Perspectives on JAMIA Informatics

White Paper ■

Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper

Charles Safran, MD, MS, Meryl Bloomrosen, MBA, W. Edward Hammond, PHD, Steven Labkoff, MD, Suzanne Markel-Fox, PHD, Paul C. Tang, MD, Don E. Detmer, MD, MA, with input from the expert panel (see Appendix A)

Abstract Secondary use of health data applies personal health information (PHI) for uses outside of direct health care delivery. It includes such activities as analysis, research, quality and safety measurement, public health, payment, provider certification or accreditation, marketing, and other business applications, including strictly commercial activities. Secondary use of health data can enhance health care experiences for individuals, expand knowledge about disease and appropriate treatments, strengthen understanding about effectiveness and efficiency of health care systems, support public health and security goals, and aid businesses in meeting customers' needs. Yet, complex ethical, political, technical, and social issues surround the secondary use of health data. While not new, these issues play *increasingly critical* and complex roles given current public and private sector activities not only expanding health data volume, but also improving access to data. Lack of coherent policies and standard "good practices" for secondary use of health data impedes efforts to strengthen the U.S. health care system. The nation requires a framework for the secondary use of health data with a robust infrastructure of policies, standards, and best practices. Such a framework can guide and facilitate widespread collection, storage, aggregation, linkage, and transmission of health data. The framework will provide appropriate protections for legitimate secondary use.

■ **J Am Med Inform Assoc.** 2007;14:1–9. DOI 10.1197/jamia.M2273.

Introduction

The American Medical Informatics Association (AMIA) convened a panel of diverse stakeholders and experts to discuss a full range of issues related to secondary use of health data. Specifically, AMIA has sought, in an open and neutral environment, to encourage a national discourse on this topic and attendant issues that will guide creation of a national framework. This report highlights the urgency and complexity of issues surrounding secondary use of health data by presenting the panel's key findings and recommendations. The report seeks to encourage public and private sector organizations engaged in health information policy formulation to emphasize the importance of secondary use of health data, and to recruit well-informed colleagues to develop the national framework. As important first steps, the panel recommends continuing dialogue, raising awareness, building collaboration, and clarifying issues. (See Table 1.) Secondary use of health data must become a priority for policymakers in the U.S. The panel's recommendations provide guidance on the compo-

Affiliations of the Authors: Past-Chairman, American Medical Informatics Association, Associate Clinical Professor of Medicine, Harvard Medical School/Beth Israel Deaconess Medical Center (CS), Boston, MA; Associate Vice President, American Medical Informatics Association (MB), Bethesda, MD; Professor, Fuqua School of Business, Duke University (WEH), Durham, NC; Director, Healthcare Informatics, Pfizer, Inc. (SL) NY, NY; Director, Data Exploration Sciences, GlaxoSmithKline (SM-F), King of Prussia, PA; Chairman of the Board, American Medical Informatics Association, Bethesda, MD, Chief Medical Information Officer, Palo Alto Medical Foundation (PCT), Palo Alto, CA; President and CEO, American Medical Informatics Association, Bethesda, MD, Professor of Medical Education, University of Virginia (DED), Charlottesville, VA.

The American Medical Informatics Association (AMIA) would like to acknowledge and thank the organizations that generously supported the project. Anchor Sponsors included GlaxoSmithKline, Lockheed Martin, and Pfizer. Supporting Sponsors included GE Healthcare, IBM, Intelligent Medical Objects (IMO), Medstat, and RemedyMD. The American Medical Informatics Association (AMIA) would like to acknowledge the contributions of the many individuals and organizations that helped to plan and convene this meeting and to develop the resulting paper. Douglas Barton, W. Ed Hammond, Steve Labkoff, and Suzanne Markel-Fox served as members of the Steering Committee. They were actively involved in and provided valuable input to all aspects of the planning processes. Remarks by David Brailer (as the National Coordinator for Health Information Technology) and presentations from Doug Barton (Lockheed Martin), Blake Caldwell (Centers for Disease Control and Prevention (CDC), Nancy Davenport-Ennis (National Patient Advocate Foundation), Stan N. Finkelstein (Harvard-MIT), Melissa Goldstein (Markle Foundation, Connecting for Health), Michael I. Lieberman (GE Healthcare), Eleanor Perfetto (Pfizer), and Kevin Tabb (Stanford Hospital and Clinics) helped to shape the discussions and findings. Dasha Cohen from AMIA helped organize and coordinate logistics for the meeting; Lisa Piazza helped prepare for and facilitate the onsite discussions; Elaine Steen helped edit the report; and Freda Temple provided onsite meeting support as well as helped with production of this document.

Correspondence and reprints: Meryl Bloomrosen, MBA, American Medical Informatics Association, 4915 St. Elmo Avenue, Suite 401, Bethesda, MD 20814; Tel: (301) 657-1291; e-mail: <meryl@amia.org>.

Received for review: 09/11/06; accepted for publication: 10/09/06.

Table 1 ■ Panel Recommendations

Recommendation	Discussion
Increase transparency of data use and promote public awareness	Ongoing public policy discussions must explicitly and directly address the secondary use of health data. Conducting and managing these activities must enlist diverse stakeholders and fully disclose uses and safeguards through open and readily accessible processes.
Focus ongoing discussions on data access, use, and control—not on ownership	Consensus-building meetings encompassing a broad constituency must focus on data access and control policies and practices for secondary use of data. Focus should emphasize access and control, not ownership. Discussants should consider best approaches to risk management and mitigation.
Discuss privacy policies and security for secondary use of health data	To develop consensus on pivotal issues, public and private sector organizations advancing the use of health information should promote discussions that include a wider range of stakeholders than were engaged in this conference. Ongoing discussions must address complex issues related to private and secure secondary use of health data.
Increase public awareness of benefits and challenges associated with secondary use of health data	A wide range of interested parties, especially consumer-oriented patient and caregiver groups, should promote public education regarding benefits of EHRs and about secondary use of health data. A first step is to identify appropriate organizations and agencies that have a role to play in this effort. The aim of the education is to build public awareness and trust in secondary use of health data.
Create a taxonomy for secondary uses of health data	A taxonomy identifying possible non-clinical uses of personal health information is needed to clarify societal, public policy, legal, and technical issues. The taxonomy will support more focused, productive discussions regarding health data and their use.
Address comprehensively the difficult, evolving questions related to secondary use of health data	Questions to address encompass data transparency, consumer awareness and understanding, technical issues and challenges of identity management and user authentication, commercialization and sale of data, and oversight. The de-identification and anonymization of data merit additional attention by technical experts in authentication, de-duplication, and identity management.
Focus national and state attention on the secondary use of health data	The Panel encourages AMIA to share the findings of this meeting with all interested stakeholders, including, but not limited to, the Department of Health and Human Services (DHHS) National Committee on Vital and Health Statistics (NCVHS) and the American Health Information Community (AHIC). Additional efforts should be undertaken to formulate a roadmap that depicts multi-tiered use and re-use of health data; the roadmap should take into account all foreseeable applications and the full complexity of issues.

nents of the envisioned national framework. (See Table 2). Public and private sector stakeholders can elaborate upon these components through discussions that will produce, over time, appropriate technical safeguards and supportive public policies that further the public good. Strengthening and maintaining public trust requires ongoing transparent dialogue with our citizens concerning use of their health data.

Background

In today's data-intensive health care environment, providers generate terabytes of patient data. Laboratory auto-analyzers, pharmacy systems, and clinical imaging systems produce increasingly complex and voluminous data, augmented by data from systems supporting health administrative functions such as patient demographics, insurance coverage, financial data, etc. Clinical narrative information, captured electronically as structured data or transcribed "free text," can also be captured as digital voice dictations or scanned hand-written records. As clinicians adopt electronic health records (EHRs) as the standard for clinical practice, as a byproduct, new sources of detailed clinical information will be created. Those data, combined with existing data, will dramatically increase the breadth and depth of information available for non-clinical applications. Recent advances make it increasingly likely that human genomic data will be routinely available in the future.

While individual patients' rapid, secure electronic access to their own health information can lead to better, more efficient, and more personalized care, demands proliferate for access to, and analysis of, health data outside of clinical settings. Aggregated health data provide value to a broad range of research, quality, public health, and commercial applications. For example, carefully controlled clinical data analysis underpins the measurement of quality and safety in health care delivery. Future pay-for-performance models will likely strengthen linkages between physicians' and hospitals' performance data and reimbursement. Evidence suggests that the public health community can analyze aggregated data to facilitate early detection of emerging epidemics or bioterrorist threats. Commercial enterprises collect health care data to derive products and services that they sell to customers, including third party payers, researchers, and marketing entities.

Secondary uses of health data[1] can enhance individuals' health care experiences, expand knowledge about diseases and treatments, strengthen understanding of health care systems' effectiveness and efficiency, support public health and security goals, and aid businesses in meeting customers' needs. Yet, access to and use of health data pose complex ethical, political, technical, and economic challenges. For

[1]For purposes of this meeting, secondary use of data was defined as non-direct care use of personal health information (PHI) including but not limited to analysis, research, quality/safety measurement, public health, payment, provider certification or accreditation, and marketing and other business including strictly commercial activities.

example, to meet public health, emergency preparedness, and homeland security imperatives the federal government has initiated real-time collection of data from emergency rooms and other sources—without public dialogue, based on authority from existing public health law. Further, there are reports of the buying and selling of non-anonymized patient and provider data by the medical industry—carried out without explicit consent from patients or physicians. Such activities include pressuring or coercing patients to consent to data disclosure for use not covered by regulation, and abuses of commercially available, identifiable patient information. Although the Health Insurance Portability and Accountability Act of 1996 (HIPAA) applies to health information created or maintained by health plans, health care clearinghouses, and health care providers who engage in certain electronic transactions, there is a potential lack of protection of personal health information (PHI) when used by entities not explicitly covered by HIPAA legislation or regulations. Individuals and organizations may mistakenly perceive HIPAA to assure protection of all secondary use of PHI by users, beyond those covered entities specifically noted in HIPAA.

These issues are not new. Fresh consideration of secondary uses of health data is, however, critical. Both public and private sector organizations continue to design systems enabling secondary use of health data for applications in clinical, public health, biomedical, policy, health services research areas, as well as for other evolving public concerns, including emergency preparedness, global epidemiology, and homeland security.

Renewed public and private sector efforts promote adoption of EHRs. Related efforts focus on developing a nationwide, secure health information network that can support safe, equitable, efficient, effective, and patient-centered health care. Such initiatives include establishment of the American Health Information Community (AHIC) and contracts awarded to develop prototypic architectures for a Nationwide Health Information Network (NHIN). Recent National Institutes of Health (NIH) initiatives promote population-based studies to identify genetic and environmental causes of common illnesses. These portend potential low-cost sequencing of personal genomes in the not too distant future. The NIH Roadmap for Medical Research promotes clinical research networks and data sharing. The foregoing activities emphasize the need to re-examine implications of secondary uses of person-specific data. In July 2006, the Robert Wood Johnson Foundation (RWJF) announced *Project HealthDesign: Rethinking the Power and Potential of Personal Health Records*, a national program designed to stimulate innovation in the development of personal health record (PHR) systems. Further, the *Roadmap for Clinical Decision Support*, developed by AMIA under contract to the Office of the National Coordinator for Health Information Technology (ONC), presents a vision for an ongoing cycle of data collection, research, and new knowledge generation to strengthen clinical decision support. In addition to national initiatives such as those listed, there are myriad activities related to the secondary use of health data at state, regional, and organizational levels.

Catalyzing the Discussion of Secondary Use of Health Data

While pivotal to strategic improvements in the U.S. health system, secondary use of health data poses technical, strategic, policy, process, and economic concerns related to the ability to collect, store, aggregate, link, and transmit health data broadly and repeatedly for legitimate purposes. Thus, lack of coherent policies and standard "good practices" for secondary use of health data impedes efforts to transform the U.S. health care system. Further, growing availability of technologies supporting secondary uses, combined with data expansion, per se, heightens urgency to engage the public in a transparent dialogue. Addressing these myriad challenges ultimately requires a national framework for secondary use of health data, including a robust infrastructure of policies, standards, and best practices.

AMIA has sought, in an open and neutral environment, to further national discourse on secondary uses of health data and attendant issues in a manner that will guide creation of a national framework. AMIA convened a meeting of diverse stakeholders (i.e., the panel) to discuss the full range of such issues, including, but not limited to:

- What are the potential benefits and risks regarding the secondary use of health data?
- Who owns health data and who has the right to access the data and for what purposes?
- What are the evolving public trust issues with respect to patient consent for secondary use of health data? Do patients have the right to audit or put other constraints on the use of their data, even after anonymization?
- In light of serious public health threats such as avian flu, how does society reconcile the public good with the rights of the individuals while weighing health versus privacy considerations?
- What problems may develop as innovative technologies enhance the ability and ease of widespread data sharing and additional commercial uses?
- What can be done to address issues arising from inappropriate use and/or exploitation of data sharing?
- What regulations, legislation, and/or policies and procedures are needed to address these issues?

All stakeholders must develop sufficient understanding of the inherent benefits and risks of secondary uses of health data in order to develop effective policies and practices. This, in turn, will require ongoing discussion, education, communication, and collaboration among consumers, ethicists, health care practitioners, industry specialists, informaticians, policy makers, researchers, and others. The work of this panel, as reflected in this report, is a first step in promoting dialogue among stakeholders about the opportunities and challenges related to the secondary use of health data.

Methodology

An expert panel convened April 27–28, 2006, in the metropolitan Washington, D.C., area. A steering committee composed of a small group of experts and representatives of the major sponsors of the meeting set goals and an agenda for the meeting. The steering committee suggested potential discussants and panel participants. The 36 panel members

Table 2 ■ Components of a National Framework for Secondary Use of Health Data

Transparent policies and practices for the secondary use of health data
Focus on data control, rather than data ownership per se
Consensus on privacy, policy, and security
Public awareness
Comprehensive scope (beginning with a taxonomy)
National leadership

included representatives from health care providers, technology vendors, pharmaceutical companies, consulting firms, practitioners, researchers, government agencies, and citizen stakeholders. Appendix A (available as a JAMIA online supplement at www.jamia.org) comprises a complete list of sponsors and participants. To inform discussions, participants received background information and discussion questions before the meeting.

The panel focused on secondary uses of person-specific health data. The panel designated certain topics as outside its scope, including both truly de-identified data that cannot be re-identified to specific persons, and technical processes and procedures for achieving data de-identification. These were nevertheless considered important to the overall topic.

The meeting agenda viewed secondary use of health data from four main perspectives: the consumer; patient safety, quality, and research; public health; and industry (see Appendix B, available as a JAMIA online supplement at www.jamia.org, for complete agenda). AMIA staff and consultants served as facilitators and recorders to support the deliberations. Divided into four sessions, the first day focused on these perspectives. Each session began with two background presentations that provided an overview of the topic and identified the salient issues. Next, the entire group shared observations on the topic through plenary discussions moderated by a facilitator. Following open discussions, each of the four round tables considered previously prepared common scenarios, with associated questions intended to guide discussion. (See Appendix C for the scenarios.) Each group selected a presenter who summarized the small group's discussions, including areas of agreement and ideas for future efforts. During his address to the group at a dinner meeting that closed the first day's work, David Brailer, MD, PhD, shared insights from his experience as the National Health Information Technology Coordinator and as CEO of Care Science.

The second day began with a presentation of a synthesis of Day One discussions. This was followed by additional small group discussions and reports on the common themes of Day One, and a final round of group discussions and reports focusing on recommendations and future steps.

Definition of Terms and Abbreviations

The panel quickly recognized a need to clarify terminology in common use for the context of the meeting. For effective communication, all participants in the dialogue had to use the same vocabulary in the same way. The panel offered the following working definitions for terms used during the meeting, and agreed that further refinement of the terminology is needed (see Recommendations).

anonymized data—alteration of PHI that makes it impossible to link individuals with their data.

commercialization—the sale or resale of health data.

covered entities—The Administrative Simplification standards adopted by the U.S. Department of Health and Human Services (DHHS) under the Health Insurance Portability and Accountability Act of 1996 (HIPAA) apply to any entity that is a health care provider that conducts certain transactions in electronic form (called here a "covered health care provider"); a health care clearinghouse; a health plan. An entity that is one or more of these types of entities is referred to as a "covered entity" in the Administrative Simplification regulations.

de-identified data—the elimination of all identifiers as enumerated under HIPAA under the safe-harbor method (i.e., a patient's name, medical record number, social security number, and other data fields that directly link a patient to their data). There is potentially another approach that involves having a statistician determine that the ability (likelihood) of being able to combine data with other public sources of information and successfully identify an individual is extremely small.

electronic health record (EHR)—personal data created, developed, maintained, and/or provided by clinicians, providers, and allied health providers in direct patient care; an electronic application containing health information about individuals that is used by clinicians, providers, and allied health professionals to provide direct care for the individuals.

health data—data about or from an individual such as a person's age or serum potassium level. In aggregate, an individual's data are called personal health information (PHI).

personal health record (PHR)—an electronic application through which individuals can access, manage, and share their health information, in a private, secure, and confidential environment; personal data created, developed, maintained, and/or provided by individuals about themselves.

primary use of data—the use of PHI by the organization or entity that produced or acquired these data in the process of providing real-time, direct care of an individual.

reversibly anonymized data—the alteration of PHI in such a way that re-identification may be accomplished through access to a protected key that makes it possible to link individuals with their data only through a trusted intermediary.

secondary use of data—non-direct care use of PHI including but not limited to analysis, research, quality/safety measurement, public health, payment, provider certification or accreditation, and marketing and other business including strictly commercial activities.

Meeting Highlights

The meeting style and format, including thought-provoking scenarios and questions, prompted lively discussion of the complex issues. These ranged beyond specific situations presented in the scenarios. Meeting highlights follow below, organized by the four perspectives of the conference.

Consumer Perspective

The first session focused on the issues of privacy and security of personal health information from the consumer's point of view. Background presentations highlighted policy challenges associated with electronic health information exchange, and EHR-related consumer benefits as well as pitfalls related to privacy breaches. During the discussion period, panelists reviewed the scenario of an imaginary "Mrs. Powter" whose employer is switching employees to a new health plan to cut costs (see Appendix C). The questions raised during this discussion reverberated throughout the meeting:

Who owns the data in Mrs. Powter's personal health record?

When Mrs. Powter leaves Health Plan #1 what happens to her data?

What are the issues (e.g., data exchange standards, cost) that arise when transferring data among health plans?

What additional, secondary uses of the data should be permitted?

Should Mrs. Powter be asked for permission for each instance of usage or should she give global permission?

Small group discussions covered a variety of issues related to personal data: participants drew a distinction between data ownership and access to data; differentiated between the PHR and the EHR; raised concerns about data misuse, consent under duress, and gaps in HIPAA protections; debated relevant intellectual property issues; and considered the rights of patients versus their obligations with respect to the patient's own data.

Patient Safety, Quality, and Research Perspective

Launched by presentations about secondary uses of health data for research purposes, the second session considered challenges related to conducting research with data originally collected for another purpose (i.e., insurance claims). Panel members turned to consideration of a scenario in which, as part of a cost cutting effort, a health plan queries the company's data repository to link outcomes of therapy for hypertension to medicines prescribed as evidenced by claims data, in an attempt to determine which drugs lower blood pressure most effectively. The group discussed the limitations of the study approach and the potential shortcomings of the data as well as whether the conclusion reached by this method was valid. The panel concluded that standards were lacking for establishing levels of evidence. It further determined the need to establish explicit rules or conventions to define evidence, and to validate secondary datasets. The discussion covered complex issues related to de-identification of data, including increasingly available technical approaches for re-identification of data through dataset interlinkages.

Public Health Perspective

The panel discussed the growing use of health data for purposes of emergency preparedness, public health, epidemiology, and homeland security. The first presentation described BioSense, a CDC program to improve the nation's capabilities for real-time biosurveillance and situational awareness. The second presentation offered lessons learned by a systems integration company in developing projects involving data subject to privacy constraints. Panel members discussed a scenario in which university-based researchers attempted to gain access to a scrubbed copy of BioSense data to study quality and disparity in emergency treatment across the United States. The group considered the now-familiar issues of obtaining patient consent for downstream use of data, concerns about potential data re-identification, and the need for clear rules and safeguards for release of data. There was strong agreement on the need to inform and educate patients about all downstream uses of their data. However, there was diverse opinion regarding the most effective and practical approaches to accomplish this. Participants agreed that these topics warrant further discussion.

Industry Perspective

Major topics discussed during the session on industry perspectives included growing commercialization of health data and use of health data for business and proprietary purposes. Two industry viewpoints promoted dialogue, one from a consortium of clinicians who made their pooled data available to consortium members for quality-related research and sold the data to non-consortium researchers, and a second from the pharmaceutical industry, describing the variety of uses that it makes of aggregated data, and limitations and advantages of various data sources. The group scenario addressed collection and sale of patient data by a fictional Regional Health Information Organization's (RHIO) Chief Executive Officer (CEO), who was tasked with developing a business plan identifying a revenue stream not reliant upon federal or state funds. Panel members once again grappled with the issues of patient consent for the sale of data used for non-direct patient care purposes. They considered whether the sale of data for a specific use (medical research versus proprietary or targeted marketing) should have a bearing on the issue, and whether the situation would be different if the RHIO was funded by private sector dollars rather than by the federal government.

Major Findings and Recommendations

By design, the meeting enumerated major issues associated with secondary uses of health data as the starting point for an all-encompassing, nationwide dialogue. The panel's findings and recommendations, presented below, form topics to guide AMIA's further collaborative efforts and activities.

Finding 1: Secondary use of health data is widespread. The presentations and discussions, as well as the literature (see Appendix D for a selected bibliography), document widespread, growing secondary use and re-use of health data. Such uses occur in both public and private sectors for proprietary, research, and monitoring purposes with less than comprehensive regulation. Participants agreed that, in most instances, providers, physicians, and their patients are generally unaware of this development, despite the growth and success of a multimillion-dollar industry based on the sale of health and heath-related data. Further, while HIPAA requires many health care providers and health insurers to obtain additional documentation before disclosing person-specific health information, and to closely scrutinize requests for access to health information for secondary purposes, such as for research, HIPAA rules only address the use and disclosure of health information by "covered enti-

ties" (i.e., health care providers, health plans, and clearinghouses).

Recommendation 1: Increase transparency of data use and promote public awareness. Ongoing public policy discussions must explicitly and directly address the secondary use of health data. Conducting and managing these activities must enlist diverse stakeholders and fully disclose uses and safeguards through open and readily accessible processes.

Finding 2: The focus needs to be data access and control, not data ownership. Group consensus was that focusing on "ownership" diverts attention from needed development of sound policies and practices. Participants acknowledged that responsibility for ensuring privacy and safeguarding patient data applies across the diverse continuum of data users. Technical advances enable creation of many databases that are now maintained, updated, used, and re-used for multiple purposes, including those outside direct patient care. Technology has also enabled easy transmission of such data. Despite HIPAA requirements regarding the de-identification of data and adherence to data use agreements, there is the potential for the re-identification of patients and providers through the linkage of disparate databases. A need exists to further explore and explicitly address issues of health data access and control throughout data life cycles. Extensive discussions covered the need to develop policies for secondary uses of health data, recognizing that such policies will be complex.

Recommendation 2: Focus ongoing discussions on data access, use, and control—not on ownership. Consensus-building meetings encompassing a broad constituency must focus on data access and control policies and practices for secondary use of data. Focus should emphasize access and control, not ownership. Discussants should consider best approaches to risk management and mitigation.

Finding 3: Critical issues include patient privacy and public trust. Use of person-specific patient data for purposes other than direct patient care and public health is not well understood and is poorly monitored. This raises numerous ethical, technical, economic, and procedural concerns. The sense of the meeting participants was that too few safeguards exist that adequately address secondary uses of health data. Further discussions about informed consent must clarify how data uses for specific purposes can remain in compliance with federal, state, and local laws. Health data uses not covered by privacy regulations, including uses of data obtained via coerced or compelled consent, can erode public trust and might potentially hinder the public good. Some panel members asserted that development and execution of patient choice options involving explicit authorization for use of their own data (opting in/opting out) provides the only adequate means to mitigate patient privacy issues. Participants acknowledged that no "single unified patient (consumer) perspective" exists. Consumers will view the issue in many possible ways—assuming they are informed about it. Thus, substantial variation in consumer viewpoints will make issues related to patient (consumer) consent and choice complex.

Recommendation 3a: Discuss privacy policies and security for secondary use of health data. To develop consensus on pivotal issues, public and private sector organizations advancing the use of health information should promote discussions that include a wider range of stakeholders than were engaged in this conference. Ongoing discussions must address complex issues related to private and secure secondary use of health data.

Recommendation 3b: Increase public awareness of benefits and challenges associated with secondary use of health data. A wide range of interested parties, especially consumer-oriented patient and caregiver groups, should promote public education regarding benefits of EHRs and about secondary use of health data. A first step is to identify appropriate organizations and agencies that have a role to play in this effort. The aim of the education is to build public awareness and trust in secondary use of health data.

Finding 4: Technological capabilities to merge, link, re-use, and exchange data outpace establishment of policies, procedures, and processes to do so ethically and legally. Increasingly complex issues arise from advancing technical capabilities. Meeting participants did not agree on technical issues such as whether data can be truly anonymized, or what are the preferred methodologies for "identity management." There is a need to build consensus around working definitions of secondary health data uses, and to develop clearer understanding of strengths and limitations of using specific types of health data. Defining secondary uses for health data must also envision the potential impact of future EHR evolution, as well as advances in communications capabilities and forthcoming biomedical research, such as large scale, population-based genomic studies that generate vast amounts of personal genetic information.

Recommendation 4a: Create a taxonomy for secondary uses of health data. A taxonomy identifying possible non-clinical uses of personal health information is needed to clarify societal, public policy, legal, and technical issues. The taxonomy will support more focused, productive discussions regarding health data and their use.

Recommendation 4b: Address comprehensively the difficult, evolving questions related to secondary use of health data. Questions to address encompass data transparency, consumer awareness and understanding, technical issues and challenges of identity management and user authentication, commercialization and sale of data, and oversight. The de-identification and anonymization of data merit additional attention by technical experts in authentication, de-duplication, and identity management.

Finding 5: Progress requires additional attention and leadership at state and national levels. Existing efforts to develop and implement a nationwide interconnected and interoperable network infrastructure do not adequately address issues of secondary health data use. National-level leadership must obtain input from a broad range of public and private sector stakeholders in order to develop adequate policies, standards, and legal/regulatory remedies regarding the secondary use, abuse, and misuse of health data. Stakeholders include those who collect the data for primary use; those who use the data for non-clinical purposes; patients and the public; those who create policy about health data; those who inform and educate health care professionals, industry, patients, and the public; and philanthropic organizations

that support development of policy on critical health and technology issues.

Recommendation 5: Focus national and state attention on the secondary use of health data. The Panel encourages AMIA to share the findings of this meeting with all interested stakeholders, including, but not limited to, the Department of Health and Human Services (DHHS) National Committee on Vital and Health Statistics (NCVHS) and the American Health Information Community (AHIC). Additional efforts should be undertaken to formulate a roadmap that depicts multi-tiered use and re-use of health data; the roadmap should take into account all foreseeable applications and the full complexity of issues.

Conclusion

A natural byproduct of existing clinical and administrative activity is an increasing array of rich data sources and datasets. Many such resources contain personally identifiable or potentially identifiable data—i.e., the data can be re-identified after being de-identified. The increasing volume, complexity, and diversity of health care data and information systems, as well as approaches to identifying and linking datasets, pose significant problems for the future.

Panel participants estimated that a well-established multi-million-dollar business exists that utilizes secondary health data as its primary resource. However, the panel conducted no research to establish this estimate. For several decades, various organizations such as hospitals, health plans, and payers have "mined" mostly administrative claims and prescription data. In the current health care environment, an expanding, diverse array of users in the commercial research, public health, policy, and clinical and biomedical research communities seeks access to secondary health data. Widespread use of personal health data outside of the primary care setting often occurs with commercial intent as employers, payers, and insurers attempt to fulfill business and proprietary-oriented goals and objectives. Furthermore, as EHRs continue to evolve and the adoption of health information technology increases, more health data will become readily available, with predictable increased efforts to access and use these data for various non-patient care purposes.

Unfortunately, some data usages, such as by the Medical Information Bureau, are neither well regulated nor subject to citizen oversight. Many recent regional efforts to establish health information exchanges face a business challenge to provide information utilities to the community at the lowest possible cost. Although not often in public, stewards of these data exchanges and their business partners are exploring non-subscription models for revenue generation which frequently include selling clinically rich datasets to industries that already purchase surrogates for such data. In addition, the imperatives from public health and homeland security have initiated the collection of real-time data (such as emergency room data) from hospitals and other providers across the country without public dialogue. At a minimum, a public dialogue is needed.

Meeting participants agreed that the rapidly evolving nationwide efforts for more widespread health information exchange must include work to address pressing issues of secondary health data usage, as outlined in this report. The panel report lays a foundation for new dialogue about these uses, and emphasizes important roles to be played by the public and private sectors. In addition to stimulating future action, the panel's recommendations provide guidance regarding the components that should shape a national framework for secondary use of health data:

- Transparent policies and practices for the secondary use of health data;
- Focus on data control ownership rather than data ownership per se;
- Consensus on privacy, policy, and security;
- Public awareness and trust;
- Comprehensive scope (beginning with a taxonomy); and,
- National leadership.

Public and private sector stakeholders, in future discussions on the secondary use of health data, can explore these components more fully. Through creation of appropriate technical safeguards and supportive public policy, the panel believes that the secondary use of health data can further the public good. A more transparent dialogue with our citizens concerning the use of their health data is key to maintaining and strengthening the public trust, while enhancing the public's informed actions.

AMIA Board of Directors (BOD) Response and Action

By convening this expert panel and disseminating this report, AMIA has identified the topic of the secondary use of personal health information as a critical issue for the continued widespread adoption of health information technology. The AMIA BOD reviewed the paper and endorsed the panel's recommendations. The BOD anticipates that it will commit additional organizational resources to advance the work of the panel. AMIA will encourage other organizations to collaboratively pursue the recommendations and to continue this important public discourse.

Appendix C: Discussion Scenarios

A) Mrs. Powter is a 44-year-old mother of two who works for a small business and obtains health insurance for her family through her employer. Health Plan #1 provides an online personal health record (PHR) linked to a pharmacy benefits management (PBM) company. The PHR is automatically updated with claims data and medications from the PBM. She can add problems to the problem list and add medications to her medication list. A wellness program provided by Health Plan #1 asks her questions and records answers in the PHR.

Since health premiums will rise by 15%, her employer decides to switch all 15 employees to Health Plan #2.

Who owns the data in the PHR?

Is there a difference between the data that Mrs. Powter entered vs. the plan's encounter data or data from the PBM?

When Mrs. Powter leaves Health Plan #1 what happens to her data?

Who pays the cost to transfer the data between systems, presuming that is allowable: the sending health plan, the receiving health plan, or Mrs. Powter (because it's a PHR)?

From a logical viewpoint, what would be necessary (what kind of standards) in order for no additional effort to be required to transfer the data from #1 to #2?

Where should the PHR data be stored—at the PBM, at the person's computer, both, or neither?

If the sending and/or receiving systems do not conform to clinical data exchange standards, who bears the cost of transfer change? Who determines the relevant standards?

What kind of "pressures" (and by whom) should be used to encourage or enforce the required clinical data exchange standard?

What additional, secondary use of the data should be permitted? Should Mrs. Powter be asked for permission for each instance of usage, or should she give global permission?

Would the answers to these questions differ if the health plans were federally or state funded plans (under Medicare or Medicaid)?

B) A large insurance company is facing what it perceives as a very difficult period in claims expenses coming in the next few years. Its Chief Executive Officer (CEO) directs his staff to trim costs. An eager analyst in his group wants to deliver on cost savings and decides to look to the company's spending on chronic care medicines. He decides to run a series of queries from his own company's data repository attempting to link outcomes of therapy to medicines prescribed (as evidenced by claims data). As his health plan pays for both laboratory tests and prescriptions, he can link laboratory results and hospitalization data to prescribing information. He decides to look at hypertension as a diagnosis and then tries to find out which drugs lower blood pressure most effectively. His analysis complete, he reports back to his superiors about his findings, which suggest that generic medications are the only medicines that should be covered by the plan going forward.

What defines a standard of evidence from health data?

Who decides what studies demonstrate valid conclusions (i.e., is there a peer review process for making such claims)?

Should data as described above be considered "evidence"—should its use in clinical care be considered Evidence Based Medicine (EBM)?

Should there be standards of how information from studies such as this one is reported to the public? Should the data behind these findings be made available for external verification?

C) University-based researchers wanting to study quality and disparity in emergency treatment across the United States develop a sound study methodology. They receive approval from their institutional review board (IRB) and funding from a private foundation. With support from their influential senators and representatives, they approach the Centers for Disease Control and Prevention (CDC) and request a scrubbed copy of the agency's BioSense data.

Is this a legitimate tertiary use of data?—Tertiary in the sense that the original owner of the data has not been involved in making a determination of how the data should be used.

Does the patient or provider of data to CDC need to be informed or is consent required?

Can the patient/provider opt out?

What assurance is required, if any, that the tertiary use of data in the emergency treatment study conforms to the terms of the study design and any data use agreements executed between the CDC and the researchers? Who is responsible for auditing the use of data or making this determination?

Does the patient/provider have the right to inspect/review the use of the data?

D) State RHIO has been funded by AHRQ to design, build, and implement a health information exchange. The stakeholders are convened and form a governance board and appropriate working groups to use these funds wisely and well. A CEO is hired to run the RHIO and develop a business plan that does not require federal or state funding. One idea that surfaces is collecting and selling patient data.

Who owns the data? Who can use the data and for what purposes?

Who gets compensated when the data are used for non-patient care purposes?

Should patients be informed each time their data are used for non-patient care purposes and would they have the right to opt in or out?

Under what circumstances is specific patient consent required? Would the need for consent differ if the data are de-identified?

Is physicians' consent required for use of data from patients under their care?

Does the use of the data (e.g., medical research vs. identification of patients for targeted marketing of pharmaceuticals) have a bearing on the issue?

How does use of these clinical data for payment or reimbursement fit into the privacy issues? Should payers be permitted to use the data for other purposes?

To what extent can patient data be used to evaluate provider performance?

Should these data be used without patient permission for health surveillance? Should drug companies be able to use these data for drug trials? Could these data be used to help identify patients for eligibility in clinical trials or other research protocols?

Would the answers to these questions differ if the RHIO were funded by private sector dollars?

Appendix D: Selected Bibliography

Bailey, Steve. Boston Globe March 24, 2006 "Your Data For Sale?"

Canadian Institutes of Health Research. Secondary Use of Personal Information in Health Research: Case Studies, November 2003 Available at: http://www.cihr-irsc.gc.ca/e/1475.html. Accessed November 11, 2006.

Centers for Disease Control and Prevention, Healthy People 2010 Statistical Notes Number 24 July 2002.

Healthy People 2010 Criteria for Data Suppression. Available at: www.cdc.gov/nchs/data/statnt/statnt24.pdf. Accessed April 2006.

Connecting for Health, Common Framework Materials. Available at: http://www.connectingforhealth.org/commonframework/. Accessed November 11, 2006.

The Connecting for Health Architecture for Privacy in a Networked Health Environment. Summary. Available at: http://www.connectingforhealth.org/commonframework. Accessed November 11, 2006.

Consumer Reports. The New Threat to Your Medical Privacy. Available at: http://www.consumerreports.org/cro/health-fitness/health-care. Accessed April 2006.

Detmer, Don E. Your privacy or your health—will medical privacy legislation stop quality health care? Counterpoint. Int J Qual Health Care 2000;12(1):1–3.

Ferris, Nancy. Hidden Keys to Health. The medical community is sitting on mountains of e-health data that could lead to important medical discoveries. But will its value remain buried by privacy concerns and lack of funding? Published February 13, 2006. Government Health IT. Available at: http://www.govhealthit.com. Accessed April 2006.

Gostin LO. Medical countermeasures for pandemic influenza: ethics and the law. JAMA 2006;295(5):554–6.

Gostin LO, Lazzarini Z, Neslund VS, Osterholm MT. The public health information infrastructure. A national review of the law on health information privacy. JAMA 1996; 275(24):1921–7.

Hodge, JG Jr., Gostin, L.O., Jacobson PD. Legal issues concerning electronic health information: privacy, quality, and liability. JAMA 2000;283(12):1564–5.

Martin, Zack. AMGA Data Mining Project to Provide Outcomes, Benchmarks. Available at: http://www.healthdatamanagement.com. Accessed April 26, 2006.

Mosquera, Mary. Government Computer News. March 7, 2006. CMS to test feasibility of e-personal health records. Available at: http://appserv.gcn.com. Accessed April 2006.

National Committee on Vital and Health Statistics. Personal Health Records and Personal Health Record Systems. A Report Recommendation from the National Committee on Vital and Health Statistics. Washington, D.C. October 2005.

Vijayan, Jaikumar. Computerworld. Confidential patient data sent to the wrong company for 15 months. Available at: http://computerworld.com. Accessed April 2006.

Westin, Alan F. Addressing the Privacy Challenge: Health Research Using Electronic Health Records. Hackensack, New Jersey. April 2006.