



June 21, 2007

The Honorable Michael O. Leavitt  
Secretary  
U.S. Department of Health and Human Services  
200 Independence Avenue, S.W.  
Washington, D.C. 20201

Re: Harnessing the power of Federal linked data sets to improve the nation's health while protecting confidentiality and security

Dear Secretary Leavitt:

The National Committee on Vital and Health Statistics (NCVHS) is charged to advise the Department on health data, statistics and national health information policy. The public investment in health information for statistics and program management is substantial, by both Federal agencies and state governments. To more fully realize the potential of existing survey and administrative data sources for population health statistics and to inform policy, the linkage of data sets at the person-level can be extremely beneficial. However, the linkage at the person-level of two or more data sets can increase risks to privacy and may face barriers due to varying regulatory requirements for confidentiality protection by agency and data set.

In September of 2006, NCVHS held a *Workshop on Data Linkages to Improve Health Outcomes*. The goal of this workshop was to stimulate interest and identify best practices in using linkages among administrative and survey data to improve our knowledge of health outcomes for the population and for various sub-populations. In addition, the workshop was intended to assist agencies within the U.S. Department of Health and Human Services (HHS) to better meet their responsibilities for performance measurement under various laws, such as the Government Performance Results Act of 1993, by providing more comprehensive information on the status of persons participating in government programs. This letter describes our key findings and recommendations. A summary of the workshop is posted on the NCVHS website: [www.ncvhs.hhs.gov](http://www.ncvhs.hhs.gov).

HHS health agencies, other Federal agencies outside of HHS that collect information valuable in the interpretation of health statistics, and researchers outside the government using linked data sets presented a range of examples that have proven useful to federal programs and policymaking.

An important example is the recurrent finding that survey estimates of Medicaid enrollment are well below administrative estimates. The Census Bureau's Medicaid Undercount Project is linking survey estimates of Medicaid enrollment on the Current Population Survey and National Health Interview Survey with administrative enrollment estimates, including data from some state record systems, in order to evaluate the extent to which Medicaid enrollment is undercounted on

surveys and what improvements are possible. The Office of the Assistant Secretary for Planning and Evaluation, the National Center for Health Statistics (NCHS), the Robert Wood Johnson Foundation, several states, and the State Health Access Data Center at the University of Minnesota, School of Public Health are collaborating on this project.

Linking data sets (e.g., health survey and Medicare claims data) increases the detailed person-level information available and makes it more likely that some survey respondents could be identified and privacy compromised. This is a serious concern. As a result, public use versions of linked data sets may be limited or precluded and additional protections of confidentiality are required to safeguard privacy. The approach being used by both NCHS and the Census Bureau to provide access while ensuring confidentiality is the data center. NCHS operates one data center in Hyattsville, MD and recently negotiated an agreement with the Census Bureau to make NCHS data available through Census' nine data centers throughout the nation. NCHS also has an automated remote access system whereby users can submit their data runs through an e-mail based system and receive output that has been found not to present a confidentiality risk. NCHS is currently developing an expanded remote access system with greater functionality and improved protections. The increased number of data centers and improved remote access will begin to address concerns expressed by users concerning financial and geographic barriers to accessing data resulting from the requirement to be present at the data center.

The Committee learned, however, that the approval process to use either the data center or remote access can be very time consuming, sometimes requiring up to a year or more. Such delays can seriously delay or prevent timely policy analyses using linked data. Even federal employees needing to use these data for policy analysis are affected by these delays. Participants in the workshop recommended the project approval process be streamlined to minimize delays, both for analysts in government agencies and for outside users. In addition, there was significant support for developing other methods for sharing data besides using a data center either on-site or remotely.

The NCVHS wishes to commend the NCHS for expanding access to its data. The agreement with the Census and the development of its remote access system are both recognized as positive steps to increase geographic access. To further improve access, NCVHS recommends:

- NCHS should streamline the project approval process and add staff as needed to minimize access delays for data center users.

Similar to NCHS, other agencies within the Department operate data centers including the Agency for Healthcare Research and Quality, the Substance Abuse and Mental Health Services Administration and the Centers for Medicare and Medicaid Services. NCVHS also recommends:

- Agencies within the Department in addition to NCHS that sponsor research data centers should review their data access policies and procedures to promote greater and more effective use.

Traditional use of data centers has required the researcher or policy analyst to relocate to the site of the data center and submit in-person data requests and obtain statistical outputs. The center's

staff reviews data requests and statistical outputs to ensure protection of confidentiality. Progress has been made in developing remote access systems, as exemplified by NCHS. Remote access to data centers can greatly facilitate use of health data sets. NCVHS recommends:

- Agencies that sponsor data centers should take full advantage of remote access technologies that support electronic submission of data requests and return of statistical outputs.

Government analysts and outside researchers have frequently identified potentially valuable applications from linking data sets that are not currently available through data centers. Workshop participants cited significant and extensive barriers that must be addressed before some of these linked datasets could be made available to the policy and research communities. A multiplicity of laws including the Privacy Act, the Government Information Security Reform Act, the Federal Information Security Management Act, the Confidential Information Protection and Statistical Efficiency Act and the Freedom of Information Act protect Federal data. For example, Census data can only be used for the limited purposes allowed by Title 13, namely for purposes consistent with the Census Bureau's mission. Different data sets may have unique and different requirements for access making it much more difficult to create and share a linked data set. Workshop participants recognized that changes in law may be necessary to foster broader usage of linked data systems. The consideration of such changes was beyond the scope of the workshop.

Even when there are no legal barriers to data linkages, workshop participants estimated that it takes about 18 months to obtain permission to transfer data from one federal agency to another, even within DHHS. However where there are no existing legal barriers, standardization of interagency agreements was seen as having the potential to substantially reduce the time currently required to link data across agencies. The NCVHS recommends the following:

- Standardized data linkage and sharing agreements should be developed to facilitate efficient sharing of data among HHS agencies, and to the extent possible with other government agencies.

In summary, better integration and use of health data offers the promise of improved health for millions of Americans. Information is a powerful tool in the battle against disease, epidemics, and inadequate preventive care. It is also critical in achieving a more efficient and productive health care system. Having made the initial investment in gathering data—either through the time and effort involved in completing administrative records or the costs associated with surveys—greater returns on those investments can be realized by investing comparatively modest sums to make full use of those data. In particular, connecting many health data sets to other useful information offers considerable potential to understand and improve health conditions.

These recommendations are offered by the Committee to better realize the full potential of health information for improving the public's health. The Committee will monitor progress in these areas and requests periodic progress reports from the Department. The Committee will continue to work to identify ways to improve the utility and benefit of health information.

Sincerely,

/s/

Simon Cohn, M.D., M.P.H.  
Chairman, National Committee on Vital and Health  
Statistics

Cc: HHS Data Council Co-Chairs

