3-1-2012

Approaches to Using Protected Health Information (PHI) for Patient-Centered Outcomes Research (PCOR): Regulatory Requirements, De-identification Strategies, and Policy

Raj Sabharwal
AcademyHealth, raj.sabharwal@academyhealth.org

Erin Holve
AcademyHealth, erin.holve@academyhealth.org

Alison Rein
AcademyHealth, alison.rein@academyhealth.org

Courtney Segal
AcademyHealth, courtney.segal@academyhealth.org

Issue Brief

Approaches to Using Protected Health Information (PHI) for Patient-Centered Outcomes Research (PCOR):

Regulatory Requirements, De-identification Strategies, and Policy

Introduction

Over the last few years, both legislative action and substantial federal outlays have promoted the use of patient-centered outcomes research (PCOR) and comparative effectiveness research[1] (CER) as a means to generate more robust evidence on the utility and value of health care interventions. The expectation is that better evidence will support more informed treatment decisions among physicians and patients and improve patient outcomes, while also informing formulary and coverage determinations. Arguably, one key element of conducting PCOR is ensuring that data governance[2] strategies are sufficient to preserve patient privacy while still allowing data to flow between clinicians and researchers and enable research activities that can drive improvements in health and health care.

To maximize resources and ensure timely evidence generation, the newly implemented Patient-Centered Outcomes Research Institute (PCORI) and other PCOR initiatives will undertake a range of study designs including systematic reviews of existing studies, randomized clinical trials (RCTs), and observational data analyses. Of these, observational studies — which typically use existing data sources, such as claims data or registries — will be a critical component to reflect the experience of patients in "real world" settings. Furthermore, a wealth of new data resources for all types of studies are increasingly available due to investments in the nation's health information technology (HIT) infrastructure, with the potential to contribute timely data to answer pressing questions for patients, providers, caregivers, and other important decision makers.

Within this new environment of multiple data sources, expanding networks, and increased linkages across systems, researchers still encounter many limitations and practical challenges due, in part, to the limited infrastructure and regulatory apparatus available to support the use and access of protected health information (PHI). Often researchers who wish to use multiple data sources are limited to working with only a single data source at a time, or must rely on previously aggregated data that may not include recent information or that may have been developed for other purposes. Further, while federal regulations on the use and access of PHI provide accommodations for research purposes, many in the research com-

AcademyHealth

The EDM Forum is supported by the Agency for Healthcare Research and Quality (AHRQ) through the American Recovery & Reinvestment Act of 2009, Grant U13 HS19564-01.

munity feel that these regulations limit their ability to link multiple data sources and provide negative incentives for rigorous research (Nass, Levitt, and Gostin 2009). Still, these regulations should not be seen simply as barriers to research, in our current social context, which emphasizes the importance of stakeholder engagement, including respect for privacy and confidentiality of patient information.

Despite the proliferation of social networks and the willingness of individuals to freely share personal information, recent high-profile data breaches and cyber-hacking incidents have highlighted growing public concern about privacy as more data are available electronically and shared across an ever expanding network of entities. These seemingly opposing forces underscore the need for researchers to build a framework of trust with patients, consumers, clinicians, and other key stakeholders. As PCORI and others work to further engage patients and consumers in the research process, they must ensure that issues regarding data privacy are transparent and adequately addressed. Should individuals believe the privacy of their data is in doubt, they will be less inclined to participate in the research process and potentially limit the disclosure of sensitive information to their clinicians (Herdman and Moses 2006). Researchers must also continue to acknowledge their ethical requirements to protect the privacy and confidentiality of their data and in turn, their research subjects.

This brief describes current regulations regarding PHI and common challenges researchers face in using personal health information. The use of limited data sets and de-identified data, as approaches to address privacy concerns, are discussed. New infrastructure approaches that seek to meet these regulations and generate data that are more useful for CER and related studies are explored. Finally, potential policy approaches to clarifying privacy regulations and fostering studies that are more informative are discussed.

Federal Regulation of Protected Health Information

The Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule (see 45 CFR§ 164.514) provides standards for the protection and use of PHI, containing common identifiers such as names, addresses, or dates of birth. In general, the rule requires written authorization from a patient for an entity to disclose or use identifiable health information for any purpose, including health services research.

PHI may be disclosed for research after receiving a waiver of this authorization from an IRB or privacy board, based on use or disclosure that involves only minimal risk to the subjects. Situations that may provide minimal risk include adequate protection of identifiers from improper use and disclosure, the destruction of identifying information at the earliest opportunity, and written assurances that the PHI will not be reused or disclosed to others. Further exceptions to the last requirement include required disclosures by law or for research oversight, as well as for research use that could not be conducted without access to the PHI and the necessary waiver (Gunn et al. 2004).

The Privacy Rule, however, does not regulate de-identified data, for which common identifiers have been removed, and there is no reasonable basis to believe that it can be re-identified. In addition, the Rule describes two mechanisms for the de-identification of PHI, commonly referred to as the *safe harbor method* and the *statistical method* (Nass, Levitt, and Gostin 2009).

The *safe harbor method* requires the removal of 18 specific data elements related to an individual and their relatives, household members, and employers. Specific data elements include names, dates, zip codes, telephone numbers, social security numbers, email addresses, and license plate numbers, among others. In addition, entities that employ the safe harbor method must attest that they do not have actual knowledge that the remaining information can be used to re-identify an individual. An organization, however, may assign a code to an individual record to assist with future re-identification, provided the code is protected and not shared with others.

The *statistical method* allows for the release of health information provided that a statistical analysis shows that "the risk is very small that health information could be used to identify an individual subject of the information, either by itself or in combination with other available information," (see 45 CFR§ 164.514). Further, the rule states that this determination must be performed by a qualified statistician.

The Privacy Rule also describes an alternate mechanism for data de-identification, known as a *limited data set*. Similar to the safe harbor method, a limited data set does not include specific direct identifiers of an individual, but does allow for certain data related to geography and dates, which are considered important for PCOR. In addition, use of the data for the restricted purposes requires a data use agreement (DUA) between the data holder and recipient. Such DUAs typically outline the specific terms for the sharing, use, and protection of the dataset (Herdman and Moses 2006).

In practice, the statistical method is rarely used by researchers, in part because the regulations are vague and there is no standardized approach. For example, the rule does not define what is meant by a "small risk" of re-identification. This leaves the safe harbor method and the use of limited data sets as the primary pathways for researchers (Benitez, Loukides, and Malin 2010). Table 1 provides a comparison of data elements for safe harbor and limited data sets.

Table 1: Elements of De-Identified Data Sets and Limited Data Sets that are removed or may be included based on the HIPAA Safe Harbor Rule

Data Element	De-Identified Data Set	Limited Data Set
Names	Removed	Removed
Address All geographic subdivisions smaller than a state, including street address, city, county, precinct, zip code, and their equivalent geocodes	Removed (except initial three digits of zip code where more than 20,000 people live)	Removed (except city, town, state or zip code)
Dates All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death, and any data elements indicative of all ages over 89	Removed	May be included
Telephone numbers	Removed	Removed
Fax numbers	Removed	Removed
Electronic mail addresses	Removed	Removed
Web Universal Resource Locators (URLs)	Removed	Removed
Internet Protocol (IP) address numbers	Removed	Removed
Social Security Numbers	Removed	Removed
Medical Record Numbers	Removed	Removed
Health plan beneficiary numbers	Removed	Removed
Account numbers	Removed	Removed
Certificate/license numbers	Removed	Removed
Vehicle identifiers and serial numbers, including license plate numbers	Removed	Removed
Device identifiers and serial numbers	Removed	Removed
Biometric identifiers, including finger and voice prints	Removed	Removed
Full face photographic images and any comparable images	Removed	Removed
Any other unique identifying number, characteristic, or code, except codes permitted for re-identification by the covered entity	Removed	May be included

Source: Encouraging the Use of, and Rethinking Protections for De-Identified (and "Anonymized") Health Data. Center for Democracy and Technology, June 2009.

Challenges for Researchers Using De-Identified Data

Not surprisingly, there are a set of common challenges that researchers face in working with both de-identified data and limited data sets, particularly with respect to data linkage, identifying rare events, or studying sub-populations and using dates of service, which are often important for studying health outcomes. Many of these challenges are related to the removal of data elements or variables needed for statistical analyses as well as the linking and comparison of separate datasets. In addition, these issues may severely limit the ability to use a range of data sources for PCOR, which typically requires large and detailed data sources for appropriate comparisons and risk adjustments.

Data Linkage: Perhaps the greatest limitation for both mechanisms is the narrow ability to link shared data from multiple sources to an individual record. Few of the common data elements typically used for linkages are available in a de-identified data set. Some linkage of records may be possible through a combination of elements such as gender or age, but this type of linkage is not practical for larger datasets, or for datasets outside a specific catchment area. Further, while organizations may assign a code for future re-identification, limitations on sharing this code restricts linkage across entities. While limited data sets provide additional data elements that are of interest to researchers, the matching of records by these elements is often difficult in that each of the disparate data sets must include the same specific data elements.

Rare events and sub-populations: Excluding the 18 HIPAA identifiable data elements may also limit researchers' ability to assess rare events and to study issues in sub-populations. For example, the exclusion of data such as device identifiers or serial numbers may limit the usefulness of de-identified data in surveillance studies. Similarly, the safe harbor method may hinder studies of elderly populations due to restrictions on the inclusion of age data for those 89 and over, and epidemiologic studies due to limited geographic (partial zip codes) information (Benitez, Loukides, and Malin 2010).

Dates of service: Another critical issue for researchers is the exclusion of date information from de-identified data sets. Excluding dates of service may restrict

the ability to calculate common research or quality improvement measures such as days of hospitalization and the length of time between treatments. While some researchers have received limited waivers that allow for research using information with perturbed dates of service, the lack of specificity (i.e. date ranges rather than specific durations) may limit the accuracy of data between events.

Technical Infrastructure to Support Researchers' Use of Multiple Data Sources

Despite the limitations noted above and the difficulty of aggregating disparate data sets, researchers continue to emphasize and expand upon the use of PHI. In addition to the protection provided by meeting HIPAA requirements, linking multiple PHI datasets provides more complete information and limits costs, time, and feasibility issues related to primary data collection (Bradley et al. 2010).

Typically, these efforts rely on a centralized data warehouse model, which aggregates data into a large database at a single physical location. While a centralized data model allows for a single control point for queries, major drawbacks include concerns regarding the physical storage of data outside an organization's IT infrastructure, related privacy and security concerns, and the need for regular data extracts.

More recently, many researchers and organizations have advocated for the development and use of distributed data models. A distributed data network allows for secure, remote analysis of standardized and reusable data sources from multiple sites, as well as tools to use it. Benefits of a distributed system include the ability for data partners to maintain physical control and access of their data behind firewalls protected by their security processes and rules. Local control of data also facilitates better data quality control and easier consultation for researchers on specific data issues (Brown et al. 2010). In addition, sophisticated security procedures, such as data hashing[3] may allow for the linkage of records across data partners, without the exposure of PHI.

Networks, Technology, and Systems

For the purposes of this brief, the term "distributed" refers to both distributed and federated data models. The following describes the common network, electronic clinical information systems, and warehousing models.

Distributed Research Network
A distributed data network (DRN) is an approach in which data holders maintain control over their protected data and its uses (Brown, 2009). A DRN features a central portal that performs network functions, such as operations (e.g., workflow, policy rules, auditing, query formation and distribution) and security (e.g., authentication, authorization) and distributed data marts that remain under the control of the data holders (D'Avolio, Farwell, and Fiore 2010).

Federated Research network
A federated network links geographically and organizationally separate databases so that a single database query can return results from multiple databases while maintaining the privacy and confidentiality of patient data (Pace et al. 2009).

Patient Registry
A patient registry is an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serves one or more predetermined scientific, clinical, or policy purposes. A registry database is a file (or files) derived from the registry (Gliklich and Dreyer 2007).

Virtual Data Warehouse (VDW)
VDW is not a centralized data warehouse—it is "virtual", consisting of parallel databases set up identically at each site that can be easily merged across sites. These databases have been constructed by extracting data from the local electronic data systems and reconfiguring them to use standard variable names and coded values (HMO Research Network 2011).

Aggregated or Centralized Data Model
An aggregated or centralized data model copies data from original sources and brings and standardizes these data in a centralized place. The copied data can then be queried and analyzed (Diamond, Mostashari, and Shirky 2009).

Still, a distributed data model requires agreement on common data elements and maintenance procedures among partners, which can be challenging. Further, the distributed model must allow for flexibility and workarounds in navigating multiple software and data storage systems. From a governance perspective, the varying concerns, such as expectations for security and data access, as well as potential uses of data once it has been collected and stored, have the potential to be difficult to manage. In addition to developing clear policies and procedures to guide collection, storage, and use of data, the need to build a culture of trust between partners and the community is an abiding requirement of data sharing.

A number of entities have successfully used the distributed data model to facilitate the analysis of data from multiple data sources. Notable examples include the Informatics for Integrating Biology and the Bedside (i2b2) Center, the Electronic Primary Care Research Network (ePCRN), the Cancer Biomedical Informatics Grid (caBIG), and the PopMedNet system, as highlighted below.

The PopMedNet system enables the creation of distributed health data networks through open source software (Harvard Pilgrim Healthcare Institute 2010). The system was developed through an AHRQ supported grant and requires no licensing fees. Users are able to develop and securely distribute "queries" to network data partners and have data partners review, execute, and securely return the results of those queries via a secure web-based portal. Data partners exercise full control over the files they make available for querying. Network sizes range from single datasets held by only two organizations to multi-year projects encompassing multiple organizations and data resources.

The system includes two primary user applications, the Portal and the DataMart Client. The Portal is the user starting point for all information requests and controls all system communications, security, and governance policies. Data partners receive queries, process them, and securely return them to the Portal via their local DataMart Client.

The system currently supports two types of queries: 1) menu-driven queries that execute against summary tables; and 2) file distribution queries. Menu-driven queries are created by users based on a standardized query builder interface on the portal and then distributed to data partners. These queries can then be distributed and run against standardized tables that are created and maintained by the data partners. The software also supports querying against summary tables. The tables provide summary counts of individuals by period, age group, and sex. The summary counts include information on medication use, diagnoses, procedures, and the overall data partner population. The File Distribution Query allows users to securely distribute electronic files to data partners. Although any type of file can be distributed, one expected use is distribution of SAS and SQL programs and work plans to data partners who will download and execute the programs and then securely upload results based on institutional policies.

The PopMedNet system for distributed research networks is used by many projects, including the Scalable PArtnering Network for CER (SPAN). SPAN seeks to expand a distributed research network across multiple health care systems and sites, including the Kaiser Permanente system, the Group Health Collaborative in Seattle, Harvard Pilgrim Health Care in New England, Health Partners in Minnesota, the Geisinger Health Plan in Pennsylvania, and others. The network infrastructure will have the capability to conduct large CER studies using patient-reported outcomes data collected at the point of care (Agency for Healthcare Research and Quality 2009). More information on PopMedNet is available at http://www.popmednet.org/.

Policy-Based Approaches to Preserve Privacy and Confidentiality of PHI

As the technology and protocols for distributed data models and other approaches for virtual data sharing show there is great potential to advance the use of PHI for research while still protecting information and ensuring trust in the system. Possible models currently under discussion include:

- updating approaches to secure patient consent for research participation;
- creating a research 'safe harbor' for information-based research;
- developing encodable policy models for distributed research; and
- potentially revising the HIPAA Privacy Rule or the Common Rule (see 45 CFR§ Subpart A).

Updating the current consent procedures for patients that choose to make their clinical information available for research is one approach that has been promoted (Peddicord et al. 2010). In the current model, patients typically consent to be included in a single study under specific parameters, which is appropriate for clinical trials, but may limit the use of their information for observational studies. A new consent approach that allows patients to actively manage consent options electronically may allow for broader consent by patients for their clinical information to be included in research, but still under appropriate safety and privacy procedures. For example, approaches have been proposed that would allow patients to manage consent preferences through their electronic health record (EHR) or a dynamic consent management system.

The *creation of a research safe harbor* is another, arguably more ambitious approach (Peddicord et al. 2010). Using a research safe harbor, information-based research that meets specific criteria would not be bound by consent or IRB review for privacy (IRB review for ethical and scientific concerns would still be required), but rather provide specific criteria for the appropriate and ethical use of potentially identifiable data. This would include the implementation of meaningful security controls, such as unique researcher IDs, and robust password and access requirements. Researchers would be required to demonstrate compliance with the criteria and controls, and would also be subject to external audits of their procedures and risk mitigation controls.

An *encodable policy model for distributed research* provides a basis for institutions and study investigators to control access to data. The SCAlable National Network for Effectiveness Research (SCANNER) project, based out of University of California San Diego, is developing a scalable, flexible, secure distributed network infrastructure that enables near real-time CER among multiple sites. As a policy component the project is conducting a study, which includes a set of focus groups and interviews, that aims to incorporate the perspectives of all users of electronic data systems — including patients whose health care data may be incorporated into such models — about potential sharing of health information for research, and specifically about the implementation, use, and value of proposed CER networks. These findings will inform the development of an encodable and flexible policy model for distributed research. More information on SCANNER is available at http://scanner.ucsd.edu.

Most recently, many researchers and policymakers have suggested *potential changes to the HIPAA Privacy Rule*. One proposed approach would be to expand the data de-identification options available to researchers under the Privacy Rule (Center for Democracy and Technology 2009). Recognizing that different levels of data protections are appropriate for different contexts, this would include moving beyond the two common options for anonymity, to allow for additional data set options. These additional options would be linked

to specific research or operational purposes, and retain appropriate protections against re-identification.

As dialogue regarding these potential policy approaches continues, other short-term developments are likely to inform and guide the discussion. Recent *proposed changes to the Common Rule*, which governs human subjects research, seek to streamline and simplify the process (See 45 CFR Parts 46, 160, and 164). For instance, the advanced notice for the proposed rule allows for single IRB oversight for multi-site research studies. Single IRB oversight would likely minimize the burden on researchers; however, it will also require clear guidance and criteria on a number of issues including guidelines for single site IRB selection for multi-site studies, and contact information for questions, concerns or adverse events, as well as more global issues such as accountability and liability. In addition, the PCORI Methodology Committee[4] may address issues related to the use of PHI, as it seeks to balance the trade-offs between facilitating accurate and informative research and protecting patient information.

Conclusion

The strategies discussed in this brief demonstrate the challenges and emerging solutions to enable use of PHI for CER, particularly as the amount of electronic clinical data rapidly expands. As discussed, enabling access to representative data from multiple settings, while preserving individual privacy, is a key issue to address to ensure access to robust, timely data on what works best for whom, under what conditions.

As recent investments in the HIT infrastructure and CER manifest themselves, the tension between scientific discovery and data privacy will continue to grow. Regardless of the technical innovations or techniques used to combine or match data, researchers must be transparent in their uses of PHI and demonstrate how its inclusion may provide beneficial information. To truly engender public support, researchers and clinicians participating in clinical data networks should actively engage with patient and consumer stakeholders as they develop processes for the use of PHI to link multiple datasets and demonstrate how this data is protected within their systems. Engagement efforts that place emphasis on developing trusting relationships among all parties will likely help to inform policy and may even stimulate broader acceptance for the use of PHI among the general public.

About AcademyHealth

AcademyHealth is the leading national organization serving the fields of health services and policy research and the professionals who produce and use this important work. Together with our members, we offer programs and services that support the development and use of rigorous, relevant and timely evidence to increase the quality, accessibility, and value of health care, to reduce disparities, and to improve health. A trusted broker of information, AcademyHealth brings stakeholders together to address the current and future needs of an evolving health system, inform health policy, and translate evidence into action.

About the EDM Forum

The Electronic Data Methods (EDM) Forum is a three-year grant from the Agency for Healthcare Research and Quality (AHRQ) to facilitate learning and foster collaboration across a set of eleven comparative effectiveness research (CER) projects. Collectively, these projects are designed to build infrastructure and methods for collecting and analyzing prospective electronic clinical data. Specific areas of focus include the data governance, clinical informatics, and analytic issues that are crucial to the design and use of electronic clinical data for CER and PCOR. The EDM Forum, and the connected research projects, are funded by the American Recovery and Reinvestment Act (ARRA).

About the Authors

Raj Sabharwal, M.P.H., is a senior manager at AcademyHealth. He can be reached at raj.sabharwal@academyhealth.org. Erin Holve and Alison Rein are directors at AcademyHealth, and Courtney Segal is an associate at AcademyHealth.

Acknowledgements

AcademyHealth acknowledges the Agency for Healthcare Research and Quality (AHRQ) for its support of this work. The EDM Forum is supported by AHRQ through the American Recovery & Reinvestment Act of 2009, Grant U13 HS19564-01. AHRQ's mission is to improve the quality, safety, efficiency, and effectiveness of health care for all Americans. As 1 of 12 agencies within the Department of Health and Human Services, AHRQ supports research that helps people make more informed decisions and improves the quality of health care services. For more information, visit www.ahrq.gov.

Suggested Citation

Sabharwal, R., Holve, E., Rein, A, and Segal, C., "Approaches to Using Protected Health Information (PHI) for Patient-Centered Outcomes Research (PCOR): Regulatory Requirements, De-identification Strategies, and Policy," EDM Forum, AcademyHealth, March 2012.

Also see www.edm-forum.org to access publication.

Sources

Agency for Healthcare Research and Quality. Grant Summary - Scalable PArtnering Network for CER: Across Lifespan, Conditions, and Settings. [SPAN Grant Summary web page]. 2009; http://gold.ahrq.gov/projectsearch/grant_summary.jsp?grant=R01+HS19912-01. Accessed August 29, 2011.

Basic HHS Policy for Protection of Human Research Subjects. 56 FR 28012, 28022, June 18, 1991 (45 CFR§ Subpart A)

Benitez K, Loukides G, and Malin B. Beyond Safe Harbor: Automatic Discovery of Health Information and De-identification Policy Alternatives. Proceedings of the 1st ACM International Health Informatics Symposium, 2010. pp. 163-172

Bradley, CJ et al. Health Services Research and Data Linkages: Issues, Methods, and Directions for the Future. *Health Services Research*, Vol 45, No. 5, October 2010. pp.1468-1488.

Brown JS, et al. Distributed Health Data Networks: A Practical and Preferred Approach to Multi-Institutional Evaluations of Comparative Effectiveness, Safety, and Quality of Care. *Medical Care*, Vol. 48, No. 6, Suppl 1, June 2010. pp. S45-S51.

D'Avolio LW, Farwell WR, Fiore LD. Comparative effectiveness research and medical informatics. *Am J Med.* Dec 2010;123(12 Suppl 1):e32-37.

Diamond CC, Mostashari F, Shirky C. Collecting and sharing data for population health: a new paradigm. *Health Aff (Millwood)*. Mar-Apr 2009;28(2):454-466.

Encouraging the Use of, and Rethinking Protections for De-Identified (and "Anonymized") Health Data. Center for Democracy and Technology, June 2009.

Federal Coordinating Council for Comparative Effectiveness Research. Report to the President and The Congress. US Department of Health and Human Services June 2009.

Gliklich RE, Dreyer NA. Registries for evaluating patient outcomes : a user's guide. Rockville, MD: U.S. Dept. of Health and Human Services, Public Health Service, Agency for Healthcare Research and Quality; 2007.

Gunn PP et al. The Health Insurance Portability and Accountability Act Privacy Rule: A Practical Guide for Researchers. Medical Care. Vol 42, No. 4, April 2004. pp. 321-327.

Harvard Pilgrim Healthcare Institute. PopMedNet. [PopMedNet web site]. 2010; http://www.popmednet.org/. Accessed August 29, 2011.

Herdman R and Moses H, ed. 2006. Effect of the HIPAA Privacy Rule of Health Research. Washington, DC: Institute of Medicine.

HMO Research Network. Virtual Data Warehouse (VDW). Collaboration Toolkit: A guide to multicenter research in the HMO Research Network. 2011:16-20.

Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators; Notice of proposed rulemaking, 76 Federal Register 143 (26 July 2011), pp. 44512-44531.

Maro JC, Platt R, Holmes JH, et al. Design of a national distributed health data network. *Ann Intern Med.* Sep 1 2009;151(5):341-344.

Nass SJ, Levitt L, and Gostin LO, ed. 2009. *Beyond the HIPAA Privacy Rule.* Washington, DC: Institute of Medicine.

Pace WD, Cifuentes M, Valuck RJ, Staton EW, Brandt EC, West DR. An electronic practice-based network for observational comparative effectiveness research. *Ann Intern Med.* Sep 1 2009;151(5):338-340.

Peddicord D, et al. A Proposal to Protect Privacy of Health Information While Accelerating Comparative Effectiveness Research. Health Affairs, Vol. 29, No. 11, November 2010. pp. 2082-2090.

Rosenbaum, S. "Data Governance and Stewardship: Designing Data Stewardship Entities and Advancing Data Access." Health Services Research, Volume 45, Issue 5p2, pages 1442–1455, October 2010

The Health Insurance Portability and Accountability Act; Other requirements relating to uses and disclosures of protected health information. (45 CFR§ 164.514)

Endnotes

1. Comparative effectiveness research is the conduct and synthesis of research comparing the benefits and harms of different interventions and strategies to prevent, diagnose, treat and monitor health conditions in "real world" settings. – Federal Coordinating Council on CER, June 2009.

2. Data governance refers to the policies and procedures that determine how data are acquired, managed, aggregated, stored, and used. It is foundational to the conduct of rigorous research because governance structures help to establish trust. Governance plays a large part in generating confidence on the part of key stakeholders (including patients, consumers, and data partners) that there are appropriate measures in place to ensure the privacy and security of protected health information (PHI). - Rosenbaum, S. "Data Governance and Stewardship: Designing Data Stewardship Entities and Advancing Data Access." *Health Services Research*, Volume 45, Issue 5p2, pages 1442–1455, October 2010

3. Hashing algorithms are an identity management solution that maps large data sets of variable length to small data sets of a fixed length. Hashing algorithms are used to encrypt datasets so that the input cannot be determined from the output.

4. The 15 member PCORI Methodology Committee provides recommendations to the PCORI Board of Governors regarding methods for patient-centered outcomes research. This includes guidance about the appropriate use of methods in such research, methodological standards, as well as establishing priorities to address gaps in research methods or their application.