An empirical evaluation of PRAM

Discussion paper 04012

Peter-Paul de Wolf and Ilan van Gelder* [+]

* Statistics Netherlands
[+] Hogeschool van Amsterdam

Statistics Netherlands

Voorburg/Heerlen, September 2004

Explanation of symbols

.	= data not available
*	= provisional figure
x	= publication prohibited (confidential figure)
–	= nil or less than half of unit concerned
–	= (between two figures) inclusive
0 (0,0)	= less than half of unit concerned
blank	= not applicable
2003–2004	= 2003 to 2004 inclusive
2003/2004	= average of 2003 up to and including 2004
2003/’04	= crop year, financial year, school year etc. beginning in 2003 and ending in 2004

Due to rounding, some totals may not correspond with the sum of the separate figures.

Publisher
Statistics Netherlands
Prinses Beatrixlaan 428
2273 XZ Voorburg
The Netherlands

Printed by
Statistics Netherlands - Facility Services

Cover design
WAT ontwerpers, Utrecht

Information
E-mail: infoservice@cbs.nl

Where to order
E-mail: verkoop@cbs.nl

Internet
http://www.cbs.nl

ISSN: 1572-0314
Key figure: X-10
Production code: 6008304012

Statistics Netherlands

AN EMPIRICAL EVALUATION OF PRAM

PRAM (Post Randomization Method) is a disclosure control method for microdata. In 1997 it was introduced in Kooiman et al. (1997) and discussed more extensively in Gouweleeuw et al. (1998), but PRAM has not yet been applied extensively by statistical agencies. This is partly due to the fact that, even though some theoretical results exist, little practical knowledge is available on the effect of PRAM on disclosure control as well as on the loss of information it induces.

In this paper, we will try to make up for this lack of knowledge, by supplying some empirical information on the behaviour of PRAM. To be able to achieve this, some basic measures for loss of information and disclosure risk will be introduced. PRAM will be applied to one specific microdata file of over 6 million records, using several models in applying the procedure.

Keywords: Disclosure control, PRAM, information loss, disclosure risk

1 Introduction

The Post Randomization Method (PRAM) was introduced in Kooiman et al. (1997) as a method for disclosure protection applied to categorical variables in microdata files. In Gouweleeuw et al. (1998) and de Wolf et al. (1998), the method and some of its implications were discussed in more detail.

PRAM produces a microdata file in which the scores on some categorical variables for certain records are changed with respect to the scores in the original microdata file. This is usually applied to identifying variables, i.e., variables that can be used to identify the respondent that corresponds to a record. This results in a microdata file with scores on identifying variables, that, with certain probability, are incorrect scores. Hence, the risk of identification of respondents is reduced: even in case one could make a link between a record in the microdata file and an individual, the possible incorrectness of the scores yields uncertainty on the correctness of the link.

Note that PRAM can be regarded as a form of misclassification, where the so called transition probabilities (i.e., the probabilities of changing a score into another score) completely determine the underlying probability mechanism. These transition probabilities are summarized in a Markov matrix called the PRAM-matrix. Contrary to the general situation, the probability mechanism that determines the misclassification is known in case of PRAM. Since the probability mechanism is known, some statistical analyses can still be performed legitimately, be it with a slight adjustment of the standard methods. See, e.g., van den Hout (1999), van den Hout and van der Heijden (2002) and Ronning et al. (2004). A similar situation of misclassification with known transition probabilities is the case of Randomized Response (see, e.g., Warner, 1965 and Chen, 1979). In that case it has been known for some time, that unbiased estimates of population parameters like moments of the underlying distribution can be obtained as well, see e.g., Press (1968) and Kuha and Skinner (1997).

In order to let a user make legitimate inference, the transition probabilities should hence be supplied to him. On the other hand, making use of the literature on inference about misclassification mechanisms (see e.g., Kuha and Skinner, 1997), even without the exact transition probabilities a user could still perform sound analyses.

When applying Statistical Disclosure Control (SDC) methods, one has to deal with two competing mechanisms: the microdata file has to be safe enough to guarantee the protection of individual respondents but at the same time the loss of information should not be too large. For a general discussion, see, e.g., Fienberg (1994). Moreover, these competing mechanisms are often the core of the discussion of SDC methods themselves, see, e.g., Domingo-Ferrer and Torra (2001a) and Domingo-Ferrer and Torra (2001b). However, quantifying the loss of information and the level of disclosure control can be done in several ways. We will introduce some basic measures to quantify the loss of information as well as a measure to determine the level of disclosure control in case of using PRAM.

In this paper we will apply PRAM to a microdata file of 6,237,468 records and discuss the effect of applying PRAM on the amount off information loss and the level of disclosure control, using different PRAM-matrices.

In section 2 we will give a brief description of PRAM. Moreover, in this section we will introduce the notation concerning PRAM that we will be using throughout the rest of the paper. The aim of this paper is to investigate the effect of different PRAM-matrices on disclosure control as well as information loss. In section 3 we will therefore define a measure to quantify the effect on disclosure control. Section 4 contains the definitions of the measures of information loss we used in our experiments. In section 5 we will introduce a notation that we will use to define the PRAM-matrices to be used in our experiments. Both the effect on disclosure control and the effect on the amount of information, in the different experiments we performed, will be given in section 6. Finally, in section 7 we will briefly summarize the results and draw some conclusions.

2 A short description of PRAM

In this section we will briefly describe the theory involving PRAM, mainly to introduce the notation we will use throughout this paper. For details we refer to Gouweleeuw et al. (1998).

Let ξ denote a categorical variable in the original file to which PRAM will be applied and let X denote the same variable in the perturbed file. Moreover, assume that ξ, and hence X as well, has K categories, labeled $1, \ldots, K$. The transition probabilities that define PRAM are denoted as

$$p_{kl} = \mathbb{P}(X = l \mid \xi = k), \tag{1}$$

i.e., the probability that an original score $\xi = k$ is changed into the score $X = l$, for all $k, l = 1, \ldots, K$. Using these transition probabilities as entries of a $K \times K$ matrix, we obtain a Markov matrix P that we will call the PRAM-matrix, denoted by P.

Applying PRAM now means that, given the score $\xi = k$ for record r, the score X for that record is drawn from the probability distribution $p_{k1}, \ldots, p_{kK}$. For each record in the original file, this procedure is performed independently of the other records.

To illustrate the ideas, suppose that the variable ξ is gender, with scores $\xi = 1$ if male and $\xi = 2$ if female. Applying PRAM with $p_{11} = p_{22} = 0.9$ on a microdata file with 100 males and 100 females, would yield a perturbed microdata file with again, in expectation, 100 males and 100 females. However, in expectation, 10 of these males were originally female, and similarly, 10 of the females were originally male.

More generally, the effect of PRAM on one-dimensional frequency tables is that

$$\mathbb{E}\left(\mathbf{T}_X \mid \xi\right) = P^t\, \mathbf{T}_\xi, \tag{2}$$

where $\mathbf{T}_\xi = (T_\xi(1), \ldots, T_\xi(K))^t$ denotes the frequency table according to the original microdata file and $\mathbf{T}_X$ the frequency table according to the perturbed microdata file. A conditionally unbiased estimator of the frequency table in the original file is then given by

$$\hat{\mathbf{T}}_\xi = \left(P^{-1}\right)^t\, \mathbf{T}_X. \tag{3}$$

This can be extended to two-dimensional frequency tables, by vectorizing such tables. The corresponding PRAM-matrix is then given by the Kronecker product of the PRAM-matrices of the individual dimensions. Alternatively, one could use the two-dimensional frequency tables $T_{\xi\eta}$ for the original data and T_{XY} for the perturbed data directly in matrix notation:

$$\hat{T}_{\xi\eta} = \left(P_X^{-1}\right)^t\, T_{XY}\, P_Y^{-1}. \tag{4}$$

3 Measure of disclosure control

In this section we will define the measure we used to specify the effects of the different PRAM-matrices on the level of (statistical) disclosure control.

An often used rule to determine the safety of microdata files is the so called threshold rule: whenever a certain combination of scores on identifying variables occurs less than a certain threshold, that combination is considered to be unsafe. As an example consider the case that the combination of gender, occupation and age is to be checked for the threshold rule. Moreover, assume that the threshold is chosen to be 50. Then, if only 43 female surgeons of age 57 exist in the population, each record that corresponds to a female surgeon of age 57 is considered to be an unsafe record. Even though the threshold rule is defined in terms of population frequencies, in practice one often only has the sample file at hand. In that case the rule is usually applied to that sample file, with an appropriately adjusted threshold.

In case of using PRAM as an SDC-method this rule does not make any sense: since the perturbed file is the result of a probabilistic experiment, the unsafe records would vary over each realization. To deal with this problem, an alternative approach was suggested in Rienstra (2003). In that approach, the disclosure risk is considered, i.e., the probability that given a score k in the perturbed file, the original score was k as well. By Bayes rule this can be calculated using

$$R_{\text{PRAM}}(k) = \mathbb{P}(\xi = k \mid X = k) = \frac{\mathbb{P}(X = k \mid \xi = k)\mathbb{P}(\xi = k)}{\sum_{l=1}^{K} \mathbb{P}(X = k \mid \xi = l)\mathbb{P}(\xi = l)}. \tag{5}$$

Assuming that PRAM is applied to (the combination of) variable(s) ξ and using the appropriate notation, one could estimate this by

$$\hat{R}_{\text{PRAM}}(k) = \frac{p_{kk} T_{\xi}(k)}{\sum_{l=1}^{K} p_{lk} T_{\xi}(l)}. \tag{6}$$

Note that we used $T_{\xi}(k)/n$ as an estimate of $\mathbb{P}(\xi = k)$, where n is the size of the original microdatafile.

According to the traditional threshold rule, a record is considered to be safe, whenever a certain combination of scores on identifying variables occurs more than a certain threshold d. A safe record can thus be linked with at least d records in the population with the same scores. In case this is done randomly, the probability that the record is linked with the correct entity in the population is less than or equal to $1/d$. In other words, the risk of disclosure would be at most $1/d$.

In order to link the PRAM-risk to the traditional threshold rule, we suggest to use the following definition: a record is considered to be safe, whenever

$$\hat{R}_{\text{PRAM}}(k) \leq \frac{T_{\xi}(k)}{d}, \tag{7}$$

where d is the threshold used in the threshold rule for the original microdata file. Since $T_{\xi}(k)$ is an estimate of the population frequency for score k of variable ξ, the risk of linking the record in the original file randomly with one of the $T_{\xi}(k)$ records in the population would yield a total risk of $1/d$.

Note that a safe record according to the original threshold rule applied to the original file, will be considered to be safe according to this rule as well. Moreover, the number of unsafe records according to (7) only depends on the original frequencies and the PRAM-matrix that is used, i.e., is independent of the realization.

4 Measures of information loss

In this section we will briefly define the measures of information loss we will use in our experiments.

4.1 Entropy based information loss

In Domingo-Ferrer and Torra (2001a) two measures of information loss, based on entropy arguments were introduced. The first measure is defined as

$$\mathrm{EBIL}(P, \mathcal{G}) = -\sum_{r \in \mathcal{G}} \sum_{k=1}^{K} \mathbb{P}(\xi = k \mid X = l_r) \log \mathbb{P}(\xi = k \mid X = l_r), \tag{8}$$

where $\mathcal{G}$ denotes the perturbed microdata file and l_r the score of record r in $\mathcal{G}$ on variable X. Note that the probabilities $\mathbb{P}(\xi = k \mid X = l)$ are in some sense the inverse of the entries of the PRAM-matrix P. We will denote these probabilities by $p_{lk}^{\leftarrow}$.

The second, closely related measure based on entropy arguments, is given by

$$\mathrm{IL}(P, \mathcal{F}, \mathcal{G}) = -\sum_{r \in \mathcal{G}} \log \mathbb{P}(\xi = k_r \mid X = l_r), \tag{9}$$

where k_r and l_r denote the score in record r on ξ in the original file $\mathcal{F}$ and the corresponding X in the perturbed file $\mathcal{G}$ respectively. The major difference between EBIL and IL is that the latter measure makes use of both the original file and the perturbed file.

We can rewrite these measures as

$$\mathrm{EBIL}(P, \mathcal{G}) = -\sum_{l=1}^{K} \sum_{k=1}^{K} T_X(l) p_{lk}^{\leftarrow} \log p_{lk}^{\leftarrow} \tag{10}$$

and

$$\mathrm{IL}(P, \mathcal{F}, \mathcal{G}) = -\sum_{l=1}^{K} \sum_{k=1}^{K} T_{\xi,X}(k, l) \log p_{lk}^{\leftarrow}, \tag{11}$$

where $T_{\xi,X}(k, l)$ denotes the number of records with score $\xi = k$ in the original file $\mathcal{F}$ and $X = l$ in the perturbed file $\mathcal{G}$. Since intuitively $T_{\xi,X}(k, l)$ should be close to $T_X(l) p_{lk}^{\leftarrow}$, we see that EBIL and IL will not differ much whenever the number of records is large enough, relative to the number of categories K.

Using similar arguments as in the derivation of the estimator of the PRAM-risk, the probabilities $p_{lk}^{\leftarrow}$ can be estimated by

$$\hat{p}_{lk}^{\leftarrow} = \frac{p_{kl} T_\xi(k)}{\sum_{m=1}^{K} p_{ml} T_\xi(m)}. \tag{12}$$

4.2 Frequency table based information loss

Often frequency tables are calculated for certain (crossings of) variables, as a first step in investigating a microdata file. Applying PRAM obviously effects these frequency tables, whenever one of the variables to which PRAM is applied is part of such a frequency table. Therefore, some measures of information loss will be defined, based on comparison of the original frequency tables with the estimated frequency tables, using an estimate that corrects for the fact that PRAM has been applied.

4.2.1 Relative differences

To measure the effect of PRAM on frequency tables, consider the median of the relative differences between the counts in the table T_ξ based on the original file and the counts in the estimate $\hat{T}_\xi$ based on the perturbed file:

$$\mathrm{RD}_d = \mathrm{Median}\left\{ \left| \frac{T_\xi(k) - \hat{T}_\xi(k)}{T_\xi(k)} \right|, \quad k = 1, \ldots, K \right\}, \tag{13}$$

where d denotes the dimension of the frequency table. In this paper we will only consider $d = 1, 2$.

Additionally, we will calculate the maximum relative difference mRD_d. Note that the relative difference is infinite whenever $T_\xi(k) = 0$ and $\hat{T}_\xi(k) \neq 0$. In our experiments, this only occurred in case of $d = 2$, with large numbers of categories for both variables. Hence, for $d = 2$, we will calculate the maximum over all finite relative differences and count the number of occurrences of infinity.

4.2.2 Additional variance

Another way to measure information loss, is to use the additional variance introduced by applying PRAM, when estimating one-dimensional frequency tables, i.e., the variance of the estimator (3). Obviously, the conditional variance-covariance matrix of $\hat{\mathbf{T}}_\xi$ in equation (3) is given by

$$\Sigma_{\hat{\mathbf{T}}_\xi} = \mathrm{Var}(\hat{\mathbf{T}}_\xi \mid \xi) = \mathrm{Var}\left((P^{-1})^t \mathbf{T}_X \mid \xi\right) = \left(P^{-1}\right)^t \mathrm{Var}(\mathbf{T}_X \mid \xi) P^{-1}. \tag{14}$$

In Gouweleeuw et al. (1998) it is shown that

$$\mathrm{Var}(\mathbf{T}_X \mid \xi) = \sum_{k=1}^{K} T_\xi(k) V_k, \tag{15}$$

where the V_k are matrices with entries $V_k(l, j)$ given by

$$V_k(l, j) = \begin{cases} p_{kl}(1 - p_{kl}) & \text{in case } l = j \\ -p_{kl}\, p_{kj} & \text{in case } l \neq j \end{cases} \qquad l, j = 1, \ldots, K.$$

To obtain a single figure as a measure of information loss, we will use the median of the coefficients of variation of the categories of the one-dimensional frequency table. I.e., we will use

$$\mathrm{CV} = \mathrm{Median}\left\{ \frac{\sqrt{\Sigma_{\hat{\mathbf{T}}_\xi}(k, k)}}{T_\xi(k)}, \quad k = 1, \ldots, K \right\}. \tag{16}$$

Additionally, we will calculate the maximum coefficient of variation mCV over the K categories. In our experiments we have that $T_\xi(k) > 0$ for all categories k of all one-dimensional variables ξ, i.e., the coefficients of variation we consider, are all finite.

4.3 Linear regression based information loss

A second type of statistical analysis that is often used to explore a microdata file, is linear regression. Since PRAM effects categorical variables, a way to measure the loss of information, is to consider a linear regression on a categorical variable and to compare the regression coefficients estimated using the original file with those estimated using the perturbed file.

In this paper we will consider a linear regression model, with income as the dependent variable and a perturbed variable as explanatory variable. I.e., we will use the model

$$Y = \mathbb{E}\left(\sum_{k=1}^{K} \beta_k \delta(k)\right), \tag{17}$$

with Y the dependent variable income and $\delta(k)$ a dummy variable corresponding to the k-th category of variable ξ on which PRAM is applied. The regression coefficients $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)^t$ are estimated, based on the original microdata file, by

$$\boldsymbol{\beta} = \left[\text{diag}\left(T_\xi(1), \ldots, T_\xi(K)\right)\right]^{-1} \mathbf{T}_\xi^y, \tag{18}$$

where $T_\xi^y(k) = \sum_{r \in \mathcal{F}} Y_r \delta_{\xi,r}(k)$, the sum of the response on income over all records with score $\xi = k$. When PRAM is applied to ξ, the regression coefficients β_k can be estimated using

$$\tilde{\beta} = \left[\text{diag}\left(\hat{T}_\xi(1), \ldots, \hat{T}_\xi(K)\right)\right]^{-1} \left(P^{-1}\right)^t \mathbf{T}_X^y, \tag{19}$$

where $\hat{\mathbf{T}}_\xi$ is given in (3) and $\mathbf{T}_X^y$ is the analogous of $\mathbf{T}_\xi^y$ based on the perturbed file. The measure for the loss of information is then given by

$$\text{LRD} = \text{Median}\left\{\left|\frac{\beta_k - \tilde{\beta}_k}{\beta_k}\right|, \quad k = 1, \ldots, K\right\}. \tag{20}$$

Additionally, we will calculate the maximum relative difference mLRD over the K regression coefficients.

5 PRAM-matrices

In this section we introduce a notation that we will use to describe the PRAM-matrices of our experiments. We will use three basic types:

- Band matrices $nB(p; b)$, with p the value of the diagonal elements, b the bandwidth (i.e., the number of entries p_{kl} with $|k - l| < b$) and n the size of the square matrix. The probability mass $(1 - p_{kk})$ is distributed equally over the off-diagonal elements in the band. E.g., a $4B(0.6; 2)$ matrix would look like

$$\begin{pmatrix} 0.6 & 0.4 & 0 & 0 \\ 0.2 & 0.6 & 0.2 & 0 \\ 0 & 0.2 & 0.6 & 0.2 \\ 0 & 0 & 0.4 & 0.6 \end{pmatrix}$$

- Fully filled matrices, with equal off-diagonal elements, denoted by $nE(p)$, with n the size of the square matrix and p the value of the diagonal elements. E.g., a $3E(0.8)$ matrix would look like

$$\begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}$$

- Fully filled matrices, with the off-diagonal elements depending on the corresponding frequencies in the original microdata file, denoted by $nF(p)$, with n the size of of the square matrix and p the value of the diagonal elements. The off-diagonal elements are determined using a method defined in Rienstra (2003):

$$p_{kl} = \frac{(1 - p_{kk}) \left(\sum_{i=1}^{K} T_\xi(i) - T_\xi(k) - T_\xi(l) \right)}{(n-2) \left(\sum_{i=1}^{K} T_\xi(i) - T_\xi(k) \right)} \tag{21}$$

E.g., with $\mathbf{T}_\xi = (5576, 24, 632)^t$, the matrix $3F(0.6)$ would look like

$$\begin{pmatrix} 0.6000 & 0.3854 & 0.0146 \\ 0.0407 & 0.6000 & 0.3593 \\ 0.0017 & 0.3983 & 0.6000 \end{pmatrix}$$

Note that $1B(1;1) = 1E(1) = 1F(1)$ is a special case that we will denote by $\mathbb{1}$. The three basic types can be combined into block-matrices. We will denote these block matrices by Block(m; type$_1$; $\cdots$; type$_m$), with m the number of blocks and following m the matrix type for each block. Note that, using this construction, the diagonal elements of a PRAM-matrix will be constant within each block, but may vary between the blocks.

6 The empirical results

In our experiments we used one microdata file of 6 237 468 records, representing a complete population and containing the categorical variables Gender (with 2 categories), Marital status (with 8 categories), Year of birth (with 89 categories), Place of Residence (with 130 categories) and the continuous variable Income.

To check the effect of the different PRAM-matrices on the disclosure control, we will use the notion of unsafe records as given in (7), with $d = 100$. We will check two instances of combinations of identifying variables: ‘Place of Residence × Marital status × Gender’ (RMG) and ‘Place of Residence × Marital status × Year of birth’ (RMY). RMG consists of 2 080 combined categories, of which 764 occur less than 100 times in the original microdata file (i.e., are rare), whereas RMY consists of 92 560 combinations, with 25 045 rare occurrences.

To check the effect on the loss of information, we will use all the measures we introduced in section 4.

We will apply PRAM in two different ways: firstly we will apply PRAM to a single categorical variable (called univariate PRAM). Secondly we will apply PRAM to three categorical variables simultaneously and call that multivariate PRAM.

6.1 Univariate PRAM

Firstly, we will apply PRAM to one categorical variable at a time. Tables 1 and 2 show the PRAM-matrices that we used for each categorical variable.

Table 1. PRAM matrices for Gender (G) and Place of Residence (R)

Name	Description
G1	$E(0.55)$
G2, G3	$E(0.6)$
G4	$E(0.65)$
G5	$E(0.7)$
G6	$E(0.8)$
G7	$E(0.9)$
G8	$E(0.95)$

Name	Description
R1	Block(14; $\mathbb{1}$;$1E(0.8)$; $8E(0.8)$; $4E(0.8)$; $10E(0.8)$; $2E(0.8)$; $15E(0.8)$; $6E(0.8)$; $17E(0.8)$; $24E(0.8)$; $6E(0.8)$; $18E(0.8)$; $11E(0.8)$; $\mathbb{1}$)
R2	Block(14; $\mathbb{1}$;$1F(0.8)$; $8F(0.8)$; $4F(0.8)$; $10F(0.8)$; $2F(0.8)$; $15F(0.8)$; $6F(0.8)$; $17F(0.8)$; $24F(0.8)$; $6F(0.8)$; $18F(0.8)$; $11F(0.8)$; $\mathbb{1}$)

Table 2. PRAM matrices for Year of birth (Y) and Marital status (M)

Name	Description
Y1	$89B(0.6; 2)$
Y2	$89B(0.6; 3)$
Y3	$89B(0.6; 7)$
Y4	$89B(0.75; 2)$
Y5	$89B(0.75; 3)$
Y6	$89B(0.75; 21)$
Y7	$89B(0.8; 1\frac{1}{2})^*$
Y8	$89B(0.8; 2)$
Y9	$89E(0.75)$
Y10	$89F(0.75)$
Y11	Block(3; $24E(0.6)$; $61E(0.75)$; $4E(0.6)$)
Y12	Block(3; $24B(0.6; 5)$; $61B(0.75; 21)$; $4B(0.6; 2)$)
Y13	Block(3; $24F(0.6)$; $61F(0.75)$; $4F(0.6)$)
Y14	Block(3; $24E(0.6)$; $61B(0.75; 21)$; $4E(0.6)$)
Y15	Block(3; $24F(0.6)$; $61B(0.75; 21)$; $4F(0.6)$)

Name	Description
M1	Block(2; $\mathbb{1}$; $7B(0.6; 3)$)
M2	Block(2; $\mathbb{1}$; $7B(0.6; 4)$)
M3	Block(2; $\mathbb{1}$; $7B(0.8; 4)$)
M4	Block(2; $\mathbb{1}$; $7B(0.8; 5)$)
M5	Block(2; $\mathbb{1}$; $7F(0.75)$)
M6	Block(2; $\mathbb{1}$; $7F(0.8)$)
M7	Block(3; $\mathbb{1}$; $4E(0.8)$; $3E(0.6)$)
M8	Block(3; $\mathbb{1}$; $4F(0.8)$; $3F(0.6)$)
M9	Block(3; $\mathbb{1}$; $4F(0.8)$; $3E(0.6)$)

* Non-zero elements at p_{kk}, p_{kk+1}, $k = 1, \ldots, K-1$, p_{KK} and p_{K-1K}.

6.1.1 Disclosure control

To measure the effect on disclosure control, we will count the number of unsafe combinations, as defined in (7), that will be left after applying each PRAM-matrix. Obviously, in case of applying PRAM to the variable Gender, we will not consider RMY, since the number of unsafe combinations in RMY will not be changed in that case. Similarly, we won't consider RMG in case of applying PRAM to Year of birth. In Tables 3 and 4 the results for the different PRAM-matrices are given.

Note that, since the PRAM-risk only depends on the transition probabilities and the original frequencies, the two identical matrices G2 and G3 for Gender, yield the same number of unsafe combinations that will be left after applying these matrices.

Table 3. Number of unsafe combinations after applying G- and R-matrices

Matrix	RMG-unsafe (764 before PRAM)
G1	686
G2, G3	695
G4	704
G5	712
G6	729
G7	748
G8	754

Matrix	RMG-unsafe (764 before PRAM)	RMY-unsafe (25 045 before PRAM)
R1	708	20 354
R2	706	20 336

Table 4. Number of unsafe combinations after applying Y- and M-matrices

Matrix	RMY-unsafe (25 045 before PRAM)
Y1	19 840
Y2	19 789
Y3	19 395
Y4	20 644
Y5	20 609
Y6	19 453
Y7	20 856
Y8	20 867
Y9	18 659
Y10	18 653
Y11	18 191
Y12	19 518
Y13	18 182
Y14	19 518
Y15	19 518

Matrix	RMG-unsafe (764 before PRAM)	RMY-unsafe (25 045 before PRAM)
M1	462	14 380
M2	277	13 184
M3	362	18 030
M4	356	18 507
M5	230	18 086
M6	235	19 189
M7	658	17 677
M8	659	17 360
M9	658	17 360

The results on applying PRAM to Gender show that the number of unsafe combinations decreases as the transition probability p_{kk} increases. Indeed, since a large value of p_{kk} yields a high probability that an observed score equals the original score, this is what one would expect. The same effect is apparent comparing M2 with M3, M5 with M6, and Y1 with Y4 and Y8.

In most cases, increasing the number of nonzero elements in the PRAM matrix decreases the number of unsafe combinations. Moreover, in case of fully filled matrices, the exact distribution of the probability mass over the off-diagonal elements does not seem to matter much. See e.g., R1 and R2: they only differ in the distribution of the mass over the off-diagonal elements within each block, whereas the number of unsafe combinations is virtually the same.

6.1.2 Information loss

In Tables 5–8 the results concerning the measures of loss of information as given in Section 4 are given. In the columns marked '# Inf', the number of infinite relative differences is shown.

As predicted in subsection 4.1, the results show that in our experiments the loss of information according to EBIL and to IL does not differ much. Moreover, larger diagonal probabilities p_{kk} yield a smaller loss of information, according to the same measures. This is apparent comparing all the results concerning Gender, as well as the results concerning M2 and M3, M5 and M6 and Y1, Y4 and Y8.

From the results it is also clear that the stated median relative differences are quite small. However, very large maxima are found as well. These extreme values are linked with cells with very small original frequency counts: for these cells a small absolute difference can be a large relative difference. Moreover, since we used an unbiased estimate for the frequency tables, small cell counts will occasionally be estimated with negative values.

To put the number of infinite relative differences shown in the results into perspective, the number of empty cells in the frequency tables concerned, are 3 in $G \times Y$, 232 in $Y \times M$ and 2 547 in $Y \times R$.

The difference in the measures of information loss when using the identical PRAM-matrices G2 and G3 seems quite large. However, in case of the one-dimensional relative differences, e.g., both 95% confidence intervals overlap.

Increasing the number of nonzero elements in the PRAM matrix, does not have a clear effect on the measures of loss of information: in some instances of the PRAM-matrices, the loss of information increases, whereas in other cases it decreases for the same measure of loss of information. On the other hand, using one instance of a PRAM-matrix, the effect on the different measures is not the same either.

6.2 Multivariate PRAM

In order to observe the effect of applying PRAM to several categorical variables at the same time, we applied certain combinations of the previously mentioned PRAM-matrices simultaneously. The combinations we used are given in Table 9.

Table 5. Measures of loss of information for applying PRAM to Gender

					$G \times M$		$G \times Y$						
Matrix	EBIL	IL	RD_1 (%)	mRD_1 (%)	RD_2 (%)	mRD_2 (%)	RD_2 (%)	mRD_2 (%)	# Inf	CV(%)	mCV(%)	LRD(%)	mLRD(%)
G1	1 824 382	1 824 380	0.06	0.07	4.91	411	5.46	667	3	0.41	0.48	0.39	0.58
G2	1 784 638	1 784 786	0.23	0.26	0.71	78	2.00	333	3	0.20	0.24	0.15	0.29
G3	1 784 589	1 784 577	0.06	0.07	2.89	33	2.08	300	3	0.20	0.24	0.04	0.07
G4	1 717 291	1 717 559	0.14	0.16	0.91	6	1.40	225	3	0.13	0.15	0.04	0.06
G5	1 620 944	1 620 977	0.16	0.18	0.54	22	1.13	213	3	0.10	0.11	0.07	0.12
G6	1 329 685	1 329 907	0.11	0.13	0.23	33	0.56	133	3	0.06	0.06	0.06	0.07
G7	866 107	866 217	0.02	0.02	0.57	19	0.42	119	3	0.03	0.04	0.00	0.01
G8	530 104	529 517	0.00	0.00	0.16	4	0.21	34	3	0.02	0.02	0.01	0.01

Table 6. Measures of loss of information for applying PRAM to Marital status

					$M \times G$		$M \times Y$						
Matrix	EBIL	IL	RD_1 (%)	mRD_1 (%)	RD_2 (%)	mRD_2 (%)	RD_2 (%)	mRD_2 (%)	# Inf	CV(%)	mCV(%)	LRD(%)	mLRD(%)
M1	1 716 336	1 716 579	1.29	118	0.55	1 965	37.31	50 542	204	2.50	1 975	0.91	3 912
M2	1 715 156	1 714 697	2.33	3 319	0.52	5 016	53.11	18 013	204	2.01	1 762	0.36	13
M3	1 114 185	1 114 651	0.59	318	0.72	851	19.45	11 142	204	0.96	683	0.50	401
M4	1 049 117	1 048 860	0.52	84	0.57	2 169	21.40	21 438	204	0.85	2 068	0.56	3 876
M5	943 185	942 280	0.25	1 142	0.23	1 984	22.36	14 464	204	0.97	3 230	0.58	407
M6	790 313	791 340	0.18	3 477	0.54	6 151	21.20	18 589	204	0.80	2 675	0.53	70
M7	1 260 079	1 260 776	1.04	302	0.91	322	1.83	25 482	83	0.97	329	0.24	13
M8	960 797	962 267	0.08	79	0.12	838	1.55	21 665	84	0.79	776	0.66	5 588
M9	960 931	961 093	0.29	52	0.73	511	1.93	20 128	83	0.99	329	0.69	143

Table 7. Measures of loss of information for applying PRAM to Year of birth

					$Y \times G$			$Y \times M$						
Matrix	EBIL	IL	RD_1 (%)	mRD_1 (%)	RD_2 (%)	mRD_2 (%)	# Inf	RD_2 (%)	mRD_2 (%)	# Inf	CV(%)	mCV(%)	LRD(%)	mLRD(%)
Y1	2 563 506	2 563 469	0.82	545	1.45	339	3	11.36	818	148	1.43	421	0.94	2 740
Y2	3 301 373	3 301 840	0.78	91	1.12	277	3	10.12	931	149	0.99	241	0.92	519
Y3	4 408 797	4 405 751	0.70	904	0.87	2 217	3	18.76	889	212	0.92	824	0.82	624
Y4	1 985 821	1 985 020	0.45	278	0.60	130	3	3.33	348	105	0.63	164	0.59	276
Y5	2 446 568	2 446 654	0.40	109	0.64	112	3	4.58	321	131	0.58	133	0.48	98
Y6	3 682 391	3 684 799	0.81	1 211	0.78	2 548	3	23.43	5 178	232	0.67	1 033	0.84	8 291
Y7	1 350 778	1 350 261	0.41	163	0.65	100	2	1.89	343	79	0.41	102	0.35	68
Y8	1 725 504	1 725 785	0.33	69	0.63	154	3	2.45	269	95	0.51	129	0.50	149
Y9	3 591 017	3 592 735	0.55	13 356	0.65	10 232	3	29.30	11 590	232	0.70	17 791	0.21	5 028
Y10	3 582 057	3 582 004	0.51	5 955	0.57	12 419	3	35.39	18 499	232	0.70	17 894	0.96	12 957
Y11	3 856 549	3 856 311	0.58	1 085	0.77	957	3	6.40	2 944	136	0.81	1 202	0.78	950
Y12	3 751 126	3 752 656	0.39	190	0.79	293	3	6.35	3 659	136	0.67	222	0.72	232
Y13	3 853 464	3 853 803	0.36	1 450	0.69	1 182	3	5.51	1 819	136	0.82	1 231	1.16	680
Y14	3 751 233	3 754 083	0.56	1 943	0.64	1 531	3	5.57	5 067	136	0.67	1 202	0.55	503
Y15	3 751 497	3 751 564	0.45	454	0.86	839	3	5.64	4 765	136	0.67	1 231	1.34	949

Table 8. Measures of loss of information for applying PRAM to Place of Residence

					$R \times G$		$R \times Y$						
Matrix	EBIL	IL	RD_1 (%)	mRD_1 (%)	RD_2 (%)	mRD_2 (%)	RD_2 (%)	mRD_2 (%)	# Inf	CV(%)	mCV(%)	LRD(%)	mLRD(%)
R1	2 389 259	2 389 452	0.25	3	0.46	7	3.94	670	1610	0.42	5.64	0.36	1.91
R2	2 369 024	2 369 788	0.33	6	0.43	10	3.92	739	1610	0.42	5.81	0.37	3.87

Table 9. Combinations of PRAM-matrices

Name	Combination		
PRAM1	Y6	G6	M7
PRAM2	Y11	G2	M3
PRAM3	Y14	G3	M2
PRAM4	Y12	G8	M4
PRAM5	Y13	G5	M9
PRAM6	Y15	G7	M8
PRAM7	Y10	G1	M6
PRAM8	Y8	M5	R2
PRAM9	Y7	M1	R1

6.2.1 Disclosure control

In Table 10 the number of unsafe combinations after applying PRAM to three categorical variables at the same time are given. Again, only two combinations of identifying variables are considered: RMG and RMY. On average we see that applying PRAM in a multivariate way, the number of unsafe combinations that is left, is smaller compared to the univariate applications. However, we would like to stress the notion that only comparing the unsafe combinations is not fair: this should always be related the amount of information that is lost.

Table 10. Number of unsafe combinations after applying PRAM

	RMG	RMY
Name	(764 before PRAM)	(25 045 before PRAM)
PRAM1	652	14 911
PRAM2	323	13 982
PRAM3	245	10 736
PRAM4	353	15 593
PRAM5	652	13 196
PRAM6	656	14 682
PRAM7	221	15 639
PRAM8	226	13 541
PRAM9	411	11 643

6.2.2 Information loss

We will not consider all measures of loss of information in case of multivariate PRAM, but only state the results concerning the measures EBIL and IL, and the results on two-dimensional relative differences in case PRAM is applied to both variables. I.e., in case of PRAM1 (PRAM applied to Year of birth, Gender and Marital Status), we will consider the two-dimensional frequency tables $Y \times G$, $Y \times M$ and $G \times M$. Tables 11 and 12 show the numerical results.

Table 11. Measures of loss of information for combination of PRAM on Y, M and G

			$Y \times M$			$Y \times G$			$M \times G$	
Name	EBIL	IL	RD_2 (%)	mRD_2 (%)	# Inf	RD_2 (%)	mRD_2 (%)	# Inf	RD_2 (%)	mRD_2 (%)
PRAM1	5 823 916	5 825 419	62.99	39 439	232	1.44	1 841	3	2.11	672
PRAM2	6 240 467	6 241 220	126.10	16 464	232	3.69	2 824	3	3.40	3 725
PRAM3	6 680 562	6 683 416	198.05	27 607	232	2.72	4 906	3	4.34	25 977
PRAM4	4 927 252	4 927 149	97.18	17 503	232	0.77	310	3	0.70	4 088
PRAM5	5 952 874	5 932 505	18.00	22 939	148	1.73	2 086	3	2.90	2 036
PRAM6	5 198 376	5 197 464	14.67	17 967	148	0.99	962	3	1.39	1 019
PRAM7	5 752 083	5 751 979	280.55	28 459	232	7.22	225 259	3	5.70	78 263

Table 12. Measures of loss of information for combination of PRAM on Y, M and R

			$Y \times M$			$Y \times R$			$M \times R$		
Name	EBIL	IL	RD_2 (%)	mRD_2 (%)	# Inf	RD_2 (%)	mRD_2 (%)	# Inf	RD_2 (%)	mRD_2 (%)	# Inf
PRAM8	4 765 922	4 766 419	62.98	18 729	222	9.73	798	2 425	14.76	26 235	151
PRAM9	4 929 842	4 981 720	61.39	31 015	217	7.49	772	1 957	30.12	13 264	151

As we expected, the loss of information according to EBIL and IL is larger compared to the univariate results. This is due to the fact that there are many more categories to consider. Indeed, when applying PRAM to Year of Birth, Marital Status and Place of residence, there are $89 \times 8 \times 130 = 92\,560$ combinations of categories to consider. Since the definitions of EBIL and IL consist of sums of terms including $\log p_{lk}^{\leftarrow}$, this yields a large value for these measures.

If we want to compare univariate PRAM with multivariate PRAM, we will have to take both the level of disclosure control as well as the amount of information that is lost into account. We expect that applying multivariate PRAM with the same level of information loss as a univariate application, will yield a higher level of disclosure control.

The closest values for EBIL in case of univariate and multivariate PRAM are the ones corresponding to Y3 and PRAM4. If we then look at the number of unsafe RMY-combinations, we see that PRAM4 has 3 802 unsafe combinations less (about 20%), even though the EBIL value is 12% larger than in case of Y3. I.e., even though the loss of information is larger, the level of disclosure control is higher as well. Moreover, since the univariate application with Y3 has no effect on the unsafe combinations in RMG but the multivariate application with PRAM4 does have, PRAM4 outperforms Y3 in that sense as well.

Similarly, considering the information loss according to the relative differences, the overall loss of information is larger for the multivariate cases. This is not surprising: both variables in the frequency tables have been perturbed in the multivariate setting, whereas in the univariate setting only one of the spanning variables is perturbed. Hence, more cells are effected more seriously.

However, if we take, e.g., the table $G \times M$ in case of G2 and PRAM4, the median relative differences are 0.71 and 0.70 respectively, whereas the number of unsafe combinations in RMG is reduced from 695 for G2 to 353 for PRAM4. Additionally, in case of PRAM4 the number of unsafe combinations in RMY is reduced as well (from 25 045 to 15 593), whereas in case of G2 there is no effect on the number of unsafe combination in RMY. So, with more or less the same loss of information the multivariate case has a much higher level of disclosure control.

To put the number of infinite relative differences into perspective again, the number of empty cells in the original two-dimensional tables are 232 for $Y \times M$, 3 for $Y \times G$, 2 547 for $Y \times R$ and 158 for $M \times R$.

7 Summary and conclusions

PRAM is a method to deal with disclosure control when disseminating microdata. This method was introduced in 1997, but has not yet been applied extensively. This is partly due to the fact that there is little knowledge available on the effect of PRAM on disclosure control or on the loss of information it induces.

The method is defined in terms of transition probabilities, summarized in a PRAM-matrix. In this paper we investigated the effect of different distributions of the transition probabilities, on the level of disclosure control as well as on the amount of information that is lost when applying PRAM. Several instances of PRAM-matrices have been applied to a specific microdatafile, both in a univariate as well as a multivariate setting. Several measures of loss of information have been calculated along with a measure for the level of disclosure control. The different instances resulted in different effects. In most cases, increasing the number of non-zero elements resulted in a decrease of unsafe combinations. However, its effect on the measures of loss of information was not unambiguous: some measures gave rise to an increase of loss of information, whereas others yielded a decrease. This indicates that it might be desirable to let the choice of PRAM matrix (or matrices) depend on the intended use of the microdatafile. To compare the results of the univariate and the multivariate application of PRAM, we should take into account both the effect on the level of disclosure control as well as on the loss of information. Indeed, one should only compare situations with either a comparable level of disclosure control or a comparable amount of loss of information. The results indicate that it seems possible to achieve the same level of disclosure control, with a lower loss of information, when applying PRAM in a multivariate way. Or, equivalently, to achieve the same amount of loss of information, with a higher level of disclosure control.

In our experiments, we used block matrices, with equal diagonal elements within each block. An obvious alternative would be to allow for a variation in the diagonal elements within each block. These diagonal elements might be chosen depending on the disclosure risk associated with that category. However, since that risk is related to combinations of categories of several variables, this becomes quite complicated, especially when applying PRAM in a multivariate way. This is a topic for further research.

In this paper, the effect of several different PRAM-matrices on the disclosure control as well as on the information loss is discussed. Ideally one would like to use an optimal PRAM-matrix in the sense that, given a predefined level of disclosure control, the loss of information is minimized. Since our results show that this depends on the exact measure for loss of information that is used, a more generally applicable measure is asked for. A possible candidate would be to use the (Hellinger or Kullback-Leibler) distance between the empirical distribution of certain variables in the original file and the empirical distribution of the same variables in the perturbed file. This is again a topic for further research.

References

Chen, T. T. (1979). Analysis of randomized response as purposively misclassified data. In *Proceedings of the section on survey research methods*, pp. 158–163. American Statistical Association.

Domingo-Ferrer, J. and Torra, V. (2001a). Disclosure Control Methods and Information Loss for Microdata. In P. Doyle, J. Lane, J. Theeuwes, and L. Zayatz (Eds.), *Confidentiality, Disclosure and Data Acces: Theory and Practical Applications for Statistical Agencies*, Chapter 5, pp. 91–110. North-Holland: Elsevier.

Domingo-Ferrer, J. and Torra, V. (2001b). A quantitative comparison of disclosure control methods for microdata. In P. Doyle, J. Lane, J. Theeuwes, and L. Zayatz (Eds.), *Confidentiality, Disclosure and Data Acces: Theory and Practical Applications for Statistical Agencies*, Chapter 6, pp. 111–133. North-Holland: Elsevier.

Fienberg, S. E. (1994). Conflict between the needs for access to statistical information and demands for confidentiality. *Journal of Official Statistics* 10 (2), 115–132.

Gouweleeuw, J., Kooiman, P., Willenborg, L., and De Wolf, P.-P. (1998). Post Randomisation for Statistical Disclosure Control: Theory and Implementation. *Journal of Official Statistics* 14 (4), 463–478.

van den Hout, A. (1999). The analysis of data pertubed by PRAM. Delft Univsersity Press, Delft.

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 (2), 269–288.

Kooiman, P., Willenborg, L., and Gouweleeuw, J. (1997). *A method for disclosure limitation of microdata*. Research paper 9705, Statistics Netherlands, Voorburg.

Kuha, J. and Skinner, C. (1997). Categorical data analysis and misclassification. In Lyberg, Biemer, Collins, de Leeuw, Dippo, Schwarz, and Trewin (Eds.), *Survey measurement and process quality*, Chapter 28, pp. 633–670. New York: Wiley.

Press, S. J. (1968). Estimating from misclassified data. *Journal of the American Statistical Association* 63, 123–133.

Rienstra, M. (2003). *Aanzet tot een nieuwe beveiligingsregel bij gebruik van PRAM*. Internal report (in Dutch) 283-03-SOO, Statistics Netherlands, Voorburg.

Ronning, G., Rosemann, M., and Strotmann, H. (2004). Estimation of the probit model using anonymized micro data. Paper prepared for the ‘European Conference on Quality and Methodology in Official Statistics (Q2004)’, Mainz, 24–26 May 2004.

Warner, S. L. (1965). Randomized response: a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association* 60, 63–69.

de Wolf, P.-P., Gouweleeuw, J., Kooiman, P., and Willenborg, L. (1998). Reflections on PRAM. In *Statistical Data Protection*, Luxembourg, pp. 337–349. Office for Official Publications of the European Communities.