



Toward a National Framework for the Secondary Use of Health Data

Authors:

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

A report of a working conference of the:

**American Medical Informatics Association
4915 St. Elmo Avenue Suite 401
Bethesda, Maryland 20814
www.amia.org
301 657-1291**

Author Information

Charles Safran, MD, MS
Past-Chairman, American Medical Informatics Association
Associate Clinical Professor of Medicine
Harvard Medical School/Beth Israel Deaconess Medical Center

Meryl Bloomrosen, MBA
Associate Vice President
American Medical Informatics Association

W. Ed Hammond, PhD
Professor, Fuqua School of Business, Duke University

Steven E. Labkoff, MD, FACP
Director, Healthcare Informatics
Pfizer Human Health

Suzanne Markel-Fox, PhD
Director, Data Exploration Sciences
GlaxoSmithKline

Paul Tang, MD
Chairman of the Board
American Medical Informatics Association
Chief Medical Information Officer
Palo Alto Medical Foundation

Don Detmer, MD, MA
President and CEO
American Medical Informatics Association
Professor of Medical Education
University of Virginia

Acknowledgements

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. AMIA also wants 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.

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.

Toward a National Framework for the Secondary Use of Health Data Executive Summary

Secondary use of health data refers to non-direct care use of personal health information (PHI), including but not limited to analysis, research, quality and safety measurement, public health, payment, provider certification or accreditation, and marketing and other business (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 the effectiveness and efficiency of our health care systems, support public health and security goals, and aid businesses in meeting the needs of their customers. Yet, access to and secondary use of data poses complex ethical, political, technical, and social challenges. Many of the issues surrounding the secondary use of health data are not new. These issues are *however increasingly critical* and complex in light of public and private sector activities that are expanding the volume of data available to be used and the availability of tools to access that health data. The lack of coherent policies and practices for the secondary use of health data presents a significant impediment to the goal of strengthening the U.S. health care system. A national framework for the secondary use of health data that includes a robust infrastructure of policies, standards, and best practices is needed to facilitate the broad and repeated collection, storage, aggregation, linkage, and transmission of health data with appropriate protections for legitimate secondary use.

The American Medical Informatics Association (AMIA) convened a panel of diverse stakeholders and experts to discuss the full range of issues related to the secondary use of health data. Specifically, AMIA sought to further the national discourse on secondary use of health data and their attendant issues; guide the creation of a national framework for the secondary use of health data; and provide an open and neutral environment for complicated discussions. This report describes the structure of that meeting, presents the panel's key findings and recommendations, and highlights the urgency and complexity of the issues surrounding the secondary use of health data. The goal of this report is to encourage both public and private sector organizations that have a role in shaping health information policy to increase attention to and involvement in developing a national framework for the secondary use of health data.

The panel recommends continuing dialogue, raising awareness, building collaboration, and clarifying issues as important first steps in the development of a national framework for secondary use of health data. (See Box 1.) The common thread among these recommendations is that the issue of secondary use of health data must become a priority for policymakers in the U.S. In addition to providing direction for future action, the panel's recommendations provide guidance on the components that should help shape a national framework for secondary use of health data. (See Box 2). These components can be explored more fully by public and private sector stakeholders in future discussions on the secondary use of health data. With appropriate technical safeguards and supportive public policy, the panel believes that the secondary use of health data can further the public good, and that a more transparent dialogue with our citizens concerning the use of their health data is critical to maintaining and strengthening the public trust.

Box 1: Panel Recommendations

Recommendation 1: Increase the transparency of data use and public awareness. Secondary use of health data must be conducted and managed solely through the use of open and transparent processes. Diverse stakeholders should be engaged to assure that these uses are undertaken with full disclosure. Ongoing and future public policy discussions need to explicitly address the secondary use of health data more directly. AMIA is encouraged to share the findings of this meeting with a wide range of stakeholders and through various mechanisms. These include, but are not limited to, the National Committee on Vital and Health Statistics (NCVHS) and the American Health Information Community (AHIC).

Recommendation 2: Focus ongoing discussions on data access, use, and control (not on ownership). Discussion about the secondary use of health data should focus on access to and control of data for various uses, not on “data ownership” per se. Additional meetings and efforts encompassing a broader constituency must continue to focus on data access and control policies and practices in the context of secondary use of data. The discussions should include considerations of approaches for risk management and mitigation.

Recommendation 3a: Continue discussions on privacy policy and security with regard to the secondary use of health data. Public and private sector organizations involved in advancing the use of health information should be encouraged to participate in future discussions on the array of complex issues related to privacy and security of the secondary use of health data to develop consensus on pivotal issues. Ongoing discussions should include a wider range and variety of citizen, consumer, and patient stakeholders than were engaged in this conference.

Recommendation 3b: Increase public awareness efforts on the benefits and challenges associated with the secondary use of health data. A wide range of stakeholders, especially consumer-oriented groups (including patients and their caregivers) should be convened to assure that the public is better informed and educated about the benefits of EHRs and about secondary use of their data. A first step is to identify appropriate organizations and agencies that have a role to play in improving public awareness of the benefits and challenges associated with the secondary use of health data as a means of building public trust in the secondary use of health data.

Recommendation 4a: Create a taxonomy of the secondary use of health data. A taxonomy of the non-clinical uses of personal health information is needed to address the complex environment surrounding the secondary use of data. The taxonomy is also needed to further clarify the societal, public policy, legal, and technical issues, thereby supporting more productive discussions regarding the data themselves and their potential use.

Recommendation 4b: Address increasingly difficult current and evolving questions related to the secondary use of health data in a comprehensive manner. These issues include transparency of data, consumer awareness and understanding, technical and technological issues related to identity management and user authentication, commercialization and sale of data, and oversight. Additional discussion and further clarification is needed in defining the range of issues relating to de-identified data. An explicit effort is needed to clarify issues related to data anonymization, working with technical experts in authentication, de-duplication, and identity management.

Recommendation 5: Focus national and state attention on the secondary use of health data. Additional collaborative efforts need to assure that attention is focused on the issues associated with the secondary use of health data. The process should lead to the formulation of a clear roadmap to depict and identify the multi-tiered use and re-use of health data, taking into account both current and foreseeable future applications. This is essential to address the complexity that surrounds the secondary use of health data.

Box 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

Background

Today's health care environment is increasingly data intensive. Providers generate terabytes of patient data from laboratory auto-analyzers, pharmacy systems, and clinical images; these data are augmented by other systems to support traditional health system administrative functions related to patient demographics, insurance coverage, financial data, etc. Even clinical narrative information that is either hand-written or dictated and transcribed can be scanned and stored digitally. As electronic health records (EHRs) become the standard for use by clinicians, new sources of granular clinical data may be combined with existing data, resulting in a leap in the breadth and depth of information that may be available for non-clinical applications. And, recent advances make it increasingly likely that human genomic data will be routinely available in the future.

In addition to the staggering increase in volume of health data, there are increasing demands for access to and analysis of health data outside of clinical settings. For individual patients, rapid and secure electronic access to their own health information can lead to better, more efficient, and more personalized care. In the aggregate, data are also valuable for use in a broad range of applications in the research, quality, public health, and commercial spheres. For example, the measurement of quality and safety in health care delivery is based on evidence that is derived from the controlled analysis of data. In the future, pay-for-performance models will likely strengthen links between reimbursement and performance data on physicians and hospitals. There is suggestive evidence that the examination of aggregated data may facilitate early detection by the public health community of emerging epidemics or bioterrorist threats. From a commercial perspective, companies collect health care data and sell products and services based upon these data to a variety of customers including third party payers, researchers, and marketing related entities.

Secondary use of health data¹ can enhance health care experiences for individuals, expand knowledge about disease and appropriate treatments, strengthen understanding about the effectiveness and efficiency of our health care systems, support public health and security goals, and aid businesses in meeting the needs of their customers. Yet, access to and use of data poses complex ethical, political, technical, and economic challenges. For example, based upon legal authority from public health law but without public dialogue, Congress has initiated the real-time collection of data from emergency rooms and other sources to meet public health, emergency preparedness, and homeland security imperatives. Further, there are reports of ongoing buying and selling of non-anonymized patient and provider data by the medical industry without explicit consent of either patients or physicians, including pressuring or coercing patients to consent to data disclosure for use not covered by regulation, and abuse of commercially-available, identifiable patient information by entities. 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. In fact, there may be mistaken perceptions that HIPAA assures protection of all secondary use of PHI by users, beyond those uses and users (covered entities) specifically noted in HIPAA.

The issues surrounding the secondary use of health data are not new. Fresh consideration of them is, however, critical as both public and private sector organizations are focusing on the design of systems that enable secondary use of health data for clinical, public health, biomedical, policy, and health services research applications. Other efforts aim to help ensure that information is available

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."

for a growing range of evolving relevant public concerns, including emergency preparedness, global epidemiology, and homeland security.

Renewed public and private sector efforts are underway to promote the adoption of EHRs and develop a nationwide secure health information network that is supportive of a safe, equitable, efficient, effective, and patient-centered health care system. This is reflected in a variety of initiatives including the establishment of the American Health Information Community (AHIC) and the awarding of contracts to develop prototypes for a Nationwide Health Information Network (NHIN) architecture. Recently launched initiatives by the National Institutes of Health (NIH) to conduct population-based studies to identify genetic and environmental causes of common illnesses and the potential for low-cost sequencing of personal genomes in the not too distant future contribute to the need for a re-examination of issues related to the secondary use of person-specific data. The execution of the NIH Roadmap for Medical Research that includes initiatives related to promoting clinical research networks and the sharing of data is another effort propelling this discussion. 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

Secondary use of health data is pivotal to strategies aimed at strengthening the U.S. health system. Yet, the ability to collect, store, aggregate, link, and transmit health data broadly and repeatedly for legitimate secondary use poses technical, strategic, policy, process, and economic concerns. Thus, the current lack of coherent policies and practices for the secondary use of health data presents a significant impediment to the goal of transforming the U.S. health system. Further, the increased focus on improving capabilities for secondary use of health data, combined with the growing volume of data, heightens the urgency to commence transparent and public dialogue about appropriate secondary use of health data. Ultimately, a national framework for the secondary use of health data (i.e., a robust infrastructure of policies, standards, and best practices) is needed to address these myriad challenges.

To further the national discourse on the secondary use of health data and their attendant issues, to guide the creation of a national framework for the secondary use of health data, and to provide an open and neutral environment for complicated discussions, AMIA convened a meeting of diverse stakeholders (i.e., the panel) to discuss the full range of issues related to the secondary use of data. These issues include, but are not limited to:

- What are the potential benefits and risks regarding the secondary use of health data?
- Who owns the 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 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 do we reconcile the public good with the rights of the individual, of health versus privacy considerations?
- What problems may develop as innovative technologies enhance the ability and ease of widespread data sharing and additional commercial use?
- 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?

Effective secondary use of health data will require that all stakeholders develop a sufficient understanding of the inherent benefits and risks of these uses. This, in turn, will require ongoing discussion, education, communication, and collaboration among consumers, ethicists, health care practitioners, industry specialists, informaticians, policy makers, researchers, and perhaps, others as well. 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 on 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 also made suggestions about potential discussants and panel participants. The 36 panel members included representatives from various segments of the industry such as providers, technology vendors, pharmaceutical companies, consulting firms, practitioners, researchers, government agencies, and citizen stakeholders. A complete list of sponsors and participants is provided in Appendix A. Background information and discussion questions were provided to participants in advance of the meeting to help inform the discussions.

The panel focused on secondary use of person-specific health data. Data that have been truly de-identified and cannot be re-identified to specific persons was beyond the scope of the panel's discussions. Technical discussions related to the processes and procedures for achieving data de-identification were determined to be outside the scope of the panel's discussions, but important to the overall topic.

The agenda for the meeting was structured around the secondary use of data viewed from four main perspectives: the consumer; patient safety, quality, and research; public health; and industry (see Appendix B for the complete agenda.). AMIA staff and consultants served as facilitators and recorders to support the deliberations. The first day was divided into four sessions, each focused on one of these perspectives. Each session began with two background presentations that provided an overview of the topic and identified the salient issues. These presentations were followed by plenary discussions moderated by a facilitator, during which the entire group shared observations on the topic. Following this open discussion, each of the four round tables was presented with a common scenario and questions 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. David Brailer, MD, PhD, shared insights from his experience as the National Health Information Technology Coordinator and as CEO of Care Science during an address to the group at a dinner meeting that closed the first day's work.

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 that there was a need to clarify certain terms and terminology in common use in this field, in the context of the meeting. This was important to be sure that everyone participating in this dialogue was using the same vocabulary in the same way. Thus, the panel offered the following working definitions for use during the meeting and agreed that additional efforts to refine the terms and their definitions and use would be helpful (see Recommendations).

- **anonymized data** -- the 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 along with the thought-provoking scenarios and questions prompted lively discussion of the complex issues, with discussions ranging beyond the specific situations presented in the scenarios. Below, we present meeting highlights organized by the four perspectives mentioned above.

Consumer Perspective

The first session focused on the issues of privacy and security of personal health information from the point of view of the consumer. Two background presentations highlighted the policy challenges associated with electronic health information exchange, and the promise of EHRs from the consumer point of view as well as the potential breaches of privacy associated with them. 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. 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 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?

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 mis-use, 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 their own data.

Patient Safety, Quality and Research Perspective

The second session was launched by presentations on the secondary use of data for research purposes, illustrating some of the challenges in conducting research using 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 and that explicit rules or conventions for definitions of evidence and validation of secondary datasets were needed. This discussion also raised complex issues related to de-identification of data and the increasing available technical approaches to re-identification of data through linkages among datasets.

Public Health Perspective

The panel discussed the growing use of health data for emergency preparedness, public health, epidemiology and homeland security purposes. 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 operating projects that developed systems using data subject to privacy constraints. Panel members discussed a scenario in which university-based researchers were attempting to gain access to a scrubbed copy of BioSense data to study quality and disparity in emergency treatment across the U.S. The group considered the by-now familiar issues of obtaining patient consent for downstream use of data, concerns about the potential for 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 use of their data, however, there was diverse opinion on the most effective and practical approaches to accomplish this. Participants agreed that more discussion on these topics is warranted.

Industry Perspective

The growing commercialization of health data and their use for business and proprietary purposes were the major topics discussed during the session reflecting the industry perspective. Two examples of industry viewpoints were presented: the experience of a consortium of clinicians whose pooled data were made available to consortium members for research on quality and for sale to non-consortium researchers; and the variety of uses that the pharmaceutical industry makes of aggregated data, and the limitations and advantages of the various data sources. The scenario addressed the issue of the 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 that would

identify a revenue stream that did not rely on the use of 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 specific use of the data to be sold (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

The meeting was designed to enumerate the major issues associated with the secondary use of health data as a beginning point for an all-encompassing, nationwide dialogue. The panel's findings and recommendations are presented below. Pursuant to these findings and recommendations, AMIA will undertake collaborative efforts to further address the issues discussed and the activities proposed.

Finding 1: Widespread secondary use of data. As evidenced by the presentations and discussions, as well as the literature (see Appendix D for a selected bibliography), there is widespread and growing secondary use and re-use of health data throughout the public and private sectors for various proprietary, research, and monitoring purposes with less than comprehensive regulation. Participants agreed that while in most instances, providers, physicians, and their patients were thought to be generally unaware of this development, a multimillion-dollar industry based on the sale of health and health-related data has prospered and appears to be growing. 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 entities" (i.e., health care providers, health plans, and clearinghouses).

Recommendation 1: Increase the transparency of data use and public awareness. Secondary use of health data must be conducted and managed solely through the use of open and transparent processes. Diverse stakeholders should be engaged to assure that these uses are undertaken with full disclosure. Ongoing and future public policy discussions need to explicitly address the secondary use of health data more directly. AMIA is encouraged to share the findings of this meeting with a wide range of stakeholders and through various mechanisms. These include, but are not limited to, the National Committee on Vital and Health Statistics (NCVHS) and the American Health Information Community (AHIC).

Finding 2: Data access and control versus data ownership. The sense of the group was that a focus on "ownership" diverts attention from the needed development and execution of sound policies and practices. Further, participants acknowledged that the responsibility for ensuring privacy and safeguarding patient data applies to the continuum and diversity of data users. Advances in technology and the ability to transmit data have resulted in many databases that are now maintained, updated, used, and re-used for multiple purposes other than direct patient care. Notwithstanding the HIPAA requirement that patient data be de-identified and that the use of data be subject to data use agreements, the potential ability to re-identify data (unless explicitly managed to preclude this capability), and to link disparate databases to identify patients and practitioners raises growing concerns. There is a growing need to further explore and explicitly address questions concerning access and control of data throughout their life cycle. There were extensive discussions about the need to develop appropriate policies for the secondary use of health data, recognizing that such policies are critical and complex.

Recommendation 2: Focus ongoing discussions on data access, use, and control (not on ownership). Discussion about the secondary use of health data should focus on access to and control of data for various uses, not on "data ownership" per se.

Additional meetings and efforts encompassing a broader constituency must continue to focus on data access and control policies and practices in the context of secondary use of data. The discussions should include considerations of approaches for risk management and mitigation.

Finding 3: Patient privacy issues and the public trust. The use of person-specific patient data for purposes other than direct patient care and public health is not well understood or monitored and raises numerous ethical, technical, economic, and procedural issues. The sense of the meeting participants was that too few safeguards adequately address the secondary use of health data. Further discussions about informed consent regarding data use for specific purposes that are compliant with the requirements of federal, state and local laws are needed. Use not covered by privacy regulation or data obtained with coerced or compelled consent might erode the public trust and could potentially hinder the public good. Some panel members asserted that patient privacy issues would only be mitigated adequately through the development and execution of patient choices related to explicitly authorizing (opting in/opting out) use of their data. Participants acknowledged that there is no “one unified patient (consumer) perspective” and that there are many possible ways for consumers to view the issue – assuming they are informed about it. Thus, there is a substantial variation in consumer viewpoints, which makes the issues related to patient (consumer) consent and choice complex.

Recommendation 3a: Continue discussions on privacy policy and security with regard to the secondary use of health data. Public and private sector organizations involved in advancing the use of health information should be encouraged to participate in future discussions on the array of complex issues related to privacy and security of the secondary use of health data to develop consensus on pivotal issues. Ongoing discussions should include a wider range and variety of citizen, consumer, and patient stakeholders than were engaged in this conference.

Recommendation 3b: Increase public awareness efforts on the benefits and challenges associated with the secondary use of health data. A wide range of stakeholders, especially consumer-oriented groups (including patients and their caregivers) should be convened to assure that the public is better informed and educated about the benefits of EHRs and about secondary use of their data. A first step is to identify appropriate organizations and agencies that have a role to play in improving public awareness of the benefits and challenges associated with the secondary use of health data as a means of building public trust in the secondary use of health data.

Finding 4: Technological capabilities to merge, link, re-use, and exchange data are outpacing the establishment of policies, procedures, and processes. Increasingly complex issues are arising from these technical competencies and technological capabilities. Meeting participants were not in agreement on a variety of technical issues including whether data can be truly anonymized or the preferred technical and technological approaches to “identity management.” There is a need for consensus on working definitions of the secondary use of health data as well as a clearer understanding of the strengths and limitations of specific types of data and their applicability for secondary use. When defining secondary use of data, it is necessary to look ahead to the potential impacts of the future evolution of EHRs, advances in technology and communications capabilities, forthcoming biomedical research, and large scale, population-based genomic studies that will generate vast amounts of personal genetic information.

Recommendation 4a: Create a taxonomy of the secondary use of health data. A taxonomy of the non-clinical use of personal health information is needed to address the complex environment surrounding secondary use of health data. The taxonomy is also needed to further clarify the societal, public policy, legal, and technical issues, thereby supporting more productive discussions regarding the data themselves and their potential use.

Recommendation 4b: Address increasingly difficult current and evolving questions related to the secondary use of health data in a comprehensive manner. These issues include transparency of data, consumer awareness and understanding, technical and technological issues related to identity management and user authentication, commercialization and sale of data, and oversight. Additional discussion and further clarification is needed in defining the range of issues relating to de-identified data. An explicit effort is needed to clarify issues related to data anonymization, working with technical experts in authentication, de-duplication, and identity management.

Finding 5: Need for additional attention and leadership at the national and state levels.

Existing efforts to develop and implement a nationwide interconnected and interoperable network infrastructure are not adequately addressing the issues related to the secondary use of data. The development of policies, standards, and legal/regulatory remedies regarding the secondary use, abuse, and mis-use of health data requires leadership on a national level with input from a broad range of stakeholders in the public and private sectors. These 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. Additional collaborative efforts need to assure that attention is focused on the issues associated with the secondary use of health data. The process should lead to the formulation of a clear roadmap to depict and identify the multi-tiered use and re-use of health data, taking into account both current and foreseeable future applications. This is essential to address the complexity that surrounds the secondary use of health data.

Conclusion

As a natural byproduct of many existing clinical and administrative functions and datasets, there is an increasing array of rich data sources, many of which contain personally identifiable or potentially identifiable data (that is, data that can be re-identified after it has been de-identified). Thus, there is an increasing volume, complexity and diversity of health care data and information systems, as well as approaches to identifying and linking datasets.

Meeting participants estimated that the use of health data is a well-established multimillion-dollar business although it did 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 our current health care environment, health data are increasingly sought by an expanding and diverse array of users in the commercial research, public health, policy, and clinical and biomedical research communities. This widespread use of personal health data outside of the primary care setting is often with commercial intent; employers, payers, and insurers to fulfill various business and proprietary-oriented goals and objectives use health data. Furthermore, as EHRs continue to evolve and adoption of health information technology increases, health data will be more readily available. This will likely lead to increased efforts to access and use these data for various non-patient care purposes.

Unfortunately, some use of these data by various entities (such as the Medical Information Bureau) has neither been well regulated nor subject to citizen oversight. More recently, many of the regional efforts to establish health information exchanges face a business challenge to provide an information utility to the community at the lowest possible cost. Although usually unstated, the stewards of these data exchanges and their business partners are exploring non-subscription models of revenue which almost always include selling clinically rich data to those industries that already purchase surrogates

for this data. In addition, the imperatives of 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 address the pressing issues related to the secondary use of health data that are outlined in this report. The panel laid the foundation for a new dialogue about these uses and the roles of the public and private sectors with regard to them. In addition to providing direction for future action, the panel's recommendations provide guidance on 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.

These components can be explored more fully by public and private sector stakeholders in future discussions on the secondary use of health data. With appropriate technical safeguards and supportive public policy, the panel believes that the secondary use of health data can further the public good, and that 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 committing additional organizational resources to continue to advance the work of the panel and will encourage other organizations to work collaboratively to pursue the recommendations and to continue this important public discourse.

Toward a National Framework for the Secondary Use of Health Data

Appendices

White Paper People and Process

Appendix A: Participant List

Cheryl Austein-Casnoff

Director, Office of Health Information Technology
Health Resources and Services Administration
Department of Health and Human Services

Marion J. Ball, EdD

Fellow, IBM Global Leadership Initiative (GLI)
Center for Healthcare Management
Professor, Johns Hopkins University School of Nursing
Adjunct Professor, Division of Health Sciences Informatics
Johns Hopkins School of Medicine

Douglas Barton

Director, Enterprise Solutions
Lockheed Martin Integrated Systems & Solutions

David Brailer, MD

National Coordinator for Health Information Technology
Department of Health and Human Services

Laird D. Burnett

Vice President, Legal & Government Relations
Kaiser Permanente

M. Blake Caldwell, MD, MPH

Senior Advisor to the Director
Coordinating Center for Health Information and Service
Centers for Disease Control and Prevention

Janet Corrigan, PhD, MBA

President and CEO
The National Quality Forum

Kelly Cronin

Director
Office of Programs and Coordination
Office of the National Coordinator for Health Information Technology
Department of Health and Human Services

Nancy Davenport-Ennis

CEO, National Patient Advocate Foundation

Don Detmer, MD, MA

President and CEO
American Medical Informatics Association
Professor of Medical Education
University of Virginia

Louis Diamond, MD

VP and Medical Director, Thomson Medstat

Linda L. Dimitropoulos, PhD

Project Director
Health Information Security and Privacy Collaboration (HISPC)
RTI International

Margo Edmunds, PhD

Vice President
The Lewin Group
Adjunct Associate Professor
Department of Health Policy and Management
Johns Hopkins Bloomberg School of Public Health
Center for Public Health Preparedness and
Information Security Institute

Stan Finkelstein, MD

Senior Research Scientist, MIT Engineering Systems Division
Harvard-MIT Division of Health Sciences and Technology
Senior Lecturer, Harvard Medical School Department of Health Care Policy
Director of the Harvard MD/MBA Joint Program

Paul M. Gertman, MD

Chairman, US CareLink

Melissa M. Goldstein

Adjunct Assistant Professor of Health Care Sciences
and Assistant Professorial Lecturer in Health Policy
The George Washington University
Director, Health Program, The Markle Foundation

W. Ed Hammond, PhD

Professor, Fuqua School of Business, Duke University

COL Bart J. Harmon, MD, MPH

Chief Medical Information Officer
Director of Information Management.
TRICARE Management Activity
Department of Defense

Howard Isenstein

Vice President, Public Affairs & Quality
Federation of American Hospitals

Secondary Use of Health Data

Michael Z. Jones

Executive Vice President, Sales and Marketing
RemedyMD

Robert M. Kolodner, M.D.

Chief Health Informatics Officer
Veterans Health Administration

Steve Labkoff, MD

Director, Healthcare Informatics
Pfizer, Inc

Michael I. Lieberman, MD, MS

Informatics Director, Practice Solutions, GE Healthcare;
Clinical Assistant Professor,
Department of Medical Informatics & Clinical Epidemiology
Oregon Health & Science University

Suzanne Markel-Fox, PhD

Director, Strategy & Process Development
GlaxoSmithKline

Frank Naeymi-Rad PhD, MBA

CEO, Intelligent Medical Objects, Inc.

Dennis S. O'Leary, MD

President, Joint Commission on Accreditation of Healthcare Organizations

Eleanor M. Perfetto, PhD, MS

Senior Director
Payment Policy Analysis Group
US Medical - Outcomes Research
Pfizer, Inc.

David B. Pryor, MD

Senior Vice President
Clinical Excellence
Ascension Health

Lisa Piazza

Vice President
Product Development and Delivery
MG Taylor Corporation

Charles Safran, MD, MS

Past-Chairman of the Board
American Medical Informatics Association
Associate Clinical Professor of Medicine
Harvard Medical School

Kevin Tabb, MD

Chief Quality & Medical Information Officer
Stanford Hospital and Clinics

Secondary Use of Health Data

Paul Tang, MD

Chairman of the Board
American Medical Informatics Association
Chief Medical Information Officer
Palo Alto Medical Foundation

Freda Temple

Consultant, Information Management and Editorial Services

Margaret VanAmringe

Vice President for Public Policy and Government Relations
Joint Commission on Accreditation of Healthcare Organizations

P. Jon White, MD

Health IT Portfolio Manager
Agency for Healthcare Research and Quality

David Wye

Director, Trade & Regulatory Affairs
Lockheed Martin Corporation

AMIA Staff

Meryl Bloomrosen

Associate Vice President

Dasha Cohen

Meetings Manager

Appendix B: Meeting Agenda

Thursday, April 27, 2006

- 10:00 am Welcome from AMIA Chairman – Paul Tang
 Introductions of Expert Panel – Don Detmer
 Purpose of Expert Panel – Charles Safran
 Ground Rules – Meryl Bloomrosen
- 10:30 **Session 1 Citizen’s Perspective**
 Panel: Melissa Goldstein and Nancy Davenport-Ennis
 Open discussion
 Small group discussion
- 11:30 **Session 2 Research, Quality & Safety**
 Panel: Stan N. Finkelstein and Kevin Tabb
 Open discussion
- 12:30 pm Working lunch and Small Group Discussions
- 1:30 **Session 3 Public Health**
 Panel: Blake Caldwell and Douglas Barton
 Open discussion
 Small group discussion
- 3:00 **Session 4 Industry**
 Panel: Michael Lieberman and Eleanor Perfetto
 Open discussion
 Small group discussion
- 5:00 Reception
- 6:00 Dinner
 Speaker: David Brailer

Friday, April 28, 2006

- 8:00 am Continental Breakfast
- 9:00 Synthesis of Day One – Don Detmer
- 10:00 Reports and Feedback from Small Groups and Open Discussion:
- 12:00 pm Next Steps: Charles Safran
- 12:30 Thank You and Adjourn

Appendix C: Discussion Scenarios

A) Mrs. Powter is a 44 year old mother of 2 who works for a small business and obtains health insurance for her family through her employer. HealthPlan #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 HealthPlan #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 HealthPlan #2.

1. Who owns the data in the PHR?
2. Is there a difference between the data that Mrs. Powter entered vs. the plan's encounter data or data from the PBM?
3. When Mrs. Powter leaves HealthPlan #1 what happens to her data?
4. 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)?
5. 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?
6. Where should the PHR data be stored – at the PBM, at the person's computer, both or neither?
7. 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?
8. What kind of "pressures" (and by whom) should be used to encourage or enforce the required clinical data exchange standard?
9. 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?
10. 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.

1. What defines a standard of evidence from health data?
2. Who decides what studies demonstrate valid conclusions? (i.e., is there a peer review process for making such claims)?
3. Should data as described above be considered "evidence" – should its use in clinical care be considered Evidence Based Medicine (EBM)?
4. 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 U.S. 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.

1. 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.
2. Does the patient or provider of data to CDC need to be informed or is consent required?
3. Can the patient/provider opt out?
4. 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?
5. Does the patient/provider have the right to inspect/review the use of the data?

D) StateRHIO 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.

1. Who owns the data? Who can use the data and for what purposes?
2. Who gets compensated when the data are used for non-patient care purposes?
3. 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?
4. Under what circumstances is specific patient consent required? Would the need for consent differ if the data are de-identified?
5. Is physicians' consent required for use of data from patients under their care?
6. 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?
7. 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?
8. To what extent can patient data be used to evaluate provider performance?
9. 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?
10. 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
<http://www.cihr-irsc.gc.ca/e/1475.html>

Center for Societal and Legal Research
Building Privacy by Design in Health Data Systems <http://www.privacyexchange.org/>

Centers for Disease Control and Prevention

Healthy People 2010 Statistical Notes Number 24 July 2002 Healthy People 2010 Criteria for Data Suppression Accessed April 2006 www.cdc.gov/nchs/data/statnt/statnt24.pdf

Connecting for Health

Common Framework Materials
<http://www.connectingforhealth.org/commonframework/>.

The Connecting for Health Architecture for Privacy in a Networked Health Environment. Summary <http://www.connectingforhealth.org/commonframework>

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

Detmer, Don E. Your privacy or your health-will medical privacy legislation stop quality health care? Counterpoint. International Journal for Quality in Health Care 2000. Volume 12, Number 1:pp.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. Accessed April 2006. <http://www.govhealthit.com>

Gostin LO. Medical countermeasures for pandemic influenza: ethics and the law. Journal of the American Medical Association. 2006 Feb 1;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. Journal of the American Medical Association. 1996 Jun 26;275(24):1921-7.

Hodge, JG Jr., Gostin, L.O., Jacobson PD. Legal issues concerning electronic health information: privacy, quality, and liability. Journal of the American Medical Association, 2000 March 22-29; 283 (12):1564-5.

Martin, Zack. AMGA Data Mining Project to Provide Outcomes, Benchmarks. Accessed April 26, 2006 Health Data Management. <http://www.healthdatamanagement.com>

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

Secondary Use of Health Data

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. Accessed April 2006 <http://computerworld.com>

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