Journal of Official Statistics, Vol. 13, No. 4, 1997, pp. 417–434

Statistical Disclosure Control and Sampling Weights

A.G. de Waal and L.C.R.J. Willenborg[1]

Before a microdata set is disseminated by a statistical office it should be checked whether sensitive information about individual respondents could be disclosed by a potential intruder. The procedure to check whether the dissemination of a microdata set could lead to disclosure of sensitive information usually amounts to examining how much so-called (indirectly) identifying information is contained in the microdata set. In case too much identifying information is contained in the microdata set it is considered unsafe for release. When a statistical office releases a microdata set, sampling weights are usually included to facilitate analyses. A description of the auxiliary variables, their categories and the sampling method underlying the weights is usually also provided. Unfortunately, the sampling weights, innocent as they may seem, can provide additional identifying information to an intruder when they are based on identifying information that is not contained in the released microdata set. A simple idea to prevent disclosure from sampling weights would be not to publish which weight corresponds to which stratum. Surprisingly, this is not sufficient. In this article we demonstrate that in many cases an intruder will be able to determine which stratum corresponds to a specific weight given sufficient knowledge about the population.

Key words: Statistical disclosure control; sampling weights; microdata.

1. Introduction

Before microdata are disseminated by a statistical office, it should be checked whether this could lead to the disclosure of sensitive information about individual respondents. Such disclosure should be avoided to prevent the privacy of respondents being endangered. To disclose sensitive information about an individual respondent an intruder, i.e., someone who is attempting to disclose sensitive information, would first have to identify to whom a particular record in the microdata set belongs. This is called re-identification. To prevent re-identification, it is usual first to examine how much (indirectly) identifying information is contained in the microdata set. Examples of such (indirectly) identifying information are the age and domicile of a person. Directly identifying information, such as the name or the address of a respondent, should never be included in a microdata set, of course, because this would immediately lead to re-identification. If too much identifying information is contained in the microdata set it is considered unsafe for release. In that case suitable statistical disclosure control (SDC) measures must be taken. Further discussion on SDC for microdata is given by Duncan and Lambert (1989), Bethlehem et al. (1990), De Waal and Willenborg (1996), and Willenborg and De Waal (1996).

[1] Statistics Netherlands, Division Research and Development, PO Box 4000, 2270 JM Voorburg, The Netherlands.
Acknowledgments: The views expressed in this article are those of the authors and do not necessarily reflect the policies of Statistics Netherlands. The authors would like to thank an associate editor and three referees for their helpful comments which led to an improved article.

When a statistical office releases a microdata set, sampling weights are usually included to facilitate analyses. A description of the auxiliary variables, their categories and the sampling method underlying the weights is generally also provided. Unfortunately, the sampling weights, innocent as they may seem, can in some cases provide additional identifying information to an intruder when they are based on identifying information that is not contained in the released microdata set. For instance, Statistics Netherlands does not release regional information in its microdata sets that are meant for public use. However, regional information may be used to calculate the sampling weights. If an intruder could derive the regional information from the sampling weights the SDC-rules would be violated. In general, an intruder should not be able to derive identifying information from the sampling weights that may not be included in the microdata set according to the SDC-rules. The problem is not that inclusion of the sampling weights in the microdata set necessarily leads to an unacceptably high disclosure risk, but rather that the SDC-rules may become inconsistent. Sampling weights may enable an intruder to obtain identifying information he or she is not supposed to possess.

Sampling weights can be determined by means of several procedures. In this article we consider three kinds of such procedures, namely poststratification, linear weighting and multiplicative weighting (the latter is also called raking, raking ratio estimation or iterative proportional fitting). In each case, we suppose the weights are based on population information on categorical auxiliary variables. We refer to the cells in the cross-classification of these variables as strata. Details on linear weighting can be found in Bethlehem and Keller (1987) and on multiplicative weighting in Deville and Särndal (1992). As the methods to derive additional identifying information are different for poststratification on the one hand and linear and multiplicative weighting on the other hand, we distinguish between these two situations.

A simple idea for protecting disclosure from sampling weights would be not to publish which weight corresponds to which stratum. Surprisingly, this is not sufficient. In this article we demonstrate that in many cases an intruder will be able to determine which stratum corresponds to a specific weight when his or her knowledge about the population is sufficiently large.

In Section 2 we describe a method to match the sampling weights and the strata in case poststratification was used to calculate the weights. In Section 3 we describe two methods for the cases that multiplicative and linear weighting was used. The article is concluded with a short discussion in Section 4.

In the remainder of this article we assume that the sampling weights have been calculated by using identifying information that may not be included in the microdata set, and that no explicit reference to their strata is made. We also assume that an intruder knows the population frequencies that have been used to evaluate the sampling weights (almost) perfectly. This is quite a plausible assumption as information provided by the auxiliary variables is often published by the statistical agency itself. Moreover, we assume that sampling weights corresponding to different strata are different. Again this is a plausible assumption as we are dealing with real-life data. The same assumptions are made by Van Kouwen (1993), who describes the main ideas of the methods to obtain additional identifying information from the sampling weights. However, Van Kouwen does not describe how to apply his method for general numbers of categories of the auxiliary variables, but only provides some examples for specific cases. The present article is an extension of Van Kouwen (1993).

2. Poststratification

The case of poststratification is the simplest. By counting the frequency of a certain weight in the sample and by multiplying this frequency by the weight an intruder can determine the number of units in the population that belong to the poststratum that corresponds to this weight. Because the intruder has a (nearly) perfect description of the population with respect to the auxiliary variables he or she can subsequently match the weights to the poststrata.

Example 1

Suppose that two auxiliary variables A and B have been used to calculate the sampling weights. The number of categories of A is two and of B is three. We suppose that poststratification has been used and that the frequencies of the poststrata in the population are given in Table 1.

Table 1. Frequencies of the poststrata in the population

Poststratum	Frequency in the population
$A_1 \times B_1$	1,368
$A_1 \times B_2$	725
$A_1 \times B_3$	896
$A_2 \times B_1$	2,633
$A_2 \times B_2$	2,787
$A_2 \times B_3$	1,642

The weights are listed in ascending order in Table 2.

The weight of a poststratum multiplied by its corresponding frequency in the sample is by definition equal to the size of this poststratum in the population. So, if the intruder knows the frequencies of the poststrata in the population, as given in Table 1, then he or she would be able to determine which weight corresponds to which stratum. For instance, it is easy to see that weight 82.095 corresponds to stratum $A_2 \times B_3$, and weight 89.596 to $A_1 \times B_3$.

When the intruder knows the frequencies of the strata in the population only approximately, he or she will have to choose the most likely way to match the weights with the strata. If the knowledge of the intruder about the frequencies of the strata in the population

Table 2. Weights of the poststrata

Index i	Weight W_i	Frequency in sample	Weight × Frequency (rounded)
1	82.095	20	1,642
2	89.596	10	896
3	96.102	29	2,787
4	105.320	25	2,633
5	120.833	6	725
6	136.799	10	1,368

is sufficiently precise, he or she will be able to determine which stratum belongs to a specific weight. Suppose, for instance, that the intruder thinks that stratum i occurs $X_i = X_{pop,i} + \epsilon_i$ times in the population, where $X_{pop,i}$ is the actual frequency of stratum i in the population and ϵ_i is the error made by the intruder. Suppose furthermore that $-50 \leq \epsilon_i \leq 50$. In this case the intruder would still be able to match the weights to their corresponding strata correctly, because the differences between the population frequencies of different strata are larger than 100.

3. Linear/Multiplicative Weighting

For both linear weighting and multiplicative weighting the product of a weight and its frequency in the sample is generally unequal to the frequency of the corresponding stratum in the population. Such a product is usually only an approximation of the frequency of the corresponding stratum. This complicates the situation for an intruder considerably. However, because the products of the weights and their frequencies do sum up to the marginal totals in the population an intruder is in many cases still able to derive identifying information from the sampling weights. In the sequel we show how an intruder might proceed. We concentrate on two methods the intruder may apply. This does not imply, however, that these methods are the only ones. We start by considering the case of multiplicative weighting. Having examined this case we show how the results obtained can be translated into the case of linear weighting.

To demonstrate how an intruder may proceed we assume that m auxiliary variables have been used to determine the sampling weights. These auxiliary variables are denoted by V_i $(i = 1, \ldots, m)$. The number of categories of these variables is denoted by n_i $(i = 1, \ldots, m)$. The categories themselves are denoted by C_{ij} $(i = 1, \ldots, m; j = 1, \ldots, n_i)$. We assume that the variables are ordered in such a way that $n_1 \leq n_2 \leq \ldots \leq n_m$.

The two methods we examine in this section can be applied when a weight is given by

$$W_{i_1 i_2 \ldots i_m} = F^1_{i_1} \times F^2_{i_2} \times \cdots \times F^m_{i_m}$$

in the case of multiplicative weighting, or by

$$W_{i_1 i_2 \ldots i_m} = F^1_{i_1} + F^2_{i_2} + \cdots + F^m_{i_m}$$

in the case of linear weighting, where a factor F^k_i depends only on the category C_{ki}. When two weights have the same factor $F^q_{i_p}$, we say that these weights have category C_{qi_p} in common. Now we discuss the two methods to derive additional identifying information from the sampling weights.

3.1. *Method 1*

In the first method the disclosure problem is split into two parts. Firstly, the intruder determines which weights have exactly $(m - 1)$ categories in common. Secondly, the intruder determines the actual auxiliary variables and categories. Why an intruder has to determine in the first step which weights have exactly $(m - 1)$ categories in common will become clear in the second step.

The steps for this method are illustrated by applying them to an example.

Step 1

The intruder begins by listing all the different weights that occur in the microdata set. To determine which weights have exactly $(m-1)$ categories in common, the intruder evaluates the ratios of all pairs of weights. The ratios are listed, producing what we refer to as *the ratios list.*

Example 2

Suppose that three auxiliary variables A, B and C have been used. The number of categories of variable A is two, of variable B three and of variable C six. We assume that multiplicative weighting has been used.

In Table 3 the knowledge of the intruder about the frequencies of the categories of the auxiliary variables is shown.

The weights that are released are listed in ascending order in Table 4.

The ratios list is rather large – it contains 36×36 ratios – so only part is presented in Appendix A. Only ratios pertaining to Example 2 are listed.

For the moment it is convenient to denote a weight by $W_{k_1 k_2 \ldots k_m}$. A ratio in the ratios list has the following form

$$\frac{W_{k_1 k_2 \ldots k_m}}{W_{\ell_1 \ell_2 \ldots \ell_m}} = \frac{F^1_{k_1} \times F^2_{k_2} \times \cdots \times F^m_{k_m}}{F^1_{\ell_1} \times F^2_{\ell_2} \times \cdots \times F^m_{\ell_m}} \tag{1}$$

The ratios F^i_k / F^i_ℓ will be denoted by $R(i,k,\ell)$. We assume that a ratio $R(i,k,\ell)$ is different from $R(i',k',\ell')$ whenever $(i',k',\ell') \neq (i,k,\ell)$ except when $k' = \ell''$ and $k = \ell$. Moreover, we assume that $R(i,k,\ell) \times R(i',k',\ell') = 1$ if and only if $i' = i$, $k' = \ell$ and $\ell' = k$. Note that a ratio $R(i,k,\ell) = 1$ if and only if $k = \ell$.

The value of a ratio of two weights as given by (1) occurs

$$\prod_{i \in \{i \mid k_i = \ell_i\}} n_{k_i} \tag{2}$$

times in the list, where $\prod_{i \in \emptyset} n_{k_i} = 1$ by definition, because when $k_i \neq \ell_i$ for all $i = 1, \ldots, m$, the value of this ratio occurs only once.

Table 3. Frequencies of the categories of the auxiliary variables in the population. These are supposed to be known to the intruder

Category	Frequency in the population
A_1	7,480,000
A_2	7,649,000
B_1	6,572,000
B_2	7,037,000
B_3	1,520,000
C_1	2,765,000
C_2	3,570,000
C_3	3,605,000
C_4	2,549,000
C_5	1,811,000
C_6	829,000

Table 4. Weights in the sample

Index i	Weight W_i	Frequency in sample	Weight × Frequency in sample
1	94.6384	495	46,846.01
2	95.2153	703	66,936.36
3	96.0524	960	92,210.30
4	96.6379	3,368	325,476.45
5	96.8195	7,004	678,123.78
6	96.8338	15,749	1,525,035.52
7	97.4097	6,174	601,407.49
8	97.4241	12,999	1,266,415.88
9	98.1456	168	16,488.46
10	98.2806	105	10,319.46
11	98.7439	620	61,221.22
12	98.8797	233	23,038.97
13	99.0655	2,868	284,119.85
14	99.2341	940	93,280.05
15	99.6120	805	80,187.66
16	99.6694	4,626	461,070.64
17	99.8391	623	62,199.76
18	100.2193	3,599	360,689.26
19	100.4076	1,848	185,553.24
20	100.7081	13,989	1,408,805.61
21	100.7168	1,338	134,759.08
22	100.8228	4,121	415,490.76
23	101.0197	1,236	124,860.35
24	101.3220	13,385	1,356,194.97
25	101.3308	1,959	198,507.04
26	101.4374	2,495	253,086.31
27	101.5212	10,583	1,074,398.86
28	102.1400	9,652	985,855.28
29	102.2128	0	0.00
30	102.3292	1,228	125,660.26
31	102.8358	0	0.00
32	102.9530	1,643	169,151.78
33	103.0291	0	0.00
34	103.1464	12,688	1,308,721.52
35	103.6571	0	0.00
36	103.7752	12,844	1,332,888.67

The frequency of the ratios for which the weights have all categories except one, say of variable V_s, in common is given by

$$G_s \equiv \prod_{i \in \{i \mid i \neq s\}} n_i \tag{3}$$

Note that this frequency depends only on the variable for which the categories differ and not on the categories themselves.

If all the G_s's are different from the frequencies of ratios of which the weights have less than $(m-1)$ categories in common, then an intruder can determine which weights have $(m-1)$ categories in common. In this case the first step would be complete. If, moreover, all the values G_s ($s = 1, \ldots, m$) are distinct, then the intruder would even know the associated variables for these common categories. All that would be left to find out would be the actual categories.

When a G_s is equal to the frequency of a ratio of which the weights have less than $(m - 1)$ categories in common, then the intruder does not know yet which weights have $(m - 1)$ categories in common. This situation occurs when G_s can be written as

$$G_s = n_{i_1} \times n_{i_2} \times \cdots \times n_{i_t} \quad (4)$$

for some combination $(i_1, i_2, \ldots, i_t)$, where t is less than $(m - 1)$ and all i_j's are distinct. In this case s must be one of the indices $i_1, i_2, \ldots, i_t$, say $s = i_t$.

Example 2 (continued)

In our example we denote weights by W_i, i.e., by one index only. Examining the ratios list we see, for instance, that the ratio of W_1 and W_2, i.e., 0.99394, occurs $18 = n_2 \times n_3$ times in the ratios list. So, W_1 and W_2 have a category of variable B and a category of variable C in common. The ratio of W_1 and W_3, i.e., 0.98528, occurs $12 = n_1 \times n_3$ times in the list. Because $12 \neq n_2$, we know that W_1 and W_3 have a category of variable A and a category of variable C in common. In this way we can determine all weights that have a category of variable A and a category of variable C in common, and all weights that have a category of B and a category of C in common.

The situation is somewhat more difficult for W_1 and W_6. The ratios of these weights, i.e., 0.97733, occurs six times in the list. Because $6 = n_1 \times n_2 = n_3$, we cannot determine yet whether or not W_1 and W_6 have two categories in common. The same situation occurs for W_1 and W_7. The ratio of these weights, i.e., 0.97155, also occurs six times.

Because the above situation can occur, we might hope that an intruder is not always able to determine which weights have $(m - 1)$ categories in common. Unfortunately, we have the following theorem.

Theorem

Assuming that a ratio $R(i, k, \ell)$ differs from $R(i', k', \ell')$ whenever $(i, k, \ell) \neq (i', k', \ell')$ (except when $k' = \ell''$ and $k = \ell$) and that $R(i, k, \ell) \times R(i', k', \ell') = 1$ if and only if $i' = i$, $k' = \ell$ and $\ell' = k$, an intruder can always determine which weights have $(m - 1)$ categories in common.

Proof

We start by observing that weights that have a ratio that occurs $n_2 \times n_3 \times \cdots \times n_m$ times in the ratios list have $(m - 1)$ categories in common, because relation (4) cannot be satisfied because $n_1 \leq n_2 \leq \cdots \leq n_m$. The corresponding ratios can be determined. If possible we determine other weights that have $(m - 1)$ categories in common. We can do this for those G_t's for which (4) cannot be satisfied.

Now suppose there are weights that have a ratio that occurs G_p times in the ratios list (for some p), but for which the intruder cannot determine yet whether or not they have $(m - 1)$ categories in common. There may be several numbers p for which this situation occurs. The smallest number will be denoted by s. The ratios $R(i, k, \ell)$ can be determined for $i = 1, \ldots, s-1$. Suppose there are q numbers s, $s+1, \ldots, s+q-1$ such that $G_s = G_{s+1} = \cdots = G_{s+q-1}$. Let the ratios of weights that occur G_s times be denoted by R^*_α, where α is an index.

These ratios R^*_α are either equal to an $R(s+j, k, \ell)$ (for $j = 0, \ldots, q-1$) or to a product

$$\prod_{i \neq i_1, i_2, \ldots, i_{t-1}, s, s+1, \ldots, s+q-1} R(i, k_i, \ell_i) \tag{5}$$

In the latter case we have the following relation

$$G_s = n_{i_1} \times n_{i_2} \times \cdots \times n_{i_{t-1}} \times n_s \times n_{s+1} \times \cdots \times n_{s+q-1} \tag{6}$$

where $t < m - q - 1$ and all i_j's are distinct. The intruder can multiply the ratios R^*_α by the $R(i, k_i, \ell_i)$'s for $i = 1, \ldots, s-1$. We distinguish between three cases:

1. If R^*_α is equal to an $R(s+j, k, \ell)$ (for $j = 0, \ldots, q-1$), then the product occurs G_s/n_i times in the ratios list. This follows from (2) and $i \neq s+j$.
2. If R^*_α is equal to a product given by (5) and $i \notin \{i_1, i_2, \ldots, i_{t-1}\}$, then the products occur either zero, G_s or $n_i G_s$ times in the ratios list. Namely, when $i \notin \{i_1, i_2, \ldots, i_{t-1}\}$ then a factor $R(i, k'_i, \ell'_i)$ occurs in (5). So, the product of R^*_α and $R(i, k_i, \ell_i)$ occurs zero times if $k_i \neq \ell'_i$ and $\ell_i \neq k'_i$. The product occurs G_s times if $k_i = \ell'_i$ and $\ell_i \neq k'_i$, or $k_i \neq \ell'_i$ and $\ell_i = k'_i$. Finally, the product occurs $n_i G_s$ times if $k_i = \ell'_i$ and $\ell_i = k'_i$.
3. If R^*_α is equal to a product given by (5) and $i \in \{i_1, i_2, \ldots, i_{t-1}\}$, then the products occur G_s/n_i times in the ratios list. This follows from (2) and the fact that no factor $R(i, k_i, \ell_i)$ occurs in (5).

So, if the products occur zero, G_s, or $n_i G_s$ times in the ratios list, then the intruder knows that the weights of which the ratio is given by R^*_α do not have $(m-1)$ categories in common. When the products occur G_s/n_i times, then the intruder cannot decide yet whether or not the weights of which the ratio is given by R^*_α have $(m-1)$ categories in common.

So, only when $1, 2, \ldots, s-1$ are all elements of $\{i_1, i_2, \ldots, i_{t-1}\}$ the intruder will not be able to determine in this way whether or not the weights with such a ratio R^*_α have $(m-1)$ categories in common.

We claim, however, that when $1, 2, \ldots, s-1$ are elements of $\{i_1, i_2, \ldots, i_{t-1}\}$ then the weights of which the ratio is R^*_α have $(m-1)$ categories in common. Suppose they did not have $(m-1)$ categories in common. In that case relation (6) must be obeyed. Because $1, 2, \ldots, s-1$ are elements of $\{i_1, i_2, \ldots, i_{t-1}\}$, $t < m - q - 1$ and $n_1 \leq n_2 \leq \cdots \leq n_m$, we can conclude that the product $n_{s+q} \times \cdots \times n_m$, i.e., the product of the $m-(s+q)+1$ largest n_i's, is equal to a product of less than $m-(s+q)+1$ distinct n_i's. This is clearly a contradiction. Hence, $1, 2, \ldots, s-1$ cannot all be elements of $\{i_1, i_2, \ldots, i_{t-1}\}$.

We have demonstrated that the intruder can determine which ratios that occur G_s times have $(m-1)$ categories in common. Now the intruder can apply the same procedure for numbers larger than s. In this way the intruder can determine all weights that have $(m-1)$ categories in common. This concludes the proof and Step 1.

Example 2 (continued)

We multiply the ratio of W_1 and W_6 and the ratio of W_1 and W_7 by the ratios of weights that have categories of variables A and C in common and by the ratios of weights that have categories of variables B and C in common. So, we multiply the ratio of W_1 and W_6 by the ratio of W_2 and W_1, for instance. The resulting ratio, W_2 divided by W_6, equals 0.98329. This number occurs $3 = G_3/n_1$ times in the ratios list. Multiplying the ratio of W_1 and W_6 by the ratio of each pair of weights that have categories of variables B and C in common results in a ratio that occurs $G_3/n_1 = 3$ times in the list.

Similarly, when we multiply the ratio of W_1 and W_6 by the ratio of W_3 and W_1, then the resulting ratio, W_3 divided by W_6, occurs $2 = G_3/n_2$ times in the list. Multiplying the ratio of W_1 and W_6 by the ratio of each pair of weights that have categories of variables A and C in common results in a ratio that occurs $G_3/n_2 = 6$ times in the list.

According to the proof of the above theorem, we can conclude that W_1 and W_6 have two categories in common. These two categories are categories from A and B.

When we multiply the ratio of W_1 and W_7 by the ratio of W_2 and W_1, then the resulting ratio, W_2 divided by W_7, equals 0.97747. This number occurs twelve times in the list. Therefore, according to the proof of the theorem of Section 2, W_1 and W_7 do not have two categories in common.

In the above way one can determine all the weights that have two categories in common.

Step 2

After the intruder has determined which weights have $(m-1)$ categories in common he or she has to determine the actual variables and categories involved. From now on we denote weights by W_i, i.e., by one index only.

The intruder can begin by making sets of weights that have *the same* $(m-1)$ *categories* in common. This is easy. When weights W_1 and W_2 have $(m-1)$ categories in common and so do W_1 and W_3 and also W_2 and W_3, then W_1, W_2 and W_3 necessarily have the same $(m-1)$ categories in common. Because of this property it is important to determine the weights that have exactly $(m-1)$ categories in the first step of the method. The property is quite easy to prove. Suppose W_1 and W_2 have categories of all variables except variable V_a in common, and suppose that W_1 and W_3 have categories of all variables except variable V_b in common. In that case W_2 and W_3 have $(m-1)$ categories in common if and only if $V_a = V_b$. In other words, W_2 and W_3 have $(m-1)$ categories in common if and only if W_1, W_2 and W_3 have the same $(m-1)$ categories in common.

All weights that have the same $(m-1)$ categories in common are placed in the same set. Such a set of weights that have the same $(m-1)$ categories in common will be called an *equivalence class*. Note that the equivalence classes are not disjoint: each weight is an element of m equivalence classes. The total number of equivalence classes is

$$\sum_{i=1}^{m} \prod_{j \neq i} n_j \tag{7}$$

The number of elements of the equivalence class of which the weights do not have a category of variable V_i in common is n_i, i.e., the number of categories of V_i.

Example 2 (continued)

The next step in the example is to determine which weights form equivalence classes. For example, W_1 and W_3 have two categories in common and so have W_1 and W_5 and W_3 and

W_5. Therefore, W_1, W_3 and W_5 have the same two categories in common. In particular, W_1, W_3 and W_5 form an equivalence class. Likewise it can be shown that W_1 and W_2 form an equivalence class, and that W_1, W_6, W_9, W_{14}, W_{20} and W_{22} form another equivalence class.

Now suppose weights W_p and W_q are elements of an equivalence class and that W'_p and W'_q are elements of another equivalence class. If the ratios W_p/W_q and W'_p/W'_q are equal and different from one, then W_p and W'_p have the same category of a variable V_i in common. Namely, suppose that W_p and W_q have $(m-1)$ categories in common, then their ratio is given by an $R(i, k, \ell)$. The ratio of W'_p and W'_q is also given by an $R(i', k', \ell')$. We have assumed that these ratios are equal if and only if $i = i'$, $k = k'$ and $\ell = \ell'$. This implies that W_p and W'_p have category C_{ik} of variable V_i in common.

By examining all the ratios of weights in this way the intruder can determine all the weights that have the same category of V_i in common. Note that the intruder may not know the variable V_i yet. However, if the value n_i occurs only once among $n_1, n_2, \ldots, n_m$, then the intruder can infer the variable V_i. The sets of weights that have the same category of a variable in common are not disjoint.

Example 2 (continued)

Because the ratio of W_1 and W_6, i.e., 0.97733, is equal to the ratio of W_7 and W_{16}, W_1 and W_7 have a, still unknown, category of variable C in common. In this way, we can determine all the weights that have a, as yet unknown, category C_i of variable C in common. Similarly, we can determine all the weights that have a, still unknown, category A_j of variable A in common and all the weights that have an, also unknown, category B_k of variable B in common.

The intruder can evaluate the number of times that a category occurs in the population by multiplying all the weights that have the same category in common by their frequencies and taking the sum of these products. All that remains to be done is to find the categories and variables that correspond to the evaluated population frequencies by comparing these evaluated frequencies to the known frequencies in the population. This concludes the second step.

Example 2 (continued)

We can evaluate the frequencies of a category C_i of variable C in the population by multiplying the corresponding weights by their frequencies and subsequently taking the sum of these products. This can be done for all categories of C. Similarly, population frequencies of A_j and of B_k can be evaluated by multiplying the corresponding weights by their frequencies and taking the sum of these products. The results are listed in Table 5.

All that remains is to find the actual categories corresponding to the weights. Comparing Table 5 to Table 3 yields $A_x = A_1$, $A_y = A_2$, $B_x = B_1$, $B_y = B_2$, $B_z = B_3$, $C_a = C_1$, $C_b = C_2$, $C_c = C_3$, $C_d = C_4$, $C_e = C_5$ and $C_f = C_6$. So, we have matched the weights to the strata. For instance, weight 1 corresponds to categories A_1, B_1 and C_5.

Table 5. Estimated frequencies of the categories of the auxiliary variables in the population (evaluation based on microdata set)

Category	Corresponding weight indices																		Frequency in the population
A_x	1	3	5	6	9	10	13	14	15	19	20	21	22	27	29	30	33	34	7,480,000
A_y	2	4	7	8	11	12	16	17	18	23	24	25	26	28	31	32	35	36	7,649,000
B_x	1	2	6	8	9	11	14	17	20	22	24	26							6,572,000
B_y	5	7	13	16	19	23	27	28	33	34	35	36							7,037,000
B_z	3	4	10	12	15	18	21	25	29	30	31	32							1,520,000
C_a	20	24	29	31	33	35													2,765,000
C_b	6	8	10	12	13	16													3,570,000
C_c	22	26	30	32	34	36													3,605,000
C_d	14	17	21	25	27	28													2,549,000
C_e	1	2	3	4	5	7													1,811,000
C_f	9	11	15	18	19	23													829,000

3.2. *Method 2*

The second method consists of one step only. Again the ratios of pairs of weights are evaluated and listed. Now, however, only those ratios that occur G_s times for some s are listed, where G_s is given by (3). In other words, ratios of weights are listed only when these weights may have $(m-1)$ categories in common. Of course, when a G_s satisfies (4), then also ratios of weights are listed of which these weights have less than $(m-1)$ categories in common.

Now suppose the ratio W_{k_1}/W_{ℓ_1} occurs G_s times, i.e., $W_{k_1}/W_{\ell_1} = W_{k_i}/W_{\ell_i}$ for certain weights W_{k_i} and W_{ℓ_i} $(i = 1, \ldots, G_s)$. We do not know yet whether the weights W_{k_i} and W_{ℓ_i} have $(m-1)$ categories in common or not. As in step 2 of Method 1 we multiply the weights W_{k_i} by their frequencies f_{k_i} in the sample and take the sum of these products, i.e., we compute

$$\sum_{i=1}^{G_s} W_{k_i} \times f_{k_i} \tag{8}$$

Now we examine the two possible cases.

1. *The weights W_{k_i} and W_{ℓ_i} have $(m-1)$ categories in common*
   When the weights W_{k_i} and W_{ℓ_i} $(i = 1, \ldots, G_s)$ have $(m-1)$ categories in common, then the W_{k_i}'s are all the weights that have the same category C of a certain variable V in common. As in Step 2 of Method 1 the number given by (8) is equal to the (known) frequency of category C of variable V in the population. So, an intruder can determine a category corresponding to a weight W_{k_i} whenever W_{k_i}/W_{ℓ_i} have $(m-1)$ categories in common.
2. *The weights W_{k_i} and W_{ℓ_i} do not have $(m-1)$ categories in common*
   When the weights W_{k_i} and W_{ℓ_i} $(i = 1, \ldots, G_s)$ do not have $(m-1)$ categories in common, then the number given by (8) is not equal to one of the known frequencies of the categories in the population.

So, by comparing (8) to the known frequencies of the categories in the population we can conclude whether or not the weights W_{k_i} and W_{ℓ_i} have $(m-1)$ categories in common. If they do have $(m-1)$ categories in common, we can moreover determine the category corresponding to the W_{k_i}'s. In other words, by comparing (8) to the known frequencies of the categories in the population an intruder can determine which stratum a particular weight corresponds to.

When linear weighting has been used instead of multiplicative weighting almost the same two methods as described above can be applied. In fact one should only replace ratios by differences. For instance, instead of the ratios list an intruder should make a differences list.

Example 3

Suppose that three auxiliary variables A, B and C have been used. The number of categories of A is two and of both B and C is four. So, $n_1 = 2$, $n_2 = 4$ and $n_3 = 4$. We suppose that linear weighting has been used. Suppose, furthermore, that the frequencies of the categories of the auxiliary variables in the population are given in Table 6.

The weights are listed in ascending order in Table 7.

Part of the differences list is presented in Appendix B.1. Only differences less than zero that occur 8 $(= n_1 \times n_2 = n_1 \times n_3)$ or 16 $(= n_2 \times n_3)$ times are listed.

Table 6. Frequencies of the categories of the auxiliary variables in the population (known to the intruder)

Category	Frequency in the population
A_1	1,485,135
A_2	1,514,865
B_1	754,875
B_2	735,023
B_3	775,036
B_4	735,066
C_1	735,443
C_2	784,387
C_3	745,122
C_4	735,048

Table 7. Weights of the strata

Index i	Weight W_i	Frequency in sample	Weight × Frequency in sample
1	932.4877	96	89,518.82
2	933.1395	89	83,049.42
3	944.7638	97	91,642.09
4	945.4156	85	80,360.33
5	952.5411	104	99,064.28
6	959.1034	105	100,705.86
7	964.8172	97	93,587.27
8	969.4501	97	94.036.66
9	970.1019	93	90,219.48
10	971.3795	100	97,137.95
11	981.7261	109	107,008.15
12	982.3780	102	100,202.55
13	989.5035	102	100,929.36
14	995.7425	88	87,625.34
15	996.0658	93	92,634.12
16	996.3943	97	96,650.25
17	1,001.7796	87	87,154.83
18	1,008.0186	88	88,705.64
19	1,008.3419	102	102,850.87
20	1,008.6704	87	87,754.33
21	1,015.7959	90	91,421.63
22	1,022.3582	87	88,945.17
23	1,028.0720	90	92,526.48
24	1,034.6343	98	101,394.16
25	1,034.7040	88	91,053.95
26	1,035.3558	97	100,429.51
27	1,046.9801	82	85,852.36
28	1,047.6319	92	96,382.13
29	1,054.7574	78	82,271.08
30	1,061.3197	91	96,580.09
31	1,067.0335	92	98,167.08
32	1,073.5958	97	104,138.79

Examining the list in Appendix B.1 we see, for instance, that the difference $W_1 - W_{25}$ occurs eight times. So, W_1 and W_{25} have two categories in common. Because $W_1 - W_{25} = W_2 - W_{26} = W_3 - W_{27} = W_4 - W_{28} = W_5 - W_{29} = W_6 - W_{30} = W_7 - W_{31} = W_{10} - W_{32}$, weights W_1, W_2, W_3, W_4, W_5, W_6, W_7 and W_{10} have the same category in common. In this way we can determine all the groups that have the same category in common. These groups are listed in Appendix B.2. The products of the weights and their frequencies in the sample are also listed. The sum of these products for each group, also given, is equal to the frequency of a category of the auxiliary variables in the population if all the weights in the corresponding group have this category in common.

Comparing the sums of each group in Appendix B.2 to Table 6 yields that weight W_1 corresponds to categories A_1, B_4 and C_1. In a similar way the other weights can be matched to the strata.

In example 3 it is very easy to apply Method 2. If Method 2 had been applied to the second example, then some groups would have been constructed which do not have the same category in common. The sum of the products of the sampling weights in such a group and their frequencies in the sample would, however, differ from any frequency of a category of the auxiliary variables in the population. So, it is still possible to match the sampling weights to the categories of the auxiliary variables in the same way as above.

4. Discussion

In the previous sections we have shown that sampling weights can provide additional identifying information to an intruder when identifying information not contained in the released microdata set is used to calculate these weights. If this leads to an unacceptable risk of disclosure then it may be desirable to apply specific SDC measures. Two techniques to reduce the risk of disclosure caused by sampling weights could be applied. The aim of these techniques would be to prevent the successful application of the methods described in Sections 2 and 3.

The first method is to subsample the records with a relatively low weight in the microdata set and then to re-calculate the weights for the remaining records. As a consequence the weights of the remaining records with low original weights are increased. In this way one can make all the weights of the records approximately equal. In this case the weights cannot provide any additional identifying information. Subsampling leads to a loss of information, of course, because some records are not released.

The second method to reduce the risk of disclosure caused by sampling weights is to add noise to these weights. In other words, instead of releasing the true sampling weight W_i of record i the statistical office releases values $W_i' = W_i + \epsilon_i$, where ϵ_i is a random value. When noise is added to the sampling weights these weights are perturbed. Because generally all these perturbed weights will have a different value the methods described in Sections 2 and 3 cannot be applied immediately. So, obtaining additional identifying information is made (much) more difficult for an intruder. However, when much noise is added to the sampling weights the resulting perturbed weights will hardly be useful for subsequent analysis. In particular, the results may be biased. When, on the other hand, little noise is added, the weights should remain useful for analysis, but the probability that an intruder can obtain additional information from these weights will be relatively high. How much noise should be added in order to obtain both 'safe' and useful weights is a problem that remains to be solved.

Appendix A: The Ratios List of Example 2

The numbers in Columns 1 and 2 are the indices of the weights in Table 4. The number in the third column is the ratio of the two weights in Column 1 and Column 2. Only ratios pertaining to Example 2 are listed.

Index i	Index j	Ratio W_i/W_j	Index i	Index j	Ratio W_i/W_j
9	23	0.97155	11	18	0.98528
14	28	0.97155	14	21	0.98528
6	16	0.97155	26	32	0.98528
22	36	0.97155	20	29	0.98528
1	7	0.97155	1	3	0.98528
20	35	0.97155	9	15	0.98528
			22	30	0.98528
2	8	0.97733	6	10	0.98528
4	12	0.97733	17	25	0.98528
7	16	0.97733	2	4	0.98528
5	13	0.97733	8	12	0.98528
1	6	0.97733	24	31	0.98528
3	10	0.97733			
			15	19	0.99208
14	27	0.97747	21	27	0.99208
11	23	0.97747	31	35	0.99208
9	19	0.97747	4	7	0.99208
20	33	0.97747	18	23	0.99208
26	36	0.97747	12	16	0.99208
2	7	0.97747	10	13	0.99208
6	13	0.97747	29	33	0.99208
1	5	0.97747	3	5	0.99208
8	16	0.97747	32	36	0.99208
22	34	0.97747	30	34	0.99208
24	35	0.97747	25	28	0.99208
17	28	0.97747			
			14	17	0.99394
4	10	0.98329	15	18	0.99394
2	6	0.98329	21	25	0.99394
7	13	0.98329	34	36	0.99394
			19	23	0.99394
4	8	0.99193	9	11	0.99394
3	6	0.99193	6	8	0.99394
			30	32	0.99394
			13	16	0.99394
			5	7	0.99394
			22	26	0.99394
			20	24	0.99394
			1	2	0.99394
			10	12	0.99394
			3	4	0.99394
			33	35	0.99394
			27	28	0.99394
			29	31	0.99394

Appendix B.1: The Differences List of Example 3

The numbers in Columns 1 and 2 are the indices of the weights in Table 7. In the third column the difference of the weights in Columns 1 and 2 is listed. Only differences less than 0 that occur 8 or 16 times are listed.

Index i	Index j	$W_i - W_j$	Index i	Index j	$W_i - W_j$	Index i	Index j	$W_i - W_j$
2	26	−102.2163	11	19	−26.6158	2	4	−12.2761
3	27	−102.2163	3	10	−26.6157	6	10	−12.2761
4	28	−102.2163	14	22	−26.6157	9	12	−12.2761
6	30	−102.2163	25	30	−26.6157	14	18	−12.2761
7	31	−102.2163	27	32	−26.6157	15	19	−12.2761
10	32	−102.2163	1	6	−26.6157	25	27	−12.2761
1	25	−102.2163	8	15	−26.6157	26	28	−12.2761
5	29	−102.2163	18	24	−26.6157	29	31	−12.2761
						30	32	−12.2761
11	27	−65.2540	11	18	−26.2925	1	3	−12.2761
9	26	−65.2539	9	16	−26.2924	5	7	−12.2761
12	28	−65.2539	15	22	−26.2924	13	17	−12.2761
15	30	−65.2539	8	14	−26.2924	16	20	−12.2761
17	31	−65.2539	12	20	−26.2924	21	23	−12.2761
19	32	−65.2539	13	21	−26.2924	22	24	−12.2761
8	25	−65.2539	17	23	−26.2924	8	11	−12.2760
13	29	−65.2539	19	24	−26.2924			
						7	10	−6.5623
2	16	−63.2548	2	6	−25.9639	17	19	−6.5623
3	18	−63.2548	4	10	−25.9639	21	22	−6.5623
6	22	−63.2548	9	15	−25.9639	23	24	−6.5623
1	14	−63.2548	12	19	−25.9639	29	30	−6.5623
4	20	−63.2548	16	22	−25.9639	31	32	−6.5623
5	21	−63.2548	20	24	−25.9639	5	6	−6.5623
7	23	−63.2548	26	30	−25.9639	13	15	−6.5623
10	24	−63.2548	28	32	−25.9639			
						11	12	−0.6519
20	28	−38.9615	11	17	−20.0535	3	4	−0.6518
23	31	−38.9615	1	5	−20.0534	14	16	−0.6518
24	32	−38.9615	3	7	−20.0534	25	26	−0.6518
14	25	−38.9615	8	13	−20.0534	27	28	−0.6518
16	26	−38.9615	14	21	−20.0534	1	2	−0.6518
18	27	−38.9615	25	29	−20.0534	8	9	−0.6518
21	29	−38.9615	27	31	−20.0534	18	20	−0.6518
22	30	−38.9615	18	23	−20.0534			
1	8	−36.9624	2	5	−19.4016			
2	9	−36.9624	9	13	−19.4016			
4	12	−36.9624	4	7	−19.4016			
5	13	−36.9624	12	17	−19.4016			
6	15	−36.9624	16	21	−19.4016			
7	17	−36.9624	20	23	−19.4016			
10	19	−36.9624	26	29	−19.4016			
3	11	−36.9623	28	31	−19.4016			

Appendix B.2:
The Groups of Weights Containing W_1 that Have a Specific Category in Common

(Example 3 continued)

Index	Weight × Frequency
Group 1:	
1	89,518.82
2	83,049.42
3	91,642.09
4	80,360.33
5	99,064.28
6	100,705.86
7	93,587.27
10	97,137.95
Total	735,066.02
Group 2:	
1	89,518.82
3	91,642.09
8	94,036.66
11	107,008.15
14	87,625.34
18	88,705.64
25	91,053.95
27	85,852.36
Total	735,443.01
Group 3:	
1	89,518.82
2	83,049.42
5	99,064.28
6	100,705.86
8	94,036.66
9	90,219.48
13	100,929.36
14	87,625.34
15	92,634.12
16	96,650.25
21	91,421.63
22	88,945.17
25	91,053.95
26	100,429.51
29	82,271.08
30	96,580.09
Total	1,485,135.02

5. References

Bethlehem, J.G. and Keller, W.J. (1987). Linear Weighting of Sample Survey Data. Journal of Official Statistics, 3, 141–153.

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

De Waal, A.G. and Willenborg, L.C.R.J. (1996). A View on Statistical Disclosure Control for Microdata. Survey Methodology, 22, 95–103.

Deville, J.C. and Särndal, C.E. (1992). Calibration Estimators in Survey Sampling. Journal of the American Statistical Association, 87, 376–382.

Duncan, G. and Lambert, D. (1989). The Risk of Disclosure for Microdata. Journal of Business and Economic Statistics, 7, 207–217.

Van Kouwen, M. (1993). Aspects of Statistical Disclosure Control for Sampling Weights. Internal Report, Statistics Netherlands, Voorburg. [In Dutch].

Willenborg, L. and De Waal, T. (1996). Statistical Disclosure Control in Practice. Springer-Verlag, New York.

Received May 1995
Revised November 1996