Journal of Privacy and Confidentiality (2011) **3**, Number 2, 53–55

Use of Administrative Records and the Privacy-Confidentiality Trade-off

A comment on Gerald W. Gates's *How Uncertainty about Privacy and Confidentiality Is Hampering Efforts to More Effectively Use Administrative Records in Producing U.S. National Statistics*

Jennifer H. Madans*

The paper by Gerald Gates provides a comprehensive review of how privacy and confidentiality concerns have hampered the use of administrative records in the production and analyses of national statistics. Privacy and confidentiality concerns can affect multiple aspects of the process—from obtaining the needed records to providing access to micro data files. Even though the use of administrative records (either as substitutes for other sources of data or in addition to these other sources) improves the quality of statistics and is cost efficient, problems in any phase of data collection, storage, processing, or release can greatly reduce the likelihood that administrative records will become part of the core production of national statistics. Gates's paper provides an excellent summary of privacy- and confidentiality- related challenges and goes on to suggest a range of research activities to address them.

The terms privacy and confidentiality are separate concepts with privacy referring to an individual's right to decide how personal information is used and confidentiality referring to breaking the link between substantive information about an individual and the identity of that person. While the concepts are distinct, they are often intertwined. As Gates notes, while U.S. law generally is supportive of individual rights to determine how personal information can be used, there are laws permitting sharing of identifiable data without consent if confidentiality can be protected and the data will only be used for statistical purposes. However, these laws are not universal and sharing is not permitted without consent in many cases. More important is whether there is societal acceptance of this tradeoff between privacy and confidentiality for statistical purposes. This is a key issue which has not been given sufficient attention but one which could determine how administrative records can best be incorporated into the production of national statistics. The design of a research program to address best practices of incorporating administrative records must also consider policies and processes that will need to be adopted by statistical agencies as well as statutory requirements surrounding the data collected by them.

At one end of the continuum, let us assume that privacy issues can be ignored if there is absolute protection of confidentiality and if non-statistical uses of the information can be eliminated. In this case, agency efforts should be focused on maintaining the public's trust and acceptance that privacy concerns are irrelevant given absolute confidentiality and guarantees of no non-statistical use. At the other end of the contin-

*OD Co-Deputy Director / OD Associate Director for Science / OPBL Associate Director, National Center for Health Statistics, Hyattsville, MD mailto:jhm4@cdc.gov.

© 2011 by the authors

uum, let us assume that confidentiality is not an acceptable tradeoff for privacy. In such case, agency efforts should be focused on developing appropriate means for obtaining maximum participant agreement to obtain and use their administrative information for statistical purposes.

Our knowledge of the public's acceptance of the loss of control of personal information and their willingness to accept confidentiality as a tradeoff is limited—research would need to be conducted to fill this gap. If acceptance isn't widespread, it would be important to determine why this is the case and if concerns could be addressed by increased education about statistical agency activities with respect to confidentiality and the use of information for statistical purposes only, or whether modifying agency practices could increase acceptance. The importance of promising absolute protection of confidentiality would also need to be assessed. If the public feels that absolute confidentiality protection is a requirement for relinquishing control of personal information, statistical agencies will face greater challenges, and the great lengths currently employed to assure that identities cannot be disclosed may be deemed insufficient. The protection of confidentiality is not an easy or straightforward task. While much has been done to develop nondisclosure techniques, it is very difficult to guarantee that breeches will not occur. If the agreement with the public is based on maintaining absolute confidentiality, efforts to protect data will need to increase and this could have negative consequences on other agency missions such as making data widely accessible. Data are easier to protect if they remain within the statistical agency, but this will frustrate outside users. Efforts to develop alternative access mechanisms such as synthetic files or more user friendly RDCs could continue, but given current mechanisms it is possible that access to data files produced with administrative records could be reduced. If that is the case, would higher quality and lower cost national statistics justify reduced access to microdata?

Alternatively, if the public is not willing to cede control of personal data—even if the data are used only for statistical purposes and there is an absolute assurance of confidentiality—efforts will need to be focused on developing appropriate means for obtaining participant agreement to obtain and use their administrative information for statistical purposes. While there has been some research in this area, much has been directed to crafting language that will increase the probability that consent will be obtained. While the success of any program that incorporates administrative data in the production of national statistics is dependent on high consent rates, the objective of the consent should be to provide participants with the information needed to make a knowledgeable decision. Maximizing acceptance rates is acceptable only if it is based on meaningful informed consent. Research would need to be done on how to present the relevant information to the participant so that it is understandable and accurately conveys risks and benefits. However, lengthy consents, while they might include all required elements, often do not have the desired effect of informing the participant. The very length and complexity of a consent statement can result in limited understanding of what is being asked of the participant.

Consensus would need to be reached on the elements required to be included in the consent. Some information would be needed on which administrative records would be obtained, how the information obtained would be used, and how it would be protected,

but the level of specificity needed for each of these components is unknown. Would all specific administrative systems need to be listed or would a general statement be sufficient? Is it possible that consent might be dependent on the actual records to be obtained? For example, participants might be more willing to agree to the use of health records but not to the use of immigration records. Would it be necessary to be explicit about how far into the future records would be obtained or can ongoing consent be assumed? Would it be necessary to identify all types of users of the information or would a statement that the information would only be used for statistical purposes suffice? The informed consent would also provide information related to the protection of confidentiality, but how much detail about safeguards and access would be needed? The consent would need to address issues related to the probability that the participant could be identified even though processes are in place to protect confidentiality. The need for explicit consent as opposed to an "opt-out" protocol will need to be evaluated. Crafting appropriate informed consent documents will take time and should be based on sound research initiatives. Considerable work would need to be done to appropriately describe the nature of the risk to the participant. If consent is obtained for broad linkage to administrative information and any uses that fall within the statistical use category, it may be possible to be less restrictive in terms of data access.

The issues discussed above concerning the relationship between the protection of confidentiality and the elements of informed consent apply to all types of personal data collected by federal agencies, not just to the use of administrative records in the production of national statistics. Likewise, the issue of whether it is acceptable to reduce access if the public is not willing to trade control of information for assurance of confidentiality applies whether personal information is limited to that collected in surveys or through augmentation of survey data with personal data contained in administrative records. However, the tension between confidentiality and access may be heightened by inclusion of administrative records. Is the increase in data quality and reduction in costs by inclusion of administrative records worth increasing restrictions in data access? If current trends continue, confidentiality will be harder to protect with a resulting reduction in data release regardless of whether administrative records are used. The development of alternative access methods that are less restrictive will become more resource intensive. With declining budgets, resources used to develop new access systems will, to some extent, be at the expense of data collection, affecting the quality of national statistics. Would trading quality for greater access be an acceptable option? The appropriate allocation of funds to the various aspects of the production of national statistics is a policy decision that deserves careful consideration based on sound research. The potential impact that administrative records can have on the production of national statistics can be great but the challenges presented in Gates's paper also relate to basic principles of how statistical agencies interact with the public. Gates's recommendations that these issues be addressed jointly by the statistical agencies and that the decisions be informed by a rigorous joint research program merit serious consideration.