}{|C_1 \cup C_2|}$

Table 3 presents results of the disclosure risk remaining in the Census tables as defined in Section 3 after implementing the different scenarios of record swapping.

Method	EA SJ		
	1%	10%	20%
Original	84.2%	84.2%	84.2%
Random	82.3%	66.3%	49.7%
Rand/Imp	82.0%	63.4%	43.6%
Targeted	80.6%	45.9%	18.0%

Table 3: Percent Records in Small Cells of the Tables that were Swapped or Imputed

In Table 3, we see that without any disclosure control method, there is a priori protection against disclosure risk because of the imputation. For EA SJ, there were about 16% imputed records. We see almost no impact on the disclosure risk from the 1% record swapping and it is about the same as if no SDC method was applied at all. Even for the targeted record swapping at the 1% swapping rate, we obtain about an 80% chance that a small cell in the table (one or a two) is the true value. This leaves a high probability of identifying uniques in the Census tables. For the higher swapping rates (10% and 20%), we are able to bring the disclosure risk down to lower levels of disclosure risk, especially if the records to be swapped are targeted from among the records in small cells of the tables. Note that if the random record swapping as carried out for the 2001 UK Census had not included the imputed records, the disclosure risk could have been lowered as will be shown in the R-U confidentiality map in Figure 4 at no cost to the utility of the data.

Forms of rounding eliminates all small cells in the table and therefore disclosure risk is zero with respect to the re-identification of small cells. For attribute and group disclosure, zeros in the table may not be true zeros since small cells can be rounded down to zero in the rounding procedure. The disclosure risk remains however when applying the method of small cell rounding and tables can be differenced. This is because only small cells are affected by the rounding procedure and large cells are left intact. Therefore, large counts in tables that are differenced can lead to disclosive small cells. For the 2001 UK Census, disclosure by differencing was managed and minimized by allowing only one set of geographies and other variables to be disseminated. Therefore, we won't assess the disclosure by differencing problem in this analysis and assume that for the rounding methods, there is no disclosure risk and we only need to examine one dimension of the decision problem and that is data utility.

4.2 Data Utility - Measuring distortions to distributions

Utility measures that measure distortions to distributions are based on distance metrics between the original and perturbed cells. Some useful metrics were presented in Gomatam and Karr (2003). Since the basic unit of most of the Census tables is a geography, i.e. Output Areas (OA), we are interested in a measure of distortion at this level of geography. Therefore, we will calculate the distance metric for each OA separately in the tables and then take the overall average across the OA's as the utility measure for the whole table. Note that we can also look at the table as a whole and measure distortions to distributions across all the cells.

Let D^k represent a table for OA k and let $D^k(c)$ be the cell frequency c in the table. Let $|OA|$ be the number of OA's in the estimation area. The distance metrics are:

- Hellinger's Distance:

$$HD(D_{orig}, D_{pert}) = \frac{1}{|OA|}\sum_{k=1}^{|OA|}\sqrt{\sum_{c\in k}\frac{1}{2}(\sqrt{D^k_{pert}(c)}-\sqrt{D^k_{orig}(c)})^2}$$

- Relative Absolute Distance:

$$RAD(D_{orig}, D_{pert}) = \frac{1}{|OA|}\sum_{k=1}^{|OA|}\sum_{c\in k}\frac{|D^k_{pert}(c)-D^k_{orig}(c)|}{D^k_{orig}(c)}$$

- Average Absolute Distance per Cell:

$$AAD(D_{orig}, D_{pert}) = \frac{1}{|OA|}\sum_{k=1}^{|OA|}\frac{\sum_{c\in k}|D^k_{pert}(c)-D^k_{orig}(c)|}{|k|}$$ where

$|k| = \sum_c I(c\in k)$ the number of non-zero cells in the k^{th} OA

Table 4 presents results of the three distance metrics for the record swapping scenarios for EA SJ and Tables 5 and 5a the results for the rounding procedures.

Method	EA SJ		
	1%	10%	20%
Random			
HD	1.035	3.721	5.279
RAD	4.302	32.437	53.001
AAD	0.138	0.726	1.053
Rand/Imp			
HD	1.044	3.714	5.238
RAD	4.337	32.345	52.433
AAD	0.136	0.722	1.036
Targeted			
HD	1.376	4.787	6.372
RAD	6.215	43.375	63.135
AAD	0.160	0.845	1.173

Table 4: Average Distance Metrics Between Original and Perturbed Cells per OA for Record Swapping

0 HMKRG	SCA	CSCA	CRND
HD	5.272	5.279	5.416
RAD	76.804	76.824	84.641
AAD	0.629	0.630	1.021

Table 5: Average Distance Metrics Between Original and Perturbed Cells per OA for Rounding

Method	1%			10%			10%		
	SCA	CSCA	CRND	SCA	CSCA	CRND	SCA	CSCA	CRND
Random									
HD	5.383	5.390	5.524	6.313	6.299	6.421	7.228	7.226	7.311
RAD	78.630	78.687	85.546	89.446	89.478	92.848	97.599	97.570	100.003
AAD	0.745	0.746	1.074	1.119	1.119	1.251	1.337	1.335	1.418
Random/Imputed									
HD	5.390	5.393	5.524	6.305	6.302	6.406	7.173	7.183	7.297
RAD	78.636	78.636	85.474	89.162	89.152	92.740	96.836	96.986	99.381
AAD	0.745	0.745	1.073	1.114	1.114	1.245	1.315	1.318	1.403
Targeted									
HD	5.444	5.442	5.575	6.791	6.764	6.872	7.800	7.818	7.899
RAD	78.709	78.721	85.530	89.157	89.048	92.339	96.271	96.326	98.651
AAD	0.753	0.752	1.080	1.165	1.161	1.292	1.383	1.386	1.469

Table 5a: Average Distance Metrics Between Original and Perturbed Cells per OA for Rounding Combined with Record Swapping

In Table 4, we see a consistent pattern of small distance metrics for the 1% swapping rate and larger distance metrics for the 20% swapping rate. The measure of *AAD* tells us by how much the cells are perturbed on average for each OA. For example, for the random record swapping, each cell is perturbed by about 0.7 for the 10% swap and about 1.1 for the 20% swap. Similarly, for the targeted record swapping, each cell is perturbed by about 0.8 for the 10% swap and about 1.2 for the 20% swap. Between the random record swapping and the random record swapping without imputed

records, we see almost no difference in the distance metrics. The targeted record swapping has the highest distance metrics showing that more distortion occurs as the swapping rate increases.

Tables 5 and 5a show the same distance metrics for the rounding procedures. In the table, we see much higher levels of information loss based on the distance metrics for the rounding procedures and even higher distance metrics when combining rounding procedures with the record swapping. One can argue that if we assume zero disclosure risk because no small cells are left in the table and there is no risk of disclosure by differencing, then we lower the utility of the tabular outputs by combining the record swapping with the rounding procedures. This loss of utility is minimal for the 1% swap, but increases for the higher swapping rates. Note in Tables 5 and 5a that the HD metric does not pick up differences between the full random rounding and small cell rounding (controlled or not), whereas the other distance metrics are more sensitive to the perturbation of the internal cells. This is because HD is heavily influenced by small cells. As seen in the table, the difference between the independent small cell rounding (SCA) and the controlled small cell rounding (CSCA) has about the same distance metrics for all the measures and therefore this utility measure is not sensitive to the totals which will be examined next.

One problem for the rounding procedures was that different totals were obtained in tables with the same population base. This was particular problematic for users of Census tables who are mainly concerned about obtaining high quality aggregated level data for specified and non-standard geographical areas, for example school districts. The OA level tables are typically used as building blocks to construct higher level geographies. Because the tables are highly perturbed at the OA level of geography, aggregating OA data results in much information loss to the totals. In order to evaluate the range of the differences for sub-totals on specific Census target variables, we use the statistical graphing tool of a box plot on the differences between the perturbed sub-total and the original sub-total: $AD(N^k_{orig}, N^k_{pert}) = N^k_{pert}(C') - N^k_{orig}(C')$ where $N^k(C') = \sum_{c \in C'} D^k(c)$ is a subtotal in the k^{th} OA. Each box in the plot contains the inter-quartile range (between the 25th and 75th percentile) of the *AD*'s. The lower 25th percentile and the upper 25th percentile are represented by the whiskers of the box. The line in the middle of the box is the median of the *AD*'s and the dot represents the mean. The length of the box and the length of the whiskers gives an indication of how wide spread the perturbed totals are from the original totals. Figure 2 presents the box plot of the *AD's* in EA SJ based on the number of Males born in Western Europe within ten consecutive groupings of OA's for the different scenarios of record swapping.

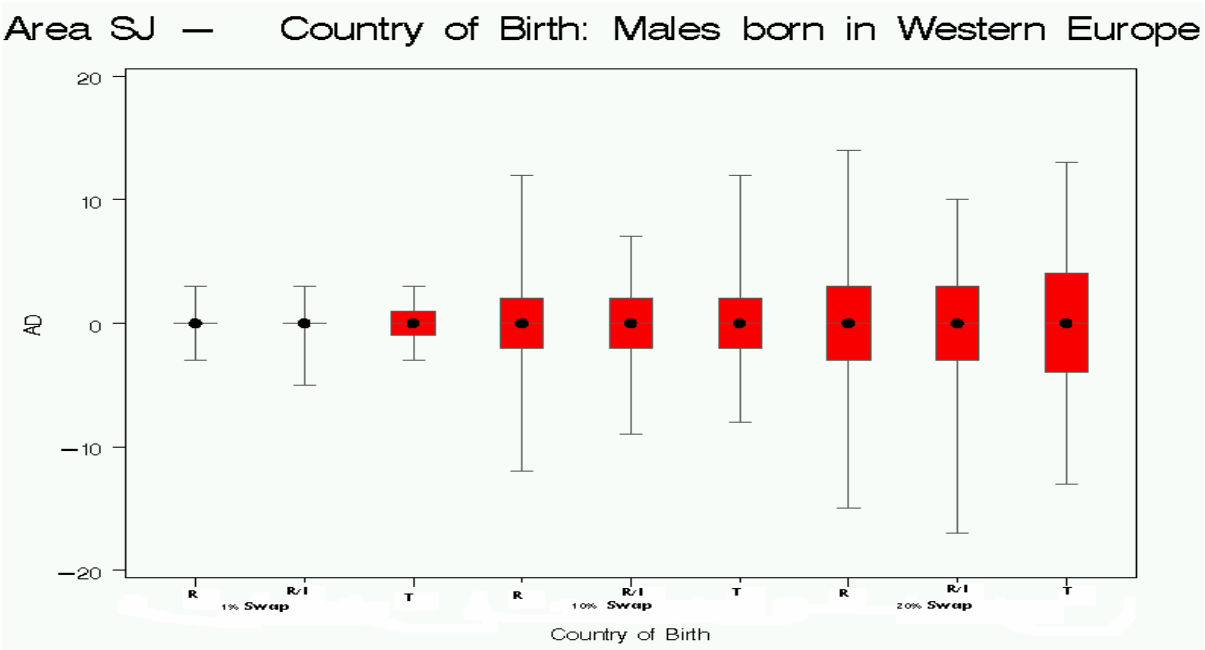

Figure 2: Box Plot of *AD*'s for the Number of Males Born in Western Europe in Ten Consecutive OA's of EA SJ for the Record Swapping Average Original Total in 10 OA's=24.6

From Figure 2 we see almost no loss of utility for the 1% swapping rate. The 10% and 20% swapping rates had more loss of utility with wide spread whiskers. For this particular example, there is a slight loss of utility between the random and targeted record swapping for each swapping rate. As shown, for this particular sub-total, the error in the total for ten consecutive OA's could be as much as ± 15, which is about 61% of the average original value. This lowers the utility of the Census data, especially since users are not able to take the perturbation error into account in their analysis. Figure 3 presents the box plot for the post-tabular methods of rounding and combined with the 10% random without imputed (R/I) and targeted (T) record swapping.

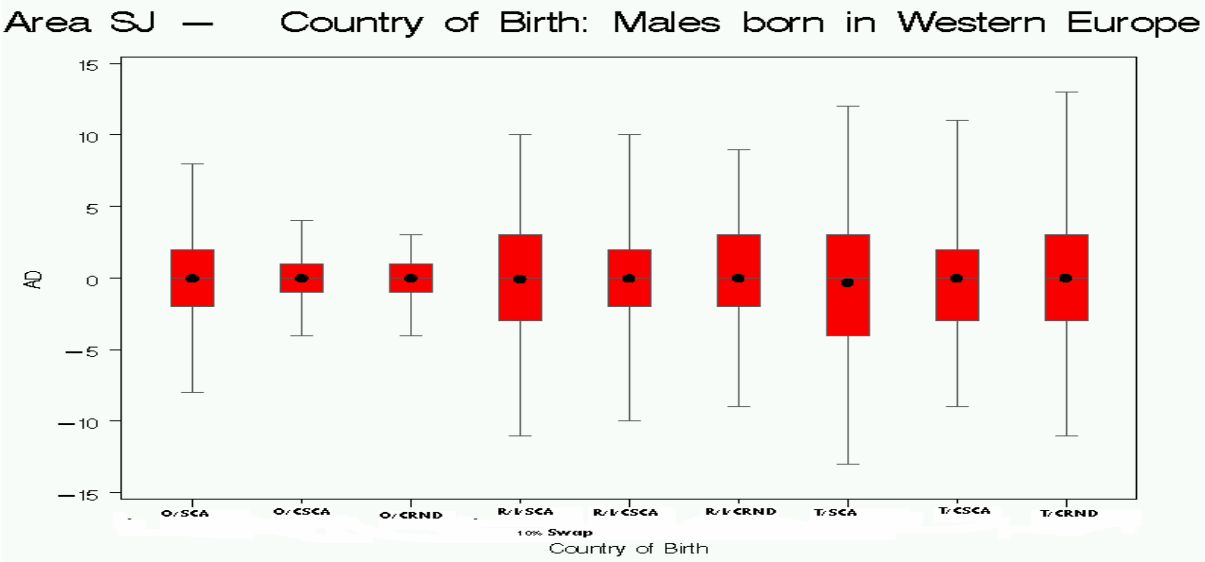

Figure 3: Box Plot of *AD*'s for the Number of Males Born in Western Europe in Ten Consecutive OA's of EA SJ for the Rounding and 10% Record Swapping Average Original Total in 10 OA's = 24.6

In Figure 3 we see that the boxes are narrower for the rounding methods on the original data compared to the rounding method when combined with record swapping. The effect of the random and the targeted record swapping on the *AD*'s is about the same. What is interesting to note is that the length of the boxes and whiskers are narrower for the semi-controlled small cell rounding compared to the independent small cell rounding. This means that there is more utility in the tables since the perturbed sub-totals do not differ from the original totals as much as the independent small cell rounding. Even the semi-controlled full rounding of all entries in the table shows slightly higher utility than the independent small cell rounding.

4.3 R-U Confidentiality Map

In Figures 4 we present an empirical R-U confidentiality map for the record swapping methods based on the risk measure *DR* and the distance metric AAD.

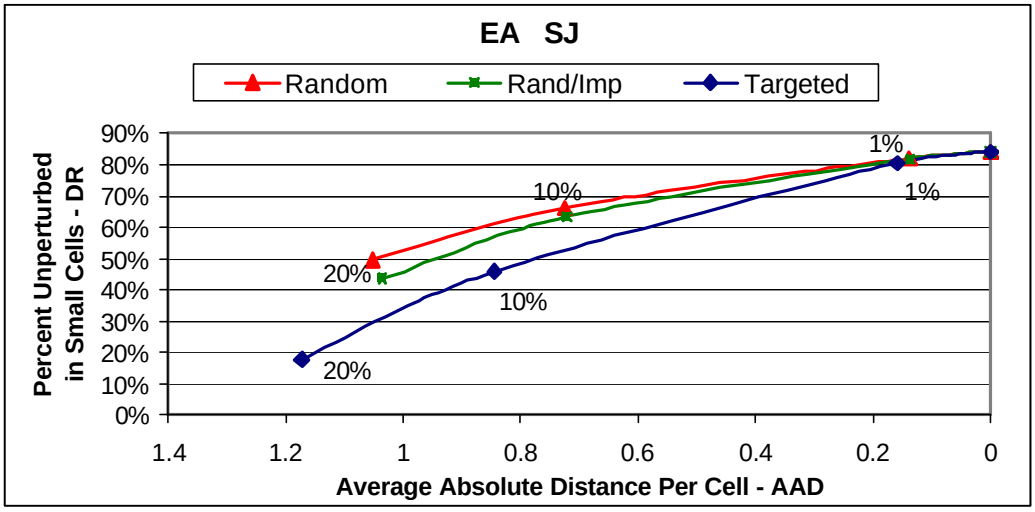

Figure 4: R-U Confidentiality Map for Estimation Area SJ

As seen from Table 3, the 1% swapping rates for all methods of record swapping have high utility but also very high disclosure risk (about 80% of the small cells in the table (ones and twos) are true values). The 10% targeted record swapping has about the same disclosure risk as the 20% random record swapping (about 45% of the small cells are true values). However, we gain much more utility in the data with the 10% targeted record swapping compared to the 20% random record swapping. This graph clearly shows that the 10% targeted record swapping is preferable for a given disclosure risk. Another conclusion from the R-U Confidentiality Map is that had the random record swapping as implemented in the 2001 Census not included the imputed records, we would have obtained higher utility in the data for the same disclosure risk.

5. Discussion

Based on the risk-utility analysis, we see that the SDC methods of record swapping and rounding used for the 2001 UK Census managed the disclosure risk. As a stand alone method, the random record swapping gives little protection against disclosure risk but could have been improved had a targeted record swapping taken place. When

combined with the small cell rounding, we obtain full protection of the Census tabular outputs taking into account that there is no risk from differencing tables because of the standard geographies and other variables that were disseminated. The loss of utility mostly resulted from the bias that occurred because of the record swapping and the fact that totals were different across tables with the same population base. In particular, the very large and sparse 2001 origin-destination tables were badly affected by the SDC methods. Utility could have been improved by placing more controls into the rounding algorithm and preserving overall totals of the tables and benchmarking.

Based on these results, it is clear that when planning for future censuses there needs to be consistent SDC methods across all of the UK Statistical Offices that disseminate Census data. The methods need to ensure that sufficient statistics (totals, averages and variances) are not compromised. Flexible table generating software should be developed so that users can design and customize their own Census tables. The SDC method would then be applied only once on the final outputted table as opposed to the standard system today where hard-copy tables are disseminated on paper or on CD's and non-standard geographies are aggregated from perturbed lower level geographies. Improved GIS systems may allow more flexible dissemination of geographies in the future and further development of the SDC tool Tau-Argus may automate better the SDC processes. Finally, SDC methods should be tailored and coordinated between the types of Census outputs, such as standard tables, microdata, origin-destination tables, in order to maximize utility while managing the disclosure risk.

6. Acknowledgements

A special thank you to Alexa Courtney for her tremendous help in preparing the data for analysis and to Joan Holland for her help in obtaining and interpreting the Census data.

References

Duncan, G., Keller-McNulty, S., and Stokes, S. (2001) "Disclosure Risk vs. Data Utility: the R-U Confidentiality Map", Technical Report LA-UR-01-6428, Statistical Sciences Group,Los Alamos, N.M.:Los Alamos National Laboratory.

Gomatam, S. and Karr, A. (2003) "Distortion Measures for Categorical Data Swapping", Technical Report Number 131, National Institute of Statistical Sciences.

Salazar, J.J., Lowthian, P., Young, C., Merola, G., Bond, S. and Brown, D. (2004) "Getting the Best Results in Controlled rounding with the Least Effort", (J. Domingo-Ferrer and V. Torra, eds.), *Privacy in Statistical Database*, Springer: New York.

Willenborg, L. and de Waal, T. (2001), *Elements of Statistical Disclosure Control*, Lecture Notes in Statistics, 155 (Springer Verlag, New York).