Comment

Janet L. Norwood

Duncan and Pearson's article provides an insightful treatment of complex issues ranging from protection of respondent rights to operational methods for protection against disclosure. The authors lay out thoughtful arguments for increasing access, review methods for masking data and provide suggestions for bringing academic analysis and data protection into a more open environment than currently exists. Their discussion is comprehensive and balanced for data collected about people. But the article would have been more useful if it had been broadened to cover data collected from business establishments where disclosure presents more formidable problems.

Duncan and Pearson recognize the tension between academics who wish to access microdata and statistical agencies who wish to protect data privacy and confidentiality. As one of the data stewards for microrecords collected by the Bureau of Labor Statistics with a pledge of confidentiality, I understood fully the concern that statistical agencies have in carrying out their responsibilities. As an economist with a real interest in academic research, I also know that data are needed for research. And I know also that modern research involves microdata, the computer and the ability to match observations. It is easy to sympathize with, indeed, to agree with, the intent of the authors. Their discussion is useful, but, unfortunately, it does not solve the dilemma we face.

It is time that we realized that the laws and customs under which we operate are somewhat contradictory in concept. Preservation of the right of privacy is a basic right in our society. The Privacy Act of 1974 as amended (5 U.S.C. 552a) prevents disclosure of records maintained on individuals while the Freedom of Information Act (5 U.S.C. 552) prevents government agencies from refusing to provide information to the public. Exceptions in the Privacy Act are designed to permit law enforcement; the Freedom of Information Act exemptions protect confidential commercial and financial information that might, at times, be useful in law enforcement. In addition, a series of laws,

Janet L. Norwood is Commissioner of Labor Statistics, U.S. Department of Labor, Washington, D.C. 20212.

judicial opinions and administrative orders affect the release of data by the major agencies in the federal statistical system. But we must also remember the statistical roots underlying the preservation of confidentiality. The respondent's belief in the agency's ability to preserve the confidentiality of the data provided tends to ensure cooperation and to enhance quality. Against this background, the bias of the Federal data steward is to withhold microdata rather than to provide it. One can, therefore, sympathize with Duncan and Pearson's intent to work within that set of biases to enhance access to microdata.

The article challenges the statistical data stewards to find new ways to provide microdata access and provides a useful review of a series of approaches that might be considered. Most of the attention in this field thus far has been given to methods for providing users with microdata that has been transformed in some way to mask the identification of respondents. One approach involves combining, deleting or altering the number of records. A second approach alters attributes within each record. And the third method adds random or deterministic noise to the microdata. Of course, all masks reduce the value of data for making statistical inferences, but the most troubling, it seems to me, is the addition of noise. The addition of random noise to microdata could produce problems for research that could be very difficult to overcome even with advanced statistical methods. Because Duncan and Pearson's purpose is to provide an overview, they pay little attention to the issue of how to determine what information is sensitive. What is acceptable risk? And how do the risks vary by the content of the data, the kind of respondent or the user of the data? As a practical matter, it seems that each data file might require a different masking technique.

Several aspects of Duncan and Pearson's vision and proposals for the future relate to the behavior of the researchers who wish to have increased access to the data. The options range from admonition ("take more responsibility"), to a code of conduct, to licensing and even to new legislation which would prescribe penalties for disclosure. Their review of these possibilities is very useful but demonstrates, I think, the difficulty we face in arriving at one set of standards to apply to all data sets and for all purposes. This imposes a considerable burden