Public Health Surveillance Data: Legal, Policy, Ethical, Regulatory, and Practical Issues

Amy B. Bernstein, ScD[1]
Marie Haring Sweeney, PhD[2]
[1] *National Center for Health Statistics, CDC*
[2] *Division of Surveillance, Hazard Evaluations & Field Studies, National Institute for Occupational Safety and Health, CDC*

Corresponding author: Amy B Bernstein, ScD, Office of Analysis and Epidemiology, National Center for Health Statistics, CDC, 3311 Toledo Road, Room 6214, Hyattsville, MD 20782. Telephone: 301-458-4700; Fax: 301-458-4031; E-mail: abernstein@cdc.gov.

In the United States, data systems are created by the ongoing, systematic collection of health, demographic, and other information through federally funded national surveys, vital statistics, public and private administrative and claims data, regulatory data, and medical records data. Certain data systems are designed to support public health surveillance and have used well-defined protocols and standard analytic methods for assessing specific health outcomes, exposures, or other endpoints. However, other data systems have been designed for a different purpose but can be used by public health programs for surveillance. Several public health surveillance programs rely substantially on others' data systems. An example of data used for surveillance purposes but collected for another reason is vital statistics data. CDC's National Center for Health Statistics (NCHS) purchases, aggregates, and disseminates vital statistics (birth and death rates) that are collected at the state level. These data are used to understand disease burden, monitor trends, and guide public health action. Administrative data also can be used for surveillance purposes (e.g., Medicare and Social Security Disability data that have been linked to survey data to monitor changes in health and health-care use over time).

Some data can be released easily to others with few or no restrictions. These include public use data sets and some regulatory or administrative data; however, these files were restricted at some point but were altered to protect respondent confidentiality and privacy. Public use data sets can be shared with everyone because they will not contain personally identifiable information (PII) and have had information removed that would allow identification of any persons. Some data cannot be released to anyone under almost any circumstances because they are highly sensitive and considered classified for security purposes. Data that allow identification of persons, either collected by surveillance programs or by other programs, can only be shared if regulation or legislation allows. PII usually is needed to identify information that allows these data to be linked to other data sets or to identify persons with a specific health condition or disease. In all but the most unusual circumstances at the level of data collection, identifiable data are maintained where public health surveillance interventions occur, usually at the local or state level.

Collaborative efforts to meet the needs of public health surveillance programs and other initiatives, programs, or objectives (e.g., information on payment, increased use of medical records, or evaluation of effectiveness of treatment) can maximize the utility of data collected. Information from disparate sources or programs often shed light on patterns that individual program data cannot. Furthermore, appropriate use of data sets collected for multiple purposes can, in some instances, be more cost effective than the collection of new data targeted at a specific condition or health event. The ability of data stewards to share with surveillance or other programs depends on several factors: 1) the rules and regulations governing how and why the data are collected and released, 2) the availability of resources to put the data into a form that can be shared, and 3) the willingness to use those resources. One method of distributing previously restricted data is to determine how to make the data unrestricted (e.g., by perturbing the data or releasing pretabulated, aggregated estimates that preserve confidentiality).

This report proposes a vision for improving access to and sharing of data useful for public health surveillance, identifies challenges and opportunities, and suggests approaches to attain the vision. This topic was identified by CDC leadership as one of six major concerns that must be addressed by the public health community to advance public health surveillance in the 21st century. The six topics were discussed by CDC workgroups that were convened as part of the 2009 Surveillance Consultation to advance public health surveillance to meet continuing and new challenges (*1*). This report is based on workgroup discussions and is intended to continue the conversations with the public health community for a shared vision for public health surveillance in the 21st century.

Vision

All data potentially relevant to public health surveillance would be harmonized across data systems, interoperable, and easily accessed by the maximum number of users in as timely a manner while protecting confidentiality and privacy of respondents.

Challenges

Constraints on data sharing of nonpublic-use (i.e., restricted) data exist. Occasionally, data stewards are reluctant to release data to others because they fear misuse of the data by those who are not well acquainted with its legal and technical limitations on use. In other cases, data stewards are not willing to share data either for political or historical reasons or because they fear that if someone else has access to their data their program's importance or visibility might be reduced. However, there are methods that can help protect against the identification of persons. For example, data perturbation is a data security technique that allows users to ascertain key summary information about the data while preventing a security breach.

Legal, Regulatory, and Ethical Limitations

All governmental data collection and release activities are governed by rules, regulations, and legislative authorizations. These include authorizing legislation. For example, Section 308 (d) of the Public Health Service Act (*2*) limits the release of the sensitive surveillance data that are either identifiable or potentially identifiable for any purpose other than the purpose for which it was supplied. The Health Insurance Portability and Accountability Act of (HIPAA) Privacy Rule (*3*) regulates the use and disclosure of individually identifiable information by health plans, providers, and other covered entities. The majority of government organizations have their own internal confidentiality restrictions on data they release. These might be more (but not less) stringent than those imposed by federal legislation and regulations. In other situations, an agency or program receives funds to collect and/or analyze specific data when it would be more efficient, or effective, for another program to do so. The funding streams and mechanisms affect how data are collected. For some private enterprises, data might not be released because they are proprietary (e.g., they might require other users to purchase the data and they are not bound by any rules or impetus requiring them to release data if they do not wish to do so).

In addition, ethical constraints not specified in legislation and regulation must be considered. Uses of data beyond those for disease-monitoring purposes should be ethically justified and meet some minimal standard for the data to be shared (*4*). For example, public health surveillance data are collected without patient consent and contain sensitive information. Even if law or regulation allows these data to be shared, uses other than those for which they are specifically collected should be considered carefully when sharing data with others. For data collection processes in which respondents have signed or signified that they consent to have their data collected, analyzed, and released, data can only be used for purposes that the respondents agreed to when consenting to provide data.

Administrative Barriers

Administrative and regulatory requirements of federal, state, and local governments can limit data sharing. Security concerns and regulations, multimode displays (e.g., displaying data both in hard copy and web-based formats), and required use of specific software for data dissemination can affect timeliness and the ability to release data. These requirements can secure data and computer systems and ensure patient, enrollee, or respondent privacy and confidentiality. However, substantial programmatic resources and financial and personnel support are necessary to implement these mandates.

Resources are often used heavily in the front-end planning, data collection, and analytic phases of public health surveillance with proportionately less focus on data dissemination and translation phases. This could be related to insufficient resources that often make data sharing and investment in data sharing enhancements a lower priority than program work. Scarce resources also might make competition for funding contentious, which can result in lack of attention to relationship building at the highest levels that, if remedied, could facilitate future data-sharing arrangements.

Processing data can result in delays for their release. Certain data collection programs do expend substantial resources on data cleaning and presentation and believe that data must be cleaned thoroughly and manipulated before they can be released and interpreted correctly by users. However, by the time data can be released, the value to public health surveillance programs might be limited if rapid response to a problem is necessary (e.g., to prevent spread of an infectious disease). On the other hand, programs that address chronic health conditions that develop slowly over time can benefit from use of data with longer, but specified, release delays.

However, by the time data can be released, they are of limited value to the public health surveillance programs that need data that are as current as possible. This is sometimes the case for

programs that want to use the data for outbreak surveillance or evaluation of new policies and programs. However, other programs can use data with longer, but specified, release delays.

Existing funding mechanisms for surveillance activities can impede the ability of federal agencies to negotiate data sharing and hinder their ability to influence how data collected can be shared with others. For example, the use of cooperative agreements can limit the ability of federal agencies to require data sharing because the funding agency might have less control over data products than when the funding mechanism is a contract that can directly specify the delivery and form of data release.

Data Incompatibilities

Data sharing can be impeded if coding, formatting, definitions, and methods differ substantially or if data are stored in incompatible formats. Resources are needed to manipulate, code, and transmit data to partners. Also, some analysis of data (e.g., analysis of trends) could be affected over time by changes in data collection, methods, and coding. These caveats often are not documented.

Data sharing can be limited by the lack of user-friendly data dissemination tools or adequate and detailed documentation and distribution**.** If data descriptions are not available, well-documented, and advertised, detailed data from federal data systems are much less likely to be used by others, including surveillance programs, to meet their specific data needs.

Data Sharing Guidance

Although policies on data sharing exist in federal and other governmental agencies, a lack of standard language and processes related to data sharing across federal programs exists, with perhaps even less standardization at state and local levels. Efforts to standardize data sharing methods have been attempted throughout the U.S. Department of Health and Human Services but have not been realized in several instances. To date, guidance by the research and policy community on matters related to data policies and procedures at the national, state, and local levels has been inconsistent.

Federal, state, and local governments have produced guidance documents on dissemination and sharing of data. For example, CDC and the Council of State and Territorial Epidemiologists developed the *CDC/ATSDR Policy on Releasing and Sharing Data* (*5*). These documents ensure that CDC routinely provides data to its partners for appropriate public health purposes while balancing privacy concerns, federal and state confidentiality concerns, proprietary interests, national security interests, or law enforcement activities (*6*). However, certain data stewards or potential data users might not be familiar with these documents. Similarly, experts and resources on how to create agreements with other agencies and nongovernmental organizations exists but are not easily located or shared (*7*).

Opportunities

Even with legal and regulatory restrictions on release and use of data sets that can be used for public health surveillance, mechanisms have been created that facilitate surveillance and other programs' ability to share data or to use other programs' data. In many cases, the possibility exists to either 1) deidentify the data, 2) obtain a subset of restricted data that complies with regulations concerning release (e.g., a perturbed data set in which data are changed before the dissemination in such a way that the disclosure risk for the confidential data is decreased but the information content is retained as far as possible, or one with small cells suppressed), or 3) develop agreements whereby data are released to others who need it for public health surveillance but who agree not to identify or contact any persons. In some cases, with sufficient cooperation or collaboration between surveillance program and data steward, surveillance programs are able to obtain at least a subset of data that meets their needs. Providing feedback or other incentives to the data stewards can encourage data sharing.

Several federal projects have been conducted successfully that share restricted data with other agencies and nongovernmental organizations. For example, CDC has an ongoing relation with the Social Security Administration and the Centers for Medicare and Medicaid Services to link their data to health-survey data and vital statistics data. Lessons learned in the process of negotiating and implementing these interagency agreements could be useful to others interested in sharing data with these agencies (*6*). However, other federal, state, and local agencies and health departments have failed to obtain access to desired administrative or survey data. Interagency agreements specify the conditions that must be met for government entities to share data and allow data to be shared under specific conditions and constraints. As more data-sharing agreements are realized, important lessons in how to share data have emerged, such as how to write data sharing agreements, how to transmit and receive data securely, and how to release shared data without violating confidentiality and privacy of respondents. Past collaborations have been the basis for model data-sharing agreements. Personnel participating in these data-sharing arrangements also can be a valuable resource for new initiatives.

Another example is the collaborative review process developed between NCHS and the National Association for Public Health Statistics and Information Systems (representing the states and territories) for review and approval of data requests involving release of restricted vital statistics files to researchers. This review by state representatives is conducted before the NCHS review and includes both federal and nonfederal requests for restricted data files for research purposes, enabling the state data owners to share oversight with NCHS in the dissemination of the restricted data.

Researchers can gain access to some restricted data by using the NCHS Research Data Center (RDC) (available from http://www.cdc.gov/rdc/), which facilitates access to detailed data files in a secure environment without jeopardizing the privacy of respondents or confidentiality of the data. Resources are necessary to operate the RDCs and to prepare the data for use within the RDCs. Numerous federal agencies, including NCHS, AHRQ, and the U.S. Census Bureau, are sponsoring RDCs through which they can allow use of confidential data. Multiple data sets can be combined in these RDC settings.

Opportunities are available for collecting new information using existing data collection systems. For example, certain federally funded, annually conducted surveillance surveys invite interested parties to propose new data elements or topic-specific modules. States and partners can add new modules to the Behavioral Risk Factor Surveillance System (BRFSS) questionnaires after review and permission of state BRFSS coordinators; questions have been added to the BRFSS to monitor new or ongoing program or policy initiatives.

Collection of informed consent at the time of data collection to permit less restricted use of respondents' data can enhance data use. If respondents do not consent to have their data used for specific purposes, including data sharing, when they are first interviewed or when their data are collected, obtaining consent at a later date is difficult.

The Office of Management and Budget mandates the use of specific questions for selected variables (e.g., race, ethnicity, and sex). This is a first step in promoting standards for data that can be used in public health surveillance. The next steps are to standardize data formats and data elements, codes, and methods across programs to meet the needs of both data collectors and surveillance programs. However, some flexibility must be maintained to ensure collection of the most accurate and appropriate data to meet the goals of the surveillance systems. With increasing emphasis on electronic data-standards development, opportunities are created to develop public/private partnerships that benefit all partners and enhance data collection and use for public health surveillance.

With the evolution of new technologies, database managers are able increasingly to share information and provide ready access (e.g., user-friendly, web-based query systems) to their program-specific data. Potential users can then familiarize themselves with the data. This allows potential data users to determine how the data can meet their specific program objectives and then pursue data-sharing arrangements for restricted data sets. This approach has been successful for sharing violent death data from the National Violent Death Reporting System. Public use data are available on the Web-based Injury Statistics Query and Reporting System (WISQARS) (available at http://www.cdc.gov/injury/wisqars) and a system is in place to request a restricted-access, detailed NVDRS data file. States and localities also have new and innovative data extraction tools, including Utah's IBIS system (http://ibis.health.utah.gov/home/ContactInformation.html), Missouri's MICA portals (http://www.dhss.mo.gov/DataAndStatisticalReports/index.html), and Boston's Health Indicators Project (http://www.preventioninstitute.org/component/jlibrary/article/id-275/127.html). One recent innovation is the HHS-wide collaboration to produce the Health Indicators Warehouse (HIW), which stores preconstructed indicator data at the national, state, hospital referral region, and county level. HIW is designed to allow users to easily download these indicators to be used for their own applications and applies suppression rules that allow previously restricted data to be accessed easily at subnational levels.

Conclusion

Improving data sharing to allow more and better data to be used to monitor the public's health is in everyone's interest. As the lead public health surveillance agency in the Federal government, the *CDC/ATSDR Policy on Releasing and Sharing Data* states: "CDC believes that public health and scientific advancement are best served when data are released to, or shared with, other public health agencies, academic researchers, and appropriate private researchers in an open, timely, and appropriate way. The interests of the public, which include timely releases of data for further analysis, transcends whatever claim scientists may believe they have to ownership of data acquired or generated using federal funds. Such data are, in fact, owned by the federal government and thus belong to the citizens of the United States" (*6*). To meet this policy, organizations that conduct public health surveillance and collect surveillance-related data should provide leadership, expertise, and service and devote sufficient resources to nurturing new data-sharing arrangements and to support existing ones. The goal is to have guidance on data release and sharing that balances the desire to disseminate data as broadly as possible with the need to maintain high standards and

protect individuals' privacy and the confidentiality of the data (*8*). Specifically, data-use agreements should be shared widely to provide models for others interested in sharing data; data sharing should be promoted by developing supportive funding mechanisms, devoting resources, fostering partnerships and centralizing support; and methods and procedures should be standardized across datasets.

References

1. CDC. Introduction. In: Challenges and opportunities in public health surveillance: a CDC perspective. MMWR 2012;61(Suppl; July 27, 2012):1-2.
2. Public Health Service Act. 30.8(d) (42 U.S.C. 242 [m]). Available at http://www.law.cornell.edu/uscode/html/uscode42/usc_sec_42_00000242-m000-html.
3. HIPAA Privacy Rule and Public Health Guidance from CDC and the U.S. Department of Health and Human Services. Available at: http://www.cdc.gov/privacyrule/Guidance/PRmmwrguidance.pdf.
4. Lee LM, Gostin LO. Ethical collection, storage, and use of public health data: A proposal for a national privacy protection. JAMA 2009;302:82–4.
5. CDC-CSTE Intergovernmental Data Release Guidelines Working Group Report. CDC-ATSDR-data release guidelines and procedures for re-release of state-provided data (January, 2005). Available at http://www.cdc.gov/od/foia/policies/drgwg.pdf.
6. CDC/ATSDR policy on releasing and sharing data. Manual GUIDE: General Administration, CDC-102. Available at http://www.cdc.gov/od/foia/policies/sharing.htm.
7. Prell M, Bradsher-Fredrick H, Comisarow C, et al. Profiles in success of statistical uses of administrative data. Available at http://familymedicine.medschool.ucsf.edu/fhop/docs/pdf/prods/ad_apx6.pdf.
8. The Privacy Act of 1974 5 U.S.C. § 552a. Available at: http://www.justice.gov/opcl/privstat.htm.

Centers for Disease Control and Prevention
MMWR
Morbidity and Mortality Weekly Report
Supplement / Vol. 61
July 27, 2012

CDC's Vision for Public Health Surveillance in the 21st Century

U.S. Department of Health and Human Services
Centers for Disease Control and Prevention

CONTENTS

The *MMWR* series of publications is published by the Office of Surveillance, Epidemiology, and Laboratory Services, Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services, Atlanta, GA 30333.

Suggested Citation: Centers for Disease Control and Prevention. [Title]. MMWR 2012;61(Suppl; July 27, 2012):[inclusive page numbers].