Comparison of Three Post-tabular Confidentiality Approaches for Survey Weighted Frequency Tables

Natalie Shlomo*, Thomas Krenzke** and Jianzhu Li**
* Social Statistics Department, University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom
E-mail Natalie.Shlomo@manchester.ac.uk
** Westat, 1600 Research Boulevard, Rockville, MD 20850, United States
E-mail TomKrenzke@westat.com JaneLi@westat.com

Received 2 September 2018; received in revised form 18 April 2019 and 5 August 2019; accepted 12 September 2019

Abstract. One of the most common forms of data release by National Statistical Institutes (NSIs) are frequency tables arising from censuses and surveys and these have been the focus of statistical disclosure limitation (SDL) techniques for decades. With the need to modernize dissemination strategies, NSIs are considering web-based flexible table builders where users can generate their own tables of interest without the need for human intervention. This has led to a shift in traditional disclosure risks of concern and a move towards inferential disclosure risk where statistical data can be manipulated and combined with other data sources to reveal sensitive information with a high degree of certainty. To protect against inferential disclosure risk, perturbative methods with more formal privacy guarantees are necessary. We examine three post-tabular confidentiality protection methods of additive random noise that can easily be applied 'on-the-fly' in a flexible table builder for generating survey weighted frequency tables: the computer science approach guaranteeing a formal privacy model called differential privacy and two SDL approaches of post-randomization and a new technique called drop/add-up-to-q. We demonstrate and compare their application in a simulation study based on survey weighted counts in tables.
Keywords: Flexible Table Builder, Survey data, Perturbation, Differential Privacy, Post-randomization

1 Introduction

National Statistical Institutes (NSIs) have been releasing tabular data arising from censuses and surveys for decades. Traditionally, the disclosure risks of concern were identity disclosure arising from small cell counts in the tables and attribute disclosure where a row/column have many zeros and only one non-zero cell count. This latter disclosure risk means that an intruder can learn something new about an individual or a group of individuals on a particular category of a variable based on an identification from the other variables defining the table. For tabular data, most of the statistical disclosure limitation (SDL) research has been based on census (whole-population) frequency tables. This is because tables generated from survey microdata have an additional layer of protection due to the random sampling under the assumption that the intruder does not have response knowledge, i.e. does not know who is selected and responded to the survey. In particular, sampling leads to uncertainty about whether zeroes that appear in the tables are structural or random and this reduces the risk of attribute disclosure.

With increasing demand for more open data and multiple releases of data products from a given dataset, this has led to growing concerns about disclosure risks that cannot be easily managed. In particular, NSIs are increasingly worried about inferential disclosure where an intruder can learn new information about individuals or a group of individuals to a high degree of certainty. For example, with multiple releases of data products from a single dataset, such as tabular data from restricted files and the release of a public-use file, intruders can manipulate and link information from the different products to reveal sensitive information. In terms of frequency tables, users can difference tables that individually may have no apparent disclosure risks but the differenced table may have high disclosure risks due to small and zero cell counts. Inferential disclosure subsumes all other disclosure risks and is becoming more of a concern for NSIs. As a consequence, they have moved to placing more controls on data access through controlled release of tabular data and licensing microdata to registered users. This has led to a shift from the SDL principle of 'safe data' to 'safe access'. However, placing more restrictions on data access is in direct contrast to government initiatives for more open and accessible data.

With demands for NSIs to release more open data and modernize their dissemination strategies, we focus in this paper on a web-based flexible table builder. Users can generate tables of interest through a user-friendly interface and define a table of interest from a set of pre-defined variables/categories using drop down lists. They can then download the table directly to their own personal computer without the need for human intervention in the process. The Australian Bureau of Statistics (ABS), Eurostat and the US Census Bureau have implemented table builders for their census dissemination although they differ in how they are set up with respect to whether pre- and/or post-tabular SDL methods are used. For example, there are general 'rules of thumb' applied in a flexible table builder with respect to minimum cell values (eg. application of k-anonymity and associated l-diversity and t-closeness rules), minimum population thresholds, the number of dimensions in a table, etc. Shlomo, et al. (2015) recognize the potential for inferential disclosure in flexible table builders and established that perturbative disclosure limitation methods are needed to protect the confidentiality of data subjects from inferential disclosure. They also found that a post-tabular protection procedure applied directly on the generated table, as opposed to a pre-tabular protection of the underlying microdata prior to generating the table, improves the utility of the data.

The research on flexible table builders have led to increasing exploration on whether the privacy guarantees offered by the definition of differential privacy (DP) established in the computer science literature can address the confidentiality protection in a flexible table builder (see: Dwork, et al. 2006, Dwork and Roth, 2014 and references therein). Rinott, et al. (2018) have addressed the implementation of a DP perturbation mechanism on a table builder for census counts from both a theoretical and applied perspective and we follow this approach here. Other examples in the computer science literature for the protection of count data are Barak, et al. (2007), Yaroslavtsev, et al. (2013) and Qardaji, et al. (2014).

In this paper, we focus again on a flexible table builder from an NSI perspective but as opposed to Rinott, et al. (2018) we use survey data as the underlying microdata. Therefore, the generated tables contain weighted survey frequency counts and not whole population counts. In other words, instead of counting the number of individuals for a given cell defined from the spanning variables of a table, we aggregate the survey weights of the individuals. Recall that survey weights are calculated by modifying the design weights (inverse of the inclusion probabilities) to account for non-response and calibration to known population totals.

As mentioned, there is an extra layer of protection when using survey data compared to census data since the number of individuals in each of the cells of the table is random. Survey weights vary across data subjects and hence for a published weighted survey count, there is uncertainty on the value of the original sample count underlying the weighted count. However, tabular data based on surveys are often more problematic than census tables because they are usually accompanied by freely available public-use microdata which are modified versions of the original restricted microdata with some variables dropped and others grouped. Combining public-use files and tables that are generated from the original restricted microdata may lead to disclosing sensitive information. For example, a public-use file will typically have no low-level geography information for individuals in the dataset but by linking to tables obtained from the original restricted microdata, it is not too difficult to obtain the sensitive geographical information, particularly for those individuals with other rare attributes. Therefore, in these situations where both public-use microdata and tables generated from restricted microdata are disseminated, there is a need to use perturbative methods on the generated tables to protect the confidentiality of data subjects.

In this paper, we compare three confidentiality protection methods based on additive random noise to be used in an online flexible table builder to generate tables containing weighted survey counts where internal and marginal cells of the tables are to be perturbed. For ease of comparison of the methods, we assume that there are no multiple hierarchical structures in the generated tables, rather only one set of marginal totals.

We note that there are other approaches for protecting tabular data in the SDL literature. These include cell suppression (Willenborg and De Waal 2001 and references therein) and its extension of controlled tabular adjustment (CTA) (Cox and Dandekar 2002) where suppressed cells are replaced by synthetic values. Another approach is controlled rounding which is a similar concept to additive random noise proposed in this paper but in this case, internal rounded cells are forced to equal the rounded marginal cells. We do not pursue these methods further since they are not condusive to an open online flexible table builder. Controlled rounding has the advantage of preserving additivity but it is a result of mathematical linear programming carried out separately on single tables and it is not possible to preserve consistency of perturbation in same cells across different tables. Like controlled rounding, cell suppression and CTA have the advantage that internal cells aggregate to marginal totals but in general the marginal totals are not perturbed. In addition, the consistency problem mentioned with controlled rounding also exists for the methods of cell suppression and CTA.

The three additive random noise perturbation methods compared here are:

- The computer science approach using a perturbation mechanism that guarantees the definition of differential privacy (DP).
- An SDL approach based on post-randomization (PRAM) which has been adapted to the case of perturbing tabular cell counts described in Shlomo and Young (2008).
- An SDL approach developed at Westat, Inc. called drop/add-up-to-q (Q) and described in Li and Krenzke (2016).

Section 2 describes the three confidentiality protection methods based on additive random noise and Section 3 a simulation study comparing the methods for a flexible table builder of survey weighted cell counts. Section 4 concludes with a discussion.

2 Confidentiality Protection Approaches

There are similarities between the three confidentiality protection methods based on additive random noise as all are based on output perturbation and carried out after the requested table has been generated. All are based on a probability mechanism M which is applied to a list A of all possible cell counts in all allowed tables that can be distributed in a flexible table builder including internal and marginal cell counts. As an example, for a microdata set with 10 variables and allowing for all 3-way tables, the list $\mathbf{A}$ of cell counts are all those in the 3-way tables (120 possible tables), the 2-way tables (45 possible tables), 1-way tables (10 possible tables) and the overall total (which is considered to be known). So there are 175 possible tables (besides the overall total) that can be requested from a flexible table builder.

Define the tables that can be generated in an online flexible table builder as a collection of cells arranged in a list $\mathbf{a} = (\mathbf{a}_1, \dots, \mathbf{a}_K) \in A$ which includes internal cells and marginal cells. Applying the probability mechanism M, M(a), we generate a set of new cell counts in the table $\mathbf{b} = (\mathbf{b}_1, \dots, \mathbf{b}_K)$ where b$\in B$ is the set of all possible outputs that can be obtained from mechanism M. We assume that cell counts are discrete and that b has the same structure as a.

Fraser and Wooton (2005) propose the use of microdata keys to preserve the consistency of perturbation across same cells that may be generated in a table builder. Each individual in the microdata underlying the table builder is assigned a random number, denoted as a 'key'. Any collection of a group of individuals formulating a single cell will also consistently have the same seed by aggregating the microdata keys. Although the perturbations are pre-determined in advance due to the consistency property, the actual perturbation is carried out at the stage that the table is generated. This is referred to as a non-interactive mechanism in the computer science literature and hence privacy budgets are set in advance and any request for the same table will not deplete the privacy budget. This contrasts with the case of an interactive mechanism such as a dynamic online query system in the computer science literature.

Regarding the additivity property for tables generated in an online table builder and assuming simple non-hierarchical structures, we have the following options:

(1) aggregate the perturbed internal cell counts to obtain perturbed marginal cell counts. This will maintain the consistency of internal cells but impact on the consistency of marginal cell counts. For example, given two tables of age crossed with different variables means that the marginal cell counts of age will differ between tables.

(2) perturb the marginal cell counts separately and tables are not additive although the consistency property will be preserved for both internal and marginal cells.

(3) perturb the marginal cells counts separately as in (2) but follow with linear programming (iterative proportional fitting (IPF) and rounding) to ensure additivity of the perturbed internal cell counts to the perturbed marginal cell count in the generated table. Note that this will now impact slightly on the consistency of internal cells across different tables due to the adjustment carried out on a single table but will maintain consistency of marginal cell counts.

We next describe the confidentiality protection approaches that will be used in the simulation study.

2.1 Differential Privacy

We first define differential privacy (DP) (Dwork et al, 2006). A mechanism M satisfies ε-differential privacy if for all neighboring lists a, a′ $\in A$ differing by one individual and all possible outputs b$\in$ B we have:

$$P(M(a) = b)/P(M(a') = b) \leq e^{\varepsilon}. \quad (1)$$

This means that little can be learnt (up to a degree of e^{ε}) by an intruder about the target individual that was dropped when moving from database a to a′. In other words, the ratio is bounded and the probability in the denominator cannot be zero. Rinott et al. 2018 propose using an exponential mechanism (McSherry, et al. 2007) based on a utility function: $u(a, b)$ described as follows:

Given a$\in$ A choose b$\in$ B with probability proportional to

$$e^{\frac{\omega}{2}u(a,b)/\Delta u} \quad (2)$$

where $\varepsilon = \frac{\omega}{2}$ is the privacy budget and the scale is defined as: $\Delta u = \max_{b\in B} \max_{a,a'\in A} |u(a, b) - u(a', b)|$ where a and a′ are neighboring databases that differ by removing one individual. We use a utility function based on the l_1 loss function: $l_1 = \sum_{k=1}^{K} |a_k - b_k|$ and $u(a, b) = -l_1$. This is a discretized Laplace distribution. Note that when we are dealing with internal cell counts of a table, the maximum difference Δu is one as an individual can only appear once. Rinott et al. (2018) proves that this perturbation mechanism M is ε-differentially private.

DP has the advantage that it provides a priori privacy guarantees under a 'worst-case' scenario where the intruder knows everything about the population except for one target individual. This definition subsumes all of the disclosure risks in SDL including inferential disclosure which in the case of an online flexible table builder is mainly caused by the ability to difference and manipulate tables.

On the other hand, NSIs are concerned about utility and one way to ensure high utility is to put a cap on the amount of perturbation to the original cell count. For example, perturbations can be capped up to ± 7. Note that for small survey cell counts, this may result in perturbed cell counts that are negative but these can be converted back to zeros without violating the privacy guarantees as any post-processing on a DP protected table will remain differentially private. The cap causes a 'slippage' in the DP definition since beyond the limits of the cap there is an unbounded ratio in (1). If however the probability of perturbing an original cell count beyond the cap is very small, for example less than 1/N where N is the size of the population, then this slippage leads to the definition of (ε, δ) - differential privacy where δ is the probability of failing to perturb beyond the cap. Therefore, there is a tradeoff between the two parameters ε and δ and NSIs can determine optimal parameters under a risk-utility sensitivity analysis.

Up till now, the focus of table builders has been on whole-population counts. DP does not distinguish between censuses or surveys and the intruder is assumed to know everything about the whole population except for one target individual. For survey microdata where only weighted cell counts are published, removing an individual in a neighboring database means that their associated survey weight is removed.

There are two ways of dealing with frequency tables of weighted survey counts:

- Perturbation carried out on sample cell counts and then the perturbed sample cell counts are used to adjust the displayed weighted survey cell counts;
- Perturbation carried out directly on the weighted survey cell counts.

Rinott et al. (2018) suggest that for survey microdata Δu should be the maximum survey weight and the perturbation should be carried out on the weighted survey cell counts. However, a large Δu leads to lower utility as the perturbation mechanism M takes on uniform probabilities and thus leads to larger perturbations. In this paper, we propose defining Δu as the average survey weight, denoted $\bar{w}$, which is feasible if the survey weights do not vary too much, for example, the relative variance of the survey weights with respect to the squared mean (the efficiency) is less than 10%. In this case, the exponential mechanism defined in (2) for the case of internal survey weighted cell counts is :

$$e^{-\frac{\omega}{2}\bar{w}\sum|a_k-b_k|/\bar{w}} \qquad (3)$$

and thus the average weight $\bar{w}$ cancels out from the numerator and denominator. This means that the perturbations can be carried out first on the sample counts. Following the DP protection of the table containing the sample counts, we then apply a post-perturbation adjustment to obtain the perturbed weighted survey cell counts. We note the important caveat that the the overall average survey weight $\bar{w}$ is known since both the sample size and the population size is known and hence the post-perturbation adjustment does not violate the principles of DP which states that any post-processing on a DP protected table will still be differentially private.

For a DP protected table, we propose the following post-processing adjustment:

Let W_k be the weighted survey cell count in cell *k* of the table and assume a perturbation of $\pm p$ on the orignal unweighted cell count according to the random draw of the DP mechanism. Then the post-perturbation adjustment for the weighted cell count *k* is :

$$W_k \pm p\bar{w}. \quad (4)$$

As mentioned, marginal cell totals can be obtained by aggregating perturbed internal cells. If however we perturb the marginal totals as well as the internal cell counts then an individual can appear multiple times in the table and $\Delta \mathrm{u}$ of (2) will increase. For example, in a 3-way table where all 2 -way tables and 1 -way tables are also perturbed than an individual can appear 2^3-1 times and therefore $\Delta \mathrm{u}$ =7. This leads to higher levels of perturbation. Note that in this case, we can either provide non-additive tables or ensure the additivity of the table by linear programming (IPF and rounding) so that perturbed internal cell counts aggregate to perturbed marginal cell counts. This post-perturbation procedure will not violate the property of DP since any post-processing on a DP protected table will still be differentially private. The linear programming however may result in a slight deterioration of the consistency property of the internal cells across tables.

Other complexities of DP are: (1) zero cells are to be perturbed (unless it is a structural zero in the table resulting from an impossible combination, such as children with an occupation of doctor); and, (2) the nature of the microdata keys that inform the consistency of the perturbation is problematic since clearly two databases a and a' differing by only one individual and only one cell affected will inform the cell to which the individual belongs if tables are differenced. This problem is typically solved through threshold rules in the table builder which does not allow dissemination of sub-populations that differ by only one individual.

2.2 Post-randomization Method

The post-randomization method under SDL for frequency counts in census tables is defined in Shlomo and Young (2008) and a similar approach is used in the ABS Table Builder (Fraser and Wooton, 2005). The post-randomisation method is similar to the DP approach in that there is a probability mechanism M that is applied to original cell counts a $\in \mathrm{A}$ to produce a set of perturbed cell counts b∈ B . The use of microdata keys ensures consistent perturbation for any single cell in a requested table.

The probability mechanism is generally developed arbitrarily to ensure a fixed perturbation variance with caps on the range of perturbations and typically the small cell values in the table are treated differently than the large cell values in the table with zeros not being perturbed.

Let M be a $L \times L$ probability transition matrix containing conditional probabilities: p_{ij}= p(perturbed cell value is j | original cell value is i) for cell values from 1 to L where L is a value beyond which all cell values will take on the same perturbation. Let t be the vector of frequencies of the cell values where the last component would contain the number of cells above L. In each cell of the table, the cell value is changed or not changed according to the prescribed transition probabilities in the matrix M and the result of a draw of a random multinomial variate u with parameters p_{ij} (j =1,2,...L) . If the j-th value is selected, value i is moved to value j. When i = j, no change occurs. Let t* be the vector of the perturbed frequencies of cell values in a single table. Then, t* is a random variable and E(t*|t)=tM. Assuming that the probability transition matrix M has an inverse, we can obtain an unbiased moment estimator of the original table as follows: $\hat{t} = t^{*}\mathrm{M}^{-1}$. Shlomo and Young (2008) place the property of invariance on the probability mechanism M to obtain a new probability mechanism M^{*} such that $t\mathrm{M}^{*} = t$. This means that the perturbed vector t* will be an unbiased estimate of t and we preserve the expected marginal distribution of a table. This approach however results in inconsistencies of same cells across different tables similar to the case of applying linear programming to preserve additivity following perturbation. To transform the original probability transition matrix into an invariant probability transition matrix, a two-stage algorithm is given in Willenborg and De Waal (2001) and also shown in Shlomo and Young (2008).

Clearly, the transformed probability transition mechanism M* depends on the data and hence violates the principle of DP. As discussed, perturbed tables can remain non-additive or linear programming applied to ensure the additivity of the table but since marginal totals are preserved in expectation, little adjustment is needed. Note that same levels of perturbation are applied to internal cell counts and marginal cell counts and there is no distinction as in the case of DP.

Similar to the discussion in Section 2.1, we can apply the perturbations on the sample cell counts and then adjust the weighted sample cell counts accordingly from the perturbed sample cell counts using the overall average survey weight in (4). Additionally we can adjust the weighted cell count by the average survey weight in cell k rather than the overall average survey weight as follows:

$$W_k + p\frac{w_k}{n_k} \quad (5)$$

where n_k is the sample size in cell k. This procedure of course would not be DP since the post-perturbation adjustment depends on the original data in cell k.

2.3 Drop/Add-up-to-q Approach

Li and Krenzke (2016) describe an SDL approach that was developed at Westat, Inc. in the USA. It starts with the perturbation of the sample cell counts as follows: First, q is defined as 1% of the cell value (rounded up to the nearest integer and capped, say at 7) to produce the perturbation vector {-q,-q+1,…-1, 0, 1,….q-1, q} so that the length of the perturbation vector varies according to the original sample cell count. Then the perturbation is carried out using a uniform distribution so that all perturbations are equally likely. In the simplest case, for q=1, the perturbation vector is {-1,0,1} and each of these possible outcomes will have a 1/3 chance of selection.

As in the post-randomization method described in Section 2.2, zeros are not perturbed. Also, since the perturbation depends on the original sample cell count, this approach is not DP. We can adjust the weighted cell counts similarly to the post-randomization described in Section 2.2 using the overall average survey weight or the average survey weight in the cell. An additional method of adjusting the survey weighted cell counts is described in Li and Krenzke (2016) and is based on utilizing the microdata underlying the table. Similar to other approaches, we can leave the tables non-additive or apply linear programming.

3 Simulation Study

In all confidentiality protection approaches, we generate a table and then perturb the sample cell counts in the first step. Following the perturbation of the sample counts in the table, we then adjust the survey weights to produce perturbed weighted sample counts. For the DP approach we adjust the weighted sample counts by the overall average survey weight. For the other SDL approaches we adjust the weighted sample counts by both the overall average survey weight and the average survey weight in the cell. For ease of comparing the confidentiality approaches, we focus only on internal cell counts where an individual appears once in list $\mathrm{a} = (\mathrm{a}_1, \dots, \mathrm{a}_K) \in \mathbf{A}$ and K is the number of internal cells in a table. We assume that marginal totals of a table are obtained by aggregating the internal cells.

We next describe the probability mechanisms M for each of the confidentiality protection approaches described in Section 2.

3.1 Differential Privacy

We use the exponential mechanism defined in Section 2.1 to perturb the internal sample counts of the table with $\varepsilon = 2$, $\Delta \mathrm{u} = 1$ and a cap of ± 7. We have set a rather high ε since our focus is on survey microdata where there is an added layer of protection due to the uncertainty of sample counts from the random sampling and we assume that intruders do not have response knowledge. This results in the perturbation vector and associated probabilities shown in Table 1 calculated as $e^{-2|u|}$ where u={-7,-6,-5,…5,6,7 } and then normalized so that the probabilities sum to 1. We note that under a full risk-utility sensitivity assessment we would vary ε and the caps, but for the purpose of comparing the three confidentiality protection approaches we use these parameter settings.

The parameter δ is determined by the probability at the cap ± 7 which in this case is equal to 0.000000633. This value is very small and therefore is an acceptable slippage for (ε, δ) –differential privacy.

3.2 Post-Randomization

For the post-randomization method, we use a similar perturbation vector to Table 1 but place it in the framework of this SDL approach. Here the small cell counts are perturbed separately to ensure that no negative values occur and in addition, the mechanism M is placed in a matrix formulation with truncations in order to carry out the transformation that ensures the property of invariance as described in Section 2.2. The truncations introduce bias into the perturbation. The small cell probability matrix is presented in Table 2.

Table 1: Differentially private perturbation mechanism probabilities for $\varepsilon = 2$, $\Delta u = 1$ and a cap of ± 7

Perturbation	Probability of Perturbation
-7	0.000000633
-6	0.000004679
-5	0.000034576
-4	0.000255486
-3	0.001887804
-2	0.013949
-1	0.10307
0	0.76159
1	0.10307
2	0.013949
3	0.001887804
4	0.000255486
5	0.000034576
6	0.000004679
7	0.000000633

Table 2: Small cell probability mechanism for post-randomization of cell counts below 6

Original values	Perturbed values						
	0	1	2	3	4	5	6
0	1	0	0	0	0	0	0
1	0.11920	0.76160	0.10308	0.01395	0.00189	0.00026	3.46E-05
2	0.01395	0.10525	0.76160	0.10311	0.01395	0.00189	0.00026
3	0.00189	0.01395	0.10337	0.76160	0.10337	0.01395	0.00189
4	0.00026	0.00189	0.01395	0.10311	0.76160	0.10525	0.01395
5	3.46E-05	0.00026	0.00189	0.01395	0.10308	0.76160	0.11920
6	5.31E-06	3.46E-05	0.00026	0.00189	0.01395	0.22227	0.76160

For large cell counts over the value of 7 we first calculate a residual value from base 15 denoted $m = mod(C_k, 15)$ where C_k is the original sample count in cell k. This determines the row defining the probability vector in Table 3. Then the perturbed value v is selected based on the random draw according to the appropriate probability vector in Table 3. The final perturbation p for the sample count in cell k on the orignal scale is obtained by: p =m- v which can be either a negative or positive value. The perturbation p is then added to the original sample cell count in cell k. Table 3 presents the probability mechanism M for large cell counts in the post-randomization method.

Table 3: Large cell probability mechanism for post-randomization of cell counts over 7

Residual from 15	Perturbed value that is subtacted from the residual if selected														
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	0.76160	0.22227	0.01395	0.00189	0.00026	3.46E-05	4.68E-06	6.33E-07	0	0	0	0	0	0	0
1	0.11920	0.76160	0.10307	0.01395	0.00189	0.00026	3.46E-05	4.68E-06	6.33E-07	0	0	0	0	0	0
2	0.01395	0.10525	0.76160	0.10307	0.01395	0.00189	0.00026	3.46E-05	4.68E-06	6.33E-07	0	0	0	0	0
3	0.00189	0.01395	0.10337	0.76160	0.10307	0.01395	0.00189	0.00026	3.46E-05	4.68E-06	6.33E-07	0	0	0	0
4	0.00026	0.00189	0.01395	0.10311	0.76160	0.10307	0.01395	0.00189	0.00026	3.46E-05	4.68E-06	6.33E-07	0	0	0
5	3.46E-05	0.00026	0.00189	0.01395	0.10308	0.76160	0.10307	0.01395	0.00189	0.00026	3.46E-05	4.68E-06	6.33E-07	0	0
6	4.68E-06	3.46E-05	0.00026	0.00189	0.01395	0.10307	0.76160	0.10307	0.01395	0.00189	0.00026	3.46E-05	4.68E-06	6.33E-07	0
7	6.33E-07	4.68E-06	3.46E-05	0.00026	0.00189	0.01395	0.10307	0.76160	0.10307	0.01395	0.00189	0.00026	3.46E-05	4.68E-06	6.33E-07
8	0	6.33E-07	4.68E-06	3.46E-05	0.00026	0.00189	0.01395	0.10307	0.76160	0.10307	0.01395	0.00189	0.00026	3.46E-05	4.68E-06
9	0	0.00E+00	6.33E-07	4.68E-06	3.46E-05	0.00026	0.00189	0.01395	0.10307	0.76160	0.10308	0.01395	0.00189	0.00026	3.46E-05
10	0	0	0.00E+00	6.33E-07	4.68E-06	3.46E-05	0.00026	0.00189	0.01395	0.10307	0.76160	0.10311	0.01395	0.00189	0.00026
11	0	0	0	0.00E+00	6.33E-07	4.68E-06	3.46E-05	0.00026	0.00189	0.01395	0.10307	0.76160	0.10337	0.01395	0.00189
12	0	0	0	0	0.00E+00	6.33E-07	4.68E-06	3.46E-05	0.00026	0.00189	0.01395	0.10307	0.76160	0.10525	0.01395
13	0	0	0	0	0	0.00E+00	6.33E-07	4.68E-06	3.46E-05	0.00026	0.00189	0.01395	0.10307	0.76160	0.11920
14	0	0	0	0	0	0	0.00E+00	6.33E-07	4.68E-06	3.46E-05	0.00026	0.00189	0.01395	0.22227	0.76160

The probability mechanisms for small and large cell counts in the post-randomisation method have the same value of $\varepsilon = 2$ as DP in Section 3.1 but the caps on the cell counts (and subsequently the values of δ) will vary depending on the original sample cell count. The probability mechanism M is not symmetric. For example, for an original cell count of 1 perturbed to 0, δ is equal to 0.119203 enforcing the rule that no negative count is allowed. This δ is very large and hence it may be possible for an intruder to gain sensitive information about a target individual. However, for the same original sample count of 1 perturbed to a higher value of say, 8, δ is equal to 0.000034576.

3.3 Drop/Add-up-to-q Approach

For the drop/add-up-to-q approach, the probability mechanism M is uniform and depends on the value q which is 1% of the original sample cell value (rounded up to the nearest integer) and capped at ± 7. For a uniform probability mechanism, this means that ε is very small and in this particular case is equal to 0.01. Therefore, compared to the other approaches, this approach seemingly has stricter privacy guarantees. However, the slippage in the form of δ can be large. For q=7, δ is 0.067 and for q=1, δ is very high and is equal to 0.333.

3.4 Generating Tables

The simulation study is carried out on two-dimensional tables and will focus on a risk and utility assessment under two different sets of tables, the first set with dependent attributes and the second set with independent attributes. In particular, we focus on Chi-square tests for independence to measure the impact of the perturbation. The tables are generated as follows:

1. Generate a population table of size 7 by 7 using a Poisson distribution with $\mu_{ij} = \eta + \alpha_i + \beta_j + \text{Const} \times \gamma_{ij}$ for row i and column j, where each of α_i, β_j and γ_{ij} are drawn by Uniform(0.5,0.5) and η =6.5. This produces population tables of approximately 44,000 individuals.
2. The independent attribute tables are generated with Const=0.02. This is to introduce slightly lower power to the Chi-square test for independence. The dependent attribute tables are generated with Const=0.2.
3. Initial survey weights are introduced into the tables that are generated by Uniform (20,40) with the mean at 30 and the relative variance with respect to the squared mean of 3.7%.
4. The sample cell counts are calculated by rounding the value obtained by dividing the population counts with the generated survey weights in step 3, and final survey weights are calculated by the population counts divided by the rounded sample cell counts. This produces tables of sample counts of approximately 1,530 individuals for an average cell size of 31.
5. On each generated table, we carry out the three confidentiality protection methods based on additive random noise on the sample counts (denoted 'DP' for differential privacy, 'PRAM' for the post-randomization method and 'Q' for the drop/add-up-to-q approach).
6. The perturbed weighted cell counts are then obtained as described in Section 2. We adjust the original weighted cell counts by the overall average weight (denoted, 'Avg') and the average weight in the cell (denoted 'Avg cell').
7. Repeat 500 times.

3.5 Risk-Utility Analysis

We first present a disclosure risk assessment of the three confidentiality approaches in Table 4. We have previously equated the DP parameters to the SDL approaches of PRAM and Q, and these appear in the first row of Table 4 for the case where the original sample cell count is 1 which means that the δ parameter is a maximal value. We also include two other disclosure risk measures that are defined in the SDL literature. The first is the average percentage of cells perturbed in the tables across the 500 generated independent or dependent tables. The second is a risk measure developed in Antal, et al. (2014) based on Information Theory. It is defined as the average across the 500 generated independent or dependent tables of the following risk measure: $RM = 1 - \frac{H(a|b)}{H(a)}$ where for a given table with K internal cells: $a = (a_1, a_2, \dots, a_K)$, the entropy is defined as $H(a) = -\sum_k \frac{a_k}{N} \log \frac{a_k}{N}$ and $\sum_k a_k = N$ and the conditional entropy of table a and the perturbed table b, $H(a|b)$, is calculated according to formula (4) in Antal, et al. (2014) as follows (with $\log(0)$ defined as 0):

$$H(a|b) = -\sum_k \frac{\min(a_k,b_k)}{N} \log\left(\frac{\min(a_k,b_k)}{b_k}\right) - \sum_k \frac{a_k - \min(a_k,b_k)}{N} \log\left(\frac{a_k - \min(a_k,b_k)}{N - \sum_k \min(a_k,b_k)}\right) - \sum_k \frac{b_k - \min(a_k,b_k)}{N} \log\left(\frac{b_k - \min(a_k,b_k)}{b_k}\right)$$

Note that if $\sum_k b_k = M$ and $M \neq N$ then we need to adjust the cell counts to have equal totals by multiplying vector a by M and vector b by N. The conditional entropy represents the uncertainty in the original table given we have observed the perturbed table. For RM=0 this implies that H(a|b)=H(a) and we do not learn any new information about the original table given we have observed the perturbed table. Therefore, the higher the risk measure, RM, the more we may infer information from the original table given the perturbed table. In Table 4, we show the RM measure for the weighted survey counts according to the weight adjustment based on the overall average survey weight. Results were similar for the case of the weight adjustment according to the average survey weight in the cell.

From Table 4, the small ε in the SDL approach of drop/add-up-to-q (Q) is indicative of the fact that a higher percentage of cells are perturbed compared to the other approaches and the risk measure RM is lower reflecting that there is more uncertainty introduced into the tables as a result of the perturbation. However, the Q approach has a very large δ for the case of an original sample cell value of 1 which means a high probability of an unbounded likelihood ratio in (1). Whilst we fixed the post-randomization method (PRAM) to be similar to differential privacy (DP) with $\varepsilon = 2$, the fact that the perturbation mechanism is not symmetric caused slightly higher levels of perturbation but again we see that PRAM has a large δ for the case of an original sample cell value of 1. Under the SDL risk measures, there is little difference on whether the tables had independent or dependent attributes.

Table 4: Risk Measures for weighted sample counts according to confidentiality protection methods (averaged over 500 replications)

Risk Measures		DP	PRAM	Q
DP parameters when original sample count=1	ε δ	2 0.00000063	2 0.1192	0.01 0.333
Percent Cells Perturbed	Independent Dependent	23.8 24.5	27.7 27.7	67.1 66.8
1-Proportion of conditional entropy (RM)	Independent Dependent	0.9891 0.9892	0.9870 0.9870	0.9849 0.9751

In Table 5, we compare the confidentiality protection approaches with respect to a range of utility measures: the average of the total sample count, total weighted sample count and the percent relative absolute difference from the true counts over the 500 generated dependent and independent tables. In addition, we calculate the Hellinger's Distance metric on each of the tables: $HD(a,b)=\sqrt{\frac{1}{2}\sum_k(\sqrt{a_k}-\sqrt{b_k})^2}$ where a_k is the original cell value and b_k is the perturbed cell value, and present the average of the Hellinger's Distance over the 500 generated dependent or independent tables.

From Table 5, all confidentiality protection approaches in both the case of dependent and independent attributes preserve the overall sample and weighted totals with differential privacy (DP) slightly outperforming post-randomization (PRAM) and drop/add-up-to-q (Q) with a smaller percent relative absolute difference. DP also has smaller Hellinger's Distances compared to PRAM and Q. We note that the DP approach is unbiased if there are no negatively perturbed sample counts which are converted back to zeros. Q is also an unbiased perturbation mechanism although given the uniform distribution of perturbing cell counts, there are more cells that are perturbed. PRAM on the other hand has a perturbation mechanism that biases the perturbation at the tail ends of the distribution. Results in Table 5 show no discernible differences for tables with dependent or independent attributes.

Table 5: Overall sample and weighted sample counts and Hellinger's distance according to confidentiality protection methods (averaged over 500 replications)

Confidentiality Protection Methods	Mean value	Standard Error	Percent Relative Absolute Difference	Average Hellinger's Distance
Dependent Sample Counts				
Original	1531.1	21.7	-	-
DP	1531.3	21.7	0.245	0.352
PRAM	1531.1	21.6	0.276	0.404
Q	1531.5	21.6	0.346	0.491
Dependent Weighted Counts				
Original	44164.7	573.0	-	-
DP Avg	44168.8	573.2	0.253	1.931
PRAM Avg	44164.6	572.8	0.285	2.224
PRAM Avg cell	44163.7	572.8	0.290	2.245
Q Avg	44176.8	573.0	0.358	2.696
Q Avg cell	44177.0	573.0	0.361	2.739
Independent Sample Counts				
Original	1522.3	21.4	-	-
DP	1522.1	21.4	0.245	0.350
PRAM	1522.3	21.4	0.289	0.401
Q	1522.5	21.4	0.337	0.491
Independent Weighted Counts				
Original	43921.2	567.3	-	-
DP Avg	43916.0	567.5	0.253	1.920
PRAM Avg	43922.1	567.1	0.297	2.206
PRAM Avg cell	43921.8	567.1	0.307	2.232
Q Avg	43927.2	567.6	0.346	2.697
Q Avg cell	43925.2	567.6	0.357	2.742

In Figure 1 we show a risk-utility confidentiality map summarizing our findings for the dependent attribute tables where the Y-axis presents the risk measure RM and the X-axis the Hellinger's Distance (in reverse order). The figure shows that DP in the upper right hand quadrant has the highest risk measure and the highest utility and Q in the lower left hand quadrant has the lowest risk measure and the lowest utility although we note there is a very small difference in scale for RM in Figure 1 .

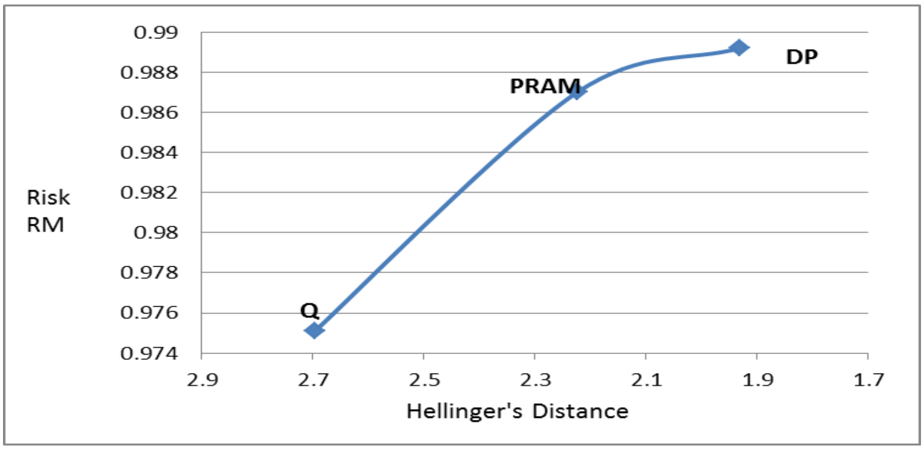

Figure 1: Risk-Utility confidentiality map on dependent attributes (Y-axis: RM=1- $(\frac{H(a|b)}{H(a)})$; X-axis: $HD(a,b)$ (reverse order))

We now turn to assessing the impact of the perturbations on statistical inference, in this case the Chi-square test for independence. Figure 2 show the chi-square statistics calculated from the weighted sample counts under the dependent attributes. Note that we do not account for any survey design features in our calculation of the chi-square statistic. We see little differences in the confidentiality protection approaches on the chi-square statistic. All p-values (not shown here) were close to zero for all confidentiality protection methods and hence there would be no change in rejecting the null hypothesis of independence under a statistical hypothesis test.

Figures 3 and 4 show the chi-square statistics and their associated p-values calculated from the weighted sample counts under the independent attributes. Here we can see that the perturbations for all confidentiality protection methods distort the independent attributes and introduce dependencies which increase the chi-square statistics and pushes p-values to be close to zero, thus we fail to reach a correct decision under our hypothesis testing for independence. DP is slightly outperforming PRAM and both are performing better than Q with the mean of the chi-square statistics closer to the true mean, although Q has less outliers and seems to be more stable. It is clear that using perturbed tables naively as if they are original tables will severely bias statistical inference. Since DP is based on a probability mechanism that is not related to the original data and is grounded in computer science cryptography, the probability mechanism is not secret and can be released to the users. Rinott, et al. (2018) show how to use DP parameters to adjust statistical inference for the case of a Chi-square test for independence and goodness of fit.

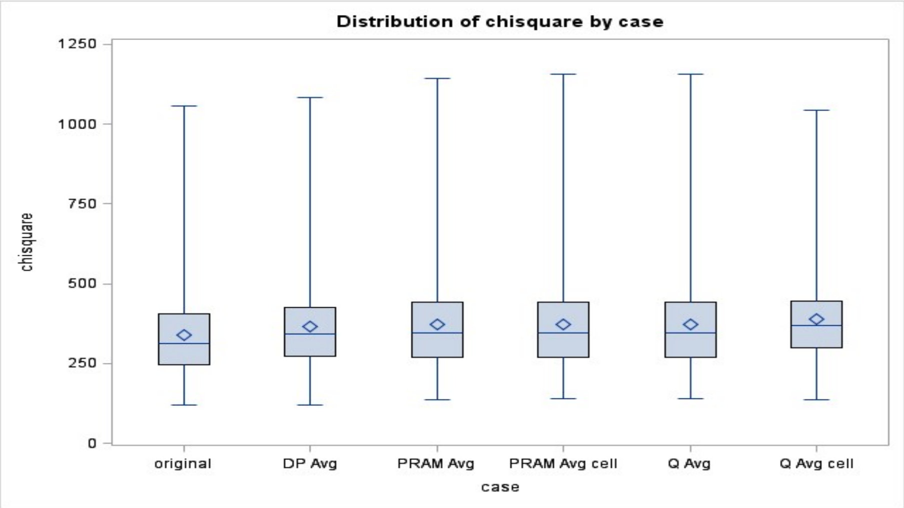

Figure 2: Chi-square statistics for tables of dependent attributes according to confidentiality protection methods (500 replications)

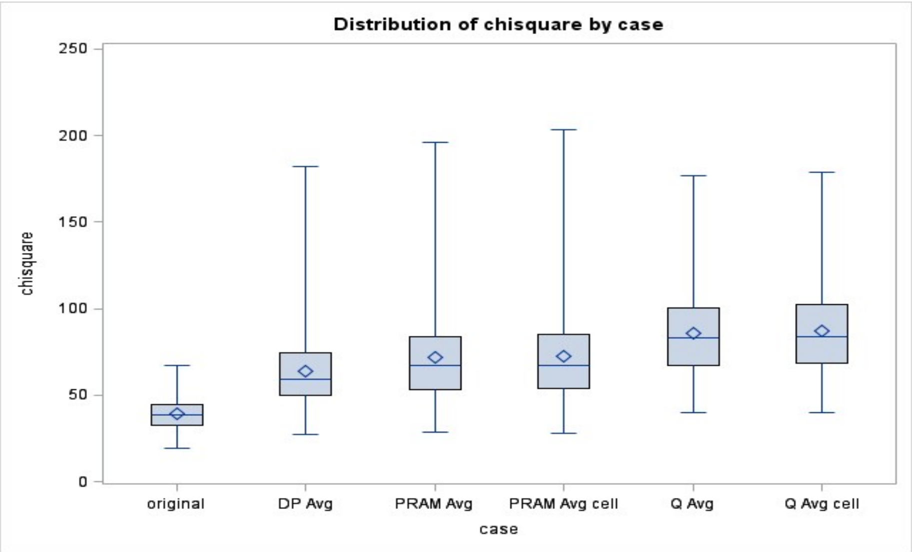

Figure 3: Chi-square statistics for tables of independent attributes according to confidentiality protection methods (500 replications)

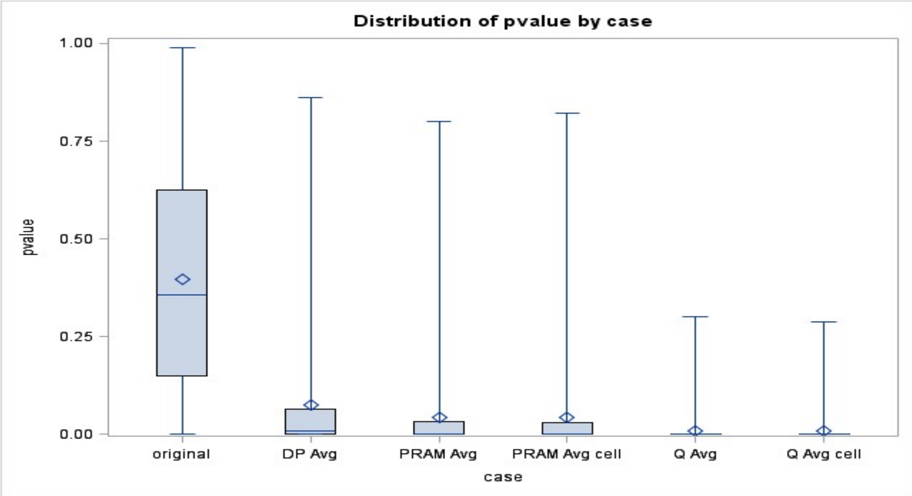

Figure 4: P-values for testing independence in tables with independent attributes according to confidentiality protection methods (500 replications)

4 Discussion

We have compared three confidentiality protection methods based on additive random noise that can be used to protect the confidentiality of tables generated from an online flexible table builder. We have shown that differential privacy (DP) can be a useful confidentiality protection method for a flexible table builder of survey weighted cell counts. The level of perturbation for the DP mechanism may be slightly lower based on the tested parameters in the simulation study compared to the SDL approaches according to some standard SDL measures of disclosure risk. However, the DP mechanism is independent of the data and has a very small parameter δ and hence a stronger guarantee of a bounded ratio in (1) and therefore may offer more protection against inferential disclosure. Furthermore, the consistency property of a flexible table builder across same cells in same domains ensures that the perturbation can be carried out under a non-interactive mechanism with a fixed privacy budget. We have seen that the utility in the DP mechanism is also higher compared to the SDL confidentiality approaches and since the perturbation mechanism is known and not secret, it can be used to compensate for the perturbation in statistical inferences. This would not be the case for the SDL confidentiality approaches where the parameters of the perturbation are not made public. More research is needed to compare the confidentiality protection methods on other tables and on other surveys where survey weights may be more variable.

There are a number of limitations to this simulation study:

- We have set a rather high privacy budget ε which we feel is justified in the case of disseminating survey weighted cell counts through a flexible table builder. This is because there is an additional layer of protection afforded by the sampling and the underlying sample cell counts of the weighted cell counts are random.
- To compare the confidentiality protection methods we did not focus on the marginal cell counts and assume that these are obtained by aggregating the internal perturbed cell counts of generated tables. Perturbing marginal cell counts separately and then applying linear programming to adjust for the additivity in the table will likely impact on the consistency property which may incur a loss of privacy budget.
- We have not dealt with complex multidimensional tables with multiple hierarchical structures. Further investigation is needed on the issue of consistency/additivity when multiple marginal totals are included in the same table requested from a flexible online table builder.
- We have seen that perturbing the sample counts in the first step and then adjusting the survey weights according to the overall average survey weight to obtain the perturbed weighted cell count provided smaller distance metrics compared to adjusting according to the average survey weight in the cell. However, this may be an artifact of the simulation study which had a generally low amount of variation in the survey weights. For larger survey weights with more variation, future work will explore the perturbation directly on the weighted survey counts.

Acknowledgements

This work was funded by Westat under contract to the National Center for Science and Engineering Statistics and the National Science Foundation.

References

[1] Antal, L., Shlomo, N. and Elliot, M. (2014) Measuring Disclosure Risk with Entropy in Population Based Frequency Tables, In Privacy in Statistical Databases 2014, (Ed. J. Domingo-Ferrer), Springer LNCS 8744 62-78.

[2] Barak, B., Chaudhuri, K., Dwork, C., Kale,S., McSherry,F., and Talwar,K. (2007) Privacy, Accuracy, and Consistency too: A Holistic Solution to Contingency Table Release. Symposium on Principles of database systems, ACM 2007 273-282.

[3] Chipperfield, J., Gow, D. and Loong, B. (2016) The Australian Bureau of Statistics and Releasing Frequency Tables via a Remote Server, Statistical Journal of the IAOS 32 53-64.

[4] Cox, L.H. and Dandekar, R.A. (2004) Disclosure Limitation Method for Tabular Data That Preserves Accuracy and Ease-of-Use, In Proceedings of the 2002 FCSM Statistical Policy Conference, Washington, DC: Office of Management and Budget 15-30.

[5] Dwork, C., McSherry, F., Nissim, K. and Smith, A. (2006) Calibrating Noise to Sensitivity in Private Data Analysis, In 3rd IACR Theory of Cryptography Conference 265-284.

[6] Dwork, C. and Roth, A. (2014) The Algorithmic Foundations of Differential Privacy, Foundations and Trends in Theoretical Computer Science 9 211-407.

[7] Fraser, B. and Wooton, J. (2005) A proposed Method for Confidentialising Tabular Output to Protect Against Differencing, In Joint UNECE/Eurostat work session on statistical data confidentiality, Geneva, Switzerland. Available at: https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2005/wp.35.e.pdf

[8] Li, J., and Krenzke, T. (2016) Confidentiality Approaches for Real-time Systems Generating Aggregated Results, Proceedings of the Survey Research Methods Section of the American Statistical Association. Available at: https://ww2.amstat.org/sections/srms/Proceedings/y2016f.html

[9] McSherry, F. and Talwar, K. (2007) Mechanism Design via Differential Privacy, In Foundations of Computer Science (FOCS'07) Proceedings of the 48th Annual IEEE Symposium IEEE 94-103.

[10] Qardaji, W., Yang, W. and Li, N. (2014) Preview: Practical Differentially Private Release of Marginal Contingency Tables, In Proceedings of the 2014 ACM SIGMOD international conference on Management of data, ACM 2014 1435–1446.

[11] Rinott, Y., O'Keefe, C., Shlomo, N., and Skinner, C. (2018) Confidentiality and Differential Privacy in the Dissemination of Frequency Tables, Statistical Sciences 33:3 358-385.

[12] Shlomo, N., Antal, L. and Elliot, M. (2015) Measuring Disclosure Risk and Data Utility for Flexible Table Generators, Journal of Official Statistics 31:2 305-324.

[13] Shlomo, N. and Young, C. (2008) Invariant Post-tabular Protection of Census Frequency Counts, In PSD'2008 Privacy in Statistical Databases, (Eds. J. Domingo-Ferrer and Y. Saygin), Springer LNCS 5261 77-89.

[14] Yaroslavtsev, G., Cormode, G., Procopiuc, C.M., and Srivastava, D. (2013) Accurate and Efficient Private Release of Datacubes and Contingency Tables, In ICDE 2013.

[15] Willenborg, L.C.R.J. and De Waal, T. (2001) Elements of Statistical Disclosure Control in Practice, Lecture Notes in Statistics 155 Springer-Verlag.