Journal of Official Statistics, Vol. 14, No. 4, 1998, pp. 337–345

Introduction to the Special Issue: Disclosure Limitation Methods for Protecting the Confidentiality of Statistical Data

Stephen E. Fienberg[1] *and Leon C. R. J. Willenborg*[2]

1. Introduction

The protection of confidentiality is both a private and a public issue. Corporations collect extensive information on customers, some of which is highly sensitive and often protected by law. Similarly government agencies collect both administrative and statistical data subject to pledges of confidentiality and as more and more administrative data are used for statistical purposes there are increased dangers of disclosure of confidential data.

At the same time, the public demand for data from statistical offices about diverse aspects of modern society seems insatiable. Other government agencies use statistical data for the allocation of funds and the monitoring of social programs, policy analysts use statistical data to do calculations. The potential effects of new legislation have to be investigated, and academic researchers are constantly looking for data to validate and extend theoretical social science models.

Even Saskia Groenwald, Otto Normalverbraucher, Jan Modaal, John Smith – or whatever their names in countries around the world: they are all showered with statistical data on crime, money, employment, and health, via the news-media to which they are exposed. And their lives are to a large extent governed by policies that are fueled by the analysis of data collected by statistical agencies of the countries in which they live. In a way, statistical agencies provide mirrors for society, and society itself is a keen user of its own images, just to gaze at in satisfaction, amazement, embarrassment or in order to improve its looks. The images are the data that a statistical office releases, after having collected and processed them.

Of course, the metaphor of mirror image is simply that, a metaphor, and while it is evocative, statistical data do not quite provide an accurate picture of reality. First, forming an image of an aspect of society is not as easy as reflecting light in a mirror. Second, the statistical image lacks detail. This lack of detail is in part a fiscal necessity, since it is more costly to employ sophisticated methodologies (or models), which produce better images

[1] Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.
[2] Statistics Netherlands, Voorburg, The Netherlands.
Acknowledgments: We thank the contributors to this special issue for their articles and pleasant cooperation throughout the review process. Without their help this issue would never have been completed. Furthermore we are grateful for the assistance provided by several reviewers and other JOS Associate Editors including: Josep Domingo-Ferrer, George Duncan, Sallie Keller-McNulty, Udi Makov, Jeroen Pannekoek, Peter Kooiman, Gordon Sande, Chris Skinner, Ton de Waal, Alan Zaslavsky, Laura Zayatz. Furthermore, several articles in this issue were prepared as the result of the European SDC project (Esprit 20462), which through its financial support gave a boost to research on disclosure limitation in Europe, while others were supported by U.S. statistical agencies, especially the U.S. Bureau of the Census and the U.S. Bureau of Labor Statistics.

but require more data. This lack of detail is also in part a statistical necessity, since too much detail would give a crisp picture in which individual entities were clearly visible. The release of such a clear image would run counter to legal strictures on the right to privacy and the widely-held view that the release of individual information does not serve a statistical purpose, because it unnecessarily exposes individual businesses, institutions, persons, households, etc., to public scrutiny. This exposure of individuals and enterprises is viewed as harmful for all parties involved and eventually for all of society. Exposed individuals and enterprises often feel threatened and either refuse to cooperate in future surveys and censuses, or corrupt their responses. As a result the statistical information collected about society would diminish in quality, thereby reducing the possibility of fathoming and – through adequate policies – improving society. It is for this reason that government statistical agencies and other organizations collecting and releasing statistical data have a strong interest in preserving the confidentiality of their respondents.

So the question arises how statistical agencies should collect, process, and release information, in the light of this concern for privacy and confidentiality. On the one hand, there is the agencies' public obligation to provide maximum information to society, while, on the other hand, the agencies must ensure that the privacy of individual entities represented in the data is sufficiently protected. Agencies that balance these two concerns walk a tight-rope, high above the public arena. The subfield of statistics that occupies itself with this form of tight-rope walking is variously referred to as statistical disclosure control, statistical disclosure limitation, statistical data protection, or statistical confidentiality. Two major reports from the U.S. Federal Committee on Statistical Methodology, in 1978 and 1994, document the progress that had been made on the topic, as does the 1993 Special Issue of the *Journal of Official Statistics* on ''Confidentiality and Data Access.'' See also the related report of a panel of the Committee on National Statistics at the U.S. National Research Council (Duncan, Jabine, and de Wolf 1993), as well as a 1992 special issue of *Statistica Neerlandica*. The articles in this special issue of JOS deal with different aspects of research on this topic.

Statistical disclosure limitation can involve several strategies:

- Reporting only a subset of the data, by selection of cases and/or variables.
- Modifying the data in some form.
- Not reporting the observed data at all, but only ''pseudo data.''

Duncan and Pearson (1991), Cox (1994), and Fienberg (1997) provide a description of some of these methodologies of data modification under the broad rubric of matrix masking, and examples of some of these methods can be found in e.g., the earlier special issue of this journal and Willenborg and De Waal (1996).

Since statistical data, gathered at government expense, are generally viewed as a public good (e.g., see Fienberg 1997), not releasing the data in some form runs counter to the principles associated with an open society and is generally viewed as an unacceptable response to disclosure risk. Virtually all agencies and organizations do not report *key* variables that can clearly identify individual respondents with certainty (sometimes referred to as *direct identifiers*). Further, many agencies report only samples of census data or even subsamples of large-scale sample surveys as another mechanism for reducing disclosure risk. Coupled with such measures, more and more agencies have begun to turn

to methodologies involving modification of statistical data, usually accompanied by a series of legal, organizational and computer security based measures. Although these latter tools are a *sine qua non* for safe data release, in this issue we have selected articles that focus on statistical disclosure limitation, which has become a sizable field of inquiry in its own right.

2. Contents of This Issue

In the statistical disclosure limitation literature it is common to focus either on microdata or on tabular data. While it has been traditional for statistical offices to release statistical data primarily in tabular form in their publications, with the widespread release of data in electronic form, more and more releases involve microdata that are then used for reanalysis by others. An alternative to large-scale releases of microdata would be for the agency to retain control of the data, and to grant access to users remotely through some type of query system, with agency control and checking for consistency with earlier releases and responses as well as for residual disclosure risk. A possible drawback of this type of access might be that the restrictions on the queries are too severe and a user cannot carry out all statistical analyses that he or she would like to do. In their article in this volume, Keller-McNulty and Unger describe a system that permits such on-line access and that combines insights from data bases and statistical disclosure limitation ideas.

Microdata contain information on individual entities and tables contain aggregate information, often in the form of (estimated) totals. Microdata are usually released as electronic data, and tables in either paper or electronic form, e.g., as spreadsheets. Tables typically contain frequency count data or magnitude data (such as turnover, income, weights, etc.). Along with a table in the form of a cross-classification, agencies usually release marginal tables for subsets of the variables. The presence of these marginal tables is one type of constraint information that is present in the released data and that is at the core of the protection of statistical data.

How exactly should a statistician approach the protection of statistical data? This question is a fundamental one, and should be the starting point for a serious study of statistical data protection. As the answer depends, among other things, on the type of data, we discuss this matter for microdata and tables separately. We first consider microdata.

In the introduction above we alluded to a type of disclosure that is usually referred to as *identity disclosure*. In this form of disclosure – which is in practice the most important form – an individual is first re-identified and on the basis of this, confidential information about this individual is gleaned from the data. As said, this is the most important form of disclosure, but by no means the only form. The establishment of someone's identity is not a prerequisite to disclosing sensitive information about this individual. It is sometimes enough to know that an individual is a member of a group of individuals, without knowing which one of these, to disclose certain information on this individual. An example of this so-called *group* or *attribute disclosure* occurs if an individual i is established as belonging to some group of individuals G, and from the file it is clear that the income of each individual in G (in a particular reference period) was larger than t. This establishes that the income of i is larger than t.

We now turn our focus to identity disclosure. In order to be able to make rational decisions about how to protect microdata, we must introduce the concept of a disclosure scenario. Such a scenario describes how an intruder might go about trying to establish the identity of one or more individuals in the microdata set, and the knowledge that he or she is assumed to have about the population. What this essentially leads to is a matching situation. On the one hand there is a microdata set M in which an intruder I is interested, and on the other hand we have one or more (imagined) data sets summarizing, in a stylized form, the intruder's information about some or all population elements. The intruder's data set(s) can be real or imaginary, existing only in his or her head. In the first case the matching is assumed to be carried out with the help of a computer, whereas in the second case this is done in the intruder's head. The intruder's files are stylized because they are assumed to be in a form which makes matching with the microdata set M possible. If the contents of M change, so do the contents of the intruder's file(s). The matching procedure is not usually spelled out in a disclosure scenario. For simplicity one can assume that some exact matching procedure is used. (In the case of the matching taking place in the head of the intruder, this procedure, however, is likely to be more sophisticated.)

An intruder is going to use identifying variables. These are variables that concern information about an individual which can be known to somebody in the ''social neighborhood'' of this individual, or that can be found in registers that an intruder might use for matching against the microdata set. In practice one does not have a clearcut decision procedure as to which variables in a microdata set are identifying and which are not. This is a matter of judgment on the part of the data protector. In order to fill in further details of a disclosure scenario one should also specify which combinations of key variables to consider. This corresponds to the combinations of variables that an intruder is assumed to possess information about, i.e., with respect to which the intruder is assumed to know individuals in the population.

A disclosure scenario should also contain the goals and incentives of an (imagined) intruder. It can be that an intruder is interested in obtaining information about particular individuals, that he or she tries to establish a disclosure in order to discredit the data provider, or that the intruder tries to show his or her own cleverness. Some of these disclosures would go unnoticed, whereas others would be publicized. In the latter case the harm to the data provider, say a statistical office, is such as should be avoided for reasons mentioned above: dropped response rates in future surveys due to a lack of public confidence in the statistical office as a guard for the privacy of respondents. Although it is evident that such ''public disclosures'' should be avoided, the same holds in fact for the ''silent disclosures,'' even though they do not create any public stir. One cannot be satisfied with the idea that such disclosures are of no direct harm to the statistical office, because there is still the respondent's interest to take care of.

Given the situation with intruder I, his or her data sets, and his or her matching procedure, the next thing to do would be to assess the disclosure risk, i.e., the probability that I would be able to establish certain re-identifications correctly; see e.g., Paass (1988), Lambert (1993) and Fienberg, Makov, and Sanil (1997). This assessment of ''correct'' and ''incorrect'' re-identifications requires a probability model. In at least one formulation the probability model is used to provide the probability of the number of disclosures in a microdata set (or alternatively, the expected number of re-identifications), or the

disclosure risk per record. The latter type of disclosure risk can be used for calculating the first type of disclosure risk, but not the other way round. A *global* disclosure risk model, i.e., one which gives the expected number of disclosures in an entire file, was first attempted by Willenborg, Mokken, and Pannekoek (1990). In recent years interest has turned to more refined, i.e., per record, disclosure risks measures, and in the present issue the article by Skinner and Holmes discusses a model for assessing such per record risk.

In practice it is not always necessary to use an elaborate probability model. Instead one can use a simple thresholding model, in which a combination of identifying values is considered unsafe if it does appear frequently enough in the population. In an extreme case one may decide that population uniques are not acceptable. Because one usually works with a sample, the number of unsafe combinations or uniques in the population has to be estimated. This is an interesting problem in its own right, and one to which several articles in the literature of statistical data protection have been devoted. Also the present issue contains two contributions on the subject, namely the article by Samuels and the article by Fienberg and Makov, which also offers a disclosure risk per record model for cross-classified categorical data that is an alternative to that of Skinner and Holmes. In view of the usual disclosure risk scenario, it is not enough to be able to make such estimates. In fact, the statistical agency needs to be able to identify to which combinations of key variables the procedure should be applied. This issue tends to be subject matter specific, although in general variables linked to geography are typically included in any list.

Applying a disclosure risk model to a combination of identifying variables requires a threshold value in addition, to distinguish safe from unsafe microdata. This value expresses the maximum risk that a statistical office is going to accept when it releases microdata. If the microdata set has an associated disclosure risk below the threshold value, it is considered safe; otherwise it is not. If one uses a frequency threshold criterion to find the unsafe combinations in a microdata file which is based on a sample of the target population, one faces the following problem. The threshold level used is at the population level and one should find an analogous one at the sample level. Pannekoek and De Waal describe several estimation techniques to achieve this in their article, including empirical Bayes approaches, and Zaslavsky and Horton explain how to embed their methods in a fully Bayesian framework.

Once we have information on disclosure risks and threshold values we in fact have the basic ingredients to distinguish ''safe'' from ''unsafe'' data, and we can use them to produce ''safe'' data through the application of data modification techniques.

It is intuitively clear that the modification of a microdata set decreases its information content and thus generates a revised dataset that should be safer than the original one. To hamper re-identification it is natural to modify (some of) the identifying variables. This modification can be achieved by the application of techniques such as global recoding (or grouping), local suppression, imputation, data swapping, etc. Such techniques are generally referred to as *statistical disclosure control techniques*. Application of these techniques can in principle be done without using specialized software, but this is quite a daunting task. It is much more preferable to take advantage of software created especially for the tasks outlined, such as μ-ARGUS (cf. Hundepool et al. 1998). In this package some of the protection techniques mentioned have been implemented (as well as some

others not mentioned), and the user can either interactively or automatically eliminate unsafe combinations. In order for the user to carry out this type of examination in an automated fashion, additional theoretical work is necessary.

In fact with an automatic mode for protecting microdata one is forced to define the goal of protection of microdata in precise terms. Of course, it is clear that the final result should be a microdata set that is safe, according to the safety rules associated with the disclosure risk model adopted. But more is needed. In the most extreme case a microdata file with all values suppressed is certainly safe, but contains no information. So what is missing is a notion of information content. The formalization of this concept is not so easy. A user of the data has usually only an intuitive notion of the information content of a microdata set. And no two users necessarily agree on the information content of the same microdata set or on how the dataset should be structured for release. Since a statistical office cannot honor every request for microdata files, it would be beneficial in deciding what data to release if the office could use a measure of information loss that counteracts the tendency to decrease detail in the categories by grouping them, to suppress information, or even to replace values in the data, using a stochastic procedure.

One can obtain inspiration for information loss measures from statistical information theory e.g., forms of entropy, and statistical estimation theory e.g., the variance or mean squared error of particular linear estimators (usually of totals). But the choice of a suitable information loss function is not easy, and it requires considerable judgment and experimenting.

Once such an information loss function has been specified, one can formulate the statistical disclosure control problem for microdata as an integer programming problem. A solution to such a problem would yield an optimum way to apply certain disclosure control techniques to a given microdata set, resulting in such a set as is considered safe. When one restricts the disclosure limitation techniques to local suppression only, one can eliminate unsafe combinations by replacing values appearing in these combinations by ''missing values.'' In the present issue De Waal and Willenborg formulate and discuss several optimum local suppression models. The ideas in this article (or rather an earlier report that appeared several years ago) were important for the development of the automatic local suppression facility in μ-ARGUS. In their article Hurkens and Tiourine carry them one step further by considering optimization models in which both global recoding and local suppression can be used. Their goal is to find the optimum mix of global recodings and local suppressions in order to find a safe microdata set without losing too much information. Most of the ideas in the Hurkens and Tiourine article have been implemented in μ-ARGUS.

The optimization problems considered in both articles are provenly difficult, and as a consequence it is not always possible to calculate the optimum solutions in a reasonable amount of time. This is not a reason for despair, however, as the choice of the target function (i.e., the loss function) is somewhat fuzzy. A good approximation may just be as valuable in practice as an optimum solution.

The disclosure protection measures for microdata in the articles by De Waal and Willenborg and by Hurkens and Tiourine are non-perturbative, in the sense that they cannot cause inconsistencies in the data. They differ from techniques such as noise addition and data swapping, and also from the techniques studied in the three articles

by Anwar and Defays, by Gouweleeuw, Kooiman, Willenborg, and De Wolf and by Fienberg, Makov, and Steele in the present issue. The Anwar-Defays article is concerned with a method referred to as microaggregation, which can also be viewed as a rounding technique. This method is only defined for quantitative variables. The Gouweleeuw et al. article discusses a method called the post randomization method (abbreviated as PRAM). There is a link between PRAM and the randomized response method proposed for the collection of confidential information by Warner (1965). PRAM can be viewed as a technique that transposes the noise adding technique for quantitative data to categorical data. For categorical data one cannot add something, but one can replace a value by another value. PRAM does this, using a well-defined random mechanism. Because certain parameters governing a PRAM application can be made publicly available, it is possible to correct certain estimators for the perturbation process. This makes the method an appealing one in the class of perturbative techniques. The literature on randomized response is available when it comes to analyzing microdata that have been subjected to PRAM, and in his discussion of this article Sande provides a number of relevant references and other background.

In their article, Fienberg, Makov, and Steele consider a general strategy for data perturbation, motivated by the bootstrap and multiple imputation methods from the literature on statistical estimation, and suggested earlier by Rubin and by Little in their contributions to the 1993 Special Issue of this journal. They also propose a particular procedure for cross-classified tables of counts based on the exact distribution of a contingency table given a set of marginal totals. The discussion of the Fienberg et al. article by Kooiman reflects the perspective of a statistical office.

In general one can say that these perturbative techniques are somewhat more difficult to apply than the non-perturbative ones, and that they require care in quantifying the reduction of disclosure risk. Some have also argued that the potential of introducing inconsistencies in the data may be somewhat unattractive to some statistical agencies and to some users, but the advocates of such methods note that they are likely to lead to more extensive data release and thus provide users with expanded access to data.

Fienberg, Makov, and Steele's contribution spans the domain of microdata files and tabular releases, and they contrast their specific methodology for contingency tables with two other methods used for tabular data, cell suppression and data swapping. In tables or cross-classifications, the goal of most agencies is to protect ''sensitive'' cells. A cell is typically deemed to be sensitive if it corresponds to a small number of individuals or enterprises, or in the case of weighted tables, a small number of individuals or enterprises constitute a large fraction of the cell total (see e.g., Federal Committee on Statistical Methodology 1994 or Willenborg and De Waal 1996). In the case of a sensitive cell, many agencies choose to suppress the cell value; however, due to the presence of constraints such as marginal totals and non-negativity of cell values, one is able to determine upper and lower bounds for each sensitive cell. Without extra protective measures the intervals associated with such bounds for many sensitive cells may be of zero width, and hence the original value would be known. Thus, one way out of this problem is to suppress additional cell values (corresponding to non-sensitive cells). Finding a (weighted) minimum of other cells to suppress in a table such as to protect the sensitive ones (by providing a large enough protection interval for each of these cells) is referred

to as the secondary cell suppression problem and is known to be notoriously difficult (cf. Kelly, Golden, and Assad 1992). Because this problem is so difficult, one usually has to be satisfied with sub-optimum solutions, found by applying heuristic methods (cf. Fischetti and Salazar 1998). In practice this requires specialized software. Several such programs have been written at the major statistical offices in the world. One package for dealing with the secondary cell suppression problem and of recent origin is τ-ARGUS (cf. Hundepool et al. 1998). Sande and Kirkendall in their article describe and compare other packages of American and Canadian origin for cell-suppression in cross-classified tables.

Another solution to protect the sensitive cells in a table is to add noise to all cell values, of sensitive and non-sensitive cells alike even in the marginal tables, in order to mask the original sensitive values. Non-sensitive values are perturbed, in order to maintain additivity in the perturbed table, which is often a requirement. A particular way of adding noise to a table is rounding. In the present issue there are two articles on applying perturbative techniques to tables: the article by Evans, Zayatz, and Slanta is on noise addition where the structure of the noise is linked to the overall structure of the sampling frame for the data, and the article by Fischetti and Salazar is on controlled rounding, a non-random perturbation approach, in the presence of marginal constraints. The optimization model for the rounding problem is so general as to allow general linear constraints among the cell values, i.e., not only those derived from the presence of the marginals.

3. References

Committee on Maintaining Privacy and Security in Health Care Applications of the National Information Infrastructure. (1997). For the Record: Protecting Electronic Health Information. National Academy Press, Washington, DC.

Cox, L.H. (1994). Matrix Masking Methods for Disclosure Limitation in Microdata. Survey Methodology, 20, 165–169.

Duncan, G.T. and Pearson, R.W. (1991). Enhancing Access to Microdata While Protecting Confidentiality. Statistical Science, 6, 219–239.

Duncan, G.T., Jabine, T.B., and de Wolf, V.A. (eds.) (1993). Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics. Panel on Confidentiality and Data Access, Committee on National Statistics, National Academy Press, Washington, DC.

Federal Committee on Statistical Methodology (1978). Report on Statistical Disclosure and Disclosure-avoidance Techniques. Statistical Policy Working Paper 2. Subcommittee on Disclosure-Avoidance Techniques. U.S. Department of Commerce, Washington, DC.

Federal Committee on Statistical Methodology (1994). Report on Statistical Disclosure Limitation Methodology. Statistical Policy Working Paper 22. Subcommittee on Disclosure Limitation Methodology. Office of Management and Budget, Executive Office of the President, Washington, DC.

Fienberg, S.E. (1997). Confidentiality and Disclosure Limitation Methodology: Challenges for National Statistics and Statistical Research. Background paper commissioned by Committee on National Statistics. Available electronically at http:/lib.stat.cmu.edu/www/cmu-stats/.

Fienberg, S.E., Makov, U.E., and Sanil, A.P. (1997). A Bayesian Approach to Data Disclosure. Optimal Intruder Behavior for Continuous Data. Journal of Official Statistics, 13, 75–89.

Fischetti, M. and Salazar, J.-J. (1998). Modeling and Solving the Cell Suppression Problem for Linearly-constrained Tabular Data. Proceedings of the Statistical Data Protection '98 Conference, IOS Press, Luxembourg, in press.

Hundepool, A., Willenborg, L., Wessels, A., Van Gemerden, L., Tiourine, S., and Hurkens, C. (1998). μ-ARGUS user's manual. Department of Statistical Methods, Statistics Netherlands, The Netherlands.

Hundepool, A., Willenborg, L., Van Gemerden, L., Wessels, A., Fischetti, M., Salazar, J.-J., and Caprara, A. (1998). τ-ARGUS User's Manual. Department of Statistical Methods, Statistics Netherlands, The Netherlands.

Journal of Official Statistics (1993). Special issue: Confidentiality and Data Access. 9, 269–607.

Kelly, J.P., Golden, B.L., and Assad, A.A. (1992). Cell Suppression: Disclosure Protection for Sensitive Tabular Data. Networks, 22, 397–417.

Lambert, D. (1993). Measures of Disclosure Risk and Harm. Journal of Official Statistics, 9, 313–331.

Paass, G. (1993). Disclosure Risk and Disclosure Avoidance for Microdata. Journal of Business and Economic Statistics, 6, 487–500.

Statistica Neerlandica (1993). Special Issue: Proceedings of the International Symposium on Statistical Disclosure Avoidance. Volume 46, No. 1.

Warner, S.L. (1965). Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. Journal of the American Statistical Association, 60, 63–69.

Willenborg, L.C.R.J., Mokken, R.J., and Pannekoek, J. (1990). Microdata and Disclosure Risks. Proceedings of the 1990 Annual Research Conference. U.S. Bureau of the Census. Washington D.C.: U.S. Department of Commerce, 167–180.

Willenborg, L. and De Waal, T. (1996). Statistical Disclosure Control in Practice. Lecture Notes in Statistics, Vol. 111. New York: Springer Verlag.