The papers by Cox and by Gastwirth, Krieger and Rubin (GKR) are both about disclosure-avoidance techniques. Cox's paper is about a specific technique - cell suppression - used by the Census Bureau to avoid disclosure in tabulations of economic census data. The GKR paper is somewhat broader in scope. Although the title refers to "summary data", their formulation also covers the release of microdata for statistical purposes, as long as appropriate confidentiality conditions are met. They discuss 3 disclosure-avoidance techniques - grouping, rounding and contamination. The results of grouping can be presented either in the form of tabulations or of microdata. In the latter case, the record for an individual will identify only the class interval in which he falls. Sometimes means and variances for groups or class intervals are also released.

As GKR point out, the choice of suitable disclosure-avoidance techniques requires a trade-off. The data user wants information, with as much detail and precision as possible. Data subjects, whether persons or businesses, require protection against disclosure of information which can be associated with them. The requirement for such protection derives from laws and regulations (although these are not usually specific about the degree of protection required), from assurances given to individuals who supply data, and from agency policies. Ideally, evaluation of any proposed disclosure-avoidance technique should be based on quantitative analysis of the resulting loss of information and of the degree of protection provided.

My primary interest, in reviewing these two papers, has been to see to what extent the authors have provided the information needed to make such an evaluation of the disclosure-avoidance techniques they discuss. This may not be entirely fair, because the authors may not have intended to make this type of evaluation. No criticism of the technical quality of the papers is intended; they are both well-written and make useful contributions to our knowledge of disclosure-avoidance techniques. Nevertheless, it is essential that disclosure-avoidance techniques not be applied in a mechanical fashion, without a critical assessment of the extent to which they meet their objectives.

Cox's paper does not discuss how loss of information resulting from cell suppression limits analysis based on the data. The n respondent, k percent rule with specific values of n and k is taken as given; no alternatives are discussed.

There is evidence from other sources that some users of Census economic data feel handicapped by the disclosure-avoidance techniques used. Two appropriate references are Statistical Working Paper 2, Report on Statistical Disclosure and Disclosure-Avoidance Techniques (U.S. Department of Commerce, 1978, Chapter V) and the invited

address given at these meetings last year by Professor F. M. Scherer (1977).

Cox does discuss the amount of protection afforded by the suppression method used. However, if I represented a particular business asked to respond to an economic census, I would still have difficulty knowing just how much protection would be given to my responses. There are two reasons for this.

1. Cox's analysis is in terms of the protection afforded to cell totals for unpublished cells and to totals for the n largest respondents for published cells. The analysis could and should be extended to the degree of protection given to individual respondents. In doing so, it would be necessary to analyze separately the amount of protection against external and internal disclosure, and in the latter case, to indicate whether the potential "attacker" is a single member of the cell or a coalition (see Department of Commerce, 1978, Chapter III).

2. A more difficult problem is that the Census Bureau cannot reveal the specific values of n and k it uses, because to do so would reduce the amount of protection provided by the cell-suppression procedures used. Thus, respondents know that they are getting some protection, but not how much.

Whether the economic census respondent gets enough protection is partly a technical issue that can be clarified by the type of analysis suggested. Beyond that, it is a political issue. I personally believe that some of the economic census constraints should be relaxed in the interest of providing more complete and detailed data for economic policy decisions. This might require legislation, depending on the extent of the change.

Notwithstanding any reservations one might have about the suitability of the particular disclosure-avoidance techniques used, Larry Cox deserves considerable credit for his substantial contributions to the establishment of the theoretical structure and the development of computer systems to apply cell-suppression procedures to the economic censuses. What is now done effectively, at moderate cost with computers used to be a large-scale time-consuming clerical operation, and it is quite likely that because of the complex structure of the tabulations, many complementary disclosures were overlooked.

The paper by Gastwirth, Krieger and Rubin is primarily concerned with how the disclosure-avoidance techniques they discuss - grouping, rounding and contamination - affect the estimates of various statistics, such as means, variances and correlation coefficients, which users might want to derive from the data.

Much is known about the univariate case, and this topic is admirably dealt with. Unfortunately,

it is not of much practical interest. If a microdata file to be released has no identifiers and only one variable, there is usually not much that can be disclosed about any individual, unless the distribution contains a few extreme or unusual values, in which case grouping may deprive users of some essential information.

As indicated by GKR, little work has been done on the effects of grouping and related techniques on analyses of multivariate data. Thus, in those situations where multivariate analysis based on microdata files is called for, we know little about what information may be lost as a result of grouping, rounding or contaminating the data.

What about the other side of the tradeoff? GKR do not say much about the protection afforded to individual respondents or data subjects when these techniques are used. Their "C1" condition is that not too much information about individuals should be recoverable from the data released to the analysts. How much is too much? This is not an easy question to answer.

I am concerned about the assumptions made by GKR that C, the data collector, can have all kinds of identifiable data, but that A, the analyst, cannot; or that (in the matching case) C1 and C2 can have certain data with identifiers but that neither can have the union of these sets. Must we necessarily assume that C is well-intentioned and will observe appropriate security precautions, but that A will be careless, or will do his best to identify individuals and use data about them for non-statistical purposes?

Obviously, the answer to this question depends on who C and A (or C1 and C2) are, and how they plan to use the data. It also depends on what sanctions can be applied to them in the event of intentional or inadvertent disclosure of information about individuals.

The fact that data about identifiable individuals is recoverable from tabulations or microdata does not mean that it will be recovered and used for non-statistical purposes. It is the latter that we really want to prevent. In some cases it may be possible to do this through the establishment of effective sanctions and the use of security techniques such as encryption with no loss of information to the analyst.

We must clearly do whatever is necessary to obey the laws pertaining to disclosure, and to keep our bargains with respondents. However, let's not go overboard and uncritically apply disclosure-avoidance techniques without evaluating their effects, both in terms of loss of information to users and in terms of additional protection for data subjects.

REFERENCES

Scherer, F. M.

1977 - "Statistics for Governmental Regulation". Special address invited by the President, Annual Meeting of the American Statistical Association.

U.S. Department of Commerce

1978 - Statistical Working Paper 2. "Report on Statistical Disclosure and Disclosure-Avoidance Techniques." Prepared by Subcommittee on Disclosure-Avoidance Techniques, Federal Committee on Statistical Methodology, Office of Federal Statistical Policy and Standards.