PUBLISHED BY THE INTERNATIONAL SOCIETY FOR DISEASE SURVEILLANCE

Advances in Disease Surveillance

VOLUME 7
NUMBER 2
DATE 2009

Expert Meeting on Privacy, Confidentiality, and Other Legal and Ethical Issues in Syndromic Surveillance

Report from an International Society for Disease Surveillance Consultation Washington DC, October 4–5, 2007

Michael A. Stoto,[1] James X. Dempsey,[2] Atar Baer,[3] Christopher Cassa,[4] P. Joseph Gibson,[5] and James W. Buehler[6]

[1] Department of Health Systems Administration, School of Nursing and Health Studies, Georgetown University, Washington DC.
[2] Center for Democracy and Technology, Washington DC.
[3] Public Health, Seattle and King County, Seattle, WA.
[4] Massachusetts Institute of Technology, Cambridge, MA.
[5] Health and Hospital Corporation of Marion County, Indianapolis, IN.
[6] Epidemiology Department and Center for Public Health Preparedness and Research, Rollins School of Public Health, Emory University, Atlanta, GA.

For syndromic and related public health surveillance systems to be effective, health departments need access to a variety of types of health data. Since the development and implementation of syndromic surveillance systems in recent years, health departments' experience in gaining access to personal health information for syndromic surveillance has been mixed. Although the HIPAA Privacy Rule permits health care providers to disclose protected health information without patients' consent to public health agencies for authorized purposes, some health care providers have cited HIPAA in refusing to provide data for syndromic surveillance. Beyond HIPAA, a variety of federal, state, and local public health laws enable, restrict, and otherwise influence the sharing of health information between health care providers and public health agencies for surveillance, as well as research, purposes. To address these issues in the context of syndromic surveillance practice, an expert meeting was convened to (a) share experiences regarding privacy, confidentiality, and other legal and ethical issues; (b) clarify how these legal and ethical issues enable or constrain data sharing; and (c) identify approaches to protect privacy and confidentiality.

Rooted in the principle that state and local governments have inherent and broad powers to protect the health, safety and welfare of the people, public health agencies have the authority to collect personal health information, including the power to compel disclosure of such data under certain conditions. The precise limits of these public health powers have never been defined, however, and are limited by and must be balanced with the right of privacy. Although existing laws and regulations provide little clarity on how to balance disclosure risks and potential benefits of public health actions that reasonably follow from surveillance, guidance can be found in the principles known as the "Fair Information Practices." Two specific fair information practices—specification of purpose and limitation on secondary use—are especially critical for syndromic surveillance.

Considering these principles should help health officials to determine what level of health information detail, and thus what level of potential ability to identify individuals, is needed for the intended public health purpose, and whether surveillance will lead to effective public health action. The principles also suggest a number of strategies for dealing with concerns that the public or health care providers may have about sharing data with public health agencies: (a) improve communication with the public about how data are used to safeguard the

population's health and how confidentiality is protected; (b) further develop approaches to sharing data in aggregate or de-identified form that have capability to link back to identified data when necessary; (c) further develop statistical approaches to anonymization and aggregation; and (d) conduct evaluation research and case studies that demonstrate utility and clarify how privacy and confidentiality are protected.

INTRODUCTION

For syndromic and related public health surveillance systems to be effective, state and local health departments and the Centers for Disease Control and Prevention (CDC) need access to a variety of types of health data. Since the development and implementation of syndromic surveillance systems in recent years, health departments have gained varied levels of access to personal health information for inclusion in these systems. A variety of federal, state, and local laws enable, restrict, and otherwise influence the sharing of health information between health care providers and public health agencies for surveillance, as well as research, purposes. Some health care providers have expressed reluctance or refused to provide identifiable data for syndromic surveillance to health departments (1), citing state privacy laws or the Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule (2). Although the HIPAA Privacy Rule permits health care providers to disclose protected health information without patients' consent to public health agencies for authorized purposes, it does not supersede state laws that provide greater protection of individual privacy (2,3).

The use of individuals' health information for syndromic surveillance poses challenging questions regarding the interpretation and future development of ethical and legal standards for public health practice and research. While the practice of syndromic surveillance extends the longstanding tradition of public health surveillance as an essential element of public health practice (4), it raises in a new light equally longstanding questions about governments' authority to collect and use health information (5). As the practice of syndromic surveillance evolves, it is in the national interest to clarify the conditions under which health information can be shared, the ways that privacy and confidentiality can be protected, and the ways that local, state, and federal public health agencies can legally, ethically, and effectively exercise their respective responsibilities to detect, monitor, and respond to public health threats.

CONSULTATION OBJECTIVES AND PROCESS

To address these issues, Georgetown University's O'Neill Institute for National and Global Health Law, with support from the International Society for Disease Surveillance (ISDS), convened 17 experts in syndromic surveillance and related areas in Washington, DC, on October 4–5, 2007. The meeting focused on the development, use, and evaluation of syndromic surveillance. Its objectives were (a) to share experiences regarding privacy, confidentiality, and other legal and ethical issues; (b) to clarify how these legal and ethical issues enable or constrain data sharing; and (c) to identify approaches to protect privacy and confidentiality. The participants are listed in Appendix I, and the consultation agenda appears in Appendix II. This report summarizes the proceedings of the consultation. The participants had an opportunity to comment on earlier drafts, but do not necessarily agree with every conclusion in this report. Moreover, this report does not represent the official position of the participants' agencies or organizations or the ISDS.

BACKGROUND AND ASSUMPTIONS

For the purpose of the consultation, the participants chose to define "syndromic surveillance" broadly to include near real-time acquisition and use of nonclinical or prediagnostic health data for population health monitoring purposes. In this sense, *prediagnostic* refers to information on symptoms and stages of a patient's illness before a diagnosis is established. Because these systems emphasize timeliness between the health-related events and when trends in these events can be monitored, they typically use data that are routinely collected and stored electronically and automate the process of data collection, management, and analysis to detect unusual patterns such as unexpected numbers of cases.

Participants held a range of views on the effectiveness of syndromic surveillance for achieving different goals. It was not the purpose of the consultation to address, let alone resolve, questions of effectiveness. However, the legal justification for different methods of surveillance is rooted in the presumption that the information obtained will lead to effective public health action, so questions of effectiveness are relevant. In this context, participants noted that the purpose of syndromic surveillance in public health practice is shifting from an emphasis on the early detection of bioterrorism or other infectious disease outbreaks to other uses, including identification of potential cases of notifiable disease, surveillance for noninfectious conditions such as chronic diseases or injuries, and "situational awareness." Such uses include following up potential cases when aberrant trends are detected, monitoring the course of outbreaks, monitoring seasonal illnesses such as influenza and viral gastroenteritis, and preparing for surveillance of pandemic influenza.

It is important to note that the practical and legal justification for surveillance is based on public health agencies'

responsibility to monitor and to take action when necessary to protect the community's health. Such actions comprise a spectrum of public health activities, ranging from individual case follow-up to program evaluation and policy development. These varying uses of surveillance result in varying levels of need for detailed information, for personal identifiers, and for timely data. For example, in some instances, data may be needed daily to detect infectious disease outbreaks and respond in a timely way. In other instances, annual data may be sufficient to guide programs, allocate resources, or set policies.

Effective use of syndromic data sometimes requires that public health epidemiologists have some capacity to "drill down" or examine the data in more detail when aberrant trends are detected. This need derives from an important attribute of syndromic surveillance: in order to assure that systems have sufficient sensitivity to detect events of public health significance, the systems inherently generate a sizeable number of statistical "false alarms" due to random variation in the data. Epidemiologists are able to dismiss some of these alerts by interpreting them in the context of their knowledge and experience, including reviewing the syndromic data in greater depth or considering the data in the context of other information. However, when these analytic methods are insufficient and the initial assessment of the alert raises sufficient concern, it may be necessary to identify individuals in order to conduct further investigations.

Health departments represented at the meeting receive patient-level data from local hospitals in a wide variety of ways: (a) without any patient identifiers; (b) with the hospitals' medical record number; (c) with a code generated specifically for such reporting; and (d) with a combination of these methods. Participants noted that in some public health departments epidemiologists have direct access to hospital information systems, particularly if the hospital is managed by a public health agency, enabling them to conduct preliminary follow-up assessments of alerts without troubling hospital staff. In one jurisdiction, data are shared through a preexisting Regional Health Information Organization (RHIO). In another, complete identified data are shared for patients with certain pre-identified conditions, and the list of conditions can be expanded through a formal approval process. For those hospitals that provide detailed clinical and diagnostic data to the CDC's BioSense program, CDC has the capacity to conduct detailed follow-back assessments with data maintained at CDC (6,7).

In practice, a variety of concerns can discourage or even prevent the sharing of data for syndromic surveillance purposes. Most prominent among them are concerns about patient privacy and confidentiality that are reflected in various federal, state, and local laws and regulations and in the HIPAA Privacy Rule (2,3), and in varying interpretations of these laws and regulations, by health care providers and public health agencies. CDC technical guidance and stipulations attached to the funding also shape health departments' privacy and confidentiality practices. In addition, data owners may be willing to share data with their local or state health department but may have concerns about potential subsequent access and use by others, including researchers, agencies aside from public health departments, or the federal government. The personnel or other operational costs of sharing data, as well as proprietary concerns and turf issues, may also discourage data owners from sharing information with public health agencies.

Local health departments also express concern about information overload, including the large numbers of false positives alerts that nevertheless require preliminary investigations, and about their ability to interpret surveillance reports for higher-level authorities. The sense that the effectiveness of syndromic surveillance is not yet proven, especially for early detection of bioterrorist attacks, can also discourage sharing data.

LEGAL AUTHORITIES FOR SYNDROMIC SURVEILLANCE

At its core, public health surveillance stands on firm legal footing, rooted in the principle that the state governments have inherent and broad sovereign power (commonly called the "police power") to provide for the health, safety, and welfare of the people. It has long been recognized that this power includes the authority to collect personal information, including the power to compel disclosure of health data for sufficiently justifiable purposes. The precise limits of the police power have never been defined. Specifically with respect to information collection and disclosure, the police power is limited by constitutional protection of the right to privacy in one's personal medical information. The scope of this right to information privacy also remains unsettled. Typically, the purpose of a data collection program is balanced against the loss of privacy it may entail. Though the U.S. Supreme Court has upheld reporting requirements generally (7), the absence of specific guidance from the Court makes it difficult to determine how to strike the balance between the potential benefits of public health actions and risks to privacy in public health surveillance systems like syndromic surveillance. In these circumstances, public health officials must use their authority judiciously in order to maintain public support and ultimately legislative approval to retain this authority (8,9).

Compounding the questions inherent in the concept of privacy and its codification in the HIPAA Rule, public health officials are dealing with rapid changes in the nature of health care and public health itself. First, there is a trend toward the digitization of medical records and the ability to transfer large quantities of data by electronic means and to analyze them by sophisticated statistical techniques, creating the opportunity for public health to access more data, in

greater detail. These capabilities may generate both greater public concerns about privacy and greater public expectations about the ability of health departments to automatically detect unusual disease trends and monitor population health status.

A second trend is the growing public health emphasis on an expanding array of noninfectious conditions, such as chronic diseases, disabilities, injuries, or reproductive health outcomes and their behavioral or environmental antecedents. Surveillance for noninfectious conditions has long been part of modern public health practice, but the growing emphasis on health promotion and prevention of a large spectrum of diseases has expanded the scope of what data are of value to public health programs and policies (10).

A third trend is the increasing role of the federal government in public health. Unlike the states, the federal government does not have general police powers, and its authority to collect public health data rests on less comprehensive grounds, such as national defense, immigration, regulation of interstate commerce, the ability to attach conditions to federal spending, and border control. However, in recent years, particularly since 9/11, heightened concerns with bioterrorism and naturally emerging microbial threats have accelerated the expansion of federally managed and federally funded public health activity. Some of these initiatives involve the transfer of large volumes of data to the federal government, bypassing the traditional role of local and state health departments in managing the collection of surveillance data and transferring it to a federal database for analysis and possible action. Notably, CDC's BioSense program involves the transfer of large volumes of data directly from participating hospitals to the federal government, bypassing the traditional role of local or state health departments in managing the collection of surveillance data, and transferring it to a federal database for analysis and possible action (6).*

A wide variety of federal, state, and local laws and regulations authorize and regulate public health surveillance activities as well as seek to protect privacy and confidentiality (8). In addition to the federal constitutional right to privacy, there are state common-law rules, statutes, and some state constitutional provisions that protect rights of privacy in medical information. The laws of many states generally forbid physicians and other care providers from disclosing a patient's identifiable information without the patient's consent. For this reason, specific state legislation may be required to authorize state public health agencies to collect and use identifiable personal information for public health purposes. As noted earlier, the constitutionality of these laws depends upon whether the state's reason for collecting the data outweighs individuals' reasonable expectations of privacy in the information. Many of these laws date from long before the introduction of electronic data systems or growing federal involvement in public health surveillance; it is unclear whether the legislatures intended the kinds of data sharing contemplated by today's programs. Indeed, syndromic surveillance per se is the subject of only a few very recent and specific laws that specifically authorize or mandate participation in local or state syndromic surveillance networks. Elsewhere, public health agencies have adopted regulations or initiated syndromic surveillance under the umbrella of existing public health reporting laws.

Specific laws setting out detailed reporting requirements for syndromic surveillance may make it easier for data owners to share data with public health authorities. However, in the absence of such specific laws—or in the political process needed to develop a reporting law for syndromic surveillance—health departments and health care providers are understandably left with a lot of questions because of differences in what is meant by "syndromic surveillance," differing federal, state and local responsibilities, and confusion about the HIPAA Privacy Rule. As a result, the current legal framework for syndromic surveillance does not provide clear guidance to public health practitioners. There is a need for a rational framework that can guide policymakers and public health professionals as they necessarily and appropriately seek to balance disclosure risks against potential benefits of public health actions that reasonably follow from surveillance and against competing public health priorities and broader societal interests.

The HIPAA Privacy Rule and Syndromic Surveillance

Although a variety of laws and regulations both enable and restrict public health surveillance, health officials, policy makers, and others typically focus their attention on the HIPAA Privacy Rule (11). This rule generally governs the disclosure of personally identifiable health information by hospitals and other health care providers (called "covered entities"). HIPAA provides an important addition to the laws governing health privacy in the United States, but it does not supersede (preempt) applicable state laws that provide greater privacy protection. The HIPAA Privacy Rule has also been the subject of considerable confusion and controversy. The complexity of the rule and the ambiguities surrounding key terms often lead to differences of opinion about what is permissible and what is not.

A key attribute of the HIPAA Privacy Rule is that it applies only to disclosures of "individually identifiable health information." While there is debate about what is "identifiable,"

*While the BioSense program still includes hospitals that report detailed clinical data directly to CDC, efforts to recruit additional hospitals to report in this manner has been superseded by a strategy to obtain less detailed information from existing local or state syndromic surveillance networks, restoring the traditional pathway for health information flow to CDC, and to a longer-term strategy that extends the use of decentralized surveillance methods (6).

the rule clearly contemplates disclosures of de-identified data without patient authorization. However, because surveillance parameters such as the date of health care visits, the Zip code or country of residence are considered as "identifiers" by HIPAA, there is a wide range of syndromic surveillance functions that could not be exempt from HIPAA on this basis alone.

In the public health sphere, controversy has centered on the rule's exception permitting disclosure without patient authorization of individually identifiable data for "public health activities and purposes." Specifically, the HIPAA rule allows a "covered entity" such as a hospital to disclose protected information to

> (i) A public health authority that is authorized by law to collect or receive such information for the purpose of preventing or controlling disease, injury, or disability, including, but not limited to, the reporting of disease, injury, vital events such as birth or death, and the conduct of public health surveillance, public health investigations, and public health interventions; or, at the direction of a public health authority, to an official of a foreign government agency that is acting in collaboration with a public health authority (12).

This exception is not limited to disclosures of so-called "notifiable diseases," the specified diseases or conditions that health care providers are required by law to report to public health authorities. Rather, disclosure can be made for patients with other conditions in circumstances where public health officials have the authority to investigate unusual cases or clusters of disease that may herald or represent a broader threat to public health.

This provision is qualified in that such release of information is limited to "the minimum necessary information to accomplish the intended public health purpose of the disclosure" (10), except for disclosures that are mandated by state or local law. In the absence of a state or local law requiring disclosure, the "minimum necessary" standard introduces some uncertainty. For purposes of syndromic surveillance, public health agencies have defined this minimum in various ways, with varying degrees of acceptance by hospitals and other health care providers. In particular, the minimum necessary determination depends on a range of factors, including the purpose and scope of the program (is it focused on bioterrorism, looking for patterns indicative of an infectious disease epidemic, or collecting data on any problem or condition that might be of public health concern?) and the possible public health actions that might be taken in response to the information (will some of these actions require public health officials to contact individual patients or ask clinicians to re-identify individual patients to obtain further information?).

In an effort to define "individually identifiable," the HIPAA Rule provides different ways to ensure that data are sufficiently stripped of identifying information so that they can be disclosed without patient authorization. The first, sometimes referred to as "the safe harbor," is to strip the data of 18 specified kinds of information that could be used to re-identify individuals (13). A second, less extreme approach (intended to make data available for research purposes) is to strip out all obvious identifiers, while leaving in some information such as Zip Code of the place of residence and age group. The resulting data are called a "limited data set," which can be disclosed if the covered entity obtains satisfactory assurance, in the form of a "data use agreement," that establishes who is permitted to use the limited data set and for what purposes, and that the recipient will not use it for other purposes and will safeguard it appropriately (11). Under either of these approaches, the Privacy Rule permits a covered entity to assign to, and retain with, the de-identified health information, a code or other means of record re-identification if that code is not derived from or related to the information about the individual and is not otherwise capable of being translated to identify the individual.

Another issue facing public health officials is whether some data collection and assessment activities should be considered research. Research at institutions receiving federal research funds must conform to federal regulations that protect human subjects in research and require institutional review board (IRB) approval. If a public health activity is research, then the question also arises as to whether the public health or research exemptions in HIPAA apply (2,11,14). This distinction is sometimes unclear or debatable (15,16), as was evident from the differing views of meeting participants. As a result, the participants did not attempt to clarify or redefine this boundary in the context of syndromic surveillance development, operation, or evaluation. However, to the extent that surveillance is conducted for research purposes rather than for authorized public health functions, it is likely to be subject to the consent requirements of applicable laws. In other words, agencies that conduct surveillance must be alert to the purpose for which they collect data, because the purpose and use of the data—research versus public health practice—largely determine what laws apply.

FAIR INFORMATION PRACTICES

In the absence of definitive court decisions or law fully specifying what data can be collected and used for different purposes, those working in syndromic surveillance can seek guidance from the principles known as the fair information practices (FIPs)." The FIPs are a widely accepted privacy framework, first developed in the 1970s by an Advisory Committee to the U.S. Department of Health, Education and Welfare, and later refined and expanded. They have been embodied in privacy or "data protection" law worldwide, serving as the basis for directives of the European Union, influential guidelines of the Organization for Economic Cooperation and Development, and an emerging privacy

framework in Asia. In the United States, the FIPs can be found in the federal Privacy Act (17), which is generally applicable to all U.S. government systems of records, and in other federal privacy laws. Most importantly, the FIPs served as the basis for the HIPAA Privacy Rule, although that rule is so complicated that it is hard to discern the framework.

There are various formulations of FIPs, but we find the following 10 to be most comprehensive and useful as a framework for considering the privacy issues raised in syndromic surveillance practice. The relevance and implications for the FIPs for syndromic surveillance are discussed in the next section.

1. **Notice (or openness).** An entity (in this case, both the entity that engages with the patient and the public health authority) should disclose when it is collecting data and specify what data it is collecting, through a published notice and wherever possible on an individual basis. Sometimes notice is combined with individual choice or consent, which may be expressed through an opt-out or an opt-in. Public health uses, however, may not be subject to a requirement of individual consent, at least not when disclosure is required by law.
2. **Purpose specification.** The notice should specify the purpose for which data is being collected. This important principle impels the data collector to think through in advance what are all the purposes for which the data will or could be used.
3. **Collection limitation.** The entity collecting data should collect no more than is relevant and necessary for the specified purpose.
4. **Retention limitation.** An entity should retain information in personally identifiable form no longer than is necessary for a specified purpose.
5. **Use and disclosure limitation.** An entity should not use or disclose information for purposes other than those specified at the time the information was collected. In other words, uses and disclosures should be tied back to the purpose specification. An entity should obtain individual consent before disclosing data for uses that depart from the purpose specification.
6. **Data quality.** An entity holding personal information is responsible for ensuring that it is timely, accurate, and complete.
7. **Security.** An entity holding personal information should take reasonable measures to protect it against loss or unauthorized destruction, disclosure or alteration.
8. **Access to one's own records.** An individual should be able to obtain a copy of his or her records. Sometimes this is more broadly defined as "individual participation" and may include the element of individual control or consent.
9. **Redress.** The subject of data should have the right to challenge inaccurate data, to correct mistakes and to obtain redress for abuse.
10. **Accountability.** Any privacy system should have a system for enforcement of the foregoing principles; there should be monitoring and auditing of information flows, and it should be clear who is responsible for enforcing the rules.

The FIPs are not absolutes, but rather serve as a framework or roadmap to navigate through the confusion about the HIPAA Rule and the challenges posed by rapid changes in public health. The FIPs can serve as the basis for a "concept of operations," clarifying what information is being collected, for what purpose, with whom will it be shared, how long will it be retained, how accurate and reliable is the information, how will the data be secured against loss or unauthorized access, and how individuals will know the basis for decisions affecting them and be able to respond to mistakes. The FIPs can also guide the drafting of a local ordinance requiring designated entities to submit certain data, and can help legislative staff and public health officials to write a "legislative history" showing why it was justified and how its implementation will be accomplished. Such a history can provide public health authorities a ground for defending policies should they be challenged in court. Likewise, the FIPs can help in drafting data use agreements, providing in essence a checklist of issues that must be addressed and helping allocate responsibilities among participants.

Considerations for Syndromic Surveillance

The relative importance of specific principles outlined in the FIPs differ in the context of information management systems that are geared to providing health care services, reimbursement, or insurance versus syndromic surveillance systems. We have outlined several considerations for the interpretation of the FIPs as they relate to syndromic surveillance.

First, syndromic surveillance is essentially a secondary use of data collected in the process of providing health care to individuals, and the FIPs must be interpreted in this context. For example, given that syndromic data are generally collected from institutions rather than individuals, the practice of "notice" must be applied in that same framework. While public health uses of syndromic data are not subject to individual consent, the notice principle could guide syndromic surveillance practitioners to educate data sources about the purpose and scope of data collection, including descriptions of instances when investigations into public health threats may prompt follow-back to individual patients, with the involvement of health care providers where appropriate. Broadly, public health data collection purposes and procedures should be transparent to the communities that public health agencies serve.

Second, it is important to recognize that, as understanding is growing about the types of events and situations where syndromic surveillance is more or less useful, the purpose

of these systems have shifted over time. Systems that were at one time designed for bioterrorism detection or as early warning systems might now be used for situational awareness. Indeed, such flexibility is a recognized desirable attribute. The FIPs relating to purpose specification and collection limitation should not be so rigid as to restrict these systems from evolving to meet the standard of practice; however, they could serve to guide practitioners to communicate with data providers when the purpose of their systems has shifted over time, and periodically reevaluate, as appropriate, the necessity of collecting certain data elements to support that specified purpose.

For infectious diseases, the following criteria can help to determine how much data to collect and how to use the data. Although no single factor is definitive, the following should be considered: (a) the extent to which the disease in question is transmissible from person to person; (b) disease severity; (c) the extent to which identification and reporting of people with the condition benefits those individuals; (d) the extent to which identification and reporting of people with the condition is effective in controlling spread in the population (eg, treatment may also lessen transmissibility); (e) the value of implementing such interventions sooner rather than later; (f) the vulnerability of the affected population; and (g) the sensitivity of the information reported (10,18). Similar considerations may shape assessments of the level of detail needed in to allow for "drill down" assessments in syndromic data following statistical alerts.

Third, the principle of "retention limitation" may have limited relevance for syndromic surveillance, if the information received by public health agencies is generally not traceable to individuals without collaboration from health care providers, where such assessment is merited and justifiable under state laws. To the extent that syndromic records maintained by health departments contain record numbers that can be used to link back to patient records, consideration should be given to the duration of time that such numbers are needed.

Similarly, the interpretation of the "use and disclosure limitation" has different implications for health care settings, where medical records serve the primary purpose of enabling individual patient care versus public health departments, which legitimately use those data for the secondary purpose of health surveillance without obtaining patient consent. This FIP cautions public health agencies against "tertiary" uses and could also help guide them in describing how they might share syndromic surveillance data with third parties, such as research investigators, for example, by providing data in aggregate form and enacting strict assurances which prevent the potential identification of individuals. This FIP could also guide health departments to communicate with, and seek permission from, data providers when considering use of the data that shifts from the original intent.

Finally, the FIPs describing "data quality," "access to one's own records" and "redress" have special considerations in the context of syndromic surveillance. "Data quality" does not mean absolute accuracy, but rather a level of adequacy likely to produce reliable outcomes in the particular context. Even though health departments have limited, if any, control over the quality of the data captured from individual patients, they should consider whether the data they are acquiring is suitably accurate for that purpose and whether there are ways to obtain more accurate data. Furthermore, any public health intervention that might result in a direct interaction with a patient as the result of a syndromic surveillance record would necessarily occur through collaboration with the data provider, who must account for its disclosures to public health unless a data use agreement is signed. In terms of access, health departments are appropriate points of contact for individual access to one's own records only to the extent that their data is reasonably accessible by individual identifiers.

DEALING WITH CONCERNS ABOUT SHARING DATA

Consideration of the FIPs suggests a number of strategies for dealing with concerns about sharing data between health care providers and with public health agencies.

First, it is important to improve communication with the public about how data are used to safeguard the population's health and how confidentiality is protected. The latter includes distinguishing between identifiable patient-level data and aggregated and processed information such as disease incidence rates and the role of different types of information in detecting, characterizing, or monitoring potential or actual outbreaks of disease in a community. It should also be noted that approaches to sharing data vary in the degree to which the data are de-identified, and that the need for identified data varies with the intended application and whether it will be used at the local, state, or federal level.

The need for individual-level data is most clear at the local level, where it allows for "drill down" assessments of statistical alerts and follow-back to individuals when required. Individual-level data also allow for substantial flexibility in analyses, justifying such data collection even when individual-level public health interventions are unlikely. Sometimes, however, aggregate data alone are sufficient to guide public health actions, particularly at the federal level. For instance, aggregate data could be analyzed for anomalies that would trigger collection of identified individual level data. At the point when there is evidence of an emerging event of concern, accessing personal information may be important for identifying and containing the threat. Alternatively, in the later stages of an influenza pandemic, when tracking individual patients is no longer feasible or has diminished value, aggregate data can be evaluated to identify the populations that are most heavily affected and to help distribute resources

accordingly, without knowing which particular patients are infected.

However, even at the state and federal levels, analyses of data to examine risk factors for disease or characterize trends, while taking into account potentially confounding factors, may require access to individual-level data (but not necessarily as many data elements, as are needed for local uses). Other multi-state or national-level uses, however, may not require individual-level data, as illustrated by the ISDS' proof-of-concept project for collating aggregate syndromic surveillance data for influenza surveillance (see http://www.syndromic.org/projects/DiSTRIBuTE.htm).

Second, public health authorities and other stakeholders should further develop approaches to sharing data in aggregate or de-identified form that have "drill-down" or "break the glass" capability when it is necessary, such as examining specific chief complaints that contribute to a statistical alert in a syndrome category or patient demographic attributes to assess whether the events that contributed to an alert were related. As described above, consultation participants described the current norm of syndromic surveillance practice that allows this by collecting individual-level data without specific identifiers or with a code or record number where the link to patients' identity is retained by participating health care providers.

Changing the process so that aggregates are created by each participating hospital could ease confidentiality concerns, but could result in substantial increases in requests from public health agencies for hospitals to conduct assessments that could otherwise be done by health department staff when statistical alerts occur, as they often do. Other options include using a regional data repository, collecting data only from individuals with certain syndromes (with a process for approval of an expanded list when necessary), and a direct connection to provider data systems. The operating policies of such systems need to specify what data are shared on a regular basis, what can be shared when there is an indication of a potential public health problem, and the conditions that would trigger expanded access. The optimal approach is to retain data where they were created, allowing access as needed. However data access is arranged, practitioners at the meeting stressed the importance of keeping its interpretation as close to the source of data as possible, in other words ensuring that those who know the data systems and local health care situations best are involved in conducting and interpreting the analyses.

Third, statistical approaches to anonymization and aggregation should be explored and further developed to mitigate privacy concerns. For example, spatial data can be highly identifying and thus pose a special confidentiality threat. Even geographical data published in low-resolution maps can reveal exact or nearly exact patient home addresses (19). While individual-level data and spatial visualizations may be needed by a local health department to conduct its epidemiologic investigations, anonymized, or even data aggregated by day and age group may be almost as useful in many routine applications (20), and may be sufficient for surveillance at the national level, where the primary objective may be to summarize spatial and temporal trends, and where drill-down capability to an individual level may not be necessary to support that objective.

Anonymization can be accomplished, for instance, by aggregating data with the same Zip code or age group, removing portions of those fields (eg, the last two digits of Zip code and the month and day of birth), aggregating nearest neighbors, and randomly skewing geocoded data. Another recent advance includes population-density adjusted Gaussian anonymization (19,20), which is designed to increase anonymity (at a cost of decreased sensitivity and specificity of cluster detection). Research suggests that existing algorithms can add anonymity to a dataset in a way that does not appreciably alter the detection performance of clustering and machine learning algorithms (20,21). These methods and their potential tradeoffs may be worthy of further exploration in the context of sharing local data with national surveillance systems. Questions to address include the level of anonymization that best balances public health utility and privacy protection, and whether these approaches can assuage concerns yet still provide useful surveillance data.

Finally, since the justification for any particular surveillance program should be linked to public health uses and actions, there is a need for additional evaluations that demonstrate the utility of syndromic surveillance for various purposes and clarify how privacy and confidentiality are protected across the stages of event recognition, assessment, and response.

ACKNOWLEDGMENTS

This meeting was supported in part by the O'Neill Institute for National and International Health Law at Georgetown

REFERENCES

1. Drociuk D, Gibson J, Hodge J. Health information privacy and syndromic surveillance systems. *MMWR*. 2004;53 (Suppl):221–5.
2. HIPAA Privacy Rule, 45 CFR § 164.
3. HIPAA Privacy Rule, 45 CFR § 160.203.
4. Buehler JW. Surveillance. In: Rothman KJ, Greenland S, Lash T, eds. *Modern Epidemiology*. 3rd edition, Philadelphia, PA: Lippincott, Williams and Wilkens; 2008:459–80.
5. Bayer R, Fairchild A. Surveillance and privacy. *Science*. 2000;5498:1898–99.

6. Centers for Disease Control and Prevention. *Biosense, 2008*. CDC Web site. http://www.cdc.gov/biosense.
7. Whalen v. Roe, 429 U.S. 589 (1977) (upholding New York law requiring prescriptions of certain drugs to be reported).
8. Mariner WK. Mission creep: public health surveillance and medical privacy. *Boston Univ Law Review.* 2007;87:347–95.
9. Gostin LO. Public health law: power, duty, restraint. London, England: University of California Press; 2000:287–304.
10. Stoto MA. Public health surveillance in the 21st century: achieving population health goals while protecting individuals' privacy and confidentiality. *Georgetown Law Journal*. 2008;96:703–19.
11. Centers for Disease Control and Prevention. HIPAA privacy rule and public health: guidance from the CDC and the US Department of Health and Human Services. *MMWR*. 2003;52(S-1):1–12.
12. HIPAA Privacy Rule, 45 CFR §164.512(b)(1)(i).
13. HIPAA Privacy Rule, 45 CFR § 164.514 (e).
14. Council of State and Territorial Epidemiologists (CSTE). Public health practice vs. research: a report for public health practitioners including cases and guidance for making distinctions. CSTE Web site. http://www.cste.org/pdffiles/newpdffiles/CSTEPHResRptHodgeFinal.5.24.04.pdf. Published May 24, 2004.
15. MacQueen KM, Buehler JW. Ethical issues in HIV, STD, and TB public health practice and research: results of a workshop. *Am J Public Health*. 2004;94:928–31.
16. Fairchild A. Dealing with Humpty Dumpty: research, practice and the ethics of public health surveillance. *J Law, Med & Ethics.* 2003;31:615.
17. The Privacy Act of 1974. 5 U.S.C. §552a.
18. Centers for Disease Control and Prevention. Updated Guidelines for Evaluating Public Health Surveillance Systems. *MMWR*. 2001;50(RR13):1–35.
19. Brownstein JS, Cassa CA, Mandl KD. No place to hide: reverse identification of patients from published maps. *N Engl J Med*. 2006;355:1741–42.
20. Brownstein JS, Cassa CA, Kohane IS, Mandl KD. An unsupervised classification method for inferring original case locations from low-resolution disease maps. *International Journal of Health Geographics.* 2006;5:56.
21. Cassa CA, Wieland SC, Mandl KD. Re-identification of home addresses from spatial locations anonymized by Gaussian skew. *International Journal of Health Geographics*. 2008;7:45.

APPENDIX I

List of Participants

Atar Baer, PhD, Public Health-Seattle & King County

Benjamin E. Berkman, JD, MPH, O'Neill Institute for National and Global Health Law, Georgetown University

James Buehler, MD, Department of Epidemiology and Center for Public Health Preparedness and Research, Rollins School of Public Health, Emory University, and Georgia Division of Public Health

Christopher Cassa, PhD, Harvard/MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology

James Dempsey, JD, Center for Democracy and Technology

P. Joseph Gibson, MPH, PhD, Health & Hospital Corporation of Marion County (Indianapolis)

Julia E. Gunn, RN, MPH, Boston Public Health Commission

Melissa Higdon, School of Nursing and Health Studies, Georgetown University

W. Gary Hlady, MD, MS, Medical Epidemiologist for Emergency Preparedness, California Department of Health Services

Joseph Lombardo, Johns Hopkins Applied Physics Laboratory/ESSENCE

Wendy K Mariner, JD, LLM, MPH, Boston University School of Public Health

John Monahan, O'Neill Institute for National and Global Health Law, Georgetown University

Elizabeth Rea, Toronto Public Health and University of Toronto

Robert Rolfs, Utah Department of Health

Henry Rolka, Centers for Disease Control and Prevention

Mark Smith, MD, MedStar Health

Amy Sonricker, MPH, International Society for Disease Surveillance

Michael A. Stoto, PhD, Professor of Health Services Administration and Population Health, School of Nursing and Health Studies, Georgetown University

APPENDIX II

Agenda for Thursday, October 4, 2007

9:00–9:15 Welcome and meeting objectives

9:15–9:30 Introductions

9:30–11:00 Public health practice perspective; presentations by
- Julia Gunn, Boston Public Health Commission
- Atar Baer, Public Health-Seattle & King County
- Joe Gibson, Health & Hospital Corporation of Marion County
- Gary Hlady, California Department of Health Services
- Elizabeth Rea, Toronto Public Health

11:00–11:15 Break

11:15–12:15 Issues in sharing health-related data for syndromic surveillance; presentations by
- Chris Cassa, Harvard/MIT Division of Health Sciences and Technology
- Mark Smith, Department of Emergency Medicine, Georgetown University School of Medicine & Washington Hospital Center
- Joe Lombardo, Johns Hopkins Applied Physics Laboratory
- Henry Rolka, CDC

12:15–1:00 Lunch provided in meeting room

1:00–2:00 Discussion of practice issues
- What data are needed for effective syndromic surveillance? How does this vary by target condition, stage of the surveillance process (eg, outbreak detection, investigation, case tracking, situational awareness etc.), and other factors?
- What concerns prevent or discourage hospitals and other data owners from providing syndromic surveillance data to public health? Patient privacy concerns? Proprietary concerns? HIPAA regulations? Personnel or other costs for sharing data? Other concerns?
- What are effective strategies for dealing with the concerns of data owners about sharing information syndromic surveillance data with public health? Can technical approaches to de-identification assuage concerns yet still provide useful surveillance data? What level of aggregation best balances utility and privacy protection? Can data be shared in aggregate form but with "drill-down" capability when necessary?

2:00–3:00 Legal perspective; presentations by
- Jim Dempsey, Center for Democracy and Technology
- Wendy Mariner, Boston University

3:00–3:15 Break

3:15–5:00 Discussion of legal authorities and ethical responsibilities
- What are the legal authorities that enable syndromic surveillance at the local, state, and federal level? Is more specific authority needed?
- What does the HIPAA Privacy Rule actually imply for syndromic surveillance? In what circumstances does the public health practice clause allow syndromic surveillance? Is specific legal authority necessary? Are data use agreements (DUA) necessary? Is personal health information (PHI) that is "de-identified" according to HIPAA standards useful for syndromic surveillance?
- How do other federal, state and local laws and regulations enable, restrict, and otherwise influence the ability to share data for public health surveillance purposes?
- How does the application of the HIPAA privacy rule and other laws and regulations depend on whether data are being used for research as opposed to public health practice? In this context, how are distinctions made between research, practice, and evaluation of public health practice? Is some new form of ethical review called for?
- If syndromic surveillance data are shared with law enforcement and intelligence agencies, or used for other nonpublic health purposes (or perceived by the public as being used for these purposes), how with that affect the public's confidence in public health and public health's ability to function?

Agenda for Friday, October 5, 2007

9:00–10:30 Discussion of issues from previous day

10:30–10:45 Break

10:45–12:00 Identification of areas of consensus and where further research is needed

12:00–12:30 Dissemination plans