DISCUSSION

Nancy L. Spruill, Office of the Assistant Secretary of Defense
for Force Management & Personnel

I really enjoyed reviewing the four papers given in this session. These papers include some of the best written, clearly laid out ones that I've read. They do a good job of discussing the complex issues in privacy protection and disclosure avoidance. They also present some good new ideas for avoiding disclosure and enhancing data analysis when compared to existing techniques.

I just want to quickly review each paper and then ask a question of each of the presenters to begin the discussion.

MARCH - NORRIS PAPER

Disclosure Avoidance Techniques in the Canadian Censuses of Population and Agriculture

This paper is clearly written and looks at disclosure issues for the Census of Population -- where small frequency is the main problem -- and the Census of Agriculture -- where the problem comes from one or two respondents contributing almost all the information in a cell.

The Census of Population provides data both in tables and in building blocks so that the user can produce tables. The masking techniques used are random rounding to the base 5 and suppression. The Census of Agriculture provides data in tables but does not accommodate specific user requests. The masking technique is a customized system of cell suppression (and complementary cell suppression) because of small sample sizes or one or two dominate farms.

Question: Suppose a data intruder gets out of George and Diane's paper and slips into Canada. Can this intruder design multi-custom tables for data from the Census of Population that will lead to disclosure problems not seen in the individual tables? How do your disclosure avoidance techniques of random rounding and suppression protect against multi-custom tables "designed" by an intruder?

DUNCAN - LAMBERT PAPER

The Risk of Disclosure for Microdata

This is a good, well-written paper that analyzes the risk of disclosure for several cases. It looks at two kinds of disclosure -- identity and attribute -- and two kinds of data "intruders" -- an uninformed outsider and an informed insider. A unique and useful thing about this paper is the introduction of a loss function that describes the intruder's goals with respect to the data. With a known loss function, the data releaser can test the amount of risk for a proposed data release. Also, he or she can modify the masking techniques to minimize the risk.

I especially appreciated the numerical examples given in this paper. I had a much better understanding of the issues involved from these examples and from the discussion of what a data releaser can do to "dissuade linking" and, hence, to dissuade disclosure. I found this discussion very informative.

What I'd like to see is more examples looking at different loss functions. These examples should include:

1) translating the results for the data releaser to what he or she can do to provide protection; and

2) translating the results for the data user on how he or she can get better information when doing analyses using the released data.

Question: Each data intruder has his or her own loss function. How do the data releasers provide protection against all of these threats?