A Quantitative Assessment of Data Confidentiality and Data Utility to Create Anonymized Census Microdata in Japan

Shinsuke Ito[*], Naomi Hoshino[**], Fumika Akutsu[***]

[*] Faculty of Economics, Chuo University, 742-1 Higashinakano, Hachioji, Tokyo, 192-0393 Japan,
E-mail: ssitoh@tamacc.chuo-u.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 Statistics Bureau of Japan, and National Statistics Center.
[**] National Statistics Center, 19-1 Wakamatsu-cho, Shinjuku-ku, Tokyo, 162-8668 Japan,
E-mail: nsaitou2@nstac.go.jp
[***] Statistics Bureau of Japan, 19-1 Wakamatsu-cho, Shinjuku-ku, Tokyo, 162-8668 Japan,
E-mail: fakutsu@soumu.go.jp

Abstract: Anonymized microdata for seven types of official statistics are currently made available under the Statistics Act in Japan, including Anonymized microdata from the 2000 and 2005 'Population Census' conducted by the Statistics Bureau of Japan. For almost all official statistics, only one type of Anonymized microdata is provided.

Several empirical studies on the effectiveness of disclosure limitation methods for official microdata, such as microaggregation, additive noise and data swapping have been conducted by the National Statistics Center of Japan in order to promote a broader use of anonymized official microdata. In addition, the Statistics Bureau and the National Statistics Center are conducting empirical studies to assess data utility and disclosure risk as part of their preparations to provide anonymized official microdata from the 2010 Population Census.

This paper quantitatively assesses data utility and data confidentiality for anonymized data generated from original microdata from the Population Census. This research aims to contribute to the provision of different types of Anonymized microdata e.g. with more detailed geographical information. Different types of Anonymized microdata will allow researchers from a variety of fields including economics, sociology, demography, geography etc. to conduct more detailed statistical analysis based on official statistics in Japan.

1 Introduction: Anonymized Census Microdata in Japan

Japan's Statistics Act was revised in April 2007 – the first major revision in sixty years – with the objective of promoting the development and use of official statistics, and thereby contributing to the development of the national economy and enhancement of the living standards of the citizens. The 'Master Plan Concerning the Development of Official Statistics' was established based on the Statistics Act and introduced a "secondary usage" system that includes the production and provision of tailor-made tabulations and Anonymized microdata ('Anonymized microdata' with a capital "A" are defined as individual data 'that is processed so that no particular individuals or juridical persons, or other organizations shall be identified' (Article 36 of the Japanese

Statistical law)). This was the starting point for the production and provision of anonymized census microdata Japan.

According to the 'Guidelines for the Production and Provision of Anonymized Microdata', data provision is limited to requests that fall within the following objectives: to contribute to the development of academic research, to contribute to the development of advanced education or to contribute to the advancement of Japan as a member of the international community and the development of the international economic community. In order to ensure data confidentiality and data utility, the Statistical Commission considers the classification of survey items, choice of anonymization methods, and timing of data provision as well as other factors when planning the production and provision of Anonymized microdata.

The Statistics Bureau has been providing Anonymized census microdata since 2013. The data is made available five years after each census, so data from the 2000 and 2005 census is currently available. Various disclosure limitation methods such as sampling (at a sampling rate of 1%), recoding, top (bottom) coding, and data deletion are applied to the data before it is made available.

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 with the objective of promoting a broader use of anonymized official microdata (Ito and Murata (2011), Ito and Hoshino (2012, 2013, 2014)). The Statistics Bureau and the National Statistics Center are currently conducting empirical research to prepare the provision of Anonymized microdata from the 2010 Census.

In order to collect reference information for the preparation of Anonymized microdata from the 2010 Census, the Statistics Bureau conducted research into user needs based on the needs of current users, and collected information on the processes and procedures for providing Anonymized microdata in other countries including the United States, United Kingdom, Australia, New Zealand, and Canada in order to better understand the needs of non-users.

This research found that in other countries users are usually given access to different types of microdata files, and offered several options to access the data including online downloads, research data centers, remote access and others, while anonymization methods also vary from country to country.

Small area results are a very important type of microdata from a user perspective, as the detailed geographical information allows researchers from a variety of fields including economics, sociology, demography and geography to use microdata for detailed statistical analysis. Currently, geographic information contained in the Anonymized census microdata for 2000 and 2005 is limited to prefecture level, and the amount of available data is progressively restricted for smaller areas. For data from

the 2010 census, the Statistics Bureau is researching ways to produce anonymized microdata that includes small area results while maintaining confidentiality.

This paper aims to suggest an approach for the creation of such anonymized small area microdata in Japan. Towards this objective, a quantitative assessment of data confidentiality was conducted using the concepts of 'threshold of data confidentiality' and 'allowable population unique ratio'. Second, data utility was assessed based on entropy-based measures in order to identify combinations of recoding and top coding below the 'threshold for data confidentiality.' Third, data utility and data confidentiality were compared for different sampling rates in order to identify the optimal combination of recoding and sampling.

2 Quantitative Assessment of Data Confidentiality Based on the "Allowable Threshold"

In the U.K., Samples of Anonymised Records (SARs) from the 1991 Population Census are produced and released. For this data, the level of detail for geographical classification, categories of household and individual attributes as well as the sampling rate are determined based on the 'thresholding rule' (Dale (1995), Marsh *et al*. (1994)). For the creation of Small Area Microdata (SAM) from the 2001 Population Census, Tranmer *et al*. (2005) conducted an empirical analysis on the disclosure risk for microdata that contains more detailed geographical information, and compared it to the disclosure risk for 1991 Individual SARs.

In Japan, achieving data confidentiality on the same level as Anonymized microdata from the 2000 and 2005 Population Census would enable the provision of a broader range of Anonymized data including data with more detailed geographical information. By calculating the thresholds for data confidentiality, possible combinations of geographical classification, categories of household, individual attributes and sampling rate can be determined. One way to determine the threshold is to compare data confidentiality for anonymized data (anonymized data with a lower-case "a" is defined as microdata to which disclosure limitation methods have been applied as part of this research) to that for Anonymized microdata ('Anonymized microdata' with a capital "A" are defined as official microdata 'that is processed so that no particular individuals or juridical persons, or other organizations shall be identified'). As test data, anonymised data with more detailed geographical information were created from Japanese census microdata, and a quantitative assessment of data confidentiality was conducted.

Three sets of data from the 2000 Population Census – each containing a different number of records – were used as test data. The first set was created based on more than 500,000 records of individual data from a certain geographic area within a specific Japanese prefecture. This area is referred to as "Area A". The second set of data was created based on more than 100,000 records of individual data from another geographic

area within the same prefecture. This area is referred to as "Area B". The third set of data was created based on more than 50,000 records of individual data from a third geographic area within the same prefecture. This area is referred to as "Area C".

Anonymized test data with standard levels of geographical information was created based on head of household records from Area A. Anonymized test data with more detailed geographical information was created for both individual records and head of household records for Areas B and C. Table 1 shows the characteristics of the anonymized test data for Areas A, B, and C.

The ratio of population uniques was calculated for both original data and anonymized data from all areas using the following 10 variables:

Gender (2 categories for both original data and anonymized data)

Marital Status (5 categories for both original data and anonymized data)

Nationality (12 categories for original data, 2 categories for anonymized data)

Type of (Work) Activity (9 categories for original data, 6 categories for anonymized data)

Occupation (10 categories for original data, 8 categories for anonymized data)

Type and Tenure of Dwelling (9 categories for original data, 6 categories for anonymized data)

Type of Building and Total Number of Floors (5 categories for original data, 4 categories for anonymized data)

Age

Employment Status

Industry

For the variables of gender, marital status, nationality, type of (work) activity, occupation, type and tenure of dwelling, type of building and total number of floors recoding was applied identically to Anonymized microdata. Age, employment status and industry were recoded and/or top coded based on the following patterns:

<u>Age (9 patterns)</u>

(1) One-year age brackets

(2) One-year age brackets and top coding for 85 years and above

(3) One-year age brackets and top coding for 75 years and above

(4) Five-year age brackets

	Unit	Number of Records	Anonymized Data
Area A	Household	250,000 records	Anonymized Data A
Area B	Individual	140,000 records	Anonymized Individual Data B
	Household	50,000 records	Anonymized Household Data B
Area C	Individual	80,000 records	Anonymized Individual Data C
	Household	20,000 records	Anonymized Household Data C

Table 1 The Characteristics of Anonymized Test Data for Areas A, B and C.

(5) Five-year age brackets and top coding for 85 years and above (the same categories as for Anonymized microdata)

(6) Five-year age brackets and top coding for 75 years and above

(7) Ten-year age brackets

(8) Ten-year age brackets and top coding for 85 years and above

(9) Ten-year age brackets and top coding for 75 years and above

Employment status (3 patterns)

(1) Categories from original data (8 categories)

(2) Categories from Anonymized microdata (6 categories)

(3) 4 Categories

Industry (3 patterns)

(1) Categories from original data (14 categories)

(2) Categories from Anonymized microdata (10 categories)

(3) 4 Categories

Population uniques were calculated for all 81 possible combinations of the patterns for age, employment status and industry, and a quantitative assessment of data confidentiality was conducted based on the following steps:

(1) The ‘decrease rate of population uniques’ for Area A was calculated for the above key variables as the ratio of population uniques for the recoded categories from the original data relative to the population uniques for the original categories from anonymized data.

(2) The ‘allowable population unique ratio’ was calculated by multiplying the population unique ratios for Areas B and C with the ‘decrease rate of population uniques’ for Area A. This ratio was used as the ‘threshold of allowable data confidentiality’.

(3) Recoding and top coding were applied to age, employment status and industry for both household data and individual data from Areas B and C. Anonymized individual and anonymized household data was created for each of the 81 patterns of recoding and top coding. The population unique ratios for anonymized individual data and anonymized household data for Areas B and C were calculated and compared to the ‘allowable population unique ratio’.

Table 2 contains the results of this comparison. For individual anonymized data from Area B (referred to as “anonymized individual data B”), the population unique ratio calculated based on the categories of key variables from the original data for Area A is 13.46%, while the population unique ratio calculated based on the recoded categories of key variables from anonymised data for Area A is 4.20%. Therefore the ‘decrease rate of population uniques’ is approximately 31%. Multiplying this ‘decrease rate of population uniques’ with the population unique ratio for individual data and household data, the ‘allowable population unique ratio’ is 5.30% and 8.35% respectively for individual records and household records for Area B, and 5.77% and 9.78% respectively for individual records and household records for Area C.

Table 3 contains the patterns of recoding and top coding for anonymised individual data B for which the population unique ratio is smaller than the allowable population unique ratio. The number of patterns is 42. The number of patterns of recoding and top coding for which the population unique ratio is smaller than the allowable population unique ratio is 36 for anonymized household data B, 42 for anonymized individual data C, and 24 for anonymized household data C. This result indicates that for anonymized data consisting of individual records, area size does not impact the allowable combinations of recoding.

3 Quantitative Assessment of Data Utility Based on Entropy

Calculating information loss using entropy-based measures in order to assess data utility of quantitative attributes was first proposed by Kooiman *et al*. (1998) and Domingo Ferrer and Torra (2001). De Waal and Willenborg (1999) calculated entropy-based measures of information loss for anoymized data that was created using recoding.

Based on this research, a quantitative assessment of data utility was conducted for combinations of recoding and top coding for key variables for Areas B and C where the population unique ratio is lower than the allowable population unique ratio. Entropy-based measures of information loss were used to determine the optimal combinations of recoding and top coding for which information loss is lowest.

	Population Unique ratio		Decrease Rate of Population Uniques
	Original Categories	Recoded Categories	
Area A	13.46%	4.20%	31.20%

	Anonymized Data	Population Unique Ratio for Original Categories	Allowable Population Unique Ratio
Area B	Anonymized Individual Data B	16.97%	5.30%
	Anonymized Household Data B	26.72%	8.35%
Area C	Anonymized Individual Data C	18.47%	5.77%
	Anonymized Household Data C	31.31%	9.78%

Table 2 Population unique ratio and threshold of data confidentiality for Areas A, B, and C.

Table 4 contains the entropy-based information loss for anonymized individual data B. Among the combinations with a population unique ratio smaller than the "allowable population unique ratio", the pattern of five year age brackets, original categories for industry, and three categories for employment status shows the lowest information loss.

Figure 1 shows the R-U confidentiality map based on population unique ratios which are lower than the "allowable population unique ratio", and information loss for individual records and household records for Area B. This result illustrates that patterns with higher population unique ratios generally tend to have lower information loss.

4 The Relationship between Sampling and Recoding from a Perspective of Data Confidentiality and Data Utility

Sampling (or resampling) is used as a disclosure limitation method for the creation of Anonymized microdata. As the sampling rate impacts data confidentiality and data utility, a quantitative assessment of data confidentiality and data utility for different sampling rates was conducted as part of this research. The following steps were used:

(1) UUSU rates (ratio of records which pertain to the population unique and sample unique to the number of records which pertain to sample unique) were calculated for 1% sampled anonymized data A using the same key variables that were used for calculating population uniques. The allowable UUSU rate was defined as the 'threshold of data confidentiality' for the sampled data.

(2) In the same way, UUSU rates were calculated based on a sampling rate p % (p=1, 2 ,3 ,4 ,5 ,6 ,7 ,8 ,9 ,10) for anonymized individual data and anonymized household data for Areas B and C.

(3) The combinations of sampling rate and recoding for which the UUSU rate is lower than the allowable UUSU rate were determined and the highest possible sampling rate that meets the threshold for data confidentiality was selected.

One-year age brackets	One-year age brackets and top coding for 85 years and above	One-year age brackets and top coding for 75 years and above	Five-year age brackets	Five-year age brackets and top coding for 85 years and above	Five-year age brackets and top coding for 75 years and above	Ten-year age brackets	Ten-year age brackets and top coding for 85 years and above	Ten-year age brackets and top coding for 75 years and above	Employment Status			Industry			Population unique ratio
									8 Categories	6 Categories	4 Categories	14 Categories	10 Categories	4 Categories	
								*			*			*	1.79%
						*					*			*	1.79%
							*				*			*	1.82%
						*				*				*	2.44%
								*		*				*	2.45%
							*			*				*	2.46%
						*			*					*	2.61%
								*	*					*	2.62%
							*		*					*	2.63%
					*						*			*	2.73%
				*							*			*	2.84%
			*								*			*	2.88%
								*			*		*		2.92%
						*					*		*		2.93%
							*				*		*		2.96%
								*			*	*			3.06%
						*					*	*			3.07%
							*				*	*			3.09%
					*					*				*	3.63%
				*						*				*	3.75%
			*							*				*	3.79%
					*				*					*	3.86%
						*				*			*		3.92%
								*		*			*		3.92%
							*			*			*		3.94%
				*					*					*	3.98%
			*						*					*	4.02%
						*				*		*			4.06%
								*		*		*			4.06%
							*			*		*			4.08%
						*			*				*		4.14%
								*	*				*		4.14%
							*		*				*		4.16%
						*			*			*			4.28%
								*	*			*			4.28%
							*		*			*			4.30%
					*						*		*		4.36%
				*							*		*		4.49%
					*						*	*			4.52%
			*								*		*		4.53%
				*							*	*			4.65%
			*								*	*			4.69%

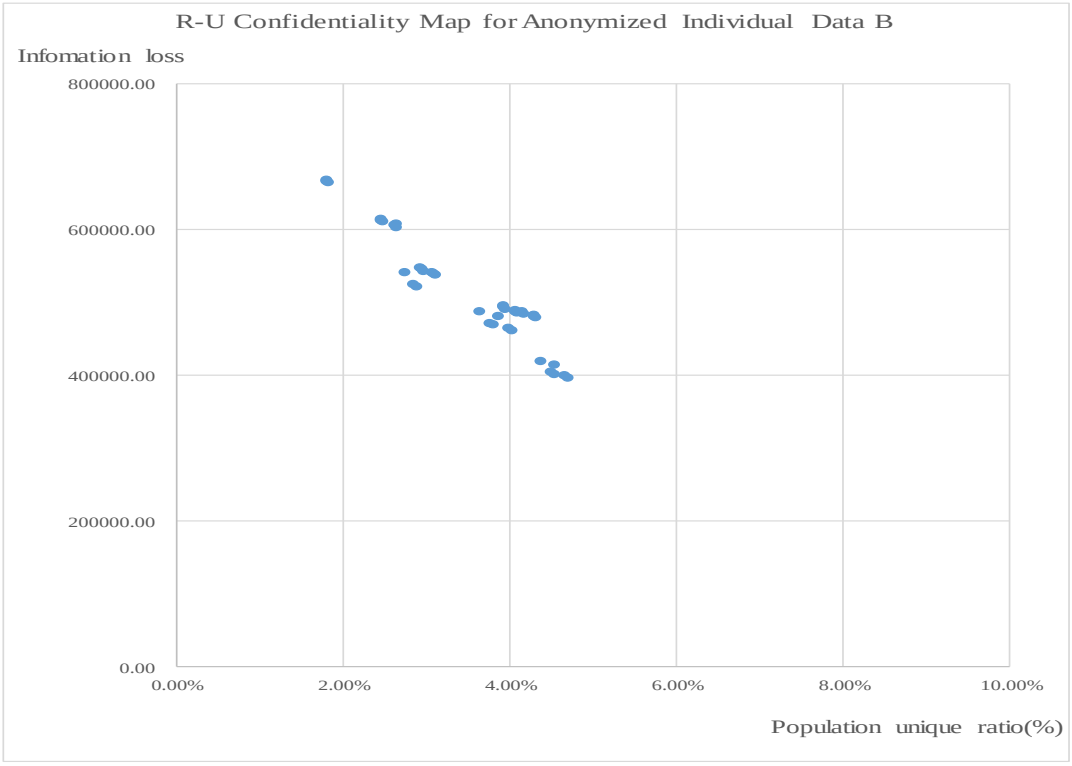
Note '*' denotes the combination of recoding and top coding selected in this research.

Table 3 Anonymized Individual Data B: Patterns for recoding and top coding for which the population unique ratio is lower than the allowable population unique ratio.

One-year age brackets	One-year age brackets and top coding for 85 years and above	One-year age brackets and top coding for 75 years and above	Five-year age brackets	Five-year age brackets and top coding for 85 years and above	Five-year age brackets and top coding for 75 years and above	Ten-year age brackets	Ten-year age brackets and top coding for 85 years and above	Ten-year age brackets and top coding for 75 years and above	Employment Status			Industry			Information Loss
									8 Categories	6 Categories	4 Categories	14 Categories	10 Categories	4 Categories	
								*			*			*	668591.46
						*					*			*	666693.51
							*				*			*	664538.06
						*				*				*	612928.28
								*		*				*	614847.03
							*			*				*	610772.87
						*			*					*	605439.05
								*	*					*	607366.01
							*		*					*	603287.21
					*						*			*	540950.79
				*							*			*	525165.57
			*								*			*	522368.46
								*			*		*		547181.28
						*					*		*		545273.81
							*				*		*		543126.77
								*			*	*			541829.90
						*					*	*			539921.07
							*				*	*			537773.51
					*					*				*	487572.89
				*						*				*	471797.63
			*							*				*	469002.16
					*				*					*	480136.36
						*				*			*		493147.09
								*		*			*		495069.76
							*			*			*		491000.77
				*					*					*	464369.69
			*						*					*	461575.65
						*				*		*			487897.07
								*		*		*			489820.15
							*			*		*			485750.22
						*			*				*		486183.40
								*	*				*		488113.27
							*		*				*		484042.27
						*			*			*			480957.72
								*	*			*			482887.81
							*		*			*			478816.05
					*						*		*		419892.82
				*							*		*		404145.94
					*						*	*			414594.79
			*								*		*		401353.95
				*							*	*			398848.47
			*								*	*			396057.25

Note '*' denotes the combination of recoding and top coding selected in this research.

Table 4 Anonymized Individual Data B: Information loss for patterns of recoding and top coding for which the population unique ratio is lower than the allowable population unique ratio.



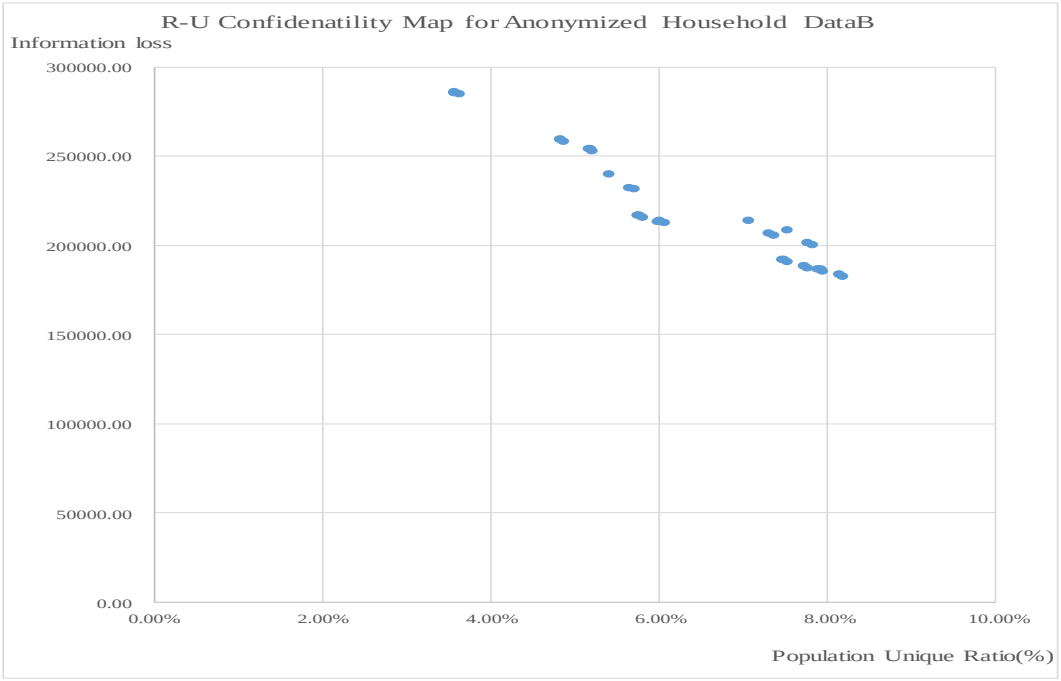


Figure 1 R-U confidentiality map for anonymized individual data and anonymized household data for Area B.

The UUSU rate for anonymized data A was 12.32 %. This rate was set as the allowable UUSU rate.

Table 5 contains the UUSU rates for anonymized individual data B at different sampling rates. These results show that for all 42 patterns, the UUSU rate is lower than the allowable UUSU rate. For a 3% sampling rate, there were 6 patterns of recoding for which the UUSU rate was lower than the allowable UUSU rate. Table 6 shows the combinations of recoding and top coding for anonymized data of individual records and household records for Area B and C for which the UUSU rate is lower than the allowable UUSU rate. For anonymized individual data C, there were 18 among a total of 42 patterns of recoding for which the UUSU rate was lower than the allowable UUSU rate, whereas for a 2% sampling rate there were only 3 patterns for which the UUSU rate was lower than the allowable UUSU rate.

These results indicate that by using the 'allowable UUSU rate' as an additional threshold for data confidentiality and for 3% sampled data for individual records from Area B, a combination of ten year age brackets and top coding for over 85 years old for age, three categories for employment status and the original categories for industry are the optimal combination of all recoding patterns used in this research.

5 Conclusion

This paper uses anonymized official microdata created from Japanese Population Census data to assess data confidentiality and data utility based on the two thresholds of 'allowable population unique rate' and 'allowable UUSU rate'. The results show that for individual data it is possible to create anonymized microdata with more detailed geographical information while identifying the combination of recoding about categories of individual and household attributes to maintain data confidentiality at the same level as currently provided Anonymized census microdata.

It is hoped that this research will contribute to the provision of different types of Anonymized microdata that will allow researchers from a variety of fields including economics, sociology, demography, geography etc. to conduct more detailed statistical analysis based on official statistics in Japan.

Note

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

One-year age brackets	One-year age brackets and top coding for 85 years and above	One-year age brackets and top coding for 75 years and above	Five-year age brackets	Five-year age brackets and top coding for 85 years and above	Five-year age brackets and top coding for 75 years and above	Ten-year age brackets	Ten-year age brackets and top coding for 85 years and above	Ten-year age brackets and top coding for 75 years and above	Employment Status			Industry			UUSU Rate		
									8 Categories	6 Categories	4 Categories	14 Categories	10 Categories	4 Categories	Sampling Rate of 1%	Sampling Rate of 2%	Sampling Rate of 3%
								*			*			*	5.26%	7.91%	11.65%
						*					*			*	6.00%	7.93%	11.30%
							*				*			*	6.48%	7.90%	11.41%
						*				*				*	6.68%	9.47%	13.47%
								*		*				*	6.05%	9.47%	13.61%
							*			*				*	7.10%	9.44%	13.44%
						*			*					*	6.61%	10.63%	14.44%
								*	*					*	6.01%	10.47%	14.57%
							*		*					*	7.02%	10.60%	14.40%
					*						*			*	6.67%	10.75%	13.62%
				*							*			*	7.54%	10.89%	13.56%
			*								*			*	7.47%	10.98%	13.64%
								*			*		*		8.09%	10.34%	12.40%
						*					*		*		8.66%	10.43%	12.28%
							*				*		*		8.84%	10.43%	12.36%
								*			*	*			7.99%	10.48%	12.38%
						*					*	*			8.53%	10.55%	12.24%
							*				*	*			8.70%	10.55%	12.31%
					*					*				*	7.91%	11.79%	15.66%
				*						*				*	8.68%	11.89%	15.57%
			*							*				*	8.61%	11.97%	15.72%
					*				*					*	8.06%	12.34%	16.04%
						*				*			*		9.93%	12.56%	14.86%
								*		*			*		9.58%	12.51%	14.93%
							*			*			*		10.09%	12.56%	14.83%
				*					*					*	8.82%	12.46%	15.95%
			*						*					*	8.75%	12.53%	16.10%
						*				*		*			9.79%	12.65%	14.78%
								*		*		*			9.48%	12.62%	14.87%
							*			*		*			9.95%	12.65%	14.75%
						*			*				*		9.87%	13.31%	15.72%
								*	*				*		9.57%	13.13%	15.72%
							*		*				*		10.04%	13.31%	15.69%
						*			*			*			9.73%	13.26%	15.71%
								*	*			*			9.46%	13.09%	15.74%
							*		*			*			9.89%	13.26%	15.68%
					*						*		*		10.34%	13.56%	15.18%
				*							*		*		11.02%	13.75%	15.18%
					*						*	*			10.27%	13.63%	15.31%
			*								*		*		10.94%	13.81%	15.32%
				*							*	*			10.95%	13.80%	15.31%
			*								*	*			10.87%	13.86%	15.44%

Note '*' denotes the combination of recoding and top coding selected in this research.

Table 5 Individual Anonymized Data B: Patterns for recoding and top coding for which the UUSU rate is lower than the allowable UUSU rate at different sampling rates.

Anonymized Data	Sampling Rate									
	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
Anonymized Individual Data B	42	21	6	0	0	0	0	0	0	0
Anonymized Household Data B	10	0	0	0	0	0	0	0	0	0
Anonymized Individual Data C	18	3	0	0	0	0	0	0	0	0
Anonymized Household Data C	3	3	3	0	0	0	0	0	0	0

Table 6 Anonymized Individual Data and Anonymized Household Data for Areas B and C: The number of patterns for recoding and top coding at different sampling rates for which the UUSU rate was lower than the allowable UUSU rate.

References

Dale, A(1995) "Samples of Anonymised Records from the 1991 Census for Great Britain" *IASSIST Quarterly*, pp.5-12.

De Waal, T. and Willenborg, L. (1999) "Information Loss Through Global Recoding and Local Suppression", *Netherlands Official Statistics (special issue on SDC)*, Vol.14, pp.17-20.

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.

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.

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.

Ito, S. and Hoshino, N.(2013) "Assessing the Effectiveness of Disclosure Limitation Methods for Census Microdata in Japan" Paper presented at Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Ottawa, Canada, pp.1-10.

Ito, S. and Hoshino, N.(2014) "Data Swapping as a More Efficient Tool to Create Anonymized Census Microdata in Japan", Paper presented at Privacy in Statistical Databases 2014, Ibiza, Spain, pp.1-14.

Kooiman, P., L. Willenborg and J. Gouweleeuw (1998) "PRAM: A Method for Disclosure Limitation of Microdata", Research Paper, No. 9705, Statistics Netherlands, Voorburg.

Marsh, C., Dale, A., Skinner, C.(1994) "Safe Data versus Safe Settings: Access to Microdata from the British Census", *International Statistical Review*, Vol.62, No.1, pp.35-53.

Tranmer, M., Pickles, A., Fieldhouse, E., Elliot, M., Dale, A., Brown, M.(2005) "The Case for Small Area Microdata", *Journal of Royal Statistical Society A*, Vol.168, pp.29-49.