# Assessing the Effectiveness of Disclosure Limitation Methods for Census Microdata in Japan

Shinsuke Ito* and Naomi Hoshino**

* Faculty of Economics, Meikai University, 1 Akemi Urayasu, Chiba, 279-8550 Japan,
E-mail: ssitoh@meikai.ac.jp
Shinsuke Ito is a research fellow at the National Statistics Center and conducts research on disclosure limitation methods for microdata in co-ordination with officials at the National Statistics Center.
** National Statistics Center, 19-1 Wakamatsu-cho, Shinjuku-ku, Tokyo, 162-8668 Japan,
E-mail: nsaitou2@nstac.go.jp

**Abstract:** Following the revision of the Statistics Act, Anonymized microdata from official statistics have been released in Japan since April 2009. Currently, six types of Anonymized microdata from Japanese official statistics such as the 'Survey on Time Use and Leisure Activities' and the 'Employment Status Survey' are available. In addition, there are plans to release Anonymized microdata from the 'Population Census'. It is hoped that these developments will promote the secondary use of official statistics in Japan.

Several empirical studies on the effectiveness of disclosure limitation methods such as microaggregation, additive noise, and data swapping for official microdata have been conducted by the National Statistics Center, Japan. However, empirical studies to assess data utility and disclosure risk of anonymized official microdata are required for further official microdata releases.

Based on individual data from the Population Census, this paper explores the effectiveness of disclosure limitation methods including non-perturbative methods and perturbative methods, and their potential for the creation of anonymized official microdata in Japan.

## 1 Introduction

In Europe and North America, several types of census microdata are released. In the United States, Public Use Microdata Sample (PUMS) from the Population Census have been released since the 1960 Census. In the United Kingdom, several types of Anonymized microdata such as Sample of Anonymised Records (SARs) and Small Area Microdata (SAM) are released. In Japan, following the revision of the Statistics Act, Anonymized microdata from official statistics have been released since April 2009[1]. Currently, six types of Anonymized microdata from official statistics such as the 'Survey on Time Use and Leisure Activities' and the 'Employment Status Survey' are available, and there are plans to release Anonymized microdata from the 'Population Census' in the future.

[1] 'Anonymized microdata' are defined as individual data 'that is processed so that no particular individuals or juridical persons, or other organizations shall be identified' (Article 36).

The National Statistics Center of Japan has conducted several empirical studies on the effectiveness of disclosure limitation methods such as microaggregation, additive noise, and data swapping for official microdata (Ito and Murata (2011), Ito and Hoshino (2012) etc.). However, empirical studies to assess data utility and disclosure risk for anonymized official microdata[2] are required to support decisions towards further releases of official microdata (e.g. anonymized official microdata for small area analysis).

This paper explores the effectiveness of non-pertubative and perturbative disclosure limitation methods with regard to data utility and disclosure risk, and assesses their potential for the creation of anonymized official microdata.

## 2 Methodology

Disclosure limitation methods are classified into non-perturbative methods and perturbative methods (Willenborg and De Waal (2001)). Non-perturbative methods include recoding, suppression, top-coding and bottom-coding. Perturbative methods include noise addition, data swapping, rounding, microaggregation and PRAM (Post Randomisation Method) (Domingo-Ferrer and Torra (2001), Willenborg and De Waal (2001) etc.). In this research, anonymized official microdata were created using both perturbative and non-perturbative methods. Data utility and disclosure risk for this data were then determined in order to assess and compare the effectiveness of these disclosure limitation methods for the creation of anonymized official microdata.

Data from the 2005 Population Census was used for this research. This data consists of approximately 100,000 records of individual data from a specific geographic area (referred to as "Area A"). The anonymized official microdata used in this research were created from this data through the following steps:

First, the categories "type of activity" and "employments status" were recoded, and the one-year age brackets for "age" were recoded to five-year age brackets. The "number of household members" was top-coded to 'eight and more' and "age (five-year age brackets)" was top-coded to '85 or more'. Second, sampling at the rate of 10% was performed. Third, data swapping was performed.

The detailed process for data swapping was as follows: First, the number of sample uniques was calculated based on the following key variables:

Relationship to the Household Head (13 categories)

Gender (2 categories)

Age (Five-year age brackets) (19 categories) (recoded and top-coded)

[2] Individual data for which disclosure limitation methods have been applied as part of this research are referred to as 'anonymized official microdata' in this research.

Marital Status (5 categories)

Nationality (13 categories)

Type of (Work) Activity (6 categories) (recoded)

Employment Status (4 categories) (recoded)

Industry (19 categories)

Occupation (10 categories)

Type and Tenure of Dwelling (9 categories)

Type of Building and Number of Stories (5 categories) (30 sub-categories in the case of 'apartment house or flat')

Second, records that correspond to unique cells for the various combinations of the 11 key variables were selected as target records for data swapping. In order to determine the degree of priority for data swapping, cross-tabulation was conducted for all combinations of key variables. The number of times a specific record corresponds to a unique cell for every combination of cross-tabulations was calculated, and this score was added to every record in the test data. Records for which the score was high were regarded as 'risky' records with a higher priority for data swapping (Elliot *et al*. (2002)).

Third, targeted data swapping and random data swapping were performed for records with a score of 1 or higher. Targeted data swapping was performed for records that correspond to the top p% (p=1, 2, 3, 5, 10, 20, 30) of the group, and was performed in order of descending score. Random data swapping was performed based on a swapping rate p for target records randomly selected from records with a score of 1 or higher. Partner records for a different area to "Area A" that is referred to as "Area B" (approximately 5,000 records) were selected from the donor file[3].

Fourth, the distance between each target record and all donor file records was calculated, and the nearest record in the donor file was swapped. In case of multiple records with identical distances, the partner record was randomly selected from among these records.

Based on the distance calculated, record linkage between target records and donor file records was conducted. Details of the record linkage technique (Domingo-Ferrer and Torra (2001)) used in this research are as follows:

The degree to which target records and donor file records match was determined. For nominal variables except age, the score was calculated as follows: (1) The score is 1 if the values of the key variables in the target records match the values in the donor file records, and otherwise it is 0. (2) This score was then divided by the number of categories for the key variable. For ordinal variables, the score was calculated as

[3] This process is basically identical to Ito and Hoshino (2012).

follows: (1) The values of the target records were subtracted from the values of the donor file records. (2) The absolute values of these results were divided by the number of categories. This score was calculated for each of the above 11 key variables. The results were then added to calculate the distance between target records and donor file records.

## 3 Matching Anonymized Census Microdata with External Data

A key concept when it comes to data confidentiality for official microdata is the disclosure risk of personal information included in the microdata. Identification occurs if a person with access to an identification file and a microdata file conducts one-to-one matching based on the key variables included in both files, and is thereby able to identify a matched record as referring to a specific person ((Bethlehem *et al*. (1990), Müller *et al*. (1995)). Empirical studies to assess the risk of direct identification of individuals are frequently conducted in countries where anonymized official microdata are released.

There are several empirical studies on matching official microdata with external data. In Germany, research that involves the matching of data from the 1987 German Microcensus survey with data from the 1987 Kürschners Deutscher Gelehrtenkalender has been conducted to quantify factual anonymity (Müller *et al*. (1995)). In the United Kingdom, empirical research that involves matching individual SAR from the 1991 Population Census with microdata from the General Household Survey has been conducted (Dale and Elliot (2001)).

This research matches anonymized official microdata created from Japanese Population Census data with individual data from the 'Housing and Land Survey' in order to assess the disclosure risk. The data from the 2008 Housing and Land Survey used in this research consists of approximately 10,000 records of individual data, and was selected so that it covers the same areas as the Population Census. A 10% sample was taken from the Population Census data, and for this data targeted data swapping or random data swapping were performed. From the results, records from the Population Census that are sample uniques based on the below combinations of key variables were selected, and then matched to records from the Housing and Land Survey that are identical sample uniques. The following combinations of key variables that are covered in both the Population Census and in Housing and Land Survey were selected for this purpose, and categories were adjusted as far as possible:

Case 1: Prefecture ID Number, City ID Number, Gender, Age (Five-Year Age Brackets), Marital Status, Number of Household Members

Case 2: Prefecture ID Number, City ID Number, Gender, Age (Five-Year Age Brackets), Marital Status, Number of Household Members, Type of Building, Type and Tenure of Dwelling

Case 3: Prefecture ID Number, City ID Number, Gender, Age (Five-Year Age Brackets), Marital Status, Number of Household Members, Type of Building, Number of Stories of Building, Type and Tenure of Dwelling

Tables 1 to 6 present the results of data matching between the anonymized official microdata created from Population Census data (referred to as "Population Census anonymized microdata") and individual data from the Housing and Land Survey. The number of sample uniques for Population Census anonymized microdata increases for higher numbers of key variables. For example, in Case 1 (targeted data swapping) the number of sample uniques for Population Census anonymized microdata is approximately 900, whereas in Case 3 (targeted data swapping) the number of sample uniques for Population Census anonymized microdata is between 2000 and 2400. This result applies to both targeted and random swapped data.

The number of sample uniques for Population Census anonymized microdata decreases for higher swapping rates. This might be due to sample uniques having been converted to groups with a minimum size of 2 as part of the data swapping. Again, this result applies to both targeted and random swapped data.

The percentage of matched records among total sample uniques for Population Census anonymized microdata is around 20% in Case 1 (targeted and random swapped data), whereas the percentage of matched records among total sample uniques is around 4% in Case 3. This is due to the fact that survey respondents for 'number of stories of building' in the Population Census and the Housing and Land Survey are different, and therefore this value is frequently left blank in the Population Census data. As a result, the percentage of matched records among total sample uniques for Population Census anonymized microdata in Case 3 is lower than in Case 1 or Case 2.

While the percentage of swapped records among the total of matched records increases for higher swapping rates, even for a swapping rate of 30% this percentage is around 60% in Case 3. This demonstrates that swapped records from the Population Census anonymized microdata do not completely overlap with matched records from the Housing and Land Survey, i.e. only part of the records from the Population Census anonymized microdata that match with records from the Housing and Land Survey were actually swapped.

The results of this analysis also show that the percentage of truly matched records[4] among records that are sample uniques is between 1% and 2% and therefore very low, while the percentage of truly matched records among the total of matched records is between 10% and 30%. These results demonstrate that the disclosure risk for anonymized official microdata created from Population Census data can be considered low if individual data from the Housing and Land Survey is used as external data for

[4] In this research, truly matched records among total sample uniques from the Population Census were checked whether they originated from the same survey unit area based on internal documents of National Statistics Center of Japan.

| Swapping Rate | Number of Sample Uniques from the Population Census | Number of Sample Uniques from the Housing and Land Survey | Number of Matched Records | | Number of Truly Matched Records | | Percentage of Matched Records to Sample Uniques from the Population Census | Percentage of Truly Matched Records to Sample Uniques from the Population Census |
|---|---|---|---|---|---|---|---|---|
| | | | Matched Records | Number of Records Which Are Swapped | Truly Matched Records | Number of Records Which Are Swapped | | |
| 1% (100rcd) | 933 | 866 | 195 | 3 | 20 | 0 | 20.90% | 2.14% |
| 2% (200rcd) | 934 | 866 | 195 | 5 | 20 | 0 | 20.88% | 2.14% |
| 3% (300rcd) | 931 | 866 | 194 | 6 | 20 | 0 | 20.84% | 2.15% |
| 5% (500rcd) | 930 | 866 | 194 | 11 | 20 | 1 | 20.86% | 2.15% |
| 10% (1000rcd) | 930 | 866 | 191 | 21 | 19 | 1 | 20.54% | 2.04% |
| 20% (2000rcd) | 909 | 866 | 185 | 39 | 18 | 2 | 20.35% | 1.98% |
| 30% (3000rcd) | 898 | 866 | 180 | 55 | 17 | 3 | 20.04% | 1.89% |

Note: Results are the averages of the values calculated for each of the 10 files of sampled data.

**Table 1** Results of Matching of Anonymized Official Microdata Created from Population Census Data with Individual Data from the Housing and Land Survey (Targeted Data Swapping, Case 1)

| Swapping Rate | Number of Sample Uniques from the Population Census | Number of Sample Uniques from the Housing and Land Survey | Number of Matched Records | | Number of Truly Matched Records | | Percentage of Matched Records to Sample Uniques from the Population Census | Percentage of Truly Matched Records to Sample Uniques from the Population Census |
|---|---|---|---|---|---|---|---|---|
| | | | Matched Records | Number of Records Which Are Swapped | Truly Matched Records | Number of Records Which Are Swapped | | |
| 1% (100rcd) | 1,989 | 1,750 | 341 | 4 | 51 | 1 | 17.14% | 2.56% |
| 2% (200rcd) | 1,979 | 1,750 | 342 | 11 | 51 | 1 | 17.28% | 2.58% |
| 3% (300rcd) | 1,967 | 1,750 | 341 | 16 | 50 | 1 | 17.34% | 2.54% |
| 5% (500rcd) | 1,952 | 1,750 | 339 | 28 | 49 | 3 | 17.37% | 2.51% |
| 10% (1000rcd) | 1,913 | 1,750 | 336 | 53 | 44 | 4 | 17.56% | 2.30% |
| 20% (2000rcd) | 1,820 | 1,750 | 322 | 98 | 38 | 7 | 17.69% | 2.09% |
| 30% (3000rcd) | 1,740 | 1,750 | 311 | 135 | 31 | 8 | 17.87% | 1.78% |

Note: Results are the averages of the values calculated for each of the 10 files of sampled data.

**Table 2** Results of Matching of Anonymized Official Microdata Created from Population Census Data with Individual Data from the Housing and Land Survey (Targeted Data Swapping, Case 2)

| Swapping Rate | Number of Sample Uniques from the Population Census | Number of Sample Uniques from the Housing and Land Survey | Number of Matched Records | | Number of Truly Matched Records | | Percentage of Matched Records to Sample Uniques from the Population Census | Percentage of Truly Matched Records to Sample Uniques from the Population Census |
|---|---|---|---|---|---|---|---|---|
| | | | Matched Records | Number of Records Which Are Swapped | Truly Matched Records | Number of Records Which Are Swapped | | |
| 1% (100rcd) | 2,392 | 2,388 | 114 | 2 | 34 | 0 | 4.77% | 1.42% |
| 2% (200rcd) | 2,379 | 2,388 | 111 | 4 | 32 | 0 | 4.67% | 1.35% |
| 3% (300rcd) | 2,365 | 2,388 | 110 | 8 | 30 | 0 | 4.65% | 1.27% |
| 5% (500rcd) | 2,337 | 2,388 | 108 | 14 | 27 | 1 | 4.62% | 1.16% |
| 10% (1000rcd) | 2,274 | 2,388 | 102 | 25 | 22 | 2 | 4.49% | 0.97% |
| 20% (2000rcd) | 2,133 | 2,388 | 94 | 43 | 15 | 3 | 4.41% | 0.70% |
| 30% (3000rcd) | 2,002 | 2,388 | 84 | 54 | 10 | 3 | 4.20% | 0.50% |

Note: Results are the averages of the values calculated for each of the 10 files of sampled data.

**Table 3** Results of Matching of Anonymized Official Microdata Created from Population Census Data with Individual Data from the Housing and Land Survey (Targeted Data Swapping, Case 3)

| Swapping Rate | Number of Sample Uniques from the Population Census | Number of Sample Uniques from the Housing and Land Survey | Number of Matched Records | | Number of Truly Matched Records | | Percentage of Matched Records to Sample Uniques from the Population Census | Percentage of Truly Matched Records to Sample Uniques from the Population Census |
|---|---|---|---|---|---|---|---|---|
| | | | Matched Records | Number of Records Which Are Swapped | Truly Matched Records | Number of Records Which Are | | |
| 1% (100rcd) | 934 | 866 | 196 | 2 | 20 | 0 | 20.99% | 2.14% |
| 2% (200rcd) | 933 | 866 | 196 | 4 | 20 | 0 | 21.01% | 2.14% |
| 3% (300rcd) | 935 | 866 | 195 | 6 | 19 | 0 | 20.86% | 2.03% |
| 5% (500rcd) | 934 | 866 | 194 | 9 | 20 | 0 | 20.77% | 2.14% |
| 10% (1000rcd) | 930 | 866 | 195 | 21 | 19 | 1 | 20.97% | 2.04% |
| 20% (2000rcd) | 920 | 866 | 192 | 40 | 18 | 3 | 20.87% | 1.96% |
| 30% (3000rcd) | 897 | 866 | 184 | 57 | 17 | 3 | 20.51% | 1.90% |

Note: Results are the averages of the values calculated for each of the 10 files of sampled data.

**Table 4** Results of Matching of Anonymized Official Microdata Created from Population Census Data with Individual Data from the Housing and Land Survey (Random Data Swapping, Case 1)

| Swapping Rate | Number of Sample Uniques from the Population Census | Number of Sample Uniques from the Housing and Land Survey | Number of Matched Records | | Number of Truly Matched Records | | Percentage of Matched Records to Sample Uniques from the Population Census | Percentage of Truly Matched Records to Sample Uniques from the Population Census |
|---|---|---|---|---|---|---|---|---|
| | | | Matched Records | Number of Records Which Are Swapped | Truly Matched Records | Number of Records Which Are Swapped | | |
| 1% (100rcd) | 2,004 | 1,750 | 343 | 4 | 52 | 0 | 17.12% | 2.59% |
| 2% (200rcd) | 1,998 | 1,750 | 342 | 9 | 51 | 1 | 17.12% | 2.55% |
| 3% (300rcd) | 1,993 | 1,750 | 343 | 15 | 51 | 1 | 17.21% | 2.56% |
| 5% (500rcd) | 1,984 | 1,750 | 339 | 22 | 48 | 2 | 17.09% | 2.42% |
| 10% (1000rcd) | 1,956 | 1,750 | 339 | 50 | 44 | 3 | 17.33% | 2.25% |
| 20% (2000rcd) | 1,887 | 1,750 | 322 | 91 | 39 | 6 | 17.06% | 2.07% |
| 30% (3000rcd) | 1,766 | 1,750 | 310 | 133 | 32 | 8 | 17.55% | 1.81% |

Note: Results are the averages of the values calculated for each of the 10 files of sampled data.

**Table 5** Results of Matching of Anonymized Official Microdata Created from Population Census Data with Individual Data from the Housing and Land Survey (Random Data Swapping, Case 2)

| Swapping Rate | Number of Sample Uniques from the Population Census | Number of Sample Uniques from the Housing and Land Survey | Number of Matched Records | | Number of Truly Matched Records | | Percentage of Matched Records to Sample Uniques from the Population Census | Percentage of Truly Matched Records to Sample Uniques from the Population Census |
|---|---|---|---|---|---|---|---|---|
| | | | Matched Records | Number of Records Which Are Swapped | Truly Matched Records | Number of Records Which Are Swapped | | |
| 1% (100rcd) | 2,407 | 2,388 | 116 | 1 | 35 | 0 | 4.82% | 1.45% |
| 2% (200rcd) | 2,400 | 2,388 | 114 | 3 | 34 | 0 | 4.75% | 1.42% |
| 3% (300rcd) | 2,393 | 2,388 | 113 | 5 | 34 | 0 | 4.72% | 1.42% |
| 5% (500rcd) | 2,385 | 2,388 | 112 | 10 | 32 | 1 | 4.70% | 1.34% |
| 10% (1000rcd) | 2,348 | 2,388 | 108 | 21 | 27 | 1 | 4.60% | 1.15% |
| 20% (2000rcd) | 2,259 | 2,388 | 98 | 39 | 19 | 2 | 4.34% | 0.84% |
| 30% (3000rcd) | 2,048 | 2,388 | 84 | 49 | 13 | 4 | 4.10% | 0.63% |

Note: Results are the averages of the values calculated for each of the 10 files of sampled data.

**Table 6** Results of Matching of Anonymized Official Microdata Created from Population Census Data with Individual Data from the Housing and Land Survey (Random Data Swapping, Case 3)

data matching.

## 4 Comparison between Data Utility and Disclosure Risk for Swapped Data

Disclosure risk exists not only for matching with external data, but also for records that are special uniques. To determine the extent of this disclosure risk, it is necessary to assess the effectiveness of the data swapping that has been performed. For this, calculating data utility and disclosure risk is required.

In this research, data utility is defined as the average absolute distance per tabulation cell, and therefore an indicator of distance that measures distortion to the distribution based on Shlomo *et al*. (2010). The indicator of Data utility (DU) is given as:

$$DU = \frac{\sum_c \left| T^P(c) - T^O(c) \right|}{n_T} \tag{1}$$

$T^O(c)$ is the cell frequency in the tabulation using original data and $T^P(c)$ is the cell frequency in the tabulation using swapped data, where $n_T$ is the number of cells in the tabulation.

According to Shlomo *et al*. (2010), the indicator of disclosure risk (DR) is given as:

$$DR = \frac{\sum_c I\left(T^O(c) = 1, T^P(c) = 1\right)}{\sum_c I\left(T^O(c) = 1\right)} \tag{2}$$

$\sum_c I\left(T^O(c) = 1\right)$ is the number of unique cells contained in the tabulation using original data. Also, $\sum_c I\left(T^O(c) = 1, T^P(c) = 1\right)$ is calculated as the number of unperturbed unique cells in the tabulation.

DU and DR were calculated based on all possible two-variable combinations of the key variables.

Figure 1 presents the R-U confidentiality map created based on the average values of DU and DR for both targeted data swapping and random data swapping. The results show that DU tends to increase as the swapping rate increases. Also, DU tends to be higher for targeted data swapping than for random data swapping. This indicates that data utility for targeted data swapping is lower than for random data swapping. DR tends to be lower for higher swapping rates, and also tends to be lower for targeted data swapping than random data swapping. This indicates that disclosure risk for targe-

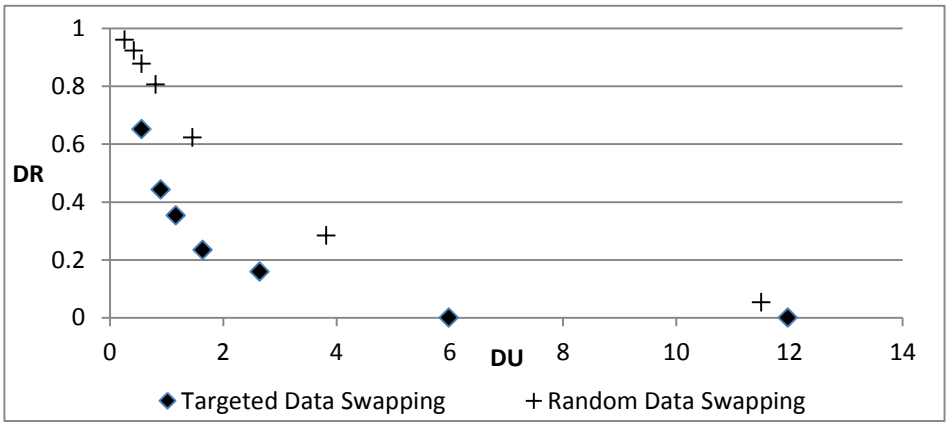


**Fig 1** R-U Confidentiality Map with Data Utility (DU) and Disclosure Risk (DR)

Note: Results are the averages of the values calculated for each of the 10 files of sampled data.

ted data swapping is lower than for random data swapping.

The results also show that for targeted data swapping at a swapping rate of 5% DR is lower than DR for random data swapping except at a swapping rate of 30%. On the other hand, DU for swapped data using targeted data swapping at a swapping rate of 5% is higher than DU for swapped data using random data swapping at a swapping rate of 10% or less. This suggests that there is not one single optimum swapping methodology or swapping rate, but rather that both should be selected based on the desired threshold of DU or DR.

## 5 Conclusion

This paper assesses the effectiveness of disclosure limitation methods for official microdata by matching anonymized official microdata created from Japanese Population Census data with individual data from the Japanese Housing and Land Survey. The results show that the rate of true matching is low, which indicates that disclosure risk for matching with external data is also low.

This paper also assesses data utility and disclosure risk for data swapping based on the R-U map. The results show that disclosure risk for targeted data swapping is lower than for random data swapping, whereas data utility for random data swapping is overall higher than that for targeted data swapping. Therefore, a balance between data utility and disclosure risk is important when creating anonymized official microdata.

This research provides an approach for determining the most effective disclosure

limitation method for a particular data set, and thereby has the potential to help minimize disclosure risk for Japanese official microdata. It is hoped that the results from this research will contribute to the creation of anonymized official microdata in Japan.

## Note

The opinions expressed in this paper do not necessarily reflect those of organizations to which the authors belong or the National Statistics Center.

## References


Bethlehem, J. M., Keller, W. J., Pannekoek, J.(1990) "Disclosure Control of Microdata", *Journal of American Statistical Association*, Vol. 85, pp.38-45.

Dale, A. and Elliot, M. (2001) "Proposal for 2001 Samples of Anonymized Records: An Assessment of Disclosure Risk", *Journal of the Royal Statistical Society*, Series A, Vol.164, No.3, pp.427-447.

Domingo-Ferrer, J. and Torra, V. (2001) "Disclosure Control Methods and Information Loss for Microdata", Doyle *et al.* (eds.) *Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies*, Elsevier Science, Amsterdam, pp. 91-110.

Elliot, M., Manning, A. M., Ford, R. W. (2002) "A Computational Algorithm for Handling The Special Uniques Problem", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 10, No.5, pp.493-509.

Ito, S. and Murata, M. (2011) "Quantitative Methods to Assess Data Confidentiality and Data Utility for Microdata in Japan", Paper presented at Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Tarragona, Spain, pp.1-10.
http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2011/20_Japan.pdf.

Ito, S. and Hoshino, N.(2012) "The Potential of Data Swapping as a Disclosure Limitation Method for Official Microdata in Japan: An Empirical Study to Assess Data Utility and Disclosure Risk for Census Microdata" Paper presented at Privacy in Statistical Databases 2012, Palermo, Sicily, Italy, pp.1-13.

Müller, W., Blien, U., Wirth, H.(1995) "Identification Risks of Micro Data: Evidence from Experimental Studies", *Sociological Methods and Research*, Vol.24, No.2, pp.131-157.

Shlomo, N., Tudor, C., Groom, P. (2010) "Data Swapping for Protecting Census Tables", Domingo-Ferrer, J. and Magkos, E.(eds) *Privacy in Statistical Databases UNESCO Chair in Data Privacy International Conference, PSD 2010 Corfu, Greece, September, 2010 Proceedings*, Springer, pp.41-51.

Willenborg, L. and De Waal, T. (2001) *Elements of Statistical Disclosure Control*, Springer, New York.