between privacy and access concerns, while maintaining high levels of participation and truthfulness in surveys—without public support.

I believe the time has arrived for the statistical system to "go public" with confidentiality and data access issues. This is based on my opinion that both the policy debate and the development of technical solutions to disclosure protection problems within the statistical system have matured sufficiently to be analyzed and discussed at a general level. This was not the case 10 years ago. There are dangers to raising these issues, however, that must be kept in mind: By raising the issue, it can be made salient in a way that frightens the public; also, we run the risk of confusing people with arcane, inconclusive or contradictory technical and legal information (thereby eroding their confidence in a different way).

The approach should be to communicate the importance of reliable and varied statistics to the society and the economy, while instilling confidence that individual respondents have rights, the protection of which is the bedrock of the statistical system. The approach should first be made through influence groups: advocates for privacy; groups representing the disadvantaged (e.g., the homeless) or those at risk (e.g., AIDS); those concerned with the rights of individuals (e.g., ACLU); the press; and those concerned with the political process. Avenues to many of these groups exist already within the normal workings of the statistical system. Well- and nontechnically articulated arguments need to be developed and discussed with these groups, leading perhaps to experiments of one kind or another, and, ultimately, consensus and change. If these groups are convinced and, to a degree, become advocates for the statistical system on issues of privacy and data access, I believe the support of the public at large will follow.

Within these deliberations, it is important to maintain a focus on legislative issues. Laws do not prescribe how statistical agencies are to design questionnaires and samples, estimate parameters or edit questionnaires and impute missing or faulty data; yet regarding confidentiality, many laws are absolute, one-side in assigning penalties and, although written in the absence of technical information, exert a driving influence on agencies' confidentiality practices. Most agree that responsibility for disclosure protection should, like the data, be shared between the data provider and the data user. This strikes me as an issue easily understood and potentially supportable by outside groups.

The Duncan–Pearson article does a good job of presenting the mounting issues faced by the statistical system along the data confidentiality/data access front. It is readable outside the statistical community, and that is important if we are to broaden the discussion, as I suggest be done.

Comment

Sallie Keller-McNulty

I would like to commend Duncan and Pearson for their contribution on the very important topic of data access and confidentiality. I am pleased that Statistical Science has had the foresight to publish such an article, and I hope that many researchers will read and react to the material. I have no disagreements with the opinions expressed in this manuscript, but I would like to bring more attention to a few points that were made.

First, I would like to comment on the various ways that disclosure has been conceptualized. In particular, attention has been focused on inferring

Sallie Keller-McNulty is Associate Professor of Statistics, Kansas State University, Manhattan, Kansas 66506.

attribute values. I contend that, with the database systems as a data storage and access medium, we need to also be concerned with the direct disclosure of relationships between attributes. We have been conditioned to view data as a file or rectangular array where the columns represent attributes, the rows represent data records (one for each respondent), and the entries within a row represent attribute values. Attribute disclosure is conceptualized as inferring an element of this array. In this setup, hypotheses about the relationships between attributes are validated through analysis of the data. In a database, relationships among attributes are contained in the schema, or logical structure, of the system. Relationships as well as attribute values are considered objects (i.e., encapsulation of values with their semantic meanings). Disclosure in a database system can be defined as inferring an