

Statistical Disclosure Control Using Post Randomisation: Variants and Measures for Disclosure Risk

Ardo van den Hout¹ and Elsayed A.H. Elamir²

This article discusses the post randomisation method (PRAM) as a method for disclosure control. PRAM protects the privacy of respondents by misclassifying specific variables before data are released to researchers outside the statistical agency. Two variants of the initial idea of PRAM are discussed concerning the information about the misclassification that is given along with the released data. The first variant concerns calibration probabilities and the second variant concerns misclassification proportions. The article shows that the distinction between the univariate case and the multivariate case is important. Additionally, the article discusses two measures for disclosure risk when PRAM is applied.

Key words: Calibration; information loss; misclassification; PRAM.

1. Introduction

The post randomisation method (PRAM) is discussed in Gouweleeuw et al. (1998) as a method for statistical disclosure control (SDC). When survey data are released by statistical agencies, SDC protects the identity of respondents. SDC tries to prevent that a user of the released data can link the data of a respondent in the survey to a specific person in the population. See Willenborg and De Waal (2001) for a general introduction to SDC including PRAM.

There is a close link between PRAM and randomised response (Warner 1965), a method to ask sensitive questions in a survey. An early version of PRAM based on randomised response is discussed in Rosenberg (1979). Differences and similarities between randomised response and PRAM are discussed by Sande (1998) and Van den Hout and Van der Heijden (2002).

The idea of PRAM is to misclassify some of the categorical variables in the survey using fixed misclassification probabilities and to release the partly misclassified data together with those probabilities. For example, variable X , with categories $\{1, \dots, J\}$, is misclassified into variable X^* . The survey containing X^* but not X is released together with conditional probabilities $\mathbb{P}(X^* = k|X = j)$ for $k, j \in \{1, \dots, J\}$. In this way PRAM introduces uncertainty into the data: the user of the released data cannot be sure that the information is original or perturbed due to PRAM and it becomes harder to

¹ Department of Methodology and Statistics, Faculty of Social Sciences, Utrecht University, Heidelberglaan 1, 3584 CS Utrecht, the Netherlands. Currently at MRC Biostatistics Unit, Cambridge, UK. Email: ardo.vandenhout@mrc-bsu.cam.ac.uk

² College of Business, Department of Management and Marketing, University of Bahrain, P.O.Box 32038, Kingdom of Bahrain. Email: sali@buss.uob.bh

establish a correct link between a respondent in the survey and a specific person in the population. Since the user has the misclassification probabilities his or her analysis can be adjusted by taking into account the perturbation due to PRAM.

When SDC is used, there will always be a loss of information. This is inevitable since SDC tries to determine the information in the data that can lead to the disclosure of the identity of a respondent, and eliminates this information before data are released. It is not difficult to prevent disclosure, but it is difficult to prevent disclosure *and* release data that are still useful for statistical analysis. Applying SDC means searching for a balance between disclosure risk and information loss.

When the information in the data that can lead to a disclosure is contained in categorical variables, PRAM is an alternative to SDC techniques such as global recoding (see Willenborg and De Waal 2001, Section 1.8) and data swapping (see Dalenius and Reiss 1982 and Fienberg and McIntyre 2005).

This article discusses two ideas to make PRAM more efficient with respect to the balance between disclosure risk and information loss. First, the article discusses the use of *calibration probabilities*

$$\mathbb{P}(\text{true category is } j | \text{category } i \text{ is released}) \quad (1)$$

in the analysis of released data and compares this with using *misclassification probabilities*

$$\mathbb{P}(\text{category } i \text{ is released} | \text{true category is } j) \quad (2)$$

The idea of using calibration probabilities is discussed by De Wolf et al. (1997), who refer to the discussion of calibration probabilities in the misclassification literature; see e.g., Kuha and Skinner (1997). We will elaborate the discussion and show that the advantage of calibration probabilities is limited to the univariate case. Secondly, the article shows that information loss can be reduced by providing *misclassification proportions* along with the released data. These proportions inform about the actual change in the survey data due to the application of PRAM. (Probabilities (1) and (2) inform about the expected change.) Additionally, the article discusses two measures for disclosure risk when PRAM is applied. The first is a general measure presented in Elamir and Skinner (2003) as an extension of the measure introduced by Skinner and Elliot (2002). The second measure links up with the SDC practice at Statistics Netherlands. Simulation results are given to illustrate the theory.

The outline of the article is as follows. Section 2 provides the framework and the notation. In Section 3, two distinctions are made that are important when PRAM is discussed. Section 4 summarises frequency estimation for PRAM data. Section 5 discusses the use of calibration probabilities. In Section 6 we introduce the use of misclassification proportions. Section 7 discusses measures for disclosure risk. Section 8 presents some simulations, and Section 9 is the discussion.

2. Framework and Notation

In survey data we distinguish between *identifying variables* and *nonidentifying variables*. Identifying variables are variables that can be used to reidentify individuals represented in the data. These variables are assumed to be categorical, e.g., gender, race, place of

residence. We assume that the sensitive information of respondents is contained in the non-identifying variables, (see Bethlehem et al. 1990), and that we want to protect this information by applying PRAM to (a subset of) the identifying variables.

The notation in this article is the same as in Skinner and Elliot (2002). Units are selected from a finite population U and each selected unit has one record in the *microdata sample* $s \subset U$. Let n denote the number of units in s . Let the categorical variable formed by cross-classifying (a subset of) the identifying variables be denoted X with values in $\{1, \dots, J\}$. Let X_i denote the value of X for unit $i \in U$. The *population frequencies* are denoted

$$F_j = \sum_{i \in U} I(X_i = j), \quad j \in \{1, \dots, J\}$$

where $I(\cdot)$ is the indicator function: $I(A) = 1$ if A is true and $I(A) = 0$ otherwise. The *sample frequencies* are denoted

$$f_j = \sum_{i \in s} I(X_i = j), \quad j \in \{1, \dots, J\}$$

In the framework of PRAM, we call the sample that is released by the statistical agency the *released microdata sample* s^* . Note that unit $i \in s^*$ if and only if $i \in s$. Let X^* denote the released version of X in s^* . By *misclassification* of unit i we mean $X_i \neq X_i^*$. The *released sample frequencies* are denoted

$$f_k^* = \sum_{i \in s^*} I(X_i^* = k), \quad k \in \{1, \dots, J\}$$

Let P_X denote the $J \times J$ transition matrix that contains the conditional misclassification probabilities $p_{kj} = \mathbb{P}(X^* = k | X = j)$, for $k, j \in \{1, \dots, J\}$. Note that the columns of P_X sum up to one. The distribution of X^* conditional on s is the J -component finite mixture given by

$$\mathbb{P}(X_i^* = k | i \in s) = \sum_{j=1}^J \mathbb{P}(X_i^* = k | X_i = j) \mathbb{P}(X_i = j | i \in s), \quad k \in \{1, \dots, J\}$$

where the component distributions are given by P_X and the component weights are given by the conditional distribution of X . The conditional distribution of X in sample s is assumed to be multinomial with

$$\pi_j = \mathbb{P}(X_i = j | i \in s) = \frac{1}{n} f_j, \quad j \in \{1, \dots, J\} \tag{3}$$

This conditional distribution does not take into account possible correlations with non-identifying variables.

3. When Discussing PRAM

When discussing PRAM, it is important to keep in mind two distinctions. The first distinction is whether data perturbed by PRAM are discussed from the viewpoint of the statistical agency or from the viewpoint of the user of the data. Note that the statistical

agency has all the information, i.e., the original data, the perturbed data and the misclassification probabilities. Using this information, the statistical agency can compute the calibration probabilities, the misclassification proportions, and the calibration proportions, as will be shown in Sections 5 and 6. It is for the statistical agency to decide which information about the misclassification is released together with the perturbed data.

Taking the viewpoint of the user, analysing PRAM data means analysing misclassified data where information about the misclassification is at hand. In the original idea of PRAM, perturbed data are released together with the misclassification probabilities (see Kooiman et al. 1997). It is important to realise that in this situation misclassification probabilities can be considered as known parameters – a distinctive difference with misclassification in for example social statistics where misclassification probabilities have to be estimated (see e.g., Kuha and Skinner 1997). The same reasoning applies to the release of calibration probabilities: if the statistical agency decides to release the probabilities, the user can treat them as known parameters.

When this article discusses efficiency concerning the analysis of PRAM data it will be a discussion from the viewpoint of the user: which kind of information about the misclassification is the most efficient in the analysis of PRAM data? When this article discusses measures of disclosure risk it will be a discussion from the viewpoint of the statistical agency: how shall one measure disclosure risk when PRAM is applied?

The second important distinction when discussing PRAM is whether PRAM is applied independently to a selection of variables or is applied to a Cartesian product. In the present article, we choose to apply PRAM independently to a selection of identifying variables. Information about the misclassification is released per variable. This procedure is not without drawbacks, as will be described in the discussion.

When PRAM is applied independently and there is a transition matrix per variable, the misclassification can be described with respect to Cartesian products. Both the statistical agency and the user can construct transition matrices of Cartesian products. For instance, consider identifying variables X_1 , with categories $\{1, \dots, J_1\}$ and X_2 , with categories $\{1, \dots, J_2\}$, and the cross-classification $X = (X_1, X_2)$, i.e., the Cartesian product of X_1 and X_2 . Since PRAM is applied independently, we have

$$\mathbb{P}(X^* = (k_1, k_2) | X = (j_1, j_2)) = \mathbb{P}(X_1^* = k_1 | X_1 = j_1) \mathbb{P}(X_2^* = k_2 | X_1 = j_2) \quad (4)$$

for $k_1, j_1 \in \{1, \dots, J_1\}$ and $k_2, j_2 \in \{1, \dots, J_2\}$. In matrix notation, we have $P_X = P_{X_1} \otimes P_{X_2}$, where \otimes is the Kronecker product. Note that when one of two variables is not perturbed by PRAM, the transition matrix of that variable is the identity matrix.

For users, a transition matrix of a Cartesian product is important when estimating for example a bivariate frequency distribution. For statistical agencies, transition matrices of Cartesian products become important when disclosure risk is investigated.

4. Frequency Estimation for PRAM Data

When PRAM is applied and some of the identifying variables are misclassified, standard statistical models do not apply to the released data since these models do not take into account the perturbation. This section shows how the misclassification can be taken into

account in frequency estimation. This section does not contain new material, but is preliminary to the subsequent sections.

We have $E[\underline{f}^*|\underline{f}] = P_X \underline{f}$, where $\underline{f} = (f_1, \dots, f_J)^t$ and $\underline{f}^* = (f_1^*, \dots, f_J^*)^t$ is the stochastic vector of the released sample frequencies. An unbiased moment estimator of \underline{f} is given by

$$\hat{\underline{f}} = P_X^{-1} \underline{f}^* \tag{5}$$

(see Kooiman et al. 1997). In practice, assuming that P_X is nonsingular does not impose much restriction on the choice of the misclassification probabilities. A sufficient condition for P_X to be nonsingular is $p_{ii} > 1/2$ for $i \in \{1, \dots, J\}$, since in that case P_X is strictly diagonally dominant (see Horn and Johnson 1985, p. 349). An additional assumption is that the dimensions of \underline{f} and \underline{f}^* are the same.

The variance of (5) equals

$$V[\hat{\underline{f}}|\underline{f}] = P_X^{-1} V[\underline{f}^*|\underline{f}](P_X^{-1})^t = P_X^{-1} \left(\sum_{j=1}^J f_j V_j \right) (P_X^{-1})^t \tag{6}$$

where V_j is the $J \times J$ covariance matrix of two released values given the original value j , i.e.,

$$V_j(k_1, k_2) = \begin{cases} p_{k_2 j}(1 - p_{k_2 j}) & \text{if } k_1 = k_2 \\ -p_{k_1 j} p_{k_2 j} & \text{if } k_1 \neq k_2 \end{cases} \quad \text{for } k_1, k_2 \in \{1, \dots, J\}$$

(see Kooiman et al. 1997). The variance can be estimated by substituting \hat{f}_j for f_j in (6), for $j \in \{1, \dots, J\}$.

The variance given by (6) is the extra variance due to PRAM and does not take into account the sampling design. The formulas for the latter are given in Chaudhuri and Mukerjee (1988) for multinomial sampling and compared to (6) in Van den Hout and Van der Heijden (2002); see also Appendix A.

5. Calibration Probabilities

Literature concerning misclassification shows that calibration probabilities (1) are more efficient in the analysis of misclassified data than misclassification probabilities (2) (see the review article by Kuha and Skinner 1997, Section 28.5.1 and the references therein). The idea of using calibration probabilities for PRAM is mentioned in De Wolf et al. (1997) (see also Willenborg and De Waal 2001, Section 5.5.4). The following compares the use of calibration probabilities with the use of misclassification probabilities as far as PRAM is concerned. We will emphasize the multivariate case, since this is relevant when PRAM is applied. As far as we know, the multivariate case with respect to calibration probabilities and PRAM has not been discussed in the literature.

The $J \times J$ matrix with calibration probabilities of the univariate variable X is denoted by \underline{P}_X . Using Bayes's theorem and the conditional distribution of X given by (3), the entries

\bar{p}_{jk} of \bar{P}_X are

$$\mathbb{P}(X_i = j | X_i^* = k, i \in s) = \frac{p_{kj} f_j}{\sum_{j_0=1}^J p_{kj_0} f_{j_0}}, \quad j, k \in \{1, \dots, J\} \quad (7)$$

where p_{kj} are the entries of P_X . Matrix \bar{P}_X is again a transition matrix; each column sums up to one. It follows that

$$\mathbf{f} = \bar{P}_X \mathbb{E}[\underline{\mathbf{f}}^* | \mathbf{f}] \quad (8)$$

An unbiased moment estimator of \mathbf{f} is therefore given by

$$\tilde{\mathbf{f}} = \bar{P}_X \mathbf{f}^* \quad (9)$$

In general, $\bar{P}_X \neq P_X^{-1}$, which can be proved by showing that $\bar{P}_X P_X$ is not equal to the identity matrix. A more intuitive explanation is that \bar{P}_X changes when the survey data change, whereas P_X can be determined independently from the data and hence does not necessarily change when the data change. The variance of (9) is given by (6) where P_X^{-1} is replaced by \bar{P}_X and f_j is estimated by \tilde{f}_j , for $j \in \{1, \dots, J\}$.

In the remainder of this section we compare estimators (5) and (9). The first difference is that (5) might yield an estimate where some of the entries are negative, whereas (9) will never yield negative estimates (see e.g., De Wolf et al. 1997).

Secondly, Estimator (9) is more efficient than (5) in the univariate case. This is already discussed in Kuha and Skinner (1997). Consider the case where X has two categories. Let us say we want to know $\pi = \pi_1 = \mathbb{P}(X_i = 1 | i \in s)$. Let $\hat{\pi}$ be the estimate using P_X and $\tilde{\pi}$ the estimate using \bar{P}_X . The efficiency of $\hat{\pi}$ relative to $\tilde{\pi}$ is given by

$$\text{eff}(\hat{\pi}, \tilde{\pi}) = \frac{V[\tilde{\pi}]}{V[\hat{\pi}]} = (p_{11} + p_{22} - 1)^2 (\bar{p}_{22} - \bar{p}_{21})^2 < 1 \quad (10)$$

So $\tilde{\pi}$ is always more efficient than $\hat{\pi}$. An important difference from the general situation of misclassification is that in the situation of PRAM, matrices P_X and \bar{P}_X are given and do not have to be estimated. Comparison (10) is therefore a simple form of the comparison in Kuha and Skinner (1997, Section 28.5.1.3.).

The third comparison is between the maximum likelihood properties of (5) and (9). Assume that X_1, \dots, X_n are independently multinomially distributed with parameter vector $\boldsymbol{\pi} = (\pi_1, \dots, \pi_J)^t$. In the framework of misclassification, Hochberg (1977) proves that Estimator (9) yields an MLE. When (5) yields an estimate in the interior of the parameter space, the estimate is also an MLE. See Appendix A for the maximum likelihood properties of (9) and (5). Note that the likelihood function corresponding to (9) is different from the likelihood function corresponding to (5), since the information used is different. This explains why both can be MLEs despite being different estimators of \mathbf{f} .

The fourth comparison is with respect to transition matrices of Cartesian products and is less favourable for (9). It has already been noted that $P_{X_1} \otimes P_{X_2}$ is the matrix with misclassification probabilities for the Cartesian product $X = (X_1, X_2)$ (see 4). Analogously, given \bar{P}_{X_1} and \bar{P}_{X_2} the user can construct matrix $\bar{P}_{X_1} \otimes \bar{P}_{X_2}$. However, this

matrix does *not* necessarily contain calibration probabilities for X . Note that

$$\begin{aligned} & \mathbb{P}(X_i = (j_1, j_2) | X_i^* = (k_1, k_2), i \in s) \\ &= \frac{p_{k_1 j_1} p_{k_2 j_2} \mathbb{P}(X_i = (j_1, j_2) | i \in s)}{\sum_v^{J_1} \sum_w^{J_2} p_{k_1 v} p_{k_2 w} \mathbb{P}(X_i = (v, w) | i \in s)} \end{aligned} \tag{11}$$

It follows that $\bar{P}_X = \bar{P}_{X_1} \otimes \bar{P}_{X_2}$ when X_1 and X_2 are independent. In general, this independence is not guaranteed and since the user of the released data does not have the frequencies of X , he or she cannot construct \bar{P}_X . In other words, there is no general analogue to (4) for calibration probabilities and when $\bar{P}_{X_1} \otimes \bar{P}_{X_2}$ is used, results will be biased when X_1 and X_2 are correlated.

The fifth and last comparison is with respect to the creation of subgroups. Consider the situation where a user of the released data creates a subgroup by using a grouping variable that is not part of X . When the number of categories in the subgroup is smaller than J , Estimate (9) is not well-defined. When the number of categories is equal to J , Estimate (9) is biased due to the fact that (8) does not hold. Note with respect to (8) that the frequencies that are used to construct \bar{P}_X are the frequencies in the whole sample which will differ from the frequencies in the subgroup. The Estimator (5) is still valid for the subgroup.

Since calibration probabilities contain information about the distribution of the sample s , they perform better than misclassification probabilities regarding the univariate case. However, in a multivariate setting this advantage may disappear. Section 8 presents some simulation results.

6. Misclassification Proportions

Matrices P_X and \bar{P}_X inform about the expected change due to PRAM. As an alternative we propose to use transition matrices that inform about the actual change due to an application of PRAM. These matrices contain proportions and will be denoted P_X° and \bar{P}_X° . Matrix P_X° contains *misclassification proportions* and \bar{P}_X° contains *calibration proportions*. This section shows how P_X° and \bar{P}_X° are computed and discusses properties of these matrices.

We start with an example. Suppose that X has categories $\{1,2\}$. Assume that applying PRAM yields the cross-classification in Table 1. From this table it follows that the proportion of records with $X = 1$ that have $X^* = 1$ in the released sample is $300/400 = 3/4$ and that the proportion of records with $X^* = 1$ that have $X = 1$ in the original sample

Table 1. Classification by X^* and X

X^*	X		Total
	1	2	
1	300	200	500
2	100	400	500
Total	400	600	1,000

is $300/500 = 3/5$. Analogously we get the other entries of

$$P_X^\circ = \begin{pmatrix} 3/4 & 1/3 \\ 1/4 & 2/3 \end{pmatrix} \quad \text{and} \quad \overleftarrow{P}_X^\circ = \begin{pmatrix} 3/5 & 1/5 \\ 2/5 & 4/5 \end{pmatrix}$$

For the general construction of P_X° and $\overleftarrow{P}_X^\circ$, let the cell frequencies in the cross-classification X^* by X be denoted c_{kj} , for $k, j \in \{1, \dots, J\}$. The entries of the $J \times J$ transition matrices with the proportions are given by

$$p_{kj}^\circ = \frac{c_{kj}}{f_j} \quad \text{and} \quad \overleftarrow{p}_{jk}^\circ = \frac{c_{kj}}{f_k^*}$$

where $k, j \in \{1, \dots, J\}$.

It follows that $\mathbf{f}^* = P_X^\circ \mathbf{f}$ and $\mathbf{f} = \overleftarrow{P}_X^\circ \mathbf{f}^*$. This is the reason to consider the matrices with the proportions more closely, since it is a great improvement on (5) and (9). Note that when the user of the released sample has P_X° or $\overleftarrow{P}_X^\circ$, he or she can reconstruct Table 1.

Conditional on \mathbf{f} , P_X° and $\overleftarrow{P}_X^\circ$ are stochastic, whereas P_X and \overleftarrow{P}_X are not. In expectation P_X° equals P_X , and $P_{X_1}^\circ \otimes P_{X_2}^\circ$ equals $P_{X_1} \otimes P_{X_2}$ (see Appendix B). However, since \mathbf{f}^* is a realisation of a stochastic process and $f_k^* = 0$, for a $k \in \{1, \dots, J\}$, is possible, the expectation of $\overleftarrow{P}_X^\circ$ does not exist. Nevertheless, an approximation shows that $\overleftarrow{P}_X^\circ$ will be close to \overleftarrow{P}_X (see Appendix B).

There is a setback with respect to the use of proportions for Cartesian products and this is comparable to the problem mentioned in the previous section. Given $P_{X_1}^\circ$ and $P_{X_2}^\circ$ the user can construct $P_{X_1}^\circ \otimes P_{X_2}^\circ$ for $X = (X_1, X_2)$. However, $P_{X_1}^\circ \otimes P_{X_2}^\circ$ does *not* contain proportions as defined above. Note that the user does not have the cross-classification of X and X^* , so he or she cannot derive the proportions in P_X° . The same holds for $\overleftarrow{P}_X^\circ$. The optimal use of misclassification proportions is thereby limited to the univariate case.

Since misclassification proportions contain information about the actual perturbation due to PRAM, we expect them to perform well also in the multivariate case. Section 8 discusses a multivariate example.

7. Disclosure Risk

There are several ways to measure disclosure risk: see e.g., Skinner and Elliot (2002) and Domingo-Ferrer and Torra (2001). This section discusses two measures for disclosure risk with respect to PRAM. Section 7.1 discusses an extension of the general measure of disclosure risk introduced by Skinner and Elliot (2002). Section 7.2 introduces a measure that links up with the scenario of spontaneous recognition. In Section 7.3 the two measures are compared.

7.1. The Measure θ

The following describes how the general measure for disclosure risk introduced in Skinner and Elliot (2002) can be extended to the situation where PRAM is applied before data are released by the statistical agency. When a disclosure control method such as PRAM has been applied, a measure for disclosure risk is needed to quantify the protection that is offered by the control method. Scenarios that may lead to a disclosure of the identity of a

respondent are about persons that aim at disclosure and that may have data that overlap the released data. A common scenario is that a person has a sample from another source and tries to identify respondents in the released sample by matching records. Using the extension of the measure in Skinner and Elliot (2002) we can investigate how applying PRAM reduces the disclosure risk.

Under simple random sampling Skinner and Elliot (2002) introduced the measure of disclosure risk $\theta = \mathbb{P}(\text{correct match}|\text{unique match})$ as

$$\theta = \frac{\sum_j I(f_j = 1)}{\sum_j F_j I(f_j = 1)}$$

where the summations are over $j = 1, \dots, J$. The measure θ is the proportion of correct matches among those population units which match a sample unique. The measure is sample dependent and a distribution-free prediction is given by

$$\hat{\theta} = \nu n_1 / (\nu n_1 + 2(1 - \nu)n_2)$$

where ν is the sampling fraction, $n_1 = \sum_j I(f_j = 1)$ is the number of uniques and $n_2 = \sum_j I(f_j = 2)$ is the number of twins in the sample (see Skinner and Elliot 2002). Elamir and Skinner (2003) extended θ for the situation where misclassification occurs. The extension is given by

$$\theta_{mm} = \frac{\sum_{i \in s} I(f_{X_i} = 1, X_i^* = X_i)}{\sum_j F_j I(f_j = 1)}$$

and its distribution-free prediction is given by

$$\hat{\theta}_{mm} = \nu \sum_j I(f_j = 1) p_{jj} / (\nu n_1 + 2(1 - \nu)n_2)$$

where p_{jj} is the diagonal entry (j, j) of the transition matrix P_X which describes the misclassification (see Elamir and Skinner 2003).

Section 8 presents some simulation results with respect to the measure θ before applying PRAM, and θ_{mm} after applying PRAM.

7.2. Spontaneous Recognition

Data can be released in several ways under various conditions. One way is the releasing of detailed survey data under contract, i.e., data are released to bona fide research institutes that sign an agreement in which they promise not to look for disclosure explicitly, e.g., by matching the data to other data files. In this situation, SDC only concerns the protection against what is called *spontaneous recognition* (see Willenborg and De Waal 2001, Section 2.9.1). This section introduces a measure for disclosure risk for PRAM data that is specific to the control for spontaneous recognition.

Controlling for spontaneous recognition means that one should prevent that certain records attract attention. A record may attract attention when a low dimensional combination of its values has a low frequency. Also, without cross-classifying, a record may attract attention when one of its values is recognized as being very rare in the

population. Combinations of values with low frequencies in the sample are called *unsafe combinations*.

Note that applying PRAM causes two kinds of modifications in the sample that make disclosure more difficult. First, it is possible that unsafe combinations in the sample change into apparently safe ones in the released sample; second, it is possible that safe combinations in the sample change into apparently unsafe ones. Since misclassification probabilities are not that large (in order to keep analysis of the released sample possible) and the frequency of unsafe combinations is typically low, the effect of the first modification is negligible in expectation. The second modification is more likely to protect an unsafe combination j when there are a lot of combinations k , $k \neq j$, which are misclassified into j . This is the reason to focus, for a given record i with the unsafe combination of scores j , on the calibration probability

$$\mu_j = \mathbb{P}(X_i = j | X_i^* = j), \quad i \in s$$

When there are hardly any k , $k \neq j$, misclassified into j , this probability will be large, and, as a consequence, the record is unsafe. Note that combinations with frequency equal to zero are never unsafe. Calibration probabilities are also used by De Wolf and Van Gelder (2004), who investigate disclosure risk per record after PRAM is applied.

Assume that the number of unsafe combinations in the original sample is C and that μ_1, \dots, μ_C are computed. One way to obtain an overall disclosure risk measure is to use $\mu = \max\{\mu_1, \dots, \mu_C\}$, and this is the choice that is made in the example in Section 8.2. This choice is sensitive to outliers: if a relatively small number of records lead to a high maximum, one might consider deleting these records from the sample, or use local suppression of values of identifying variables in these records.

Measure μ is a simplification since it ignores possible correlation between X and other variables in the sample. Note that X will be a Cartesian product and that the statistical agency can compute the calibration probability μ using (7) since the agency has the frequencies of X .

7.3. Comparing the Measures θ and μ

Measure θ is a measure for an entire data set, whereas measure μ is a measure for a single unsafe record. The following compares both measures and makes some remarks about using the measures in practice.

The intuition behind θ is that when sample uniques have large corresponding population frequencies, the disclosure risk is small, i.e., θ will be close to zero. In other words, if all sample uniques are population uniques, θ will be one. As a referee pointed out, θ is sensitive to sample uniques that have large corresponding population frequencies. When there are several sample uniques, one of which has a large corresponding population frequency, θ will be close to zero. This situation is not likely to occur when the sample design is reasonable and the sample size is not too small. Note, however, that when the sample size increases, the probability that a sample unique is a population unique also increases. With respect to the usage of θ in complex sample designs, see Skinner and Elliot (2002, Section 4).

What is stated about θ holds for θ_{mm} although the misclassification softens the conclusions. For instance, if all sample uniques in the original data are population uniques and PRAM is applied, θ_{mm} will not be equal to one, but close to one.

The intuition behind μ is that protection is offered by safe combinations that change into combinations that are unsafe in the original data set. Whether sample uniques in the original data actually need this protection because of corresponding population uniques is neglected by μ .

When it comes to PRAM in practice, the statistical agency has to make decisions about which disclosure scenario is relevant and whether it is wise to release a large data set. Note that releasing data under contract combined with the scenario of spontaneous recognition is a rather specific form of releasing survey data (compare Willenborg and De Waal 1996, Section 3.3, where SDC policies of various statistical offices are summarised). Under the scenario of spontaneous recognition, a large data set is an advantage, as will be illustrated in Section 8.2. Under the more general scenario where users may actually aim at disclosure, a large data set is a disadvantage since the probability that a sample unique is a population unique increases when the sample size increases, as will be illustrated in Section 8.1.

It is clear that additional research into some aspects of θ_{mm} and μ is still needed. For instance, it would be interesting to know whether measure μ can be combined with per record measures of disclosure risk for data that have not been perturbed (see e.g., Mokken et al. 1992 and the discussion on reidentification risk in Willenborg and De Waal 2001, Section 2.5). A problem is that applying PRAM is not independent with respect to the information that is used to determine the reidentification risk and hence that it is not beforehand clear how to combine reidentification risk and μ .

8. Simulation Results

The objective of this section is to illustrate the theory in the previous sections and to investigate disclosure risk and information loss for different choices of misclassification parameters. The population is chosen to consist of units with complete records in the British Household Survey 1996–1999. We have $N = 16,710$ and we distinguish five identifying variables with respect to the household owner: Sex (S), Marital Status (M), Economic status (D), Socio-Economic Group (E), and Age (A), with number of categories 2, 7, 10, 7, and 8, respectively. In the following we consider simple random sampling without replacement from the population where the sample fraction ν is equal to 0.05, 0.10 or 0.15. The three samples are denoted s_1 , s_2 and s_3 and have sample sizes 836, 1,671 and 2,506, respectively.

The transition matrices used to apply PRAM to the selected variables are mostly of a simple form and are parameterised by $p_d \in (0, 1)$. The idea is as follows. When PRAM is applied, the diagonal probabilities in the transition matrices are fixed and equal to p_d for all selected variables. In the columns, the probability mass $1 - p_d$ is equally divided over the entries that are not diagonal entries. A more sophisticated construction of the transition matrices can reduce the disclosure risk further. An example of this fine-tuning will be given.

8.1. Disclosure Risk and the Measure θ

The following discusses disclosure risk by comparing the measure θ before PRAM is applied with the measure θ_{mm} after PRAM has been applied (see Section 7.1). The identifying variables are described by $X = (S, M, D, E, A)$ with $J = 7,840$ possible categories.

Table 2 presents simulation results using simple random sampling without replacement using different sampling fractions ν and different choices of p_d . Given a choice of ν and p_d , each simulation consists of drawing a sample and applying PRAM to the sample. Since the population is known, measures θ , θ_{mm} , and predictions $\hat{\theta}$, $\hat{\theta}_{mm}$ can be computed in each simulation. Note that θ and $\hat{\theta}$ reflect the risk before applying PRAM and θ_{mm} and $\hat{\theta}_{mm}$ reflect the risk after applying PRAM. The number of simulations is 100 and the means of the computed and predicted measures are reported in Table 2.

It is clear from Table 2 that applying PRAM will reduce the risk. For example, when $p_d = 0.80$ and $\nu = 0.10$, applying PRAM reduces the risk from $\theta = 0.166$ to $\theta_{mm} = 0.055$. When p_d decreases, disclosure risk decreases too, as one might expect. Note that disclosure risk increases when sample size increases. In a larger sample, a record with a unique combination of scores is more likely to be a population unique and therefore the danger of a correct match is higher.

8.2. Disclosure Risk and Spontaneous Recognition

The following illustrates the measure μ for disclosure risk for spontaneous recognition that is discussed in Section 7.2.

Only combinations of three identifying variables are assessed with respect to disclosure control for spontaneous recognition. By doing so we follow the rule of thumb that is also used by Statistics Netherlands: a recognition of a combination of values of more than three variables is not spontaneous any more. So there are ten groups to consider. We will discuss only one of them, namely the group defined by $X = (M, D, E)$. The number of categories of X is 490. The measure for disclosure risk is given by

$$\mu_{(m,d,e)} = \mathbb{P}((M, D, E) = (m, d, e) | (M^*, D^*, E^*) = (m, d, e))$$

Table 2. Simulation results of disclosure risk measures for $X = (S, M, D, E, A)$ before and after applying PRAM with p_d

p_d	ν	$\bar{\theta}$	$\bar{\hat{\theta}}$	$\bar{\theta}_{mm}$	$\bar{\hat{\theta}}_{mm}$
0.95	0.05	0.084	0.087	0.061	0.065
	0.10	0.151	0.147	0.112	0.110
	0.15	0.217	0.215	0.165	0.166
0.90	0.05	0.087	0.086	0.047	0.051
	0.10	0.148	0.157	0.083	0.087
	0.15	0.213	0.216	0.123	0.127
0.80	0.05	0.087	0.090	0.028	0.031
	0.10	0.166	0.151	0.055	0.054
	0.15	0.213	0.221	0.065	0.064

for those combinations of values (m, d, e) that have frequency 1 in sample s_1, s_2 or s_3 . Note that when PRAM is not applied, $\mu = 1$. Other approaches are possible. For example, the threshold to decide whether a record is unsafe – a frequency count equal to 1 above – could be made dependent on the sample size.

Table 3 presents results with respect to the maximum of μ when PRAM is applied to M, D and E , independently. Given a choice of p_d , PRAM is applied only once to the chosen sample. With respect to $X = (M, D, E)$ the number of unique combinations in s_1, s_2 and s_3 are 48, 44, and 53, respectively.

We draw two conclusions from the results. First, the results illustrate that the probability p_d matters, as one might expect. Second, the results show that the size of the sample is important. In order to protect an unsafe combination j , it is necessary that there are a lot of combinations that can change into j due to PRAM. Note that this is the other way around compared to the measure θ where a larger sample size causes a higher disclosure risk. This difference shows that different concepts of disclosure induce different methods for disclosure control.

The following introduces a method to fine-tune a transition matrix and shows that this can help to diminish the disclosure risk. The idea is to adjust one or more columns in the transition matrix of each variable that is part of an unsafe combination. Consider P_{X_1} where variable X_1 has J_1 categories. The column that is chosen first corresponds to the category of X_1 with the highest frequency in sample s , say column j . Let furthermore k be the number that corresponds to the category of X_1 with the lowest frequency in s . The columns of P_{X_1} that are not column j are constructed as explained at the beginning of this section: p_d on the diagonal and $(1 - p_d)$ equally divided over the other entries. Column j is fine-tuned by

$$p_{ij} = \begin{cases} p_d & \text{if } l = j \\ (1 - p_d)/\eta & \text{if } l = k, \\ (1 - p_d)/(\eta(J_1 - 2)) & \text{if } l \neq j, k \end{cases} \quad (12)$$

for $l \in \{1, \dots, J_1\}$ and $\eta > 1$. The idea here is that when we choose η close to 1, the category with the highest frequency has a relatively high probability of changing into the category with the lowest frequency. Assuming a link between an unsafe combination and a low frequency in the original sample, this idea explicitly supports the concept of PRAM:

Table 3. Maximum of μ for values of (M,D,E) with frequency 1 when applying PRAM to M, D and E

	p_d					
	0.95	0.90	0.85	0.80	0.70	0.60
Sample						
s_1 with $n = 836$	0.94	0.86	0.76	0.65	0.43	0.24
s_2 with $n = 1,671$	0.94	0.85	0.74	0.61	0.36	0.18
s_3 with $n = 2,506$	0.93	0.83	0.69	0.55	0.31	0.15

an unsafe combination c is after PRAM protected by creating new combinations c from combinations that have high frequencies in the original sample.

In the same way other columns in P_{X_1} can be fine-tuned. For example, the second column chosen is the column that corresponds to the category of X_1 with the second highest frequency in sample s , and the chosen row is now the row that corresponds to the category of X_1 with the second lowest frequency in sample s .

A different method to fine-tune a PRAM transition matrix is described in Willenborg and De Waal (2001, Section 5.5), where two optimisation models are defined that maximise the diagonal entries given restrictions, some of which are formulated with respect to the desired flow from safe combinations of scores to unsafe combinations. The reason we opt for the method described above is that it links up with the p_d -parameterisation, it is simpler than the optimisation approach, and it does not need extra optimisation software.

Table 4 presents results for sample s_3 when the transition matrices of M , D , and E are fine-tuned. The advantage of fine-tuning the transition matrices is dependent on the data and on the size of the sample. One can see that the idea works e.g., if $p_d = 0.80$, fine-tuning can decrease the maximum of μ from 0.55 to 0.34.

Even after using fine-tuning, the maximum of μ is still quite large. Additional simulations, not reported, show that the maximum of μ decreases rapidly when sample size is increased. The conclusion and advice are: determine a largest tolerated μ and check all combinations of three identifying variables and use fine-tuning. The protection offered by PRAM depends on p_d , but also very much on the sample size.

8.3. Bivariate Frequency Estimation

Using simulation, this section discusses two examples concerning efficiency in frequency estimation when PRAM is applied. The first example is about the efficiency of misclassification probabilities and calibration probabilities, the second is about the efficiency of misclassification probabilities and misclassification proportions.

In the multivariate situation calibration probabilities do not always work well. To illustrate this, the following example is about bivariate frequency estimation concerning sample s_3 and variable $X = (S, E)$ that has 14 categories. The χ^2 -test of independence between S and E yields 529.55, where $df = 6$ and the p -value < 0.001 . It is this lack of independence between the variables that causes calibration probabilities to perform badly.

Table 4. Maximum of μ for values of (M, D, E) with frequency 1 in sample s_3 when using fine-tuning for all three variables

	p_d				
	0.95	0.90	0.85	0.80	0.70
No fine-tuning	0.93	0.83	0.69	0.55	0.31
Fine-tuning 1 column where $\eta = 1.001$	0.92	0.80	0.66	0.51	0.28
Fine-tuning 2 columns where $\eta = 1.001$	0.91	0.74	0.57	0.42	0.20
Fine-tuning 3 columns where $\eta = 1.001$	0.90	0.72	0.52	0.34	0.15

The example consists of $B = 10$ simulations. In each simulation, PRAM is applied independently to S and E where both the 2×2 transition matrix P_S and the 7×7 matrix P_E are constructed using the one parameter parameterisation given by $p_d = 0.85$. After each application of PRAM, the bivariate frequency distribution of $X = (S, E)$ is estimated twice. The first estimation uses the observed distribution of $X^* = (S^*, E^*)$ and $P_S \otimes P_E$, and estimates are denoted $\hat{f}_{1b}, \dots, \hat{f}_{14b}$, for simulation $b \in \{1, \dots, B\}$. The second estimation uses $X^* = (S^*, E^*)$ and $\bar{P}_S \otimes \bar{P}_E$, and estimates are denoted $\tilde{f}_{1b}, \dots, \tilde{f}_{14b}$, $b \in \{1, \dots, B\}$. The plots (a) and (b) in Figure 1 contain the points (f_{jb}, \hat{f}_{jb}) and (f_{jb}, \tilde{f}_{jb}) , respectively, for $j \in \{1, \dots, 14\}$ and $b \in \{1, \dots, B\}$.

Figure 1 shows that the misclassification probabilities perform better, i.e., in general, the points (f_j, \hat{f}_j) are closer to the identity line than the points (f_j, \tilde{f}_j) . The variance is less when $\bar{P}_S \otimes \bar{P}_E$ is used, but the figure shows that in that case estimates are biased. Violating the independence assumption regarding the use of $\bar{P}_S \otimes \bar{P}_E$ has severe consequences.

Misclassification proportions are close to misclassification probabilities in the above example. Compare for instance

$$P_S = \begin{pmatrix} 0.85 & 0.15 \\ 0.15 & 0.15 \end{pmatrix} \quad \text{and} \quad P_S^\circ = \begin{pmatrix} 0.854 & 0.154 \\ 0.156 & 0.156 \end{pmatrix}$$

A simulation can be used to investigate the performance of misclassification probabilities versus misclassification proportions. The example consists of $B = 1,000$ simulations. The simulation of applying PRAM to S and E is the same as above but this time the second estimation of the bivariate frequency distribution of $X = (S, E)$ uses the observed distribution of $X^* = (S^*, E^*)$ and $P_S^\circ \otimes P_E^\circ$. The study compares using $P_S \otimes P_E$ versus using $P_S^\circ \otimes P_E^\circ$ by looking at the average point estimates of the sample frequencies, the root mean squared errors (RMSE), and the actual coverage percentages (ACPs) of 95% confidence intervals. The ACP is the percentage of the replicated perturbed samples for

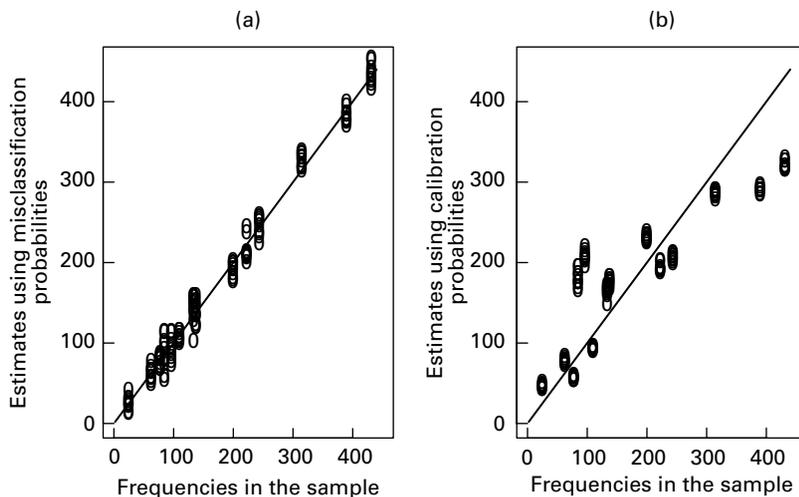


Fig. 1. Estimating frequencies of $X = (S, E)$ after applying PRAM to S and E in sample s_3 with $p_d = 0.85$ in ten simulations. (a) Using misclassification probabilities. (b) Using calibration probabilities

which the estimated confidence interval of the estimated frequency covers the actual frequency in the original sample.

Table 5 shows that misclassification proportions perform better than misclassification probabilities since the root mean squared error (RMSE) is smaller for the proportions. Table 5 also shows that in both situations the average point estimates are close to the true value and furthermore that the variance dominates the bias. The mean value of ACP when using $P_S \otimes P_E$ equals 95.1, and the mean value of ACP when using $P_S^\circ \otimes P_E^\circ$ equals 98.1. (A paired t -test yields a p -value < 0.001 .) When $P_S^\circ \otimes P_E^\circ$ is used, the estimation is conservative in the sense that the estimated confidence intervals show higher-than-nominal rates of coverage. This indicates that the estimation can be improved upon. Nevertheless, the results show that there is no false sense of precision associated with using the misclassification proportions in a bivariate situation. Although $P_S^\circ \otimes P_E^\circ$ does not contain the misclassification proportions of the Cartesian product $X = (S, E)$, it still produces more precise estimates than does $P_S \otimes P_E$.

9. Discussion

The article shows that the analysis of PRAM data is more efficient when misclassification proportions are used instead of misclassification probabilities. Calibration probabilities and calibration proportions work fine in the univariate case, but cause serious bias in the multivariate case. Since in most situations the user of PRAM data will be interested in multivariate analysis, it seems wise not to release calibration probabilities or calibration proportions together with the PRAM data. Given the results in this article, we advocate adapting the original idea of PRAM and releasing the misclassification proportions instead of the misclassification probabilities.

Table 5. Sample frequencies and average point estimates w.r.t. $X = (S, E)$ for sample s_3 and 1,000 simulated samples, where $p_d = 0.85$. (Root MSE and actual coverage percentage within parentheses.)

Sample frequencies	Average point estimates (RMSE; ACP)	
	Using $P_S \otimes P_E$	Using $P_S^\circ \otimes P_E^\circ$
$f_{(1,1)} = 76$	75.9 (10.8; 94.8)	75.9 (8.9; 97.7)
$f_{(1,2)} = 242$	241.4 (14.5; 95.8)	241.5 (12.3; 98.2)
$f_{(1,3)} = 132$	132.0 (14.4; 93.4)	132.2 (12.5; 96.9)
$f_{(1,4)} = 83$	83.2 (15.0; 94.6)	83.3 (12.4; 98.2)
$f_{(1,5)} = 430$	430.4 (16.9; 95.8)	430.3 (13.2; 99.0)
$f_{(1,6)} = 198$	198.3 (14.8; 96.2)	198.4 (12.8; 98.5)
$f_{(1,7)} = 61$	60.3 (11.0; 95.1)	60.3 (9.5; 97.5)
$f_{(2,1)} = 23$	23.0 (10.1; 94.8)	23.1 (8.9; 97.5)
$f_{(2,2)} = 136$	136.9 (14.3; 95.3)	136.5 (12.3; 98.0)
$f_{(2,3)} = 221$	220.5 (14.9; 95.1)	220.8 (12.5; 97.8)
$f_{(2,4)} = 388$	387.8 (16.6; 93.7)	387.7 (12.4; 99.2)
$f_{(2,5)} = 95$	94.8 (15.7; 94.9)	94.7 (13.2; 97.9)
$f_{(2,6)} = 313$	312.7 (15.9; 96.1)	312.6 (12.8; 98.7)
$f_{(2,7)} = 108$	108.8 (11.7; 95.7)	108.7 (9.5; 98.0)

Given that releasing misclassification proportions makes PRAM more efficient with respect to information loss, it is still an open question how this works out when PRAM is compared with other SDC methods (see Domingo-Ferrer and Torra 2001). It might be worthwhile to state that PRAM was never meant to replace existing SDC methods. Working with PRAM data and taking into account the information about the misclassification in the analysis might be quite a burden for some researchers. However, when researchers are interested in specific details in data, details that might disappear when e.g., global recoding is used, PRAM can be a solution. Note that PRAM is statistically sound. Data are perturbed, but information about the perturbation is at hand. Although estimates will have extra variance due to the perturbation, they will be unbiased.

An option that takes away some of the problems of users is to use data perturbed by PRAM as source data in a remote access situation. This idea was suggested by a referee. Users outside the statistical agency submit simple queries such as tabulations via the World Wide Web and receive output that has already been adjusted with respect to the perturbation. The advantage is that when the remote access system is hacked, source data are still protected against disclosure.

Since the misclassification proportions provide more information about the original sample than the misclassification probabilities, one should consider the question whether providing these proportions increases the disclosure risk. Since the privacy protection that is offered by PRAM is at the record level, we do not think that disclosure risk increases when misclassification proportions are released. With these proportions, sample frequencies of the identifying variables can be deduced, but these frequencies are not sensitive information. Note also that when one works with the measures for disclosure risk discussed in Section 7, the risk does not change when misclassification proportions are released.

With respect to applying PRAM in practice, there are problems that are not discussed in this article. For example, how shall one select the identifying variables? And how shall one choose threshold values for the measures of disclosure risk? Another problem in practice – one that we want to discuss briefly – is the situation where variables are strongly correlated and where applying PRAM introduces inconsistencies. A simple example taken from Gouweleeuw et al. (1998) is the situation where we have the variable Gender and a variable that denotes whether or not the respondent is or has been pregnant. When PRAM is applied independently to Gender, inconsistencies will be hard to prevent.

Inconsistencies might lead to a higher disclosure risk. An inconsistent record attracts the attention of persons that aim at disclosure since the inconsistency may give an indication with respect to the privacy protection of the underlying original record. We can prevent inconsistencies by applying PRAM to the Cartesian product of variables and carefully choosing the entries of the PRAM matrix of this product. In general, however, the problem of strongly correlated variables is difficult. When there are a lot of high-dimensional variables, a reasonable PRAM matrix of the Cartesian product might be difficult to determine and its size might be impractical. Furthermore, when a user of the released data wants to analyse a subset of the correlated variables, the misclassification information for that subset requires additional estimation. As an example, consider the categories $1 \equiv (m, np)$, $2 \equiv (m, p)$, $3 \equiv (f, np)$ and $4 \equiv (f, p)$ of $X = (G, P)$, where G denotes

Gender and P denotes whether or not the respondent is or has been pregnant. A possible transition matrix for X is given by

$$P_X = \begin{pmatrix} 0.8 & 0 & 0.1 & 0.1 \\ 0 & 1 & 0 & 0 \\ 0.1 & 0 & 0.8 & 0.2 \\ 0.1 & 0 & 0.1 & 0.7 \end{pmatrix}$$

in which case the misclassification of G is correlated with P . Since

$$\begin{aligned} \mathbb{P}(G^* = f | G = f) &= \mathbb{P}(G^* = f | P = p, G = f) \mathbb{P}(P = p | G = f) \\ &+ \mathbb{P}(G^* = f | P = np, G = f) \mathbb{P}(P = np | G = f) \end{aligned}$$

the user of the released data has to estimate the bivariate distribution of X in order to estimate the univariate misclassification of G .

We do not know a general solution for the inconsistency problem. However, when data are released to a group of users with specific research questions, applying PRAM to Cartesian products of variables might be an option. This will only work, when the data structure is not too complex and when the statistical agency is willing to attune the released data to the specific research questions. This is basically the same idea as described in Bycroft and Merret (2005), where stratification is used to preserve multivariate distributions.

In the situation where data are released to a group of users with different research questions, a combination of applying PRAM and data editing might be considered. Assume that the original microdata consist of consistent records. First, within the group of identifying variables, one selects those that have the lowest correlation with the other variables in the microdata. Second, PRAM is applied independently to the selected variables. Third, when the number of newly created inconsistencies is relatively small, PRAM data are made consistent using data editing software such as SLICE (De Waal, 2001). This will destroy the aforementioned statistical soundness of PRAM, but when the number of inconsistencies is small the damage might be kept within limits. Shlomo and De Waal (2005) investigate similar ideas and apply PRAM while minimising the number of newly created inconsistencies.

Appendix A

The following derives the maximum likelihood properties of (5) and (9). The reasoning is the same as in Hochberg (1977), but simpler, since in the PRAM situation calibration probabilities do not have to be estimated. Also, we show that the reasoning applies both to (5) and to (9).

Assume that the frequency vector of X_1, \dots, X_n is multinomially distributed with parameter vector $\boldsymbol{\pi} = (\pi_1, \dots, \pi_J)^t$, where $\pi_j > 0$ for $j \in \{1, \dots, J\}$, and $\sum_{j=1}^J \pi_j = 1$. Consider the transformation $\boldsymbol{\pi}^* = P\boldsymbol{\pi}$, where P is a $J \times J$ transition matrix, i.e., columns sum up to one and $p_{kj} \geq 0$ for $k, j \in \{1, \dots, J\}$. Assume that P is nonsingular. Let the distribution of X^* be given by $\mathbb{P}(X^* = k) = \pi_k^*$, for $k \in \{1, \dots, J\}$.

It follows that $X_1^*, X_2^*, \dots, X_n^*$ are multinomially distributed with parameters n and $\boldsymbol{\pi}^*$. Indeed, $\pi_k^* = p_{k1}\pi_1 + \dots + p_{kJ}\pi_J > 0$ for $k \in \{1, \dots, J\}$ and

$$\sum_{k=1}^J \pi_k^* = \left(\sum_{l=1}^J p_{l1} \right) \pi_1 + \dots + \left(\sum_{l=1}^J p_{lJ} \right) \pi_J = 1$$

The likelihood L^* for $\boldsymbol{\pi}^*$ and observed $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_n^*)^t$ is well-known. Let $\mathbf{f}^* = (f_1^*, f_2^*, \dots, f_J^*)^t$ denote the observed cell frequencies. The MLE is given by $\boldsymbol{\pi}^* = \mathbf{f}^*/n$ and has covariance matrix $\Omega = [\text{Diag}(\boldsymbol{\pi}^*) - \boldsymbol{\pi}^*(\boldsymbol{\pi}^*)^t]/n$, where $\text{Diag}(\boldsymbol{\pi}^*)$ is the diagonal matrix with the diagonal entries given by the elements of $\boldsymbol{\pi}^*$.

Next we can use the invariance property of maximum likelihood (see e.g., Mood et al. 1985, p. 285). Define the transformation $g(\boldsymbol{\pi}^*) = P^{-1}\boldsymbol{\pi}^*$. With respect to (5) maximum likelihood properties are proved by taking $P = P_X$ and obtaining $\hat{\boldsymbol{\pi}} = g(\hat{\boldsymbol{\pi}}^*) = P_X^{-1}\hat{\boldsymbol{\pi}}^*$. With respect to (9), the misclassification design is described by $\boldsymbol{\pi} = \bar{P}_X\boldsymbol{\pi}^*$, so $P = \bar{P}_X^{-1}$ and the MLE is given by $\hat{\boldsymbol{\pi}} = g(\hat{\boldsymbol{\pi}}^*) = \bar{P}_X\hat{\boldsymbol{\pi}}^*$. Since g has a first order derivative, the covariance matrices can be obtained using the delta method (see e.g., Agresti 1990, Chapter 12).

Appendix B

Let P_{kj}° denote the stochastic variable of the kj -th entry of P_X° and C_{kj} the stochastic number of observations in the kj -th cell in the cross-classification X^* by X . It follows that C_{kj} has a binomial distribution with parameters f_j and p_{kj} . Consequently, $\mathbb{E}[P_{kj}^\circ|\mathbf{f}] = \mathbb{E}[C_{kj}/f_j|\mathbf{f}] = f_j p_{kj}/f_j = p_{kj}$ and in expectation P_X° equals P_X . Since $C_{k_1j_1}$ and $C_{k_2j_2}$ are independent given \mathbf{f} , it follows that $\mathbb{E}[P_{k_1j_1}^\circ P_{k_2j_2}^\circ|\mathbf{f}] = p_{k_1j_1} p_{k_2j_2}$. So, in expectation, $P_{X_1}^\circ \otimes P_{X_2}^\circ$ equals $P_{X_1} \otimes P_{X_2}$.

Let $C_{k+} = \sum_{j=1}^J C_{kj}$. We define $\bar{P}_{jk} = C_{kj}/(C_{k+} + \varepsilon)$ where ε is a small positive value. Using the delta method (see e.g., Rice 1995, Section 4.6), we obtain

$$\mathbb{E}[\bar{P}_{jk}|\mathbf{f}] \approx \frac{\mathbb{E}[C_{kj}|\mathbf{f}]}{\mathbb{E}[C_{k+}|\mathbf{f}]} + \frac{1}{\mathbb{E}[C_{k+}|\mathbf{f}]^2} \left(V[C_{k+}|\mathbf{f}] \frac{\mathbb{E}[C_{kj}|\mathbf{f}]}{\mathbb{E}[C_{k+}|\mathbf{f}]} - \rho \sqrt{V[C_{kj}|\mathbf{f}]V[C_{k+}|\mathbf{f}]} \right)$$

where $\mathbb{E}[C_{kj}|\mathbf{f}] = f_j p_{kj}$ and ρ is the correlation between C_{kj} and C_{k+} . From this we see that the difference between $\mathbb{E}[\bar{P}_{jk}|\mathbf{f}]$ and \bar{p}_{jk} will be small when $V[C_{k+}|\mathbf{f}]$ is small and $\mathbb{E}[C_{k+}|\mathbf{f}]$ is large.

10. References

Agresti, A. (1990). *Categorical Data Analysis*. New York: Wiley.

Bethlehem J.G., Keller, W.J., and Pannekoek, J. (1990). Disclosure Control of Microdata. *Journal of the American Statistical Association*, 85, 38–45.

Bycroft, C. and Merret, K. (2005). Experience of Using a Post Randomisation Method at the Office for National Statistics. United Nations Economic Commission for Europe Work Session on Statistical Data Confidentiality.

Chaudhuri, A. and Mukerjee, R. (1988). *Randomized Response, Theory and Techniques*. New York: Marcel Dekker.

Dalenius, T. and Reiss, S.P. (1982). Data Swapping: A Technique for Disclosure Control. *Journal of Statistical Planning and Inference*, 6, 73–85.

- De Waal, T. (2001). SLICE: Generalised Software for Statistical Data Editing. In Bethlehem J.G. and van der Heijden P.G.M. (eds). *Proceedings in Computational Statistics*, Physica-Verlag, New York, 277–282.
- De Wolf, P.-P., Gouweleeuw, J.M., Kooiman, P., and Willenborg, L.C.R.J (1997). Reflections on PRAM. *Statistical Data Protection Conference Proceedings*, Lisbon, Eurostat, Luxembourg, 337-349.
- De Wolf, P.-P. and Van Gelder, I. (2004). An Empirical Evaluation of PRAM. Discussion Paper 04012, Voorburg/Heerlen, September: Statistics Netherlands.
- Domingo-Ferrer, J. and Torra, V. (2001). A Quantitative Comparison of Disclosure Control Methods for Microdata. In Doyle, P., Lane, J., Zayatz, L. and Theeuwes, J. (eds). *Confidentiality, Disclosure, and Data Access, Theory and Practical Application for Statistical Agencies*, Amsterdam: North-Holland.
- Elamir, E.A.H. and Skinner, C.J. (2003). Disclosure Risk Assessment for Microdata: The Treatment of Measurement Error. Forthcoming.
- Fienberg, S.E. and McIntyre, J. (2005). Data Swapping: Variations on a Theme by Dalenius and Reiss. *Journal of Official Statistics*, 21, 309–323.
- Gouweleeuw, J.M., Kooiman, P., Willenborg, L.C.R.J., and De Wolf, P.-P. (1998). Post Randomisation for Statistical Disclosure Control, Theory and Implementation. *Journal of Official Statistics*, 14, 463–478.
- Hochberg, Y. (1977). On the Use of Double Sampling Schemes in Analyzing Categorical Data with Misclassification Errors. *Journal of the American Statistical Association*, 72, 914–921.
- Horn, R.A. and Johnson, C.R. (1985). *Matrix Analysis*. Cambridge: Cambridge University Press.
- Kooiman, P., Willenborg, L.C.R.J., and Gouweleeuw, J.M. (1997). PRAM, A Method for Disclosure Limitation of Microdata. Research paper no. 9705, Voorburg/Heerlen: Statistics Netherlands.
- Kuha, J.T. and Skinner, C.J. (1997). Categorical Data Analysis and Misclassification. In Lyberg, L., Biemer, P., Collins, M., De Leeuw, E., Dippo, C., Schwarz, N. and Trewin, D. (eds). *Survey Measurement and Process Quality*, New York: Wiley.
- Mokken, R.J., Kooiman, P., Pannekoek, J., and Willenborg, L.C.R.J. (1992). Disclosure Risks for Microdata. *Statistica Neerlandica*, 46, 49–67.
- Mood, A.M., Graybill, F.A., and Boes, D.C. (1985). *Introduction to the Theory of Statistics*. Auckland: McGraw-Hill.
- Rice, J.A. (1995). *Mathematical Statistics and Data Analysis*. Belmont: Duxbury Press.
- Rosenberg, M.J. (1979). *Multivariate Analysis by a Randomised Response Technique for Statistical Disclosure Control*. Ph.D. Dissertation, University of Michigan, Ann Arbor.
- Sande, G. (1998). Comment on “Post Randomisation for Statistical Disclosure Control, Theory and Implementation” by Gouweleeuw et al. *Journal of Official Statistics*, 14, 479–484.
- Shlomo, N. and De Waal, T. (2005). Preserving Edits when Perturbing Microdata for Statistical Disclosure Control. S3RI Methodology Working Papers, M05/12 Southampton Statistical Sciences Research Institute.
- Skinner, C.J. and Elliot, M.J. (2002). A Measure of Disclosure Risk for Microdata. *Journal of the Royal Statistical Society, Series B*, 855–867.

- Van den Hout, A. and Van der Heijden, P.G.M. (2002). Randomized Response, Statistical Disclosure Control and Misclassification, A Review. *International Statistical Review*, 70, 269–288.
- Warner, S.L. (1965). Randomized Response: A Survey Technique for Eliminating Answer Bias. *Journal of the American Statistical Association*, 60, 63–69.
- Willenborg, L.C.R.J. and De Waal, T. (1996). *Statistical Disclosure Control in Practice*. New York: Springer.
- Willenborg, L.C.R.J. and De Waal, T. (2001). *Elements of Statistical Disclosure Control*. New York: Springer.

Received February 2004

Revised August 2006