Implementing Multiple Evaluation Techniques in Statistical Disclosure Control for Tabular Data

Amang Sukasih, Donsig Jang, John Czajka
Mathematica Policy Research
1100 1st Street, NE, 12th Floor, Washington, DC 20002-4221

Abstract

Statistical agencies that disseminate their data to the public often require the implementation of Statistical Disclosure Control (SDC) to protect their data. For tabular data, the procedures include identifying sensitive cells and applying SDC techniques to protect the table. Several methods exist for both identifying sensitive cells and protecting the table cells and agencies may implement different methods depending on each situation and need. This paper will identify each of these methods and present techniques that can be used in evaluating protected tables prior to publishing them.

Key words: Disclosure risk; sensitivity rules; cell suppression; concentration rules; optimization.

1. Disclosure Limitation in Tabular Data

When statistical data are disseminated either in the form of tabular data or public use microdata, the data producer often needs to protect the confidentiality of the respondents who provided the information. Confidential information may include identity of the respondents as well as information about them. The Federal Committee on Statistical Methodology (FCSM), in their Statistical Policy Working Paper 22 (2005), summarized three types of data disclosure discussed in Duncan et al. (1993, pp. 23-24) as follows:

"Disclosure relates to inappropriate attribution of information to a data subject, whether an individual or an organization. Disclosure occurs when a data subject is identified from a released file (identity disclosure), sensitive information about a data subject is revealed through the released file (attribute disclosure), or the released data make it possible to determine the value of some characteristic of an individual more accurately than otherwise would have been possible (inferential disclosure)."

To avoid such disclosure, data producers develop rules and procedures to protect confidentiality, and implement these rules to their tables or microdata files prior to publishing the tables or releasing the data. Confidentiality protection rules may vary from data to data and from agency to agency—or even from table to table within the same agency or data source.

This paper focuses exclusively on disclosure limitation for tabular data (not microdata). We will also discuss some SDC techniques that include specifications to identify sensitive cells and how to protect those sensitive cells as well as provide examples using

fictitious tables. Note that the term "statistical disclosure limitation" or "statistical disclosure avoidance" appears in the literature as well.

1.1. Identifying Sensitive Cells

To protect the confidentiality of individual respondent information in tabulation data, data producers have to first identify potentially unsafe or sensitive cells in the tables and then protect these same cells. In the SDC framework, these sensitive cells are called primary cells and common primary cells are usually those with small sample sizes. Such cells will have higher disclosure risk than cells with larger sample size. Therefore, one of the most widely used techniques to identify primary cells is to apply a threshold rule to the cell frequencies. If a cell frequency—that is, the sample count or in some applications the corresponding population estimate—is below this threshold, then the cell is deemed sensitive.

Other techniques, which do not depend on the number of respondents in a cell, include identifying the contribution of individual magnitude data to the aggregate value within the cell. For example, the (n,k) rule or dominance rule identifies a sensitive cell as a one with a small number (n or fewer) of respondents contributing a large percentage (k percent or more) of the total cell magnitude. Other rules, such as the *p*-percent and the *pq* rules, have been developed to address the disclosure risks that arise when an intruder knows the value of an individual contribution (typically the largest or second largest) and can use this information in combination with a reported total to estimate the value reported by another contributor with a certain degree of precision. In the p-percent rule, if the second largest contributor can guess the largest contribution to the cell by subtracting his or her own value from the cell total, and the resulting value is smaller than $(100 + p)$ percent of the largest contribution, then the cell is considered sensitive. The *pq* rule is an extension of the p-percent rule, and takes into account that the second largest contributor may be able to estimate from external sources the aggregate contribution of the smaller contributors, thereby allowing an even closer estimate of the largest contributor's contribution. Table 1 summarizes the most common rules used to identify sensitive cells. Suppose a cell value X composed of N contributors denoted by $x_1, x_2, \ldots, x_N$ sorted from the largest to the smallest, where $X = x_1 + x_2 + \ldots + x_N$.

Table 1: The Most Common Rules to Identify Sensitive Cells

Rule	Definition of sensitive cell
Threshold or minimum frequency rule	Cell frequency is smaller than threshold value=
(*n*,*k*) or dominance rule	The sum of *n* largest values is greater than (*k*/100)% of the cell total.
p% rule	Cell total minus the two largest contributions is smaller than (*p*/100)% of the largest contribution.
pq rule	(*q*/100)% of the cell total minus the two largest contributions is smaller than (*p*/100)% of the largest contribution. Note: *p*% rule is a special case of *pq* rule with *q*=100%.

Once sensitive cells are identified, there are approaches to protect their contents, including redesigning the table to combine categories (collapsing cells), suppressing the cells, and rounding or perturbing the cell values. Another approach is to make changes to the microdata—for example, some form of perturbation of individual data values or swapping data between observations in different domains—prior to tabulation. The objective of all of these approaches is to eliminate sensitive cells.

1.2. Protecting Sensitive Cells

Once the table format has been fixed and the cell values tabulated (with no further table redesign, recoding of categories, or collapsing of cells), and the sensitive cells have been identified, the table can be protected by implementing the SDC techniques, including perturbation, cell suppression, or control tabular adjustment.

1.2.1. Perturbation

In these techniques, the true cell values are protected by either rounding (up or down) the cell values to a specific base, or perturbing the cell values by adding or multiplying with some chosen value. The goal of protection is that the cell can still be published but the intruder no longer finds the true value in the published table. This paper will not discuss methods in this group. Readers can see Federal Committee on Statistical Methodology (2005) and Hundepool et al. (2010) for more details in this area.

1.2.2. Cell Suppression

In this method, sensitive cells are simply dropped/suppressed (not published) to protect confidentiality. Cells identified as sensitive based on the sensitivity rules discussed previously and then dropped are called the primary cells. However, simply dropping the values of the sensitive cells will not completely protect them when marginal totals of these cells are published, because an intruder may recalculate the dropped values by way of simple subtraction. Therefore, to completely protect sensitive cells, one or more nonsensitive cells (called secondary or complementary cells) must be suppressed as well. The most common way to do this is that for each primary suppressed cell, there should be at least one secondary suppressed cell in the same row and one secondary suppressed cell in the same column. Note, however, that for each suppressed primary cell, there are many possible choices of secondary cells. Also, it may still be possible for the intruder to compute a range (feasibility or sensitivity interval) in which the suppressed cells lie. This is motivation to find secondary cells that maximize disclosure limitation and minimize information loss. The method to address this objective becomes more complicated and involves solving linear programming (LP) problems. Two common methods for secondary cell suppression are discussed below.

1.2.2.1. Hypercube Method

For an *n*-dimensional table with hierarchical structure, this method subdivides the table into a set of *n*-dimensional sub-tables without substructure. For each of these simple tables without hierarchical structure, if we consider secondary cell suppression where in each row and in each column there has to be exactly one secondary suppressed cell, nevertheless, there are still many possible patterns of secondary suppressed cells. The

SDC task is then to check whether the sensitivity interval is wide enough and calculate the loss of information for each pattern of secondary cell suppression.

Successively, for each primary suppression in the current sub-table, all possible hypercubes with this cell as one of the corner points are constructed. A cell in a simple *n*-dimensional table without substructure cannot be disclosed exactly if the cell is contained in a pattern of suppressed, nonzero cells, forming the corner of a hypercube. By solving LP problems, the suppression can choose a secondary cell suppression pattern that optimizes sensitivity interval and loss of information constraints. A heuristic approach that does not need LP optimization can be used; the computation can be done by generating all candidates of *n*-dimensional hypercubes and selecting the one with minimum loss of information. Willenborg and de Waal (1996) provide detailed information on how the hypercube method for secondary cell suppression works.

1.2.2.2. Modular/HiTaS

This technique is also a heuristic approach that implements LP optimization to choose secondary cells. Such a procedure breaks down the hierarchical table into several non-hierarchical tables, protects them using LP-solver, and then composes a protected table from the smaller tables. Detailed information on how this method works can be found in Hundepool et al. (2011) and de Wolf (2002).

2. Choosing Disclosure Control Method

2.1. Agency Practices

SDC procedures are designed with the intention of ensuring that the risk of disclosing confidential information about identifiable individuals, businesses, or other units is very small. Through the application of such procedures, federal agencies and their contractors who release statistical tables or microdata files achieve the protection required either by law or agency policy. Appendix Table A1presents practices applied to tabular data (not necessarily for businesses or corporations) by selected federal statistical agencies as reported in the "Federal Committee on Statistical Methodology, Statistical Policy Working Paper 22" (December 2005). Not only are the methods implemented to identify sensitive cells different across agencies, but parameter values for the same method are also different across agencies. For example, several agencies use the threshold method. Some agencies use the number of units 3 as the threshold value, while others used 4, 5, or 10, and some agencies do not disclose this parameter.

To protect sensitive cells, most agencies perform cell suppression. In addition, most agencies also suppress additional/secondary cells to avoid reconstruction of primary cells by arithmetic calculation. However, there are variations in how the agencies perform the secondary cell suppression. Some agencies implement a simple rule; for example, by simply selecting the smallest nonzero cell among those available, while others implement more systematic secondary cell suppression; for example based on linear programming.

The "FCSM Statistical Policy Working Paper 22" (2005) provides recommended practices for federal agencies in performing SDC for tabular and microdata. Specifically for tables of frequency count data, the FCSM working paper recommends that the entity

producing the data research several methods available to compare and evaluate these methods in terms of data protection and usefulness of the resulting data product. The paper also recommends not revealing suppression parameters to the public. Also, the data producer may redesign its tables by combining categories or collapsing cells, controlling tabular adjustment, or applying cell suppression or perturbation methods to the microdata prior to tabulation.

For tables of magnitude data, the paper recommends using only subadditive disclosure rules (*p*-percent, *pq*, *n* threshold, and (*n*, *k*) rules) for defining sensitive cells, where the *p*-percent or *pq*-ambiguity rules are preferred. In addition, as in protecting tables of frequency count data, for tables of magnitude data, it is not recommended to reveal suppression parameters to the public; and the data producer may also redesign their tables by combining categories or collapsing cells, applying cell suppression, controlled tabular adjustment, or perturbation methods to the microdata prior to tabulation. Lastly, applying cell suppression and auditing of tabular data is a necessity (we will discuss this in the next section).

2.2. Empirical Examples

Below we present a fictitious table taken from Daalmans and de Waal (2010) that presents magnitude data parsed by two variables, say sector and establishment size. In practice, the sector variable can be code from industry classification system or geography, and the total row for sector variable may be a higher level in code system hierarchy. The establishment size can be based on asset, number of employees, or other size variable.

Table 2: Company Total Assets by Company Size and Sector (Unprotected Table)

Sector	Establishment size			Total
	1	*2*	*3*	
a	160	380	340	880
b	40	80	60	180
c	610	800	270	1,680
Total	810	1,260	670	2,740

Suppose cell *a1* (sector *a*, size group 1, with cell value 160) contains three establishments with individual magnitude data 155, 4, and 1, respectively. Under threshold 3+ rule, cell *a1* is not considered as sensitive and can be published. However, under the dominance/(*n*,*k*) rule with *n*=1, *k*=60%, cell *a1* is sensitive, as well as under the *p*-percent rule with *p*=20%, and this cell needs to be suppressed. In this cell, if one of smaller contributors to the cell or a coalition of smaller contributors subtracts themselves from the cell total, then the magnitude data of the largest contributor can be estimated with certain precision.

Given cell *a1* is being suppressed by simply dropping the cell value from the published table (primary suppression), other cells (at least another cell within the same row, and another cell within the same column) need to be suppressed as well to avoid recalculation

of the primary suppressed cell by subtraction. An example of protected table could be as shown in Table 3:

Table 3: Company Total Assets by Company Size and Sector (Protected Table)

Sector	Establishment size			Total
	1	*2*	*3*	
a	P	C	340	880
b	C	C	60	180
c	610	800	270	1,680
Total	810	1,260	670	2,740

P = Primary suppressed cell
C = Complementary suppressed cell

3. Evaluating Protected Table

After the table producer evaluates several alternatives of SDC methods for their tables and decides and finalizes the protected tables, it is recommended to evaluate and run an audit for these protected tables prior to publishing the tables. In this paper, we discuss the following evaluation methods: (a) evaluating information loss, (b) running other SDC methods utilizing suppression history, (c) auditing individual suppressed cell using sensitivity intervals, (d) auditing aggregation of suppressed cells, and (e) utilizing risk measure based on relative contribution.

3.1. Evaluating Information Loss

A practical consideration in releasing a protected table would be to balance data confidentiality and data quality. This is always a trade off faced in the SDC area, where overprotection could lead to higher loss of information. On the other hand, the use of less suppression to avoid too much loss of information may widen the room for disclosure risks. In evaluating the quality of the published table, the table producer can approach this task from the estimation point of view; that is, by evaluating aggregate suppressed data (magnitude and/or the frequencies) relative to the population or original (unprotected) table. In addition, the evaluation can be carried out by comparing loss of information resulted from using several different SDC methods. A rule can be set up based on, for example, a minimum number of suppressed cells or a minimum total cell values suppressed. In this case, a method that provides good balance between data confidentiality and data quality may be chosen. For the discussion in this area, the reader can refer to Duncan et al. (2001).

3.2. Auditing Individual Cell Using Sensitivity Intervals

When a protected table is being released, an intruder can still make a guess of the suppressed cell value by constructing bounds (called sensitivity interval) for each suppressed cell. Therefore, the protection should have a property that the bounds for the sensitivity interval of any sensitive cell cannot be used to deduce an individual respondent contribution too closely according to the sensitivity rule employed. In

evaluating the table, the first step in auditing the individual suppressed cell is to construct the sensitivity interval, which represents the upper and lower bounds of a cell's true value. Based on linear combination of published cells, we can set linear equations. Then, the sensitivity intervals are obtained by solving linear programming of constraint equations. For the example of suppression in Table 3, the linear equation and constrains are as follows:

Objective: minimize and maximize $a1$, $a2$, $b1$, $b2$ subject to

$$a1 + a2 = 540$$
$$a1 + b1 = 200$$
$$a2 + b2 = 460$$
$$b1 + b2 = 120$$
$$a1, a2, b1, b2 \geq 0$$

The resulted sensitivity intervals for the suppressed cells $a1$, $a2$, $b1$, and $b2$ are, respectively, $80 \leq a1 \leq 200$, $40 \leq a2 \leq 460$, $0 \leq b1 \leq 120$, $0 \leq b2 \leq 120$. To decide whether a suppressed cell is safe or not, we compare the sensitivity interval to the protection level interval (Hundepool et al. 2010) that is calculated according to the sensitivity rule used. The protection level measures the safety bounds provided by the sensitivity rule implemented. The formula for each sensitivity rule is given in Table 4.

Table 4: Sensitivity Rule and Its Upper Protection Level

Sensitivity rule	Upper protection level
$(1,k)$	$100/k * x_1 - X$
(n,k)	$100/k * (x_1 + x_2 + \ldots + x_n) - X$
$p\%$	$p/100 * x_1 - (X - x_1 - x_2)$
Pq	$p/q * x_1 - (X - x_1 - x_2)$

If the distance between the upper bound of the feasibility interval and the true value of a sensitive cell is below the upper protection level computed according to the formulas in the above table, then this upper bound could be used to estimate individual contributions of the sensitive cell too closely according to the safety rule. For example, for cell $a1$ that is protected under the 20 percent rule, the upper protection level = (0.2 * 155) – 1 = 30, while distance of the upper bound sensitivity interval to $X = 200 - 160 = 40$. Because the distance of the upper bound sensitivity interval to the cell value X is larger than the upper protection level, the suppressed cell $a1$ is therefore considered safe.

The Disclosure Audit Software (DAS), which is available for constructing sensitivity intervals, was developed through the inter-agency effort of the FCSM in collaboration with Confidentiality and Data Access Committee (CDAC) (www.fcsm.gov/committees/cdac/DAS.html). Among its various features, DAS can audit a table of up to five dimensions and run under the SAS system. However, to run DAS, the computer needs to have special SAS modules: SAS/ACCESS to PCFF (PC File Formats), SAS/Connect, SAS/OR, and SAS/FSP, which may not be available under

standard SAS. A table producer who has access to linear programming solver such as *LPSolve* package in *R* software can program to calculate sensitivity intervals using *R*.

3.3. Auditing Individual Cell Using Other Methods Utilizing Suppression History

The evaluation in Section 3.a (Evaluating Information Loss) is used to compare results by utilizing several different sensitivity rules independently. Another possible evaluation is to run a technique that combines several methods by running these different methods consecutively. That is, one can protect a table using the first sensitivity rule chosen, then he or she can run another (different) method on top of this protected table by keeping the cell suppression history as the input in running the second method. To perform this kind of evaluation, the table producer may need to use software that can take cell protection history as input, such as Tau-Argus (Hundepool et al. 2011).

3.4. Auditing Aggregation of Suppressed Cells

Daalmans and de Waal (2010) demonstrated that disclosure auditing based on *pq* rule may not be sufficient when aggregations of suppressed cells still have disclosure risk. For example, in Table 3, suppose cell *b1* ($X = 40$) consists of $x_1 = 28$, $x_2 = 10$, $x_3 = 2$. Recall that in this table, cell *a1* ($X = 160$) is the primary sensitive cell under the 20 percent rule with $x_1 = 155$, $x_2 = 4$, $x_3 = 1$; and cells *a2*, *b1*, *b2* are secondary (non-sensitive) cells, and these four cells have been suppressed for protection. An intruder can combine/merge suppressed cells to produce aggregate cells that reveal cell total (because marginal totals are given). Now, if an intruder combined the first two rows (sector *a* and *b*) into one row (as shown in Table 5) and happens to know the value of second largest contributor in cell *ab1* (which is 28) and uses this value to guess the largest contributor, then he or she will get $200 - 28 = 172$, which is 11 percent of 155 (largest contributor). So, Table 3 is still not safe because, under the 20 percent sensitivity rule, the intruder still has the ability to come too closely to guessing the largest contributor.

Table 5: Company Total Assets by Company Size and Sector (Rolled Protected Table)

Sector	Establishment size			Total
	1	*2*	*3*	
ab	S	S	400	1,060
c	610	800	270	1,680
Total	810	1,260	670	2,740

S = suppressed cell

Daalmans and de Waal (2010) developed theorems for table auditing based on aggregation of suppressed cells:

> "If all contributions to an aggregation cell are sufficiently protected according to the *pq* rule, all contributions to individual cell values involved in that particular aggregation cell are also sufficiently protected. Aggregations that only involve non-sensitive cells are non-sensitive."

Daalmans and de Waal (2010) determined operational criteria that a table is safe if and only if all aggregations of suppressed cells are safe, on the basis of the same sensitivity rule that is applied to separate cells. As a consequence, for disclosure auditing, it is not necessary to apply a sensitivity measure to individual contributions to cells involved in aggregations. In addition, it is not necessary to check whether all possible aggregations are sensitive or not; the table producer has to consider only the aggregates that involve at least one sensitive cell. Based on this, then the table producer just needs to determine the most sensitive aggregation. Then, if that aggregation is safe, the table is safe. This can reduce the time for auditing; however, it requires a complex computation, which is solving a Mixed-Integer Programming (MIP) problem (see Daalmans and de Waal 2010).

3.5. Evaluation Using Conditional Entropy

Domingo-Ferrer and Torra (2002), through the following counter example, showed that sensitivity rules based on concentration such as (*n*,*k*), *p*% and *pq* rules, have contradictive behavior: a non-sensitive cell has more disclosure risk than a sensitive cell. Suppose the agency implements the sensitivity rule (*n*=1, *k*=60%). Under this sensitivity rule, a sensitive cell with total value 100 containing contributors $x_1 = 61$, $x_2 = 20$, $x_3 = 19$ is deemed sensitive. Another cell with a total value of 100 containing contributors $x_1 = 59$, $x_2 = 40$, $x_3 = 1$ is deemed as a non-sensitive cell. If the second largest respondent knows the total cell 100, and is interested in estimating the contribution of the largest respondent, the estimate in the sensitive cell is $100 - 20 = 80$, which is within 31 percent; and in the non-sensitive cell is $100 - 40 = 60$, which is within 1.7% (much closer). This is the flaw that the cell declared non-sensitive by the rule that allows better inference than the cell declared sensitive.

Domingo-Ferrer and Torra (2002) and the papers cited therein showed that a disclosure risk measure based on relative contribution, such as the use of conditional entropy that measures the concentration of contributions, can properly measure a disclosure risk, in the sense that the value of measurement for a non-sensitive cell is smaller than that for a sensitive cell.

4. Conclusion

It is necessary to evaluate protected tables before publishing them. This paper discusses several techniques to evaluate protected tabular data to ensure that the tables are sufficiently protected before releasing them to the public. In general, the agency producing tables should consider several different disclosure risks and intruder scenarios to test their protection. The best practice is to implement several different sensitivity rules and evaluation methods. Yet, some of them may not be easily available due to computational challenges, especially for a large tabulation system where the computational system (hardware and software) has been established and changes will potentially impact production process. Nevertheless, the agency may choose technique(s) that can be easily integrated with their current system.

Appendix

Table A1: Statistical Disclosure Control Practices by Federal Statistical Agency in the USA

Agency	Magnitude Data	Frequency Data
Economic Research Service	(n, k), (1,.6) 3+	Threshold Rule 3+
National Agricultural Statistics Service	(n, k), p-percent, Parameters Confidential	1+ Not Sensitive for Est. Surveys
Bureau of Economic Analysis	p-percent	1+ Not Sensitive for Est. Surveys
Bureau of the Census	p-percent, Parameters Confidential, Noise Addition	Data Swapping, Access Query System Rules, Threshold Rule
National Center for Education Statistics	Data Swapping, Data Coarsening, Accuracy Standards/Threshold Rule 3+	Data Swapping, Data Coarsening, Accuracy Standards/Threshold Rule 3+
Energy Information Administration	(n, k), pq, Parameters Confidential	Threshold Rule, Accuracy Standards
National Center for Health Statistics	(n, k), (1,.6)	Threshold Rule 4+
Agency for Healthcare Research & Quality	N/A	Threshold Rule 4+
Social Security Administration	Threshold Rule 3+	Threshold Rule, 5+ Marginals, 3+ Cells
Bureau of Justice Statistics	N/A	Threshold Rule 10+, Accuracy Standards
Bureau of Labor Statistics	(n, k), p% rule, Parameters vary by survey and data element	Minimum Number Varies by Survey
Internal Revenue Service	Threshold Rule 3+	Threshold Rule 3+
Bureau of Transportation Statistics	Varies by Data	Threshold Rule 3+
National Science Foundation	(n, k) and/or p as Appropriate	Varies by Risk

Source: Federal Committee on Statistical Methodology, Statistical Policy Working Paper 22, December 2005

References

Daalmans, J. and de Waal, T. (2010). An Improved Formulation of the Disclosure Auditing Problem for Secondary Cell Suppression. Transactions on Data Privacy, 3, 217-251.

de Wolf, P.P. (2002). HiTaS: A Heuristic Approach to Cell Suppression in Hierarchical Tables. In Inference Control in Statistical Databases, J. Domingo-Ferrer (editor). Springer-Verlag Berlin Heidelberg, 2316, 81-98.

Domingo-Ferrer, J. and Torra, V. (2002). A Critique of The Sensitivity Rules Usually Employed for Statistical Table Protection. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10, 5, 545-556.

Duncan, G.T., Jabine, T.B., and deWolf, V.A. (1993). Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics. Committee on National Statistics and the Social Science Research Council, National Academy Press, Washington, D.C.

Duncan, G.T., Fienberg, S.E., Krishnan, R., Padman, R., and Roehrig, S.F. (2001). Disclosure Limitation Methods and Information Loss for Tabular Data. In Confidentiality, Disclosure, and Data Access: Theory and Practical Application for Statistical Agencies, edited by P. Doyle, J.I. Lane, J.J.M. Theeuwes, and L.V. Zayatz. New York: Elsevier, 135-166.

Federal Committee on Statistical Methodology. (2005). Statistical Policy Working Paper 22 (second version 2005). Report on Statistical Disclosure Limitation Methodology, Statistical and Science Policy, Office of Information and Regulatory Affairs, Office of Management and Budget.

Hundepool, A., Van de Wetering, A., Ramaswamy, R., de Wolf, P.P., Giessing, S., Fischetti, M., Salazar, J.-J., Castro, J., and Lowthian, P. (April 2011). τ-ARGUS User's Manual, Version 3.5.

Hundepool, A., Domingo-Ferrer, J., Franconi, L. Giessing, S., Lenz, R., Naylor, J., Nordholt, E.S., Seri, G., and de Wolf, P.P. (January 2010). Handbook on Statistical Disclosure Control. A Network Excellence in the European Statistical System in the Field of Statistical Disclosure Control (ESSNet SDC).

Oganian, A. and Domingo-Ferrer, J. (2003). A Posteriori Disclosure Risk Measure for Tabular Data Based on Conditional Entropy. SORT–Statistics and Operations Research Transactions, 27, 175-190.

Willenborg, L. and De Waal, T. 1996. Statistical Disclosure Control in Practice. Springer Lecture Notes in Statistics. 111.