

on the statistical agency, which has guaranteed the confidentiality of the data, because it must consider policies for a range of unknown risks and uses and then police compliance by users. Establishment of a staff of "gatekeepers" could be useful, but could such a staff built essentially to service academic users be justified in the current tight budget climate?

The authors are correct in suggesting that we need to focus more attention on obtaining the informed consent of respondents. I fully support their suggestion that federal agencies need to conduct pilot studies to assess the effectiveness and respondent understanding of the statements used in data collection. In fact, the Bureau of Labor Statistics currently has underway, with IRS sponsorship, research into respondents' understanding of and reaction to the language used in confidentiality statements. Much more work needs to be done in applying the laboratory techniques that combine the cognitive sciences and survey research to assess these issues.

The focus of the Duncan and Pearson article is on data about individuals. But confidentiality problems with establishment data are much more complex than for those about people. For one thing, there are fewer establishments than there are people. Businesses can much more easily be classified into subgroups, often with a very small number of units in the groups of particular interest. In addition, a good deal of information about business establishments is available in publicly accessible files that can be matched to the federal system's file and then used to help to disclose confidential data. Moreover, the value of such data to Duncan and Pearson's data spy might be much greater than the value of the data collected about individuals—to say nothing of those who wish to use such data in prosecution and enforcement.

The risk of disclosure is also generally greater for establishment data than for data about individuals, and the stakes for the company can be quite high

when trade secrets or business practices are involved. On the other hand, some data—for example, the number of employees or the identification of major products—may not be sensitive at all to some firms but of great concern to others. There is no simple formula for determining which items are the most sensitive.

The problems involved in finding methods for improving access to microdata on establishments for research purposes are complex and difficult, but the need to find solutions is becoming increasingly necessary. Academic researchers are becoming more and more interested in the use of longitudinal microdata files on business establishments, and access to such data would clearly improve some of the public policy research. Statistical agencies have only just begun thinking about these issues, however, and much more work needs to be done.

Duncan and Pearson are quite right in pointing out that research interests and computational capabilities have led to new and more varied demand for publicly collected data. They are also quite right in pointing to the slow and somewhat negative responses from the nation's primary statistical agencies. But their suggestions, while useful, do not point the way to a quick and clear solution. We in the statistical system strongly believe that the absolute protection of confidentiality tends to assure the cooperation of respondents in voluntary surveys (and most government surveys are based on voluntary cooperation) and enhances the quality of the responses. It is true, however, that statistical agencies have not done all that they could to find ways to provide researchers with the data they need within the practical and legal constraints under which the agencies operate. The article properly challenges the nation's statistical system to revisit the confidentiality practices now in place. In doing so, it serves a valuable function. But the problems we face are real, they are complex and there is no easy and quick solution to them.

Rejoinder

George T. Duncan and Robert W. Pearson

While generously acknowledging the centrality of the themes we identify, the discussants quite rightly point to wider issues that should command our attention in the future. To give structure to these issues, we cast the discussants' insights into

a set of nested frames. The outer frame encompasses the functional effectiveness of a government statistical system in a diverse society with democratic aspirations. The middle frame delineates the nature of the data that society collects and main-