TRANSACTIONS ON DATA PRIVACY 1 (2008) 34 —52

A Measure of Disclosure Risk for Tables of Counts

Duncan Smith*, Mark Elliot**
* Confidentiality and Privacy Group
Cathie Marsh Centre for Census and Survey Research
University of Manchester
duncan.g.smith@manchester.ac.uk
** Confidentiality and Privacy Group
Cathie Marsh Centre for Census and Survey Research
University of Manchester
mark.elliot@manchester.ac.uk

Abstract. The paper describes a new method for assessing disclosure risk for tables of counts; the *subtraction - attribution probability* (SAP) method. The SAP score is the probability of an intruder recovering a 'risky' subpopulation table given a quantity of information about the individuals in a population table. The method can be applied to exact or perturbed individual tables and sets of tables. The method can also be used to compare the risk impact of different disclosure control regimes.

1 Introduction

Releases of population data can be used by so-called data intruders to glean sensitive information about individuals in the population. Disclosure occurs when a data intruder makes reliable inferences (i.e., with a high degree of confidence) about one or more population units. Statistical agencies need to guard against disclosure in order to meet their legal obligations, to safeguard respondent confidentiality and to maintain public trust. For example, lack of trust can result in individuals refusing to complete census forms or returning forms with false or missing information. Most statistical agencies are mainly concerned with the risk of an intruder identifying a population unit, although this is not a requirement for disclosure of information about the individual concerned.

The need for appropriate measures of disclosure risk has been well discussed. Many authors have indicated that such measures should as far as possible take a data intruder's perspective of the risk (see e.g. Paass 1988, Mokken et al 1992, Elliot and Dale 1999). Although intruder-based measures have been established for identification risk (Skinner and Elliot 2002), little progress has been made with generating appropriate risk metrics for the disclosure of information about members of the population in the absence of identification.[1]

This paper describes the "subtraction - attribution probability" (SAP) method which attempts to fill this gap. We first describe the disclosure risk problem of tables of counts and motivate the idea of a measure based on the presence of inferable zeroes in a given cross classification. We then show how our chosen metric (SAP) can provide an estimate of the probability of an intruder recovering a zero from a given cross-classification given that s/he holds a given amount of information about individuals within the population which the cross-classification represents. We demonstrate the method using random rounding. However, the method is compatible with any methodology which produces tables of counts with finite bounds.[2] Finally we demonstrate the use of the SAP with some simulated data.

2 Disclosure Risk

Understanding of disclosure risk has evolved over the last twenty years and there is still no unequivocal definition of the term. However, definitions of disclosure generally involve one or both of *identification* (a one to one association between a data unit and a target) and *attribution* (the association of one or more variable values with a target).

Herein, a data unit is data about a population unit contained in a dataset that is available to a data intruder[3]; a target is the population unit about which a data intruder is trying to discover information.

In some cases it is possible for an intruder to perform identification or attribution with absolute certainty. In these cases the identification or attribution is

[1] This type of disclosure risk mostly concerns population aggregates and in particular tables of counts – the topic of this paper is noteworthy that whilst there has been extensive and increasingly sophisticated work on methods for controlling disclosure risk in such aggregates (see for example Cox (1995; et al 2004); Salazar (2005, 2006, 2007); Duncan and Roherig(2004). However, all of these in effect arbitrarily define properties of the data as being "unsafe" or "protected". They do not (as we do here) explicitly attempt to measure the risk of plausible intruder behaviour as being successful and therefore, we would argue they do not measure *disclosure risk*.

[2] Arguably, the fact that the method works with perturbative disclosure control techniques at allmakes it superior to many other methods which generally are only applicable to unperturbed data..

[3] A data intruder is an individual (or organisation) who seeks to obtain information about a population unit through statistical disclosure.

termed exact. Otherwise, identification or attribution is termed approximate. Strictly speaking there will almost always be a degree of uncertainty regarding the correctness of the data, so all inferences are approximate. However, this source of uncertainty is generally ignored for disclosure risk assessment purposes, and we follow this practice here.

In this paper we address the risk of exact attribution. Previous papers have tended to concentrate on attribution stemming from identification. Fellegi (1972) considers disclosure in terms of “sufficiently narrowly defined” populations, and goes on to state that such a population may “contain only one identifiable respondent or, at least, information can be deduced from the published estimates that can be related to a particular identifiable respondent”. He then goes on to illustrate how disclosure can occur from the conditioning on known information about a target, the conditional frequency table containing only the target individual. Clearly, if an intruder can achieve this by conditioning on a subset of the variables in the tabulation, then the levels of all the remaining variables can be discovered. If the levels of the discovered variables were previously unknown to the intruder, then disclosure has taken place. Fellegi also considers that conditioning to a (sub-) population of size two can result in similar disclosure if the intruder is the other member of the conditional population. A U.S. Department of Commerce report (1978) expands this idea by considering “coalitions” of individuals within a data set who might cooperate in order to discover new information about targeted individuals. The report also considers how disclosure can take place without the requirement for identification. Their examples are reproduced below.

Table 1. Number of beneficiaries by count and race.

		Race		
County	White	Black	Other	Total
A	15	20	5	40
B	0	30	0	30

In Table 1 conditioning on a target being a resident of County B implies that the target is black. A risk of such exact disclosure exists if a marginal total (in dimension n-1) equals one of its detail cells (in dimension n). This contrasts with the example given by Fellegi which required also that the detail cell count be 1.

The U.S. Department of Commerce report contrasts this with the case when the sum of a proper subset of detail cells equals the total in the relevant margin (Table 2). The report does not define the implication that a target in County B is either Black or Other as disclosure, because the subset of Black or Other is not as narrowly defined as possible. Similarly, the report authors do not consider exact inferences regarding age as disclosive unless age is revealed to within a single year.

Table 2. Number of beneficiaries by count and race.

	Race			
County	White	Black	Other	Total
A	15	20	5	40
B	0	28	2	30

We consider this distinction to be fairly arbitrary as ethnicity can be broken down into more detailed classifications than those of the example, and any categorisation of a continuous variable such as age will involve ranges that are not as narrowly defined as possible. One approach would be to associate sensitivities with any set/range of variable levels and consider disclosure to have taken place if the sensitivity of the discovered information exceeds some predefined threshold. However, data is often collected with an unqualified assurance of confidentiality, so that it is arguable that all data should be regarded as sufficiently sensitive to warrant protection. Therefore, for the purposes of this paper we will simply consider that disclosure takes place when an intruder, by whatever means, is able to condition to a table of population counts which contains one or more zeros. This definition encompasses the two cases illustrated above, and the additional case where an intruder can infer that a particular combination of attribute levels does not apply to a target. For example, simply conditioning on a target being a member of the population in Table 2 allows the intruder to infer that the target is not White and residing in County B (although either is individually possible). So in this strict sense, we consider a risk of disclosure to be present if a table of population counts contains one or more zeros.

Skinner (1992) defines disclosure in the sense of Fellegi's example (requiring identification and attribution) as identification disclosure, whereas disclosure that does not require identification is prediction disclosure. He considers approximate disclosure in sample tables and develops an argument that identification disclosure is a necessary and sufficient condition for prediction disclosure. In this paper we are concerned only with the risk of exact disclosure in population tables. Under these circumstances it is clear that identification is neither necessary nor sufficient for attribution (prediction disclosure).

2.1 Attribution risk from low population counts

So far, we have only considered the risks of attribution as they stem from conditioning on known information relating to some targeted individual. It is implicit that the intruder is also conditioning on a target being a member of the population. But conditioning on known variable levels (or known absence of variable levels) is not the only way an intruder might attempt to condition down to a smaller, more disclosive population. The U.S. Department of Commerce report describes the possibility of disclosure stemming from coalitions, the main questions arising regarding the likely size of coalition, and the distribution of the coalition within the population. However, we note that the type of disclosure that can arise from coalitions does not require their existence. It is possible for an intruder to hold information on a number of population units, without their explicit cooperation. If they can be identified within the population, then their records can be removed from the data set, facilitating inferences regarding the residual subpopulation. Removal of a unique (count of 1) in a population table clearly leads to the presence of a zero, and a risk of attribution. Of course, records of known individuals may be removed without identification, and partially known individuals might be 'removed' from the relevant margins, placing constraints on the counts in the full cross-classification of the residual population. In essence, an intruder can use arbitrary known information about the population units in order to try to facilitate attribution. Lower counts represent a greater risk of the recovery of zeros by subtraction of known individuals.

The above requires information that can be considered external to the data set in question, and as such might not be considered an overriding issue. However, any inferences regarding a population unit require such information. Both exact identification and exact attribution require external information; at the very least an intruder must be able to condition on a target being a member of the relevant population.

2.2 Protection against attribution

Statistical agencies tend to guard against disclosure by suppressing (withholding) data or disguising the true counts by deterministic or stochastic perturbations; (see Duncan et al (2001) for a review). For example, one deterministic method is conventional rounding. A suitable non-negative odd integer is chosen as base, and each count in the cross-classification is rounded to the closest multiple of the base. Figure 2 contains the conventionally rounded, to base 3, cross-classifications corresponding to the exact cross-classification in Figure 1.

Figure 1. A 2-way cross-classification with margins.

		VAR2			
		D	E	F	
VAR1	A	1	3	0	4
	B	4	0	0	4
	C	3	2	0	5
		8	5	0	13

Figure 2. Conventionally rounded cross-classification.

		VAR2			
		D	E	F	
VAR1	A	0	3	0	3
	B	3	0	0	3
	C	3	3	0	6
		9	6	0	12

An intruder (with knowledge of the rounding scheme) can easily generate bounds on the counts in the exact 2-way cross-classification, given the corresponding rounded cross-classification. We term these *trivial* bounds as they are based solely on the rounding scheme.

Figure 3. Trivial lower bounds of Figure 2.

		VAR2		
		D	E	F
VAR1	A	0	2	0
	B	2	0	0
	C	2	2	0

Figure 4. Trivial upper bounds of Figure 2.

		VAR2		
		D	E	F
VAR1	A	1	4	1
	B	4	1	1
	C	4	4	1

Here the rounding has managed to disguise the exact value of all counts. But subtraction of a known individual in cell (A,D) would recover a zero.

It is not unusual for statistical agencies releasing perturbed cross-classifications to also release perturbed, or occasionally exact, marginal tables. The presence of marginal counts places a system of linear constraints on the counts in the full (in this case 2-way) cross-classification. Solving the system of constraints via integer linear programming methods can lead to tighter bounds than those derived solely from a full rounded cross-classification. Dobra (2002) develops a method for solving cell bounds given marginal cell counts. Although his algorithm is designed to deal with exact cross-classifications it is relatively easily extended for dealing with perturbed counts (Smith and Elliot, 2003). The release of all the rounded cross-classifications (including both 1-way margins and rounded total) in Figure 2 results in the following lower and upper bounds.

Figure 5. Non-trivial lower bounds of Figure 2.

		VAR2		
		D	E	F
VAR1	A	0	2	0
	B	3	0	0
	C	3	2	0

Figure 6. Non-trivial upper bounds of Figure 2.

		VAR2		
		D	E	F
VAR1	A	1	3	0
	B	4	1	0
	C	4	3	0

Three of the four zeros have been recovered. This stems from the fact that the trivial lower bounds for the VAR2 margin sum to 13, which is the trivial upper bound for the rounded total. Thus the total and VAR2 margin are recovered exactly. So the perturbation of the data has done little to remove the risk of attribution. Subtraction of individuals could increase the risk still further.

2.3 A measure of attribution risk

As the discussion in section 2.2 indicates: a risk of exact attribution exists if, and only if, one or more zeros, recoverable by an intruder, exist in some cross-classification. We regard this as analytically true, i.e. true by definition of the term "exact attribution". If there are no zeroes in a given cross classification – taking account of the intruder's prior knowledge about the population units represented in that cross classification - then exact attribution is logically impossible. Conversely, if once the intruder has taken account of prior knowledge about population units represented within the cross classification and there are zeroes present anywhere in the residual cross classification (once that knowledge has been accounted for) , then it only requires further knowledge regarding membership of a target within the set of population units represented in the cross classification to (at the very least) infer a combination of variable levels that do not apply to the target.[4] Without zeros, any combination of variable levels could apply. The only exception concerns structural zeros, where it would be difficult to assert that the non-applicability of an impossible combination of variable levels constituted disclosure. This is independent of whether disclosure control has been applied to the cross classification and of what disclosure control method has been used. Furthermore, the population cross-classification in question need not necessarily have been released. In general, a set of population cross-classifications can be used to place bounds on any cross-classification from which they could be derived.[5] It is enough to consider only the 'base' cross-classification with axes corresponding to the union of the variables in the released cross-classifications. Any cross-classifications over a superset of the variables in the base cross-classification contain (recovered) zeros if, and only if, the base cross-classification contains (recovered) zeros. Bounds on smaller margins can be solved, but again this is unnecessary, as any zero in a margin implies zeros in the full cross-classification.

Given the questionable distinction between inferences on the basis of the 'narrowness of definition' we propose a measure based simply on the presence of zeros in the full population cross-classification. Sensitivities are not considered for the reason given earlier, although we note that the methodology can be applied to conditional tables as easily as marginal tables, in which case we could assess risk for given population units or population cells given an assumed set of *key* variables. We also wish to take into account the additional risk stemming from intruder knowledge of the population, and to be able to apply the measure

[4] The reader may have noted that, by this definition, standard sample microdata sets would be at risk of attribution disclosure. Logically this is so, however, the further step that the intruder requires further knowledge regarding membership of a target within the set of population units represented in the cross classification is the barrier to disclosure in this case. With sample microdata, the situation we are describing here corresponds to one of *response knowledge* (i.e. where the intruder knows, through prior knowledge, that a population unit is contained within a sample). This is a special situation, which it is generally accepted is outside the realm of practical SDC. In practice, this definition would be applied to population cross classifications, typically census tables where membership of the data cross classification can be reasonably inferred from membership of the population and it is this situation with which we are primarily concerned within this paper..

[5] A special case of this is where it is possible to recover the exact counts in a full 3-way cross-classification (containing a zero) from its three distinct 2-way margins (containing no zeros).

to relatively arbitrary releases of exact and / or perturbed cross-classifications. Specifically, our chosen measure is the 'probability of recovering one or more zeros in the full cross-classification given the subtraction of a random sample of n population units'. We term this the subtraction attribution probability (SAP).

Assume we have a base table of counts of arbitrary dimension with cell counts c_i, $i = 1$ to m. Assume that an arbitrary set of perturbed marginal tables is published, each perturbed using some independent rounding scheme (i.e. each cell is perturbed independently of the others). Then each published count, x, implies a pair of constraints of the form, $l \le c$, $c \le u$, where l and u are the trivial bounds implied by the rounding scheme and c is the total of some set of cells in the base table. Dependencies between bounds might imply that there exist tighter bounds than the trivial bounds. These may be found by integer linear programming methods. The recovery of a zero by subtraction of a known sample of the population occurs if, and only if, the sample implies that $s_i = c_i = u_i'$, where s_i is the corresponding known sample count and u_i' is the corresponding value in the table of the tightest upper bounds on the base table implied by the set of all linear constraints.

In general, the probability of recovering at least one zero for some assumed level of intruder knowledge, equivalent to a random sample of size n, is:

$$SAP(n) = \frac{\sum_{s \in S} P(s \mid p) I(\sum_i s_i = n) I(0 \in u' - s)}{\sum_{s \in S} P(s \mid p) I(\sum_i s_i = n)},$$

where S is the set of all possible sample tables, p is the population table (known to the data holder), sampling is simple random sampling without replacement, subtraction of tables is elementwise, and $I(.)$ is the indicator function. i.e. It is simply the weighted proportion of sample tables that have at least one upper bound equal to a sample count.

Where data has been perturbed, the calculation of the SAP measure may be simplified computationally due to the additional constraints imposed by the perturbation. For a data release comprising of a single rounded table we have a pair of constraints, $l_i \le c_i$, $c_i \le u_i$ for each cell i. The mutual orthogonality of these pairs of constraints in R^n ensures that the trivial bounds are the tightest bounds. For a sample with corresponding sample counts, s_i, $i = 1$ to m, the SAP measure for a given sample size, n, can be calculated as follows.

2.3.1 Single rounded table

The marginal probability of recovering zeros in any set of cells with total x is simply the following Hypergeometric probability,

$$\frac{\binom{N-x}{n-x}}{\binom{N}{x}}$$ where N is the cross-classification total, $\sum c_i$.

Applying the inclusion / exclusion principle it is simple to derive an expression for the probability that at least one cell is zero given a random sample of n population units.

Let Z denote the set of all subsets of cell indices, equal to the union of the sets of n-subsets $Z(0)$, …, $Z(m)$. i.e. $Z(0) = \emptyset$, $Z(1) = \{\{1\}, \ldots, \{m\}\}$, $Z(2) = \{\{1,2\}, \{1,3\}, \ldots, \{m\text{-}1, m\}\}$, …, $Z(n) = \{\{1, \ldots, m\}\}$.

Let e.g. $c_1 + c_2$ be denoted by $c_{\{1,2\}}$.

Then,

$$SAP(n) = \sum_{i=1}^{n} \left((-1)^{i-1} \sum_{z \in Z(i)} \frac{\binom{N-c_z}{n-c_z}}{\binom{N}{n}} \right)$$

In practice many of the terms in the above summation will be equal to zero. For exact tables we have $l_i = c_i = u_i$ for all i, and all cell counts represent some risk of recovering a zero, although for a given level of risk, n, we need only consider c_z s.t. $c_z \le n$. For rounded tables we need only consider c_z s.t. $c_z = \sum_{i \in z} u_i \le n$.

Calculations for the following exact counts (with no constraints on margins) are given below.

[2,1,3]

$$SAP(1) = \frac{\binom{6-1}{1-1}}{\binom{6}{1}} = \frac{1}{6}$$

$$SAP(2) = \frac{\binom{6-2}{2-2}}{\binom{6}{2}} + \frac{\binom{6-1}{2-1}}{\binom{6}{2}} = \frac{6}{15}$$

$$SAP(3) = \frac{\binom{6-2}{3-2}}{\binom{6}{3}} + \frac{\binom{6-1}{3-1}}{\binom{6}{3}} + \frac{\binom{6-3}{3-3}}{\binom{6}{3}} - \frac{\binom{6-(2+1)}{3-(2+1)}}{\binom{6}{3}} = \frac{14}{20}$$

$$SAP(4) = \frac{\binom{6-2}{4-2}}{\binom{6}{4}} + \frac{\binom{6-1}{4-1}}{\binom{6}{4}} + \frac{\binom{6-3}{4-3}}{\binom{6}{4}} - \frac{\binom{6-(2+1)}{4-(2+1)}}{\binom{6}{4}} - \frac{\binom{6-(1+3)}{4-(1+3)}}{\binom{6}{4}} = \frac{15}{15}$$

SAP(0) is trivially zero as there are no zero counts, and *SAP*(*n*) is trivially one for all $n > 3$.

2.3.2 Single rounded table and rounded total

In this case we have an additional pair of constraints, $l_t \le \sum_i c_i$, $\sum_i c_i \le u_t$, where u_t denotes the trivial upper bound for the table total. We also have the obvious risk of subtraction where the sum of the sample counts $\sum_i s_i = u_t$, and this only occurs when $\sum_i s_i = \sum_i c_i = u_t$. But this new constraint is not mutually orthogonal to the existing constraints, and the trivial upper bounds might not be the tightest possible bounds. In this particular case the tightest possible upper bounds on any base table cell *j* is, $u'_j = \min\left(u_j, u_t - \sum_{i \ne j} l_i\right)$.

Lemma1. If $s_i = u_i$ for any $s \in S$ and any *i*, then $s_i < u'_i$ for any $u'_i < u_i$. In other words, if there is any risk for the release without rounded total, then the release of the rounded total results in no increased risk.

Proof. It would be sufficient to show that $u_t - \sum_{i \neq j} l_i > c_j$ for any *j*. Minimum upper bounds occur when $u_t = \sum_i c_i$. So assuming the tightest possible 'new' bounds we have,

$$\sum_i c_i - \sum_{i \neq j} l_i > c_j$$

$$\sum_i (c_i - l_i) > c_j - l_j$$

So the lemma is proved true apart from the case where $c_i = l_i \ \forall \ i \neq j$, where we have equality.

In this case we have,

$$u'_j = \min\left(u_j, \sum_i (c_i - l_i) + l_j\right)$$

$$u'_j = \min(u_j, c_j - l_j + l_j)$$

$$u'_j = c_j$$

The existence of a risk (without rounded total) implies that $u_i = c_i$ for at least one c_i. So, either,

1. $u_j = c_j = u'_j$ and there is no increased risk (the bound is already tight), or
2. $l_i = c_i = u_i$ for some $i \neq j$, and we have a rounding scheme that doesn't round all counts.

Lemma1 is proved for all independent rounding schemes that perturb all base table counts.□

Corollary1. If $u_t > \sum_i c_i$, then the risk with rounded table is exactly the same as the risk without rounded table.

Corollary2. If a rounded table represents zero risk, then the addition of a rounded total represents a risk if, and only if, $u_t = \sum_i c_i$. This risk pertains only to knowledge of the full table, unless exactly one cell count, say c_j, is not equal to its trivial lower bound. In that case all tables s.t. $s_j = c_j$ represent a risk.

So if $s_i = u_i$ for any $s \in S$ or $u_t > \sum_i c_i$, then we can use the algorithm for single rounded tables.

Otherwise, the above results lead to the following algorithm.

1. Construct a list containing the trivial lower bounds for the rounded base table counts (i.e. based solely on the rounding scheme).
2. Construct a corresponding list of counts for the exact cross-classification.
3. Find the sum, S, of those counts in the list of lower bounds that are equal to the corresponding count in the list of exact counts.
4. For all *n* in the range 0 to (T – S – 1) (where T is the exact cross-classification total) the SAP measure is zero.
5. For each *n* in the range (T – S) to T the SAP measure equals $\dfrac{\binom{S}{n-T+S}}{\binom{T}{n}}$

2.3.3 General table releases

It is hoped that the existing results can be further generalised to provide efficient means for calculating SAP measures for more general table releases. The current approach is to use an extended version of Dobra's shuttle algorithm to solve the initial bounds problem and then recursively generate all tables with non-zero risk (Smith and Elliot, 2003). Randomly sampling tables is an alternative approach for generating approximate SAP measures.

Calculating exact measures by generating all feasible tables tends to be very computationally expensive. Obviously, the number of feasible tables tends to increase rapidly with population size, although it also depends on the perturbation scheme, the constraints implied by margins and the assumed sample size *n* This was the initial approach used for the analysis of large numbers of sets of small tables which prompted the development of the above algorithm. But the above results suggest other efficient algorithms for other cases.

2.4 An example with simulated data.

For contractual reasons we are, at present, unable to publish an extensive SAP analysis that we have conducted on the UK Neighbourhood Statistics. But Table 3 contains some results for an analysis of a set of 1200 randomly generated 2×6 cross-classifications. The cross-classification counts were generated from a Poisson distribution with mean 2. Each cross-classification was conventionally rounded to base 5, and the exact total was conventionally rounded (to base 5) to produce a rounded total. SAP measures for each cross-classification were generated for n=0 to 24 using the algorithm outlined in Section 2.3.2. Table 3 contains the numbers of cross-classifications that had SAP scores in various ranges. SAP scores that were exactly 0 or 1 are contained in the second left and rightmost columns respectively.

The software was implemented in the Python programming language. The simulation took just under 1 minute to run on an old Athlon XP 2100 box with 1Gig RAM running Windows XP.

For n in {0,1} the SAP measure is necessarily 0 for all cross-classifications, due to the nature of the rounding scheme. For n=2 the SAP measure could be as high as 1, given a cross-classification total of 2.

The SAP measure for any individual cross-classification and value of n must be at least that for n-1, so there tends to be a migration of SAP measures from 0 to 1 as n is increased. But for cross-classifications with no relevant cells, the SAP measure is zero for all n.

Table 3 demonstrates how the risk of recovering a potentially attribute-disclosive cross-classification tends to increase with greater intruder knowledge of the population. Of course, this depends on the size of the cross-classifications, the distribution of counts and the rounding scheme. But the pattern of results shown in Table3 is reasonably close to that which the authors have found with real-world data sets. Analyses such as this can be used to help define threshold values for n for which a non-zero (or value greater than another threshold) SAP measure can be considered to constitute too great a risk for release. Similarly, analyses can be used to investigate the protection afforded by alternative perturbation schemes. Of course, any comprehensive analysis of perturbation schemes would also consider the effect of perturbation on data quality.

Table 3. Simulation results showing for 1200 randomly generated tables the banded probabilities of producing a table containing at least one zero, given subtraction of *n* randomly selected units from the tables.

SAP	=0	0-0.1	0.1-0.2	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1	=1
n=0	1200	0	0	0	0	0	0	0	0	0	0	0
1	1200	0	0	0	0	0	0	0	0	0	0	0
2	26	1173	1	0	0	0	0	0	0	0	0	0
3	26	1137	36	0	0	1	0	0	0	0	0	0
4	25	869	275	28	2	0	0	0	1	0	0	0
5	25	460	497	160	40	16	1	0	0	0	0	1
6	25	234	471	267	127	48	20	6	1	0	0	1
7	23	108	332	365	169	113	59	13	14	2	0	2
8	23	60	226	266	292	115	117	50	28	17	4	2
9	23	33	144	201	254	212	115	102	64	31	17	4
10	23	14	93	146	188	203	220	93	105	75	30	10
11	23	9	58	110	158	150	205	176	125	95	72	19
12	22	4	42	63	108	149	186	154	188	113	138	33
13	22	4	21	51	84	126	126	169	211	148	186	52
14	22	3	12	37	57	104	111	138	157	234	244	81
15	22	3	8	33	32	65	104	124	159	217	303	130
16	22	3	2	18	32	46	80	108	127	199	379	184
17	22	3	0	12	30	31	46	101	118	182	395	260
18	21	4	0	7	16	28	39	73	93	144	442	333
19	20	5	0	2	13	26	25	45	84	143	409	428
20	20	5	0	0	9	12	29	40	54	105	423	503
21	20	4	1	0	7	9	26	18	43	104	365	603
22	20	3	1	0	2	10	9	27	35	74	318	701
23	20	3	1	0	0	7	10	16	26	42	290	785
24	19	3	1	1	0	5	6	10	26	33	235	861

3 Concluding Remarks

The SAP method provides an integrated approach for assessing attribute disclosure risk for any given release of cross-classifications. It incorporates a representation of intruder knowledge and allows the same metric to be produced for both single and multiple cross-classifications, whether perturbed or unperturbed.

Computational constraints mean that comprehensive analyses of large cross-classifications, using exact measures, can be time consuming. The computational burden can be ameliorated through sampling to derive approximate SAP measures; further work is needed on producing exact measures and far more efficient algorithms have been found for certain special cases. These cases were chosen for no other reason than the fact that they are common forms of release from the Office for National Statistics; in fact, there are obvious extensions to other cases that are not detailed here.

Another important issue not addressed here is how practical the criteria definition for baseline disclosure is. In other words, is the "recovery of a zero anywhere in the full table" too strong a definition of disclosure risk? To illustrate this with an extreme example: suppose that an NSI releases a selection of tables based on an underlying full cross classification of thirty variables. If, through subtraction, an intruder could recover a zero in an interior cell of the full cross classification should this be of concern to the NSI? Establishing that no population unit in the target population had a particular combination of thirty variables would meet a literalist definition of disclosure. However, such a disclosure is unlikely to be of interest to the data intruder, nor is it likely to be of great concern to a member of the target population.

This issue is part of a more general one of when does a literal disclosure become a sensitive one. To deal with the technical side of this, the SAP measure could be adapted to take into account the maximum depth below which the NSI-user would cease to be concerned, the point at which a technical disclosure might reasonably be regarded as uninteresting. Alternatively separate statistics could be produced for each depth; although this would make interpretation less straightforward. Further work is needed here.

A final point concerns the question of what information the intruder might have. This is an empirical question which can only be dealt with through data environment analysis; Elliot and Purdam (2002). This is clearly important and therefore is a stand of urgent applied work.

Notwithstanding the above qualifications, the SAP method, as it stands, provides a complete risk measure for exact attribute disclosure, which is anchored in mirroring what a data intruder might actually do to attack a cross-classified release of aggregate counts. It can be applied to arbitrary sets of tables as long as finite bounds can be generated for cell counts, and the exact counts are available for reference. This would also include tables with suppressed cells, tables containing intervals and tables produced from controlled rounding (Salazar et al., 2004). This in itself represents an advance. The data intruder is assumed to be able to calculate bounds, whilst the data holder has access to the original data in order to calculate the measure. Where it is not possible for the intruder to calculate finite upper bounds, then exact attribute disclosure is not possible.

References

[1] Dobra, A. Statistical Tools for Disclosure Limitation in Multi-Way Contingency Tables. *Ph.D. Thesis*, Department of Statistics, Carnegie Mellon University, 2002.

[2] Duncan, G.T. Fienberg, S.E., Krishnan, R. Padman, R. and Roehrig, S. F. Disclosure Limitation and Information Loss for Tabular Data. In P. Doyle, J. I. Lane, J. M. Theeuwes, and Zayatz, L. M. (eds) *Confidentiality and Data Access: Theory and Practical Application for Statistical Agencies.* pp 135-183, 2001.

[3] Duncan, G. T. and Roehrig S. F. (2004). Cyclic perturbation: Protecting condentiality while preserving data utility in tabular data. Technical report, Carnegie Mellon University.

[4] Cox L.H. (1995) Network Models for Complimentary Cell Suppression, *Journal of the American Statistical Association* **90,** 1453-1462.

[5] Cox, L.H., Kelly, J.P., and Patil, R. (2004). Balancing Quality and Confidentiality for Multivariate Tabular Data. In J. Domingo-Ferrer, & V. Torra (Eds.), *Privacy in Statistical Databases -PSD.* Barcelona, Catalonia, Spain: Springer-Verlag.

[6] Elliot, M. J. DIS: A New Approach to the Measurement of Statistical Disclosure Risk. *International Journal of Risk Management*, 2(4) pp.39-48, 2000.

[7] Elliot, M. J., Dale, A. Scenarios of Attack: The Data Intruder's Perspective on Statistical Disclosure Risk. *Netherlands Official Statistics*. Spring, 1999.

[8] Fellegi, I.P. On the Question of Statistical Confidentiality, *Journal of the American Statistical Association*, Vol. 67, No. 337, pp.7-18, 1972.

[9] Mokken, R. J., Kooiman, W. J., Pannekoek, J., Willenborg. Disclosure Risks for Microdata. *Statistica Neerlandica*, Vol. 46. pp.49-67, 1992.

[10] Paass, G. Disclosure risk and Disclosure Avoidance for Microdata, Journal of Business and Economic Statistics 6(4) pp.487-500, 1988.

[11] Purdam, K. and Elliot, M. J. (2002) An evaluation of the availability of public data sources which could be used for identification purposes; A Europe wide perspective *CASC project Report 5D1*

[12] Salazar, J., P. Lowthian, C. Young, G. Merola, S. Bond, and D. Brown (2004). Getting the best results in controlled rounding with the least effort. In J. Domingo-Ferrer (Ed.), *Privacy in Statistical Databases*, Volume 3050 of *Lecture Notes in Computer Science*, pp.58–72. Springer.

[13] Salazar, J. J. (2005). A unified mathematical programming framework for different statistical disclosure limitation methods. *Operations Research 53*(3), 819–829.

[14] Salazar, J. J. (2006). Controlled rounding and cell perturbation: Statistical disclosure limitation methods for tabular data. *Mathematical Programming 105*(2–3), 251–274.

[15] Salazar, J. J. (2007). Integer rounding versus continuous adjustment for tabular data. Paper presented to Work session on statistical data confidentiality (Manchester, United Kingdom, 17-19 December 2007)

[16] Skinner, C.J. On Identification Disclosure and Prediction Disclosure for Microdata. *Statistica Neerlandica*, Vol. 46, No. 1, pp.21-32, 1992.

[11] Skinner, C. J., Elliot, M. J. A Measure of Disclosure Risk for Microdata, *Journal of the Royal Statistical Society Series* B, 64(4) pp.855-867, 2002.

[17] Smith, D., Elliot, M. An Investigation of the Disclosure Risk Associated with the Proposed Neighbourhood Statistics. *Report for the Office of National Statistics*, 2003.

[18] U.S Department of Commerce, Report on Statistical Disclosure and Disclosure Avoidance Techniques, *Statistical Policy Working Paper* 2, Washington, 1978.