UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Expert meeting on Statistical Data Confidentiality

26–28 September 2023, Wiesbaden

A Disclosure-Based Framework for Comparing Frequency Table Protection

Daniel P. Lupp and Øyvind Langsrud (Statistics Norway)

{dlu, oyl}@ssb.no

Abstract

For the protection of the dissemination tables from the 2021 population census, Eurostat recommended a combined use of the cell-key method and targeted record swapping. As part of a grant awarded to Statistics Norway on multi-grid geographical data, we compared this recommendation to alternative methods (in particular small count rounding) on dissemination of frequency data over multiple grid systems. This was done using Norwegian census data as a use case. In this work, we present the findings of this project, as well as discuss the comparison framework used. This framework is based on a suite of disclosure scenarios that can occur in frequency tables. Using established notions from information retrieval, disclosures are counted and evaluated for each scenario, providing measures of risk. Given an acceptable threshold for risk, methods deemed satisfactory are compared using common utility measures. Of the remaining methods, only those preserving enough utility are considered as viable protection methods.

1 Introduction

Population statistics is disseminated using multiple overlapping grid systems at various resolutions: Statistics Norway uses a national grid system, whereas another grid system is used for European census delivery. This provides multiple challenges with regards to disclosure control. The official Eurostat recommendation for the 2021 publication is a combination of targeted record swapping (TRS) along with the cell-key method (CKM). The former identifies records deemed to have a high risk of disclosure and swaps them with similar records from nearby regions, whereas the latter is a post-tabular perturbation method designed to handle such differencing attacks.

In the 2011 population census publications, Statistics Norway employed a rounding procedure described by Heldal (2017). Since then, the algorithm has been improved upon by Langsrud and Heldal (2018), and was subsequently named *small count rounding*. This method bears a resemblance to methods used by others in previous publications (for example, the UK in 2001 as described in a report by Spicer (2021)), but promises to address many of the complaints users had: specifically, small count rounding maintains additivity in a way that attempts to minimize information loss.

As part of a grant on multi-grid geographical data, Statistics Norway wished to compare the Eurostat recommendation of the combination of CKM and TRS with the small count rounding method. In order to make a rigorous comparison, we employed a comparison framework designed to compare how each method performs with respect to different kinds of disclosure. This paper presents that framework, as well as a brief summary of the results of the evaluation. Finally, we discuss ways in which the framework could be improved upon.

2 Comparison Framework

The comparison framework we present is based on common practice in statistical disclosure control: maximize utility given an acceptable level of risk. It consists of two parts: a suite of disclosure scenarios used to quantify risk, and measures for measuring loss of utility.

In general, the disclosure scenarios one chooses for the comparison may vary depending on the needs of the entity publishing the statistics. Within this framework, the only requirement posed to the disclosure scenarios are that they must be “countable": for a given disclosure scenario and a data set, one must be able to count the number of disclosures.

In the following section, we present four disclosure scenarios that capture different flavors of classical disclosures. In particular, the presented scenarios were used to evaluate different protection methods for the dissemination of the 2021 population census.

2.1 Disclosure Scenarios

We adopt a disclosure-centric approach to measuring risk. In particular, we consider the following four types of disclosure:

Ordinary attribute disclosure: When all records in a table marginal share the same attribute for a given variable, group attribute disclosure occurs. This is the case when there is only one category with a non-zero frequency within a marginal, and thus the exact category membership can be revealed. Then the cell with the non-zero frequency is considered disclosive.

Attribute disclosure when the original total is 1: Similar to ordinary attribute disclosure, but limited to marginal cells where the population total is 1. This is of particular interest in sparsely populated countries such as Norway, because one can often assume it to be known that the population total in a grid cell is 1.

Negative attribute disclosure: When no record contributing to a marginal cell has a certain attribute, negative attribute disclosure occurs. Any frequency that is zero is disclosive in the sense that no record can have that category. When the frequencies in all categories are zero, the zeros are no longer

Table 1. Example frequency table illustrating different disclosure scenarios. The first row exhibits ordinary attribute disclosure: all units in M1 are unemployed. The second row demonstrates negative attribute disclosure; no units in M2 are self-employed (indeed, all zero cells represent negative disclosures). Additionally, all non-zero cells are existence disclosures.

Municipality	**Unemployed**	**Employed**	**Self-employed**	**Total**
M1	12	0	0	12
M2	5	6	0	11

considered disclosive. Note that for two-level categorical variables the measurement of ordinary and negative attribute disclosure will in practice be the same.

Disclosure of existence: Any non-zero frequency discloses that at least one record has a certain attribute. That is, all cells with non-zero frequencies are considered disclosive.

In related work, Geyer et al. (2022) compare multiple methods based on the preservation of singletons, i.e., frequencies of 1, due in part to the increased risk of identification, but also the perceived feeling of vulnerability a unit might have even in cases where the attribute disclosure is incorrect. This approach is not a suitable measure for our study, given the methods and parameters: none of the methods allow for publication of frequencies less than 3, and hence no singleton cells are preserved. Rather, our study focuses on the possibility of actual disclosures (an attacker disclosing information about a different unit) as opposed to perceived disclosures (a unit being able to identify themselves in a data set), and the extent to which the different methods preserve real disclosures for each of the above disclosure types.

The above disclosure scenarios are intentionally broad: indeed, it is, realistically, far too restrictive to require no possibility of disclosure of any kind. The intention is not to ensure prevention of each of the disclosure scenarios, but rather the ability to measure the performance of protection methods in different situations. This provides a solid basis for deciding which method is best suited to the given publication. For example, though we considered and measured all of the above scenarios for the evaluation of the 2021 population census, the first two scenarios (attribute disclosure, and attribute disclosure where the original total is 1) were deemed far more important to protect against than the other two.

2.2 Measuring risk

A common framework can be used to assess all these types of disclosure risk, based on measures used in information retrieval and machine learning: *precision* and *recall*.

For each disclosure type as discussed in the previous section, all cells to be published can either be marked as disclosive or non-disclosive. Then we count the number of disclosive cells in the original and the perturbed data.

$$\begin{aligned} a &= \#\text{disclosive cells in original data} \\ b &= \#\text{disclosive cells in perturbed data} \\ c &= \#\text{common disclosive cells} \end{aligned}$$

That is, c is the number of disclosive cells in the original data that are still disclosive after perturbation. With these counts we can calculate precision and recall for each method as follows:

$$\begin{aligned} \text{precision} &= c/b \\ \text{recall} &= c/a \end{aligned}$$

These measures provide two different views on the protection provided by the perturbation methods considered. Intuitively, precision provides a measure for how many of the disclosures in the perturbed data set are actual disclosures, whereas recall provides a measure for how many of the disclosures in the original data set are

preserved. Thus we use these measures as the primary means of measuring the risk for each method and disclosure type.
Occasionally, we limit the calculation to selected cells of interest. For instance, we may look at certain categories that are considered more sensitive than others. Likewise, we can limit the calculations based on cell frequencies in the original or the perturbed data. However, in this case one must then keep in mind some of the measures become degenerate.

2.3 Measuring Utility

Given an acceptable level of risk, we wish to choose the method that provides the highest utility. High utility in this context means the same as low information loss. A measure of utility is therefore also a measure of information loss. In this paper we consider three measures of utility loss:

$$\text{Maximum absolute deviation} = \max_{i=1}^{n} |y_i^* - y_i|$$

$$\text{Average absolute deviation} = \frac{1}{n}\sum_{i=1}^{n} |y_i^* - y_i|$$

$$\text{Hellinger distance} = \sqrt{\frac{1}{2}\sum_{i=1}^{n}\left(\sqrt{y_i^*} - \sqrt{y_i}\right)^2}$$

Here, n is the total number of cells to be published, and the y_i's and y_i^*'s are the original and perturbed frequencies, respectively. The Hellinger distance is a common measure that provides a good overall assessment. The average absolute deviation is a number that is very easy to understand and interpret. By looking at that number, you get quick information about the degree of perturbation. In addition to overall measures, we will also make sure that there are no single deviations that are too large. Some large deviations may render the data essentially useless to some users. Therefore we also consider the maximum absolute deviation. In practice, we may also look more closely at several of the biggest deviations.

3 The Framework Applied: Evaluating 2021 Grid Data Protection Methods

The framework presented above was the basis for comparing multiple perturbative methods for the the dissemination of multi-grid data in the 2021 population census. In this section, we present a shortened summary of the findings. All the details and results will be presented in a future publication.
We begin by giving an overview over the methods and parameters used, before moving on to the evaluation.

3.1 The Perturbation Methods Considered

3.1.1 Cell Key Method. The cell-key method (CKM) introduced in Thompson et al. (2013) is a perturbative method which produces a cell's noise based on its contributing units: each record is stochastically assigned a *record key*, which are in turn used to determine a *cell key*. This cell key is used together with a reference noise table to determine a cell's noise. In this manner, two cells with exactly the same contributors are guaranteed to to be perturbed with the same noise. This is particularly beneficial for dynamic table generators such as the Australian Table Builder and the functionality in the microdata.no platform, where this feature ensures consistency of noise across multiple tables sharing cells.
CKM, in combination with targeted record swapping which we briefly discuss in the following section, are the current Eurostat recommendation for statistical disclosure control in the 2021 population census.

This method does not preserve table additivity, however it does provide a fixed bound on how much noise is added on the cell level. Extensions to the approach that preserve or fix table additivity do exist, though come with their own caveats: for example, increasing time complexity of the algorithm, or yielding non-integer frequency counts. Furthermore, the base version of the cell-key method does not perturb cells with no contributors (as the noise is dependent on its contributors). Again, extensions exist which take this into consideration, where cell keys also rely on categorical information relating to the cell (in addition to its contributing records). However, for the sake of the evaluation of the 2021 population census, we refer to the original method where cell keys are determined solely based on contributing record keys. Zero frequency cells are therefore not perturbed by CKM, and this must rather be handled by other means, such as targeted record swapping.

The evaluation is performed with noise ranging from -5 to 5, with a minimum allowed non-zero frequency of 3. The noise table is generated using the `ptable` R package (Enderle and Giessing, 2022).

3.1.2 Targeted Record Swapping. Targeted record swapping is a SDC method for protecting sets of microdata. This method aims at protecting locally unique records from disclosure by identifying unusual/unique records in a region and swapping them with similar records from neighboring regions. Common implementations of the technique rely on k-anonymity for determining similarity between records.

In our evaluation, targeted records are swapped with nearby regions according to the Norwegian national grid squares. To achieve this, we rely on the implementation in the `sdcMicro` R package (Templ et al., 2022). Our evaluation included both *random* and *targeted record swapping*. The former selects a percentage of records at random and swaps them with similar records in nearby regions (achieved by setting the `k_anonymity` parameter to 0). The latter identifies unique records and prioritizes them for swapping with similar records in nearby regions (achieved by setting the `k_anonymity` parameter to 2). In this paper, we only present the results for targeted record swapping, as it had overall better performance than random record swapping and is the official recommendation from Eurostat. Furthermore, the evaluation was run for both 1% and 10% swap rates as input parameters. However, the actual swap rates after running the method differed greatly from the input parameters: 22.8% and 25.3% respectively. Therefore for the sake of brevity, we present only the results for targeted record swapping with 1% swap rate in this paper.

3.1.3 Small Count Rounding. The small count rounding method (SCR) described in Langsrud and Heldal (2018) is a perturbation method aimed to produce consistent and additive frequency tables without small counts. The method is about changing frequencies of the inner cells, which are the microdata aggregated into frequencies. Identical rows in the microdata are replaced with a single row and a frequency value. A heuristic algorithm ensures good and fast solutions. The method is implemented in the R package `SmallCountRounding` (Langsrud and Heldal, 2022), which has been continuously updated with new functionality. In this paper we consider three variants of the method.

`SCRsimple`: This is the basic method with three as rounding base. Ones and twos in the published tables are avoided by changing a limited number of ones and twos in the inner cells to zeros and threes.

`SCRzeros`: This is similar to the method above (`SCRsimple`) an in addition inner cells with zero frequencies are treated as candidates to be rounded up. This way some zeros will be perturbed. The method requires that the inner cell data includes zeros. However, including all possible zeros is not feasible for this type of data. Instead, a limited number of random inner cells with zero frequency were added using the `Extend0` function provided by the `SSBtools` package (Langsrud and Lupp, 2022). Care was taken to avoid introducing structural zeros, such as impossible combinations of geographical areas.

`SCRforceInner`: This is similar to the method above (`SCRzeros`), but with the change that all inner cells are rounded. Thus, the feature that limits the number of inner cells to be rounded has been disabled.

Additionally, it is worth noting that in this study, we utilized the weight parameter of the algorithm. Small original frequencies were downweighted. Given the large size of the dataset, a special looping feature, known as the `PLSroundingLoop` function, was also employed.

TABLE 2. Attribute disclosure risk (ordinary and where original population is 1) measured as precision and recall.

	Precision		*Recall*	
Method	Ordinary	Total is 1	Ordinary	Total is 1
CK	61.09	100.00	67.50	26.27
CKswk2r01	42.88	58.36	42.38	58.04
SCRsimple	59.84	100.00	63.14	24.43
SCRzeros	55.95	85.05	57.54	23.97
SCRforceInner	52.92	79.48	53.15	22.48

3.2 Results

For each method and variant, we generate a perturbed data set containing all cells across all the different grid systems. For each of these data sets, we measure precision and recall according to the defined disclosure scenarios, as well as utility loss according to the measures defined in the previous section.

In order to determine the risk one considers acceptable, one must consider how to prioritize the scenarios and precision/recall measures. For this evaluation, we consider precision a more important metric than recall: intuitively, precision represents how certain an attacker can be about a disclosure. Clearly, the lower the risk measure, the better. Therefore, we automatically disqualify all methods that have 100% precision or recall: with 100% precision an attacker can know with certainty that a disclosure is real. With 100% recall an attacker would have access to all real disclosures. Though 100% recall can, in theory, be somewhat mitigated by low precision, for the sake of this evaluation we consider this problematic.

In the full evaluation, all disclosure scenarios were considered. Indeed, one can incorporate more granularity be considering certain variables or categories as sensitive. Then one can measure the risk associated to the disclosure scenarios for these. However, in the context of this article we present only the values that had the greatest impact, and illustrate the largest differences between the considered methods. Tables 2, 3, and 4 show the results of the evaluation.

Ordinary attribute disclosure: Precision performance with respect to ordinary attribute disclosure differed only slightly, as seen in Table 2. Considering this risk measure in isolation, the cell-key method combined with targeted record swapping (CKswk2r01) performed best, with a precision of 42.88%, whereas the cell-key method alone (CK) had the worst performance, with a precision score of 61.09%. All of these values can be considered acceptable, as an attacker can at most be approximately 61% sure that a disclosure is real.

Attribute disclosure where total is 1: Norway is a sparsely populated country. Therefore, many grid cells contains few people, and it is reasonable to assume that an attacker can know that a grid cell contains only one person. All methods that leave zeros unperturbed will result in 100% precision, which we deem unacceptable. This can be seen in Table 2, where the standard cell-key method (CK) as well as the simple small count rounding method (SCRsimple) have 100% precision. Of the remaining methods, a combination of the cell-key method with targeted record swapping performs best with precision at 58.36%, whereas the remaining variants of small count rounding (SCRzeros and SCRforceInner) have similar values at 85.05% and 79.48% respectively.

Disclosure of existence: Disclosure of existence is not about specific individuals, and thus in this context is only problematic when cell totals are low. This is a common occurrence is countries such as Norway. In Table 3, we see that if a grid cell's published frequency is 4 or 5 persons, the SCRzeros has 100% precision for disclosure of existence. As we deem this unacceptable for low frequency cells, we must exclude SCRzeros.

Two methods remain which have an acceptable level of risk: CKswk2r01 and SCRforceInner. We wish to determine which of the methods maintains the greatest utility. Table 4 summarizes the results of utility

Table 3. Risk of disclosure of existence measured as precision and limited to cases where the perturbed frequency has a specific value (1-6). Values less than 3 are not published, hence the first two columns are empty.

Method	1	2	3	4	5	6
CK			100.00	100.00	100.00	100.00
CKswk2r01			80.99	84.88	88.89	92.89
SCRsimple			100.00	100.00	100.00	100.00
SCRzeros			92.77	100.00	100.00	99.71
SCRforceInner			89.65			99.16

Table 4. Utility measures for each perturbative method.

Method	Maximum absolute deviation	Average absolute deviation	Hellinger distance
CK	5	0.989	886.6
CKswk2r01	13401	2.098	1245.8
SCRsimple	17	0.944	891.5
SCRzeros	15	1.008	968.6
SCRforceInner	19	1.304	1052.3

measurements. Here, CKswk2r01 performs considerably worse across the board as compared to the flavors of small count rounding, with a maximum absolute deviation of 13401, compared to 19 for SCRforceInner. Analyzing the underlying data more closely, and considering the deviation relative to the original frequency, the large deviation in CKswk2r01 is approximately 35% of the original value, which we deem unacceptable. Thus we are left with SCRforceInner as the preferred method.

4 Conclusion

In this paper, we present a general framework for comparing SDC methods for frequency table protection. It was applied on multi-grid publication of 2021 population census data in order to compare different flavors of the cell-key method, small count rounding, and targeted record swapping. The results indicate that, in this particular case in Norway, small count rounding where inner cells are forced to be rounded (SCRforceInner) is the preferred method.

The presented framework has focused on perturbative methods. However, in general the framework could be applied to comparing both non-perturbative and perturbative methods. The only condition posed to the disclosure scenarios used to measure risk are that one can count how many disclosures there are, something that is also possible for non-perturbative methods such as cell suppression. The main challenge when adapting the framework to include non-perturbative methods is in finding utility measures that are suitable for both non-perturbative and perturbative methods.

Furthermore, both the choice of utility measure and the choice of risk measure (precision and recall) can likely be fine-tuned or adapted. One could extend the utility measurements by including measures of information loss, for example Kullback-Leibler divergence or by measuring the variation of information. Regarding risk, there are multiple ways of combining precision and recall into a single measure, allowing for instance a prioritization of one over the other. This would have the benefit of a single value for comparison. However, we consciously decided not to do so, as both precision and recall provide different, orthogonal insights, and summarizing this into a single value appeared to obfuscate some of the nuance. Despite this decision, it is likely a fruitful direction of future research and experimentation.

Finally, the results of the evaluation should not be interpreted as a conclusive answer as to which method is best. The comparison framework in this paper is intended to illustrate the different effect various protection methods have given different situations. Including other methods and variants in the comparison, such as cell-key methods with perturbation of zeros or additivity modules, is an obvious candidate for future research.

References

Enderle, T. and S. Giessing (2022). *ptable: Generation of perturbation tables*. R package on github.com/tenderle/ptable Version 0.3.3.

Geyer, F., R. Tent, M. Reiffert, and S. Giessing (2022). Perspectives for Tabular Data Protection: How About Synthetic Data? In J. Domingo-Ferrer and M. Laurent (Eds.), *Privacy in Statistical Databases*, Volume 13463, pp. 77–91. Cham: Springer International Publishing. Series Title: Lecture Notes in Computer Science.

Heldal, J. (2017). The European Census Hub 2011 Hypercubes - Norwegian SDC Experiences. In *Work Session on Statistical Data Confidentiality*. Skopje, The former Yugoslav Republic of Macedonia, September 20-22 , 2017.

Langsrud, Ø. and J. Heldal (2018, 09). An algorithm for small count rounding of tabular data. Privacy in statistical databases, Valencia, Spain.

Langsrud, Ø. and J. Heldal (2022). *SmallCountRounding: Small Count Rounding of Tabular Data*. R package version 1.0.2.

Langsrud, Ø. and D. Lupp (2022). *SSBtools: Statistics Norway's Miscellaneous Tools*. R package version 1.3.4.

Spicer, K. (2021). Statistical Disclosure Control (SDC) for 2021 UK Census. In *https://uksa.statisticsauthority.gov.uk/wp-content/uploads/2020/07/EAP125-Statistical-Disclosure-Control-SDC-for-2021-UK-Census.docx*.

Templ, M., B. Meindl, A. Kowarik, and J. Gussenbauer (2022). *sdcMicro: Statistical Disclosure Control Methods for Anonymization of Data and Risk Estimation*. R package version 5.7.4.

Thompson, G., S. Broadfoot, and D. Elazar (2013). Methodology for the automatic confidentialisation of statistical outputs from remote servers at the Australian Bureau of Statistics. Joint UNECE/Eurostat Work Session on Statistical Data.