THE RISK OF DISCLOSURE FOR MICRODATA

George T. Duncan, Carnegie-Mellon University
Diane Lambert, AT&T Bell Laboratories

ABSTRACT

Statistical agencies that provide microdata for public use strive to keep the risk of disclosure of confidential information negligible. Assessing the magnitude of the risk of disclosure is not easy, however. Whether a data user or intruder attempts to obtain confidential information from a public-use file depends on the perceived costs of identifying a record, the perceived probability of success, and the information expected to be gained. In this paper, a decision theoretic framework for risk assessment that includes the intruder's objectives and strategy for compromising the database and the information gained by the intruder is developed. Two kinds of microdata disclosure are distinguished: disclosure of a respondent's identity and disclosure of a respondent's attributes as a result of an unauthorized identification. A measure of disclosure proposed by Paass (1985) is considered within the context of the model.

1. THE PROBLEM

Statistical agencies, such as the Bureau of Census and the Internal Revenue Service, have the resources and legal standing to obtain sensitive data from private individuals and firms. Disseminating that data allows reanalysis by groups with different agendas, stimulates new social, economic and scientific research, and provides information to improve forecasts and resource allocation. For many of these purposes, microdata are crucial. For example, accurate evaluation of job training programs requires detailed geographical information at the microdata level to account for local differences in welfare programs and employment opportunities (Boruch and Cecil (1979)). Additionally, without microdata longitudinal trends are notoriously difficult to infer (e.g., Boruch and Stromsdorfer (1985)).

Yet microdata cannot be released without restriction, even if obvious identifiers, such as names, are removed. Detailed geographical information, for example, may allow a firm to link a microdata record to a competitor whose privacy may then be invaded. Such disclosures may even compromise the integrity of the data disseminating agency. Pearson (1986) writes,

> It was reported [at the Social Science Research Council's Conference on Access to Public Data, November 21-22, 1985] that the identification of only one publicly released record in a file could discredit the entire data collection process, endanger the release of the data to other researchers, and potentially evoke criminal or civil sanctions of the agency that released the data.

To protect anonymity of respondents without destroying statistical information about the group, agencies often mask data before release. Withholding some variables, releasing only a sample of records, or swapping responses on some variables between records are a few examples of masking techniques. Masking cannot guarantee that it is *impossible* for a data user to identify a respondent in a microdata file, however. Paass (1985) shows empirically that even if the released microdata are subject to error and only a fraction of records is released, there is a slight, but nonzero, chance that a respondent can be identified in the released data. Consequently, as Pearson (1986) notes, federal statistical agencies that release masked microdata try to maintain an "acceptable disclosure risk level" rather than a zero risk. Realistic assessment of disclosure risk is not easy, however.

Spruill (1982, 1983, 1984) was perhaps the first to propose and apply a measure of the risk of disclosure for microdata. She suggested the following. For each "test" record in the masked file, compute the squared distance between the test record and each record in the source file. Then determine the percentage of test records that are closer to their parent source record than to any other source record. The percentage of test records that match to the correct parent record multiplied by the sampling fraction (fraction of source records released) is defined to be the risk of disclosure. A variant of Spruill's method based on nearest neighbors has been applied to income data collected by the Internal Revenue Service (Strudler, Oh and Scheuren (1986)).

2. WHAT CONSTITUTES A DISCLOSURE ?

The general issue of measuring the risk of disclosure in masked microdata and evaluating the effectiveness of masking techniques is further explored in this paper. But before disclosure can be measured, it must be conceptualized adequately. Spruill (1983), Paass (1985), and Strudler, Oh, and Scheuren (1986) equate disclosure in microdata with identification of a respondent from a released file. For them, the mere association of a respondent with a record, here called *identity disclosure*, is important and to be guarded against.

Sometimes, however, identification is important mainly because it reveals sensitive information that would not be available otherwise. Cox and Sande (1979) equate disclosure with obtaining reliable information about a respondent as a direct result of linking a record to the respondent, here called *attribute disclosure*. They write,

> If sufficiently accurate data are present for correct identification of a respondent and a good approximation of confidential data, and if it is possible to correctly associate that data with the respondent, then statistical disclosure has occurred.

Others, including the Subcommittee on Disclosure Avoidance Techniques (1978), recommend a third concept of disclosure proposed by Dalenius (1974):

> If the release of the statistic S makes it possible to determine the (microdata) value more accurately than is possible without access to S, a disclosure has taken place...

With Dalenius's concept, here called *inferential disclosure*, a disclosure occurs if the data user infers new information about a respondent from the released data, even if no released record is associated with the respondent and the new information is inexact.

Inferential disclosure is most commonly considered when tabular data are released (cf. Duncan and Lambert (1986)), but it is also appropriate for some microdata releases. For example, consider microdata masked by swapping the value of a variable for one respondent with the value of the same variable for another respondent, and then repeating the swapping for several variables and different respondents (Dalenius and Reiss, 1982). After such data swapping, the name on the transformed source record is less interesting and perhaps meaningless since the information on the record no longer corresponds to the name. Nevertheless, disclosure without identification is a problem if a firm learns a sensitive value of a competitor from the released records.

Palley and Simonoff (1986) consider a fourth type of disclosure from microdata: disclosure of confidential information about a

population or model. Population disclosure is an issue if the relationship between salary and employee characteristics, rather than just the salary of a particular employee, is confidential. Model disclosure could be an issue with the tax compliance model of the Internal Revenue Service, which is withheld from the public. Palley and Simonoff measure population and model disclosure by determining how similar a model fit to released data is to a model fit to the source data.

Identifying a respondent's record in a released file, inferring too narrowly a sensitive characteristic of a respondent from released microdata, and uncovering a proprietary model are all reasonable types of disclosure. But each type leads to concern about different records. Strudler, Oh and Scheuren (1986), Kim (1986), and Cox, Fagan, Greenberg and Hemmig (1986) consider outliers to be more disclosure prone than non-outliers. The reason is that if a respondent is an outlier on variables available to the public, then the respondent is easier to identify in released files. Moreover, once the outlying respondent is identified, sensitive characteristics included on the same record become inadvertently available. In contrast, Palley and Simonoff (1986) consider data without outliers to be more disclosure prone. Without misleading outliers, the regression of sensitive characteristics on non-sensitive characteristics can be inferred, threatening confidentiality of the model and population. Palley and Simonoff also argue that a "typical" individual's sensitive characteristics can be predicted from a good model, which is easier to obtain from data without outliers, leading to inferential disclosure.

Plainly, different interpretations of disclosure from microdata are possible and confusion is likely as long as intuition is not formalized. A framework for analyzing disclosure was established and applied to tabular data in Duncan and Lambert (1986, with discussion by Cox, Frank, Gastwirth and Roberts). Within the framework, some current ad hoc procedures were justified and others were shown to be undesirable in some circumstances. Here the framework is modified to encompass disclosures from microdata.

The disclosure limitation framework is built on uncertainty measures (equivalently, information measures) and predictive distributions. It is re-developed in Section 3 to make its relevance for microdata more apparent. Uncertainty measures for the problem of linking a record to a respondent are developed in Section 4. The method of disclosure assessment proposed by Paass is discussed in the context of our model in Section 4. In Section 5, disclosure as conceptualized by Cox and Sande is considered. This paper focuses on disclosures that involve associating a record with an identifiable respondent. Inferential disclosures that do not involve identification can also be accomodated, however. Just a change in the loss function is required.

3. THE DISCLOSURE LIMITATION FRAMEWORK

The source file consists of the records of N individuals, here called respondents, which might be firms. It can be represented by an $N \times K$ matrix $\mathbf{X}$ in which each row gives data on K attributes for one individual. Typically, there are many attributes in the source file, including some that are sensitive (such as assets or medical condition) and some that are directly related to sensitive variables (such as taxes paid). For convenience, assume that an initial column $\mathbf{X}_{\cdot 0}$ contains the identities of the respondents; $\mathbf{X}_{\cdot 0}$ might contain social security numbers, for example. Data on the ith individual is represented by the row vector $\mathbf{X}_i$. Of course, the source file might concern only a sample of respondents from a larger population. If so, represent the K attributes in the population by a matrix $\mathbf{Z}$ with K columns and $N_* > N$ rows.

Release of a microdata file $\mathbf{Y}$ constructed from $\mathbf{X}$ is planned. In the simplest case, respondents (rows of $\mathbf{X}$) are sampled and some attributes (columns of $\mathbf{X}$) are eliminated, so $\mathbf{Y}$ consists of a subset of the rows and columns of $\mathbf{X}$. Often the data are further modified. For instance, $\mathbf{X}$ may be subject to data swapping, so that elements within a column of $\mathbf{X}$ are permuted. Or m simulated records $\mathbf{X}_{N+1}, \ldots, \mathbf{X}_{N+m}$ might be added to the file for release. In any case, the released file $\mathbf{Y}$ is a transformation of the source file $\mathbf{X}$.

For its own records, the agency keeps the identifiers for the records of $\mathbf{Y}$ in an initial column $\mathbf{Y}_{\cdot 0}$ that is hidden from the data user. For example, suppose the source file is subjected to data swapping followed by sampling of respondents. The column of identifiers after the data swapping is still $\mathbf{X}_{\cdot 0}$, and the column of identifiers after data swapping and sampling contains the identifiers in $\mathbf{X}_{\cdot 0}$ that are included in the sample. If m simulated records are included with the released data, take their identifiers to be $N + 1, \ldots, N + m$.

We consider release of $\mathbf{Y}$ as a complete file. In many database applications, however, $\mathbf{Y}$ is released sequentially in response to queries. Since strategies for compromising a database accessed iteratively are different from those for compromising a fixed microdata file, we do not consider sequential release of $\mathbf{Y}$ in this paper.

The intruder is interested in one or more target variables for one or more respondents. When the intruder intends to locate a known individual, say the one with identifier x_{10}, in the released file, the target is the location of x_{10} in $\mathbf{Y}_{\cdot 0}$. When the intruder intends to learn the jth attribute of this respondent, the target is x_{1j}. The jth attribute may or may not be included in the released file. The intruder may learn x_{1j} by identifying the target's record in the released file and taking x_{1j} from the record, or by identifying a record similar to that of the target and taking its jth attribute to be the target attribute, or by applying statistical inference to all the released records, skipping the identification step. Statistical inference is necessary when the attribute of interest is withheld and only related attributes are made available. In any case, the intruder must reason from the released $\mathbf{Y}$ and other available information to learn the target.

The extent of disclosure depends on how much the user knows about the target after data release. To quantify the extent of disclosure, the information or beliefs that an intruder has about the target before and after data release must be modeled. We choose to express these beliefs by a predictive distribution (probability function). But a predictive distribution on the target alone is insufficient. Since the released $\mathbf{Y}$ does not include the column of identifiers, the intruder must consider all respondents in the source file to make use of $\mathbf{Y}$. For example, the intruder who knows whether the target is likely to be typical or in the upper quartile is more likely to be able to identify the target record than an intruder who cannot distinguish the target from other respondents. Formally, the intruder must specify a joint predictive distribution on the attributes on all the records that might be released.

Specifying joint predictive distributions can be difficult, but not necessarily impossible. Suppose, for example, that an intruder intends to identify which released record belongs to a particular respondent. If $\mathbf{Y}$ is a 5% random sample from $\mathbf{X}$, a naive intruder may believe that the probability that any respondent is included in $\mathbf{Y}$ is .05 and that each record in $\mathbf{Y}$ is equally likely to belong to any respondent. This joint predictive distribution is sufficiently detailed for some purposes (see Section 4.1). More detailed prior distributions can sometimes be derived from worst case analyses or historical data (Sections 4.2, 4.3). Prior distributions for disclosure can also be developed by analogy with prior distributions for legitimate matching of microdata records by agencies. Prior distributions for matching are discussed by Newcombe and Abbatt (1983), Smith, Newcombe, and Dewar (1983) and Kirkendall (1985).

The agency succeeds in masking the microdata file if the intruder remains sufficiently uncertain about the target after data release. Since the predictive distribution expresses the intruder's beliefs about the target, measures of uncertainty are just properties of the

intruder's predictive distribution on the target. Appropriate properties can be generated by considering the intruder's objective "learn the target" in a decision theoretic framework.

Suppose the intruder's target is t_0, which is an identifier x_{10} for identity disclosure, a characteristic x_{1j} for attribute or inferential disclosure, or a property $h(\mathbf{X})$ of the source file for population disclosure. After seeing $\mathbf{Y} = \mathbf{y}$,the intruder's current beliefs about the possible values s of the target are described by a predictive density $p(s)$. Suppose that the intruder incurs a loss $L(t, s)$ when the target is said to be t but s is correct. Since the intruder does not know with certainty which value s is correct, the decision t cannot be chosen to minimize the incurred loss L. On the other hand, after seeing $\mathbf{Y}$ the intruder has beliefs about what the correct value s of the target is, and by averaging over the possible losses in accordance with these beliefs (i.e., by weighting losses with respect to the current predictive density $p(s)$ on the target t_0), the intruder finds that deciding the target is t leads to an expected loss of $\int L(t, s)p(s)ds$. The best t for the intruder to equate with the target t_0 minimizes this expected loss. The intruder's uncertainty $U(\mathbf{y})$ about the target after seeing $\mathbf{y}$ is the minimal expected loss:

$$U(\mathbf{y}) = inf_t \int L(t, s)p(s)ds.$$

In other words, the data are protected against the intruder to the extent that the intruder's smallest expected loss (uncertainty) after seeing the data is large.

Two common examples of uncertainty functions are variance and entropy. Variance corresponds to squared error loss for a numerical target. Entropy corresponds to loss proportional to $-log(p_i)$ for a categorical target where p_i is the intruder's probability that the target lies in category i. Other uncertainty functions appropriate for microdata disclosure are studied in Sections 4 and 5. The class of uncertainty measures is large since any concave function of the predictive distribution is an uncertainty function (De Groot (1962)).

4. LINKING A RESPONDENT TO A RELEASED RECORD

In identity disclosure, the intruder intends to learn which record belongs to a particular respondent, which amounts to locating its identifier x_{10}, say, in $Y_{\cdot 0}$, which we take to contain n records. There are two sorts of decisions: either decide to associate the ith released record with the target, i.e. decide $y_{i0} = x_{10}$ for some i in {1, ..., n}, or decide there is not enough information to link any released y_{i0} to the target. In the latter case, write the decision as ϕ and call it the null link. If $\mathbf{Y}$ involves a sample of respondents from $\mathbf{X}$ and the sample does not include the target, then ϕ is the correct decision.

In this section we assume that the intruder's only objective is to locate x_{10}, not to learn a sensitive characteristic, so a possible loss function is

$$L(link, \textit{ true is } x_{10}) = \begin{cases} 0, & \text{if } link = y_{i0} \text{ and } y_{i0} = x_{10} \\ l_1, & \text{if } link = \phi,\ x_{10} \in Y_0 \\ l_2, & \text{if } link = y_{i0} \text{ for } some\ 1 \le i \le n \text{ and } y_{i0} \ne x_{10} \end{cases}$$

If a link $y_{i0} \ne \phi$ is made, the intruder expects to incur a loss of $l_2(1 - p(y_{i0}))$, where $1 - p(y_{i0})$ is the intruder's probability that the ith released record is not the target record. If the link is null, the intruder's expected loss is $l_1 \sum_{i=1}^{n} p(y_{i0})$, since $\sum_{i=1}^{n} p(y_{i0})$ is the intruder's probability that the target record has been released. If $l_1 \sum_{i=1}^{n} p(y_{i0}) < l_2(1 - \max p(y_{i0}))$, the intruder expects to lose less by not linking than by linking and so decides not to link. Since the intruder chooses the decision with the smallest expected loss, the uncertainty about the target is

$$U(\mathbf{y}) = \min\{ l_1 \sum_{i=1}^{n} p(y_{i0}),\ l_2(1 - \max_{1 \le i \le n} p(y_{i0})) \}.$$

If incorrect links are difficult for the agency to deny and damaging, just as correct links are, then the agency's goal is not to prevent incorrect links but to convince the intruder that linking is unwise. From the intruder's perspective, linking is unwise if $l_2(1 - \max p(y_{i0})) > l_1 \sum_{i=1}^{n} p(y_{i0})$. The agency cannot manipulate l_1 or l_2 when there are no fines for compromising a database accessed legally. In that case, all an agency can do to dissuade an intruder from linking is to keep $\max p(y_{i0})$ and $\sum_{i=1}^{n} p(y_{i0})$ small. By controlling $\mathbf{Y}$, the agency influences these probabilities.

By considering a range of possible intruders, the agency can determine the kinds of intruders (kinds of predictive distributions) against which the data are secure. To illustrate, we next develop predictive distributions $p(y)$ for three types of intruders: a naive outsider, an informed insider, and a more realistic, intermediate intruder with imperfect but helpful knowledge of the target. The best type of intruder from the releasing agency's perspective is the naive outsider. The worst disclosure scenario is release of unmasked sample data to an informed insider. Both extremes are perhaps unrealistic, but they help to clarify the issues. The intermediate case of an intruder with imperfect information and masked data has also been considered by Paass (1985) and Paass and Wauschkuhn (1985) in a different framework. They also consider costs and distributions, but their distributions arise from noise in the database rather than incompleteness in the intruder's knowledge about the respondents. Their framework excludes the possibility of a null link, and the focus is on the probability of correct links rather than on the probability of an attempted link, correct or not. Nonetheless, there are connections between the two approaches which are discussed in Section 4.3.

4.1 Release to a Naive Outsider

The naive outsider has no information to distinguish respondents in $\mathbf{X}$, so whatever is believed about one respondent applies equally well to any other respondent. Since only unlabeled records are released and the intruder has identical beliefs about all respondents, the outsider cannot apply the data in $\mathbf{Y}$ to distinguish the target respondent from other respondents. If a sample of n of the N records held by the agency is released, the outsider's probability that the target is the ith released record is

$$p(y_{i0}) = P[\textit{ target released }]\, P[\, y_{i0} = x_{10} \mid \textit{target released}]$$
$$= (n/N)(1/n) = 1/N.$$

Here, the intruder believes the probability the target record has been released is $\sum_{i=1}^{n} p(y_{i0}) = n/N$ and the probability any released record is a correct link is $p(y_{i0}) = 1/N$. So, to dissuade linking the agency must keep the sampling fraction n/N small and the source file size N large. The probability $p(y_{i0})$ is the same regardless of the type of sample, since the type of sample provides no information that helps the outsider to determine whether the record of the target respondent has been released. For example, knowing that firms have been sampled proportionally to size is irrelevant if the sizes of the firms are unknown to the outsider.

If $l_1 < l_2$, i.e., if the cost of not linking is less than the cost of an incorrect link, then the uninformed outsider will not link as long as the agency withholds at least one record. Hence, uninformed intruders with $l_1 < l_2$ are of little concern to the agency. Linking may be even more unfavorable if the source file is a sample from a population $\mathbf{Z}$ and the outsider does not know whether the target record is in $\mathbf{X}$. In that case, $p(y_{i0}) = 1/N_*$ where N_* is the

number of records in Z and the outsider will not link as long as $l_1 n \le l_2(N_* - 1)$. Announcing that Y contains m simulated, artificial records affords the same protection. If the m simulated records are indistinguishable (to the outsider) from the records in X, then $p(y_{i0}) = N^{-1}n/(m+n)$. Releasing Y with a fraction f of simulated records is equivalent to releasing Y for an X based on a fraction f of a population Z; in both cases $p(y_{i0}) = f/N$. That is, sampling and simulating equally confound the uninformed outsider. But transformations, such as random noise inflation, that do not affect the number of records released, do not affect the outsider's uncertainty. Being naive means there is no context for interpreting the information in Y.

In contrast, consider a malicious outsider, intent on discrediting an agency, who intends to announce that a link has been achieved without divulging the link itself. For this intruder, the cost of not linking far exceeds the cost of an incorrect link, i.e., $l_1 > l_2(N-1)/n$, and a link is claimed no matter how likely it is to be wrong.

4.2 Release to an Insider: Sampled and Simulated Data

In the worst case, an insider knows all K attributes of all N respondents in the source file and the only objective is to identify which record is the target respondent's. That is, the worst case is an insider trying to sabotage the agency by showing that a link is possible even if no additional information is gained. Suppose that the target's record is $(x_{11}, \ldots, x_{1K})$ and the first $k \le K$ attributes are released unchanged for all N records. Then

$$p(y_{i0}) = \begin{cases} 1/N(x), & \text{if } y_{ij} = x_{1j},\ j = 1,\ldots,k \\ 0, & \text{if } \textit{otherwise} \end{cases}$$

where $N(x)$ is the number of records in X equal to $(x_{11},\ldots,x_{1k})$. (So to determine $p(y_{i0})$, the insider needs to know just the number of respondents in X whose first k attributes are the same as the target's.) Uncommon records are at risk of being identified, even if they are not outliers in the usual sense.

An agency may adopt various strategies to limit disclosure. For example, only a simple random sample (without replacement) of n records may be released. Suppose the sample contains $n(x) \ge 1$ records equal to $(x_{11},\ldots,x_{1k})$. After seeing that Y has n records and $n(x)$ of them agree with the target, the insider uses the hypergeometric distribution to calculate that

$P[\textit{target}$ and $n(x) - 1$ *identical records released*$]$

$$= \begin{cases} \dfrac{\binom{N(x)-1}{n(x)-1}\binom{N-N(x)}{n-n(x)}}{\binom{N}{n}}, & \text{if } y_{ij} = x_{1j},\ 1 \le j \le k \\ 0, & \text{if } \textit{otherwise} \end{cases}$$

A simple conditioning argument leads to

$$p(y_{i0}) = P(\textit{ith record in Y is target} \mid n(x))$$

$$= \begin{cases} 1/N(x), & \text{if } y_{ij} = x_{1j},\ j = 1,\ldots,k \\ 0, & \text{if } \textit{otherwise} \end{cases}$$

The probability a null link ϕ is incorrect is $n(x)/N(x)$, which is the fraction of records like the target that are in the sample. The probability a link is incorrect is $(N(x)-1)/N(x)$. The sampling fraction n/N has no effect on the insider's uncertainty. Hence, sampling per se need not confound the insider with perfect knowledge.

Combining sampled and simulated data can help protect anonymity of respondents. Suppose that after sampling but before release m simulated records are included in Y and $m(x)$ of these agree with the target. Let n be the total number of records in Y and $n(x)$ be the total number of records in Y identical to the target. If the insider knows $m(x)$,

$$p(y_{i0}) = \begin{cases} N(x)^{-1}[n(x) - m(x)]/n(x), & \text{if } y_{ij} = x_{1j},\ 1 \le j \le k \\ 0, & \text{if } \textit{otherwise} \end{cases}$$

Usually, $m(x)$ is unknown, but if the intruder believes that $m(x)$ simulated records have been added with probability $p(m(x))$, then

$$p(y_{i0}) = \sum_{m(x)} p(m(x)) p(y_{i0} \mid m(x))$$

$$= N(x)^{-1} E(n(x) - m(x))/n(x).$$

For example, if there are 10 records like the target in the source file, the released file contains 6 records like the target, and the insider expects that 4 artificial records like the target are included in the 6, then $p(y_{i0}) = .1(6-4)/6 = .033$. Only the mean fraction of records in the sample that agree with the target and come from the source file, which is $E[n(x) - m(x)]/n(x)$, must be assessed, not an entire probability distribution for $m(x)$. The larger the fraction $E(m(x)/n(x))$of released records expected to be artificial, the larger $l_2(1 - \max p(y_{i0}))$, which is the component of uncertainty from linking. The component of the uncertainty from not linking is $l_1 N(x)^{-1}[n(x) - Em(x)]$. So the more simulated records like the target that the intruder expects, the less incentive the informed insider has to link.

4.3 Release of Masked Data to an Intruder with Some Knowledge

So far, we have considered the data in Y to be accurate and the insider to know the attributes in X accurately. Often, this is unrealistic. Even an insider is apt to know the continuous attributes of the target within some percentage rather than exactly. Or the intruder may believe that the target respondent was not entirely truthful. Or the agency may announce that the data have been transformed to protect the identities of the respondents. The masking techniques may not be revealed or may be too convoluted for even a motivated user to attempt to reverse. On the other hand, the intruder may believe that the released data are not too corrupted since the agency must preserve certain statistical features of the population in the released data. In any case, the intruder believes that the data released are only approximate.

Consequently, the intruder may be unable to specify the target attributes $\mathbf{X}_1 = (X_{11},\ldots,X_{1K})$ precisely. Instead, the intruder may specify that if the agency releases a certain continuous attribute of the target, it is certain to be within 100 percent of the correct value, likely to be within 50 percent of the correct value, and probably within 20 percent. Here we translate such statements about relative error into a predictive distribution on how the target attributes will appear if they are released. Specifically, we assume that the intruder believes that if the target attributes are released they will appear in some random position i in Y as $\mathbf{Y}_i = S(\mathbf{X}_1) = (S(X_{11}),\ldots,S(X_{1K}))$ and $S(\mathbf{X}_1)$ will have a joint lognormal distribution with parameters $\mu_1 = (\mu_{11}, \ldots, \mu_{1K})$ and Σ. Here μ_{1j} is the intruder's best guess at the log of the jth attribute of the released version of the target and Σ describes the intruder's impression of the imprecision of the data. This is not the intruder's predictive distribution on which released record belongs to the target, but the intruder's predictive distribution on how the target will appear if it is released. Lognormal distributions are chosen for mathematical convenience. Other distributions might be more appropriate, especially for discrete attributes.

Realistically, not just the target is released with error. All released records are corrupted versions of the truth. Here, we assume that each continuous log attribute of each respondent is normal. Each attribute of each respondent may have a different mean (or most probable value), but in our examples all records share the same imprecision matrix, Σ . That is, the same model of

corruption describes all records. The relationship of μ_1 to μ_i, $i = 2,\ldots,N$ describes the intruder's beliefs about the size of the target relative to the other respondents in **X**. We could take the corruption of records of different respondents to be a priori correlated, but following Paass (1985) we do not. Note that both an insider dealing with masked data and a less knowledgeable intruder with imperfect information about **X** will work with distributions on the attributes of the respondents.

Here and in Paass (1985), prior distributions must be specified for all N respondents. As a working hypothesis, the agency evaluating the extent of disclosure may assume the means or modal values of the N distributions are based on the data in **X** or on information known to be available to the public. Or the N means can be obtained by approximating the empirical distribution of the source file **X** by a parametric distribution G_θ and then taking the N quantiles of G_θ, namely $G_\theta^{-1}(1/(N+1))$. In any case, once the N means or modal values are obtained, the intruder may multiply them by a lognormal$(0, \Sigma)$ random vector representing the uncertainty around the modal value and the imprecision introduced by masking to obtain N prior densities $f_1, \ldots, f_N$ for $S(X_1),\ldots, S(X_N)$.

The result of perturbing **X** to $S(\mathbf{X})$ is that an intruder must often consider many records as candidates for matching to the target. We show this below for general N. But since the notation for general N obscures the issues, suppose first that $K = 1$, $N = 2$, and $n = 2$; that is, there is one attribute, two respondents, and both records are released. Suppose the intruder believes that the target's released attribute $S(X_{11})$ has a lognormal(0, 1) distribution and the other respondent's released attribute $S(X_{21})$ has a lognormal(2, 1) distribution. Hence, the intruder expects the target to be released as 1.65 and the other respondent to be released as 12.18. The agency releases $\mathbf{Y} = (Y_1, Y_2) = (7, 20)$. The intruder must decide whether the record with 7 belongs to the target (i.e., $S(X_{11}) = 7$) or the record with 20 belongs to the target (i.e., $S(X_{21}) = 7$) or there is not enough evidence to link the target to either record. If a link is made, the reasonable choice is that the record with 7 belongs to the target, since the target value is expected to be smaller than the other respondent value. Precisely, the probability that $Y_1 = 7$ belongs to the target is

$$p(y_{10}) = P[S(X_{11}) = 7 \mid \mathbf{Y}] = \frac{f_1(y_1)f_2(y_2)}{f_1(y_1)f_2(y_2) + f_2(y_1)f_1(y_2)}$$

where $f_i(\cdot)$ is the lognormal$(\mu_i, 1)$ density. Here, $p(y_{10}) = .89$, $p(y_{20}) = .11$, and the probability that the target has been released is $.89 + .11 = 1$. Clearly, information about the position of the target relative to the other respondents helps a knowledgeable intruder.

To continue this example, suppose only one record is released and the agency chooses the record to be released randomly from the two available. Suppose the released record contains $y_1 = 7$. Then the intruder's probability that the record with 7 is the target is obtained by thinking about the data release in two stages. First, the record to be released after masking is chosen at random from the two records in **X** and then the y assigned to that record for release is chosen according to the appropriate density. To be specific,

$$p(y_{10}) = P(\textit{target appears as } y_{10} \mid \mathbf{Y}) =$$

$$\frac{P(X_{11}\ \textit{sampled})P(S(X_{11}) = 7)}{P(X_{11}\ \textit{sampled})P(S(X_{11}) = 7) + P(X_{21}\ \textit{sampled})P(S(X_{21}) = 7)}$$

$$= \frac{.5f_1(7)}{.5f_1(7) + .5f_2(7)} = .13.$$

The probability the target has not been released is .87. The lack of y_2 greatly alters the intruder's probability that 7 is the target. Note that $p(y_{10})$ remains .13 even if the intruder believes $S(X_{11})$ and $S(X_{21})$ are dependent.

The formulas for general N, k and n are cumbersome, but they are straightforward to interpret. Suppose a random sample (without replacement) of n records out of N is released. Also, suppose that the intruder believes that if the record of the ith respondent is released, it will appear as $S(\mathbf{X}_i)$ having density $f_i(\cdot)$. Again, the data released are generated in two stages: first n records are chosen at random from the N records in **X** and then values for the released versions of these records are chosen from the appropriate densities. Hence, with $\mathbf{X}_1$ being the target, the intruder believes that the probability the first record in **Y** belongs to the target is

$$p(y_{10}) = P(S(\mathbf{X}_1) = \mathbf{y}_1 \mid \mathbf{Y})$$

$$= \frac{N^{-1} f_1(\mathbf{y}_1)P(\mathbf{y}_2, \ldots, \mathbf{y}_n \textit{ sampled from } f_2,\ldots,f_N)}{P(\mathbf{y}_1, \ldots, \mathbf{y}_n \textit{ sampled from } f_1,\ldots,f_N)}.$$

Here, $P(\mathbf{y}_2, \ldots, \mathbf{y}_n \textit{ sampled from } f_2,\ldots,f_N)$ denotes the probability of observing $\mathbf{y}_2,\ldots,\mathbf{y}_n$ when sampling one observation each from $n-1$ of the $N-1$ densities $f_2,\ldots, f_N$. Equivalently,

$$p(y_{10}) = \tag{4.1}$$

$$\frac{f_1(\mathbf{y}_1)P(\mathbf{y}_2, \ldots, \mathbf{y}_n \textit{ sampled from } f_2,\ldots,f_N)}{\sum_{i=1}^{N} f_i(\mathbf{y}_1)P(\mathbf{y}_2, \ldots, \mathbf{y}_n \textit{ sampled from } f_1, \ldots, f_{i-1}, f_{i+1},\ldots,f_N)}.$$

Once again, how likely $\mathbf{y}_1$ is to belong to the target depends on how unlikely the other $n-1$ released records are to belong to the target. The unsampled respondents affect the probability that y_{10} is the target through the insider's beliefs expressed by the N densities $f_1,\ldots, f_N$.

Equation (4.1) is consistent with the results in Sections 4.1 and 4.2. For the naive outsider, the N respondents are indistinguishable so $f_1(\mathbf{y}) = \ldots = f_N(\mathbf{y})$ for all n released records $\mathbf{y}$. Therefore, $p(\mathbf{y}_i) = 1/N$ for $i = 1,\ldots,n$. For the insider with complete, unmasked data, $f_i(\mathbf{y}) = 1$ if $(y_{i1},\ldots,y_{ik}) = (x_{11},\ldots,x_{1k})$ and 0 otherwise. Therefore, $p(y_{i0}) = N(x_{11},\ldots,x_{1k})^{-1}$ if $(y_{i1},\ldots,y_{ik}) = (x_{11},\ldots,x_{1k})$ and 0 otherwise.

Formulas simpler than equation (4.1) can be obtained by considering each released record separately without regard to whether other records are better or worse candidates for the target. These simpler formulas are important because they lend themselves to extensive empirical investigations of disclosure using statistical techniques such as discriminant analysis (e.g., Paass (1985) and Paass and Wauschkuhn (1985)). Therefore, it is worth comparing probabilities based on the full analysis (4.1) that consider all released records $\mathbf{y}_1, \ldots, \mathbf{y}_n$ simultaneously with formulas that consider each released record $\mathbf{y}_i$ separately.

As one example of a formula that treats released records in isolation, consider the following formula from Paass (1985) for the probability that the jth record is the target :

$$p_{iso}(y_{j0}) = \frac{f_1(\mathbf{y}_j)}{f_1(\mathbf{y}_j) + f_2(\mathbf{y}_j) + \cdots + f_N(\mathbf{y}_j)}, \tag{4.2}$$

where the subscript *iso* denotes that each released record j is considered in isolation from the other $n-1$ records. To illustrate the difference between $p_{iso}(y_{j0})$ and $p(y_{j0})$ from the full analysis (4.1), consider the simple example above with $N = n = 2$ and $k = 1$. There, if both records are released and the intruder considers both records then $p(y_{10}) = .89$ and $p(y_{20}) = .11$, but if both records are released and the intruder considers each record separately, as in equation (4.2), then $p_{iso}(y_{10}) = .13$ and $p_{iso}(y_{20}) = .02$. Here, the simpler formula breaks down. Even though the intruder knows that the target has to be in the released data, the probability that one of the released records is the target, as determined from formula (4.2), is .15 which is substantially less than one. Moreover, note that unlike formula (4.1), formula (4.2) is unaffected by the sampling fraction. In particular, in the above example with $N = 2$, $p_{iso}(y_{10})$ is the same regardless of whether only $\mathbf{y}_1$ is released or both $\mathbf{y}_1$ and $\mathbf{y}_2$ are released.

The discrepancy between the full analysis (4.1) and the isolated record analysis (4.2) need not diminish as the source file size N increases and the sampling fraction n/N decreases, and the isolated record analysis may lead to smaller or larger probabilities than the full analysis. For example, consider the following simple scenario. There are $N = 100$ records in the source file and a 10% sample of modified records is released, each containing one attribute ($k = 1$). The intruder believes the target X_1 is smaller than the 99 other source records. The true values $X_2, \ldots, X_{100}$ of the other 99 respondents are believed to follow a lognormal(2, .75) distribution. The intruder knows that before sampling each true value is further modified. For convenience, the intruder assumes the masking effectively multiplies X_i by a random lognormal(0, .25) inflation factor. Hence, the intruder assigns lognormal(2, 1) prior distributions to $S(X_2), \ldots, S(X_{100})$. The intruder has less precise information about how small X_1 is likely to be, but believes the most likely value of X_1 is 1. For simplicity, the intruder describes his uncertainty around the most likely value 1 by a lognormal(0, .75) distribution. If X_1 is believed to be modified before release by lognormal(0, .25) noise, just as $X_2, \ldots, X_{100}$ are, then the prior distribution of $S(X_1)$ is lognormal(0, 1). Finally, the data released are given in Table 1, along with their probabilities of being the target under the full analysis and under the isolated record analysis.

Table 1

$N = 100$, $n = 10$
f_1 = *the lognormal* (0, 1) *density*
$f_2 = \cdots = f_{100}$ = *the lognormal* (2, 1) *density*

observed	.050	.135	1.505	2.401	3.151
$p_{iso}(y)$	.968	.803	.032	.013	.007
$p(y)$	.858	.118	.000	.000	.000
observed	3.825	4.591	8.723	10.343	10.732
$p_{iso}(y)$	.005	.004	.001	.001	.001
$p(y)$	.000	.000	.000	.000	.000

As Table 1 shows, considering each record in isolation leads to larger probabilities, especially for the released records y_1, y_2. (Recall that in the earlier example with $N = 2$ and $n = 1$, considering each record in isolation led to a smaller probability.) Under the full analysis, the probability the target record is in the sample is .976. When each record is considered in isolation, $\sum_y p_{iso}(y)$ is not bounded by one and is not the probability that one of the released records is the target. As Table 1 shows, the discrepancy between the full and isolated record analysis can be large.

The isolated case formula (4.2) is approximately equal to the full analysis (4.1) if the multipliers of $f_1(\mathbf{y}_1), f_2(\mathbf{y}_1), \ldots, f_N(\mathbf{y}_1)$ in (4.1) are all nearly equal. (The multipliers are identical if $f_1 = \ldots = f_N$, but a knowledgeable intruder is unlikely to have the same beliefs about all respondents.) The multipliers of the $f_i(\mathbf{y}_1)$'s differ only in the density that is excluded as an option for $\mathbf{y}_2, \ldots, \mathbf{y}_n$. To compare these multipliers, suppose that N is large and n is small relative to N. Then the dependence between $\mathbf{y}_i$ and $\mathbf{y}_j$ induced by sampling without replacement is negligible and

$$P[\mathbf{y}_2, \ldots, \mathbf{y}_n \textit{ from } f_2 \textit{ or } \cdots \textit{ or } f_N] \approx \prod_{j=2}^{n} P[\mathbf{y}_j \textit{ from } f_2, \ldots, f_N].$$

Also,

$$p[\mathbf{y}_j \textit{ from } f_2 \textit{ or } \cdots \textit{ or } f_N]$$

$$= \sum_{i=2}^{N} P[\mathbf{X}_i \textit{ sampled}] P[\mathbf{X}_i \textit{ masked to } \mathbf{y}_j \mid \mathbf{X}_i \textit{ sampled}]$$

$$= (N-1)^{-1} \sum_{i=2}^{N} f_i(\mathbf{y}_j).$$

Therefore,

$$\frac{P[\mathbf{y}_2, \ldots, \mathbf{y}_n \textit{ from } f_1, \ldots, f_{m-1}, f_{m+1}, \ldots, f_N]}{P[\mathbf{y}_2, \ldots, \mathbf{y}_n \textit{ from } f_2, \ldots, f_N]}$$

$$\approx \prod_{j=2}^{n} \left(\frac{\sum_{i \neq m}^{N} f_i(\mathbf{y}_j)}{\sum_{i=2}^{N} f_i(\mathbf{y}_j)} \right)$$

$$= \prod_{j=2}^{n} \left(1 + \frac{f_1(\mathbf{y}_j) - f_m(\mathbf{y}_j)}{\sum_{i=2}^{N} f_i(\mathbf{y}_j)} \right)$$

Returning to equation (4.1),

$$p(y_{10}) \approx \frac{f_1(\mathbf{y}_1)}{\sum_{i=1}^{N} f_i(\mathbf{y}_1) \prod_{j=2}^{n} \left(1 + \frac{f_1(\mathbf{y}_j) - f_i(\mathbf{y}_j)}{\sum_{m=2}^{N} f_m(\mathbf{y}_j)} \right)}. \quad (4.3)$$

Thus, the isolated approach (4.2) approximates the full analysis if the differences in the product in the denominator of (4.3) are negligible compared to the divisor $\sum_{m \neq 1} f_m(\mathbf{y}_j)$. Note in particular that the quality of the approximation of (4.2) to (4.1) is not determined only by the sampling fraction n/N. How different the beliefs are about different respondents also matters.

The important questions are when is $p(y_{10})$ from equation (4.2) much smaller than $p(y_{10})$ from equation (4.1) so that the isolated case formula misses a target record $\mathbf{y}_1$ that is at risk and when is $p_{iso}(y_{10})$ much larger so that the risk of disclosure is overstated? One answer from the approximation (4.3) is that a record $\mathbf{y}_1$ at risk of disclosure is missed when $\mathbf{y}_2, \ldots, \mathbf{y}_n$ are more unlikely under the target's density f_1 than they are under some other density f_j. This is to be expected, since treating released records in isolation prevents choosing a record because all other candidates for the target are much less likely. Here, the differences in the approximation (4.3) can be negative, making the terms in the product less than one. The worst case occurs when the target is not probable under f_1 but the other records are even more improbable under the target's density f_1. The simple example with N = n = 2 and k = 1 falls into this category. There, $\mathbf{y}_1$ is unlikely under f_1 but substantially more unlikely under f_2 and $\mathbf{y}_2$ is substantially more unlikely than $\mathbf{y}_1$ under f_1. Hence, although $\mathbf{y}_1$ is not likely under f_1, the only reasonable decision is that $\mathbf{y}_1$ is the record from f_1. Note that this state arises if the intruder judges the ranking of respondents correctly but systematically misjudges the magnitude of the average response or if the masking converts "fringe" observations into outliers.

Formulas that treat records in isolation may miss records at risk, but they are much simpler to compute for large N than the full analysis (4.1) or even its approximation (4.3). Moreover, in some examples $p(y_{i0})$ from (4.1) and $p_{iso}(y_{i0})$ from (4.3) are approximately equal. Further research is needed to develop tractable means for computing (4.1) or (4.3)or reliable methods for identifying which records are overlooked or unnecessarily flagged with the isolation formula (4.2).

5. DISCLOSURE OF AN ATTRIBUTE THROUGH IDENTIFICATION

At times, the desire to learn an attribute of the target motivates the intruder to attempt to link a record to a respondent. For example, recall from Section 2 that Cox and Sande's characterization of disclosure requires identity and attribute

disclosures to occur together. In that case, the intruder's loss function has two components: (1) the loss from linking to an incorrect record and (2) the loss from incorrectly inferring the released target attribute from the record linked to the target. In this section, two variants of a two component loss function are considered. In the first, all records not belonging to the target are considered equally useless to the intruder for inferring the target. In the second, the loss incurred by the intruder from an incorrect link depends on how similar the chosen record is to the target record. Note that attribute disclosures without links to particular records are consistent with inferential disclosure but not disclosure in the sense of Cox and Sande. Inferential disclosure is considered in Duncan and Lambert (1986), but not in this paper.

We show in Section 5.1 that if the loss from an attribute disclosure based on an incorrect link is independent of how similar the target and link are, then attribute disclosure reduces to identity disclosure. We show in Section 5.2, however, that if less loss is incurred when the linked record is similar to the target record, then the records at risk of attribute disclosure are different from the records at risk of identification. Consequently, disclosure evaluations based only on the risk of identification are important but not necessarily sufficient to protect confidential microdata.

5.1 Loss Independent of the Similarity of Target and Link

Consider an intruder who intends to identify the target's record to learn at least one attribute of the target. Suppose that the target record $\mathbf{x}_1$ is released as $S(\mathbf{x}_1)$ and the intruder specifies the target attribute to be $t(\mathbf{y}_i)$, where $\mathbf{y}_i$ is the record linked to the target. If the target attribute is "read" directly from the record linked to the target, then $t(\mathbf{y}_i) = \mathbf{y}_i$. If attribute y_{ik} of the linked record is shrunk towards a value based on all $\mathbf{Y}$, then t depends on all $\mathbf{Y}$.

The loss function for the case that the same loss is incurred whenever the link is incorrect reflects two concerns: (1) a disclosure requires a correct identification and (2) a correct identification followed by an incorrect attribute inference is no worse than an incorrect identification. The first concern implies that records that are similar to the target are no more useful to the intruder than records that are dissimilar, so all incorrect identifications lead to the same loss. The second concern says that the loss from an incorrect identification is at least as large as the loss from a correct identification. The loss from a correct identification is not necessarily zero, since an incorrect inference about the target can follow a correct identification, especially if the data are masked. Take the loss from an incorrect inference with a correct identification to be $D(t(S(\mathbf{x}_1)), \mathbf{x}_1)$ where $D(\mathbf{x},\mathbf{y})$ measures the disparity between two records $\mathbf{x}$ and $\mathbf{y}$. Assume $D(\mathbf{x},\mathbf{x}) = 0$ for any $\mathbf{x}$ and $D(\mathbf{x},\mathbf{y}) \geq 0$ for all $\mathbf{x},\mathbf{y}$. To summarize, an appropriate loss function is

$$L(\mathbf{y}_i, \mathbf{x}_1) = \begin{cases} l_1 & \text{if } y_{i0} = \phi \\ D(t(S(\mathbf{x}_1)),\mathbf{x}_1) & \text{if } x_{10} = y_{i0}, \\ l_2 + D(t(S(\mathbf{x}_1)),\mathbf{x}_1) & \text{if } \textit{otherwise} \end{cases}$$

where l_2 is a positive constant.

The expected loss from a link to $\mathbf{y}_i$ is

$$E[L(\mathbf{y}_i,\mathbf{x}_1)]$$

$$= E[L(\mathbf{y}_i,\mathbf{x}_1) \mid x_{10}=y_{i0}]p(y_{i0}) + E[L(\mathbf{y}_i,\mathbf{x}_1) \mid x_{10}\neq y_{i0}](1-p(y_{i0}))$$

$$= E[D(t(S(\mathbf{x}_1)),\mathbf{x}_1)]p(y_{i0}) + E[l_2+D(t(S(\mathbf{x}_1)),\mathbf{x}_1)](1-p(y_{i0}))$$

$$= l_2(1-p(y_{i0})) + \int D(S(\mathbf{x}_1),\mathbf{x}_1)f_1(\mathbf{x}_1)d\mathbf{x}_1$$

$$= l_2(1-p(y_{i0})) + ED(t(S(\mathbf{x}_1)),\mathbf{x}_1).$$

Here, f_1 is the density that describes the released version of the target $\mathbf{x}_1$ (e.g., f_1 describes an insider's perception of the masking) and $p(y_{i0})$ is the intruder's probability that the ith released record belongs to the target (as in Section 4).

The only term that the intruder can control is $1-p(y_{i0})$, which the intruder minimizes by choosing the most probable link, just as in Section 4. So, if incorrect links are not differentiated, then records at risk in the sense of Cox and Sande are at risk in the sense of identification. The only difference is that the uncertainty is possibly higher (and the disclosure risk lower) for attribute disclosure because of the term $ED(t(S(\mathbf{x}_1)),\mathbf{x}_1)$.

5.2 Loss Dependent on the Similarity of Linked and Target Records

Now suppose linking to a wrong record similar to the target is less costly than linking to a dissimilar wrong record. Suppose the estimate of the target attribute x_{1k} is taken to be the kth attribute reported on the record linked to the target, i.e. the intruder's estimate of the target's attribute is $t(y_{ik}) = y_{ik}$. With masked data, an appropriate loss function is

$$L(\mathbf{y}_i, \mathbf{x}_1) = \begin{cases} l_1 & \text{if } y_{i0} = \phi \\ D(S(\mathbf{x}_1),\mathbf{x}_1) & \text{if } x_{10} = y_{i0}, \\ l_2 + D(S(\mathbf{x}_1),\mathbf{x}_1) + D(\mathbf{y}_i,\mathbf{x}_1) & \text{if } \textit{otherwise} \end{cases}$$

where D is a measure of disparity and l_2 is a constant. Since linking to the correct target record is important in the Cox and Sande framework, l_2 must be positive.

For the simplest illustration, take an outsider with no prior knowledge about the respondents. The outsider cannot distinguish between the N respondents in $\mathbf{X}$ and believes that all possible values of the attributes are equally plausible for all N respondents. Suppose the outsider does not know to what extent the released data are accurate, so uses the inference rule $t(y_{ik}) = y_{ik}$. Since the released value is taken at face value, the intruder must believe that no loss is incurred if the target record is identified correctly. A reasonable loss function is

$$L(\mathbf{y}_j, \mathbf{x}_1) = \begin{cases} l_1 & \text{if } y_{j0} = \phi \\ 0 & \text{if } x_{10} = y_{j0}, \\ l_2 + D(y_{jk},x_{1k}) & \text{if } \textit{otherwise} \end{cases}$$

where $l_2 > 0$ is the penalty for an incorrect link when the estimate of the attribute of interest is correct. That is, there is a penalty for incorrect linking even if the record linked to the target has the same kth attribute as the target. Suppose the agency releases all N records from $\mathbf{X}$.

If record $\mathbf{y}_j$ is linked to the target, the expected loss is $l_2(1 - p(y_{j0})) + ED(y_{jk},x_{1k})$ where the expectation is with respect to the outsider's predictive distribution on the attribute x_{1k} after seeing $\mathbf{Y}$. Since the released data are assumed to be accurate and exhaustive, the target must assume an observed value y_{ik}. Since any released record is equally likely to belong to the target, $f(x) = N^{-1}\#(y_{ik} = x)$. Additionally, $p(y_{j0}) = 1/N$ for each released record. Hence, the expected loss from linking to $\mathbf{y}_j$ is $l_2(1-N^{-1}) + N^{-1}\sum_{i=1}^{N} D(y_{jk},x_{ik})$. For example, if $D(u,v) = (u-v)^2$, then the best link is to any record whose kth attribute is closest to $\bar{y}_k = N^{-1}\sum y_{ik}$. If $D(u,v) = |u-v|$, then the best link is to any record whose kth attribute is closest to the median of $\{y_{jk}, j = 1,\ldots,N\}$. The outsider is discouraged from linking when the minimal expected loss from linking is large. Hence, if the kth attribute varies greatly across the N records, the outsider is dissuaded from linking.

What is important in this illustration is not what discourages an uninformed outsider from linking but how attempting to learn an attribute of the target rather than merely to identify the target changes an intruder's strategy. When the only goal is identification, the outsider is equally likely to choose any released record as the target (see Section 4.1.) When the goal is to learn an attribute and records close to the target are considered less

erroneous than records far from the target, the outsider rejects many records as candidates. The records are rejected not because they are unlikely links but because they have higher costs if they do not belong to the target. Hence, microdata with a low risk of identity disclosure do not necessarily have a low risk of attribute disclosure. In practice, it may not be sufficient to consider only identity disclosure if attribute disclosure is also a concern.

6. CONCLUSIONS

There are several types of disclosure for microdata: disclosure of a respondent's identity, disclosure of a respondent's attributes following a record identification, disclosure of a respondent's attributes without a record identification, and disclosure of a model. These types of disclosure differ in the intruder's objectives for trying to compromise the microdata. The framework for disclosure proposed in this paper formalizes these objectives and measures the risk of disclosure in terms of the unauthorized information gained when the microdata are released. The framework is conservative in that the intruder is expected to use the optimal strategy for compromising confidentiality. Identity disclosure and attribute disclosure following identity disclosure have been considered in detail, but there is no conceptual barrier to considering model and inferential disclosure. The framework can be used to evaluate the effectiveness of particular masking techniques, if equation (4.1) for the intruder's probability that the ith released record belongs to the target can be approximated simply. Further research should provide trustworthy, tractable approximations to equation (4.1).

ACKNOWLEDGMENTS

This paper was presented at the Third Annual Research Conference of the Bureau of the Census, March 29 - April 1, 1987, and the Joint Statistical Meetings, August 16 - 20, 1987. We thank Gordon Sande and Nancy Spruill for their comments on the paper at these conferences.

REFERENCES

Boruch, R. F. and Cecil, J. F. [1979] "Report from the United States: emerging data protection and the social sciences' need for access to data", *Data Protection and Social Science Research*, E. Mochmann and P. Muller, eds. Springer-Verlag, N.Y., pp. 104-128.

Boruch, Robert and Stromsdorfer, Ernst [1985] "Exact matching of micro data sets in social research: benefits and problems" *Record Linkage Techniques -- 1985, Proceedings of the Workshop on Exact Matching Methodologies*, Publication 1299 (2 - 86) of the Department of the Treasury, U. S. Internal Revenue Service, Statistics of Income Division, pp. 145-153.

Cox, Lawrence H., Fagan, James T., Greenberg, Brian and Hemmig, Robert [1986] "Research at the Census Bureau into disclosure avoidance techniques for tabular data". *Proceedings of the American Statistical Association, Section on Survey Research Methods* (in this volume).

Cox, Lawrence H. and Sande, G. [1979] "Techniques for preserving statistical confidentiality". *Proceedings of the 42nd Meetings of the International Statistical Institute*, Manila, December 1979.

Dalenius, Tore [1974] "The invasion of privacy problem and statistics production -- an overview". *Statistik Tidskrift* **12**, pp. 213-225.

Dalenius, Tore and Reiss, S. P. [1982] "Data swapping: a technique for disclosure control". *Journal of Statistical Planning and Inference* **6**, pp. 73-85.

DeGroot, Morris H. [1962] "Uncertainty, information, and sequential experiments". *Annals of Mathematical Statistics* **33**, 404-419.

Duncan, George T. and Lambert, Diane [1986] "Disclosure-limited data dissemination (with discussion)". *Journal of the American Statistical Association* **81**, pp. 10-28 (in this vol.).

Kim, Jay [1986] "A method for limiting disclosure in microdata based on random noise and transformation". *Proceedings of the American Statistical Association, Section on Survey Research Methods* (in this volume).

Kirkendall, Nancy J. [1985] "Weights in computer matching: applications and an information theoretic point of view". *Record Linkage Techniques -- 1985, Proceedings of the Workshop on Exact Matching Methodologies*, Publication 1299 (2 - 86) of the Department of the Treasury, U. S. Internal Revenue Service, Statistics of Income Division, pp. 189-198.

Newcombe, Howard and Abbatt, J. [1983] "Probabilistic record linkage in epidemiology". Report prepared for Eldorado Resources, Ltd., October 1983.

Paass, Gerhard [1985] "Disclosure risk and disclosure avoidance for microdata". Paper presented at the International Association for Social Service Information and Technology, May 1985.

Paass, Gerhard and Wauschkuhn, Udo [1985] "Datenzugang, Datenschutz und Anonymisierung: Analysepotential und Identifizierbarkeit von Anonymisierten Individualdaten", Oldenburg Verlag, Munchen.

Palley, Michael A. and Simonoff, J. S. [1986] "Regression methodology based disclosure of a statistical database". *Proceedings of the American Statistical Association, Section on Survey Research Methods* (in this volume).

Pearson, Robert W. [1986] "Research access to publicly collected data: a report based on a conference November 21-22, 1985 Washington, D.C." Committee on the Survey of Income and Program Participation, Social Science Research Council N.Y., N.Y. 10158.

Smith, Martha, Newcombe, Howard and Dewar, Ron [1983] "The use of diagnosis in cancer registry death clearance". Health Division, Statistics Canada (OEHRU-No. 2), April 1983.

Spruill, Nancy L. [1982] "Measures of confidentiality". *Statistics of Income and Related Administrative Record Research: 1982* Department of the Treasury, Internal Revenue Service, Statistics of Income Division.

Spruill, Nancy L. [1983] "The confidentiality and analytic usefulness of masked business microdata". *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp. 602-607.

Spruill, Nancy L. [1984] "Protecting confidentiality of business microdata by masking". The Public Research Institute, Alexandria, Va.

Strudler, Michael, Oh, H. Lock and Scheuren, Fritz [1986] "Protection of taxpayer confidentiality with respect to the tax model". *Proceedings of the American Statistical Association, Section on Survey Research Methods* (in this volume).

Subcommittee on Disclosure Avoidance Techniques (Federal Committee on Statistical Methodology) [1978] Statistical Working Paper 2, Federal Statistical Policy and Standards, U.S. Department of Commerce, Washington, D.C.