Data governance requirements for distributed clinical research networks: triangulating perspectives of diverse stakeholders

Katherine K Kim,[1,2] Dennis K Browe,[1] Holly C Logan,[1] Roberta Holm,[3] Lori Hack,[3] Lucila Ohno-Machado[4]

▶ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/amiajnl-2013-002308).

[1]San Francisco State University, Health Equity Institute, San Francisco, California, USA
[2]Betty Irene Moore School of Nursing, University of California Davis, San Francisco, California, USA
[3]Object Health, LLC, Walnut Creek, California, USA
[4]Division of Biomedical Informatics, University of California San Diego, La Jolla, California, USA

Correspondence to
Katherine K Kim,
San Francisco State University, Health Equity Institute, University of California Davis, Betty Irene Moore School of Nursing, 1900 Holloway Avenue, HSS-359, San Francisco, CA 94132, USA;
kathykim@sfsu.edu, kathykim@ucdavis.edu

Received 2 September 2013
Revised 4 November 2013
Accepted 17 November 2013
Published Online First 3 December 2013

To cite: Kim KK, Browe DK, Logan HC, *et al*. *J Am Med Inform Assoc* 2014;**21**: 714–719.

ABSTRACT

There is currently limited information on best practices for the development of governance requirements for distributed research networks (DRNs), an emerging model that promotes clinical data reuse and improves timeliness of comparative effectiveness research. Much of the existing information is based on a single type of stakeholder such as researchers or administrators. This paper reports on a triangulated approach to developing DRN data governance requirements based on a combination of policy analysis with experts, interviews with institutional leaders, and patient focus groups. This approach is illustrated with an example from the Scalable National Network for Effectiveness Research, which resulted in 91 requirements. These requirements were analyzed against the Fair Information Practice Principles (FIPPs) and Health Insurance Portability and Accountability Act (HIPAA) protected versus non-protected health information. The requirements addressed all FIPPs, showing how a DRN's technical infrastructure is able to fulfill HIPAA regulations, protect privacy, and provide a trustworthy platform for research.

BACKGROUND AND SIGNIFICANCE

With increased availability of large health datasets, networks allowing rapid advancements in knowledge[1] for comparative effectiveness research (CER), patient-centered outcomes research, and quality improvement research have emerged. A distributed research network (DRN) is one such network that allows researchers to use data from multiple institutions through controlled network functions rather than by direct integration of systems or export of datasets, thereby allowing local organizations to retain control of their own data,[2] reduce legal and privacy concerns, and better retain control over data.[3]

Several research networks have described general governance mechanisms[4]—for example, the HMO Research Network, MiniSentinel, and eMERGE. Others have designed system requirements, functions, features, or capabilities that are necessary for operation based on the input of one stakeholder group such as health system leaders.[5] Some authors have suggested that governance should include data privacy and security, audit requirements, conflicts of interest, financial strategies and sustainability plans, clear stewardship, standardization in protocols, and agreements.[2 3 6 7] There is limited federal and state regulatory and legal guidance for DRNs, except for HIPAA deidentification standards for secondary use of data.[8] Patient control is also gaining recognition in the design of networks. Surveys have shown that patients think that electronic health records protect privacy and security better than paper records,[9] but health information exchange worsens privacy and security.[10] Patients also desire granular control over data sharing, and prototype systems for managing that control are being developed.[11 12]

DRNs tend to be complex collaborations that engender interactions among multiple stakeholders, including federal and state governments, research centers and universities, commercial entities, healthcare organizations, and patient groups. Data governance requirements must relate to and satisfy these various groups politically and ethically while upholding strict technical standards for successful operation. While several networks have suggested their own practices,[13 14] there is little objective rationale for governance decisions.

OBJECTIVE

This article aims to address the lack of objective evidence for governance by contributing an example of the multi-stakeholder informed development of data governance requirements in the Scalable National Network for Effectiveness Research (SCANNER) which may be illustrative for DRN developers and participants. SCANNER is intended as a scalable and flexible DRN for managing interoperable research resources subject to governance rules. In this architecture, data remain at the originating sites, and only results of analytical processes are transmitted to a core network node. Given different access policies at participating institutions, policy enforcement is accomplished through a combination of local and network strategies, including encoding of policies in software whenever possible (figure 1).

METHODS AND MATERIALS

This project utilizes a triangulation method for developing data governance requirements by combining data from three related SCANNER projects: (1) policy analysis with experts; (2) focus groups with patients; (3) interviews with institutional leaders. This mixed method has the potential to improve internal validity in qualitative analysis,[15 16] increase completeness of data,[17] and evaluate health information technology.[18]

The first study involved policy analysis with experts and included comparison of privacy and security laws and state health information exchange guidelines and application of a Fair Information

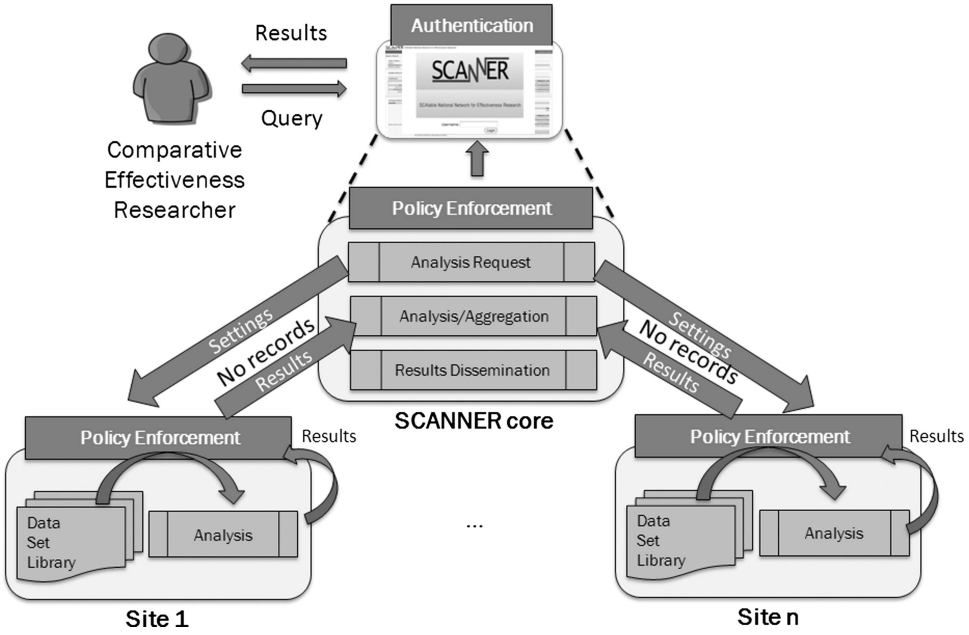

Figure 1 Example of a distributed research network conceptual architecture.

Practice Principles (FIPPs)-based framework[19] to CER use cases involving four data types: summary data consisting of counts, rates or statistical results only; deidentified data; limited dataset (LDS); and identified data. Results are reported elsewhere.[20]

In the second study, patient focus groups at three SCANNER medical centers were conducted to understand views on ethical issues in electronic data sharing. Patients expressed concerns related to altruism and personal benefit from data sharing, security, justice and the social responsibility of organizations conducting research, trust, and consent and authorization. The questions that were posed and the results have been published.[21]

Third, 15 interviews were conducted at three SCANNER institutions including two academic medical centers and one Veteran's Administration hospital. The respondents were senior staff in the institutional review board (IRB), privacy, compliance, information technology (IT) security, research informatics, and clinical research. A semistructured interview guide was developed and pretested with three members of the expert panel: a chief medical privacy and compliance officer, a research institute IRB specialist, and a health information exchange leader. The final interview guide (see online supplementary appendix) included a walkthrough of paper screenshots of the anticipated SCANNER design, and interviewees were asked for requirements that would need to be fulfilled for the organization's participation. All interviews were 1–3 h and held onsite in private offices, audio-taped, and transcribed. Individual names were removed from transcripts, but roles and institution names were retained.

Data governance requirements were generated iteratively. An initial set of legal and regulatory requirements was developed with the expert panel. Then another draft was developed from analysis of patient focus group and institutional interview transcripts. Two members of the research team independently coded potential requirements from a subset of transcripts. The initial codes were discussed in order to standardize them, and all transcripts were coded. Requirements were compiled and cross-listed by institution and role. Requirements that were identified by at least two respondents were automatically included, and all others were discussed by three researchers. Elements that were thought not necessary or 'nice to have' by respondents and those requirements identified by only one respondent were categorized as 'optional.'

RESULTS

Results of the investigation were predicated on the newly enacted policy that any entity that handles protected health information (PHI) is a business associate and is subject to some of the privacy, and all the security, rules of HIPAA.[22] With this in mind, the diverse set of 91 requirements generated from the analysis of stakeholder data was organized into two categories: (1) basic requirements for all data types; (2) data type requirements distinguishing between those for PHI and non-PHI. Table 1 details basic network requirements, which start with display of information for transparency, including a description of the network structure and governance, as well as a listing of participating institutions and studies being conducted. Additional requirements provide the ability to manage workflows based on key data elements in the network agreement, IRB protocols, and certificates of confidentiality where applicable. Access management is a multilayered process including approval of users by the participating institution supplemented by one-factor or two-factor authentication of identity, and enforcement of credentials for specific datasets according to IRB protocols. Finally, safeguards such as restrictions on devices, use of encryption, and strict segregation of datasets are required.

Table 2 summarizes selected DRN requirements by FIPPs and data type: summary, deidentified, LDS, and identified. The requirements illustrate the differences in functional requirements based on the data types. Under HIPAA, summary and deidentified data are not considered PHI, while LDS and identified data are. There are no requirements regarding summary data beyond basic sets. The requirements are cumulative going from left to right in the table —for example, requirements for identified

Table 1 Basic network requirements

Basic requirement	Description
1.0 Network information	Display of network information including leadership names, contacts, structure, monitoring, participation rules, decision-making guidelines, standard network agreement, user agreement on website for public
1.1 Institution information	Display of participating institution information including name, location on website for public
1.2 Study information	Display of study information including title, brief description, start/end date, PI name on website for public
1.3 Agreements	Availability of applicable agreements such as the DRN network agreement, IRB protocol. Display signed network agreement for users. Attributes of IRB approval: MOU/reliance/designation agreement approval—name of designee, approval date IRB protocol number IRB approval number IRB approval date Study attributes: hypothesis, reason or scientific question Study term (start, end, renewal) Data purge date (if different from study end date) Category of IRB approval (not human subjects research, exempt, expedited, full-review) Data type (deidentified, LDS, identified) PI name Key personnel names Presence of other agreements (eg, Material transfer agreement for specimen studies, Intellectual property agreement, Clinical trial agreement, certificate of confidentiality) Attributes may be linked to computable policies for data screening and study workflow management
1.4 Approved users	Creation of user accounts approved by institution Creation of report of user accounts by institution and study for monitoring and verification Attributes of user: Name Role (PI, institution system administrator, system administrator, researcher/staff) ID number (determined by institution) Official phone number Secondary phone number Email Addition of users to studies by PI authorization
1.5 Authentication and access	Authentication of user identity and credential to access specific datasets Acceptance of user agreement One-factor authentication (unique username, password) for data accessed within one institution Two-factor authentication (verification with second credential) for data accessed across institutions such as use of tokens
1.6 Data use	Enforcement of data use at granular level by role and IRB attributes Alerts to user regarding data use prohibitions and restrictions
1.7 Audit and accounting	Tracking of access by dataset, date, time, user, method of access Audit logs Searchable accounting of specific disclosures Searchable accounting of accesses by user
1.8 Patient rights	Ability to confirm patient consents and authorizations, manage changes, revocations and removals for any identified data stored in the network Management of consent for health information exchange per state law
1.9 Security	Encryption of data at rest, in transit, and in use (HIPAA *[45 CFR § 164.312 (a)(2)(iv)], current NIST* Cryptographic Module Validation Program (CMVP) level http://csrc.nist.gov/groups/STM/cmvp/index.html) Users who use portable or personal devices must be able to comply with full requirements
1.10 Data segregation	Ability to manage datasets discretely such that they are not co-mingled

DRN, distributed research network; HIPAA, Health Insurance Portability and Accountability Act; IRB, institutional review board; LDS, limited dataset; MOU, memorandum of understanding.

data subsume LDS, deidentified data requirements, and general network.

There are basic requirements that apply to the first two FIPPS, (A) 'individual access' and (B) 'correction'. Patient-level research data are usually not centrally stored in DRNs such as SCANNER, hence contact with individuals and maintenance of changes for PHI are not relevant for SCANNER. (C) 'Openness and transparency' are handled by public disclosure regarding the network and its participants and study topics via the network website. (D) 'Individual choice' principle, consent and authorization are generally handled at the institution at the time of data collection and are not under the network's responsibility. However, the network requirement is to assure that institutions attest that notification, consent or authorization was appropriately obtained. In (E) 'collection, use and disclosure limitation', the IRB attributes are used to screen for the level of identification, and assure that use is approved for the study and the user. In the case of SCANNER, there is currently no downloading or saving of identified data allowed. Requirements to segregate datasets from each other and prohibit the combination with other data help assure the data are not contaminated, satisfying the (F) 'integrity' requirement. (G) 'Accountability' is managed through a network agreement, a user agreement, and audit capabilities. Enhanced accountability for PHI (whether stored or only transmitted) is codified in an HIPAA business associate agreement and fulfillment of the required breach notification and disclosure provisions. Finally, (H) 'safeguards' are implemented through many of the requirements described above as well as additional screening of data for identifiers.

Table 2 Distributed network requirements by FIPPs and data type

Extended FIPP	Data type-specific requirements			
	1. Basic network requirements from table 1	2. Deidentified data (non-PHI)	3. Limited dataset (PHI)	4. Identified data (PHI)
A. Individual access Individuals should be provided with a simple and timely means to access and obtain their individually identifiable health information in a reliable form and format	1.8 Patient rights	Not applicable	Not applicable	Not applicable to research-only data*
B. Correction Individuals should be provided with a timely means to dispute the accuracy or integrity of their individually identifiable health information, and to have erroneous information corrected or to have a dispute documented if their requests are denied	1.8 Patient rights	Not applicable	Not applicable	Not applicable to research-only data*
C. Openness/transparency There should be openness and transparency about policies, procedures, and technologies	1.0 Network website information 1.1 Institution information 1.2 Study information			
D. Individual choice Individuals should be provided a reasonable opportunity and capability to make informed decisions about the collection, use, and disclosure of their health information			3.1 Verification that a notice of patient privacy has disclosed use of LDS for research if allowed by state law	4.1 Attestation that consent has been collected or waiver has been granted
E. Collection, use, and disclosure limitation Information should be collected, used and/or disclosed to the extent necessary to accomplish a specified purpose(s) and never to discriminate inappropriately	1.6 Data use	2.1 Verify IRB start and termination dates before allowing access to data	3.2 Enter IRB approval number, verification that LDS is allowed by protocol	4.2 Prohibition on downloading of dataset 4.3 IRB approval number, verification that identified dataset is allowed by protocol 4.4 Analysis of data conducted behind data source firewall with view-only access to results via network portal*
F. Integrity Persons and entities should take reasonable steps to ensure that health information is complete, accurate, and up-to-date to the extent necessary for the person's or entity's intended purposes and has not been altered or destroyed in an unauthorized manner	1.10 Data segregation	2.2 Prohibition on combining with other datasets for reidentification		
G. Accountability These principles should be implemented, and adherence assured, through appropriate monitoring, and other means and methods should be in place to report and mitigate non-adherence and breaches	1.3 Agreements 1.7 Audit and accounting		3.3 Verification that a valid data use agreement is in effect 3.4 Signed HIPAA business associate agreement. Investigation, reporting of breaches and accounting of disclosures	
H. Safeguards Health information should be protected with reasonable administrative, technical and physical safeguards to ensure its confidentiality, integrity, and availability and to prevent unauthorized or inappropriate access, use, or disclosure	1.4 Approved users 1.5 Authentication and access 1.9 Security and encryption	2.3 Screen for 18 identifiers and restrict publication of dataset if any are present	3.5 Screen for identifiers and restrict publication of dataset if any elements except for year and zip code are present	4.5 Disallow temporary (eg, datamarts) or long-term storage of data*

*Applicable to the current configuration of the SCANNER network in which PHI is maintained behind an institutional firewall and only results of analyses are transmitted as shown in figure 1.
FIPP, Fair Information Practice Principle; HIPAA, Health Insurance Portability and Accountability Act; IRB, institutional review board; LDS, limited dataset; PHI, protected health information; SCANNER, Scalable National Network for Effectiveness Research.

DISCUSSION

The investigation and generation of requirements for a DRN such as SCANNER yielded a large number from legal and regulatory sources, as well as the diversity of opinions expressed by stakeholders. Many of the basic network requirements were generated from the initial expert panel legal and regulatory analysis, but others were added from the patient and institutional data analysis. For example, detailed network, institutional and study information (requirements 1.0, 1.1, 1.2) was added from the patient focus group data. Similarly, enforcement of data use by IRB parameters (1.6) and the ability to search for access by user, date time, and other parameters for audit purposes (1.7) were added from institutional interview data. Attestation that consent has been collected or waiver has been granted for PHI (4.1) is an example of a requirement that was generated as a result of combining all three stakeholder perspectives. Regulation

requires that consent or a waiver is required for PHI used in research. Patient focus groups highlighted their need to consent for data sharing, and institutional interviewees required that data-sharing partners confirm that this was fulfilled.

These results may help DRN developers to construct appropriate governance, users to assess whether DRNs meet their needs, and patients to determine the trustworthiness of DRN operators.

Some requirements can be met by technology, and others can be met by contracts, attestation of users, or management supervision. The SCANNER team is actively pursuing solutions for requirements of deidentified and LDS, acknowledging that some solutions predate the development of SCANNER (eg, establishing accounts, obtaining informed consent). The remaining requirements will be placed in a development roadmap that balances stakeholder priorities, resources, and availability and potential integration of alternative tools and solutions.

There are other important data governance issues that were not addressed in this project. For example, 'anonymization' or obfuscation of data for secondary use is a critical area of investigation.[23–29] However, most DRNs expect that the data supplier would apply these tools before making the data accessible via the network. It is possible that a network could make research participant identity available depending on permissions. Like most other DRNs, SCANNER's architecture maintains PHI within the data supplier's firewall, and only results of analyses are exchanged through the network, thereby staying within the HIPAA safe harbor limits as recommended by privacy and legal experts.[20]

The requirements described here are one example of a process for integrating stakeholder perspectives into the development of data governance mechanisms in DRNs. Many of these requirements apply to other DRNs, but some that operate under different state regulations and have different purposes — for example, networks involving for-profit institutions—may have a different set of policies. This study was limited by the small number of stakeholders involved in the three projects (seven experts, 36 patients, and 15 institutional interviewees). However, we noted convergence among them, supporting the conclusion that the requirements generated are valid. Future studies using the same methodology will help to determine the applicability of our findings to other DRNs.

CONCLUSION

DRNs can protect privacy and fulfill legal/regulatory guidelines through a combination of strategies, including agreements, alerts to users of their obligations and attestations of compliance, and development of network technical capabilities for screening, monitoring, and enforcement. Generation of DRN requirements is often a process run by technical and scientific personnel with limited input from stakeholders. As DRNs seek to become more relevant to clinicians and patients and sustained by their institutions, it is prudent to engage stakeholders in developing systems that meet their needs.

Acknowledgements We thank Zia Agha, Tania Zamora, Susan Robbins, Fred Resnic, and Aziz Boxwala for assistance with interviewee recruitment, Michele Day for figure 1 and overall program management, and Jill Joseph, Jyu-lin Chen, and Robert Bell for comments on the draft manuscript.

Contributors KKK had responsibility for the conduct of the study, including performing all interviews, overseeing the analysis, and having access to the data. RH, DKB, and HCL coded and analyzed interview transcripts and verified requirements. LH verified requirements. LO-M participated in analysis. All contributed to writing and approved the manuscript.

Funding This work was funded by the Agency for Healthcare Research and Quality (AHRQ) Grant R01HS19913. KKK was supported by the Gordon and Betty Moore Foundation grant to the Betty Irene Moore School of Nursing at UC Davis.

Competing interests None.

Ethics approval The protocols for all studies were approved by the IRB at San Francisco State University, and additionally for focus groups and interviews by the healthcare institutions.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Ohmann C, Kuchinke W. Future developments of medical informatics from the viewpoint of networked clinical research. Interoperability and integration. *Methods Inf Med* 2009;48:45–54.
2 Brown J, Holmes J, Maro J, *et al*. *Design specifications for network prototype and cooperative to conduct population-based studies and safety surveillance. No 13 (Prepared by the DEcIDE Centers at the HMO Research Network Center for Education and Research on Therapeutics and the University of Pennsylvania Under Contract No HHSA29020050033I T05)*. Rockville, MD: Agency for Healthcare Research and Quality, 2009.
3 Brown JS, Holmes JH, Shah K, *et al*. Distributed health data networks: a practical and preferred approach to ulti-institutional evaluations of comparative effectiveness, safety, and quality of care. *Medical Care* 2010;48:S45–51.
4 Ohno-Machado L, Day M, Kim H, *et al*. *Standards in the use of collaborative data networks or distributed data networks in patient centered outcomes research*. Technical Report for PCORI RFAs. Washington, DC: Patient Centered Outcomes Research Institute, 2013.
5 Manion FJ, Robbins RJ, Weems WA, *et al*. Security and privacy requirements for a multi-institutional cancer research data grid: an interview-based study. *BMC Med Inform Decis Mak* 2009;9:31.
6 Grethe JS, Baru C, Gupta A, *et al*. Biomedical informatics research network: building a national collaboratory to hasten the derivation of new understanding and treatment of disease. *Stud Health Technol Inform* 2005;112:100–9.
7 Fullerton S, Anderson N, Guzauskas G, *et al*. Meeting the governance challenges of next-generation biorepository research. *Sci Transl Med* 2011;2:15cm3.
8 McGraw D. Building public trust in uses of Health Insurance Portability and Accountability Act de-identified data. *J Am Med Inform Assoc* 2013;20:29–34.
9 National Partnership for Women & Families. *Making IT meaningful: how consumers value and trust health IT*. Washington, DC: National Partnership for Women & Families, 2012:32–3.
10 Ancker JS, Silver M, Miller MC, *et al*. Consumer experience with and attitudes toward health information technology: a nationwide survey. *J Am Med Inform Assoc* 2013;20:152–6.
11 Caine K, Hanania R. Patients want granular privacy control over health information in electronic medical records. *J Am Med Inform Assoc* 2013;20:7–15.
12 Ge Y, Ahn DK, Unde B, *et al*. Patient-controlled sharing of medical imaging data across unaffiliated healthcare organizations. *J Am Med Inform Assoc* 2013;20:157–63.
13 Pencarrick Hertzman C, Meagher N, McGrail KM. Privacy by design at population data BC: a case study describing the technical, administrative, and physical controls for privacy-sensitive secondary use of personal information for research in the public interest. *J Am Med Inform Assoc* 2013;20:25–8.
14 Bradford W, Hurdle JF, Lasalle B, *et al*. Development of a HIPAA-compliant environment for translational research data and analytics. *J Am Med Inform Assoc* 2014;21:185–9.
15 Mays N, Pope C. Assessing quality in qualitative research. *BMJ* 2000;320:50–2.
16 Morse JM. Approaches to qualitative-quantitative methodological triangulation. *Nurs Res* 1991;40:120–3.
17 Shih FJ. Triangulation in nursing research: issues of conceptual clarity and purpose. *J Adv Nurs* 1998;28:631–41.
18 Ammenwerth E, Iller C, Mansmann U. Can evaluation studies benefit from triangulation? A case study. *Int J Med Inform* 2003;70:237–48.
19 Office of the National Coordinator for Health Information. Nationwide Privacy and Security Framework For Electronic Exchange of Individually Identifiable Health Information. Secondary Nationwide Privacy and Security Framework For Electronic Exchange of Individually Identifiable Health Information. 2008. http://healthit.hhs.gov/portal/server.pt/community/healthit_hhs_gov__privacy___security_framework/1173
20 Kim KK, McGraw D, Mamo L, *et al*. Development of a privacy and security policy framework for a multistate comparative effectiveness research network. *Med Care* 2013;51(8 Suppl 3):S66–72..
21 Mamo LA, Browe DK, Logan H, *et al*. Patient informed governance of distributed research networks: results and discussion from six patient focus groups. *American Medical Informatics Association Annual Symposium*; 2013.
22 McGraw D. Coping with, and Taking Advantage of, HIPAA's New Rules. Secondary Coping with, and Taking Advantage of, HIPAA's New Rules [Presentation]. 2013. http://idash.ucsd.edu/events/webinars/coping-and-taking-advantage-hipaas-new-rules

23 Gardner J, Xiong L, Xiao Y, *et al*. SHARE: system design and case studies for statistical health information release. *J Am Med Inform Assoc* 2013;20:109–16.
24 Jiang X, Sarwate AD, Ohno-Machado L. Privacy technology to support data sharing for comparative effectiveness research: a systematic review. *Med Care* 2013;51(8 Suppl 3):S58–65.
25 Ohno-Machado L. To share or not to share: that is not the question. *Sci Transl Med* 2012;4:165cm15.
26 Ohno-Machado L, Bafna V, Boxwala AA, *et al*. iDASH: integrating data for analysis, anonymization, and sharing. *J Am Med Inform Assoc* 2012;19:196–201.
27 Malin BA, Emam KE, O'Keefe CM. Biomedical data privacy: problems, perspectives, and recent advances. *J Am Med Inform Assoc* 2013;20:2–6.
28 Malin B, Benitez K, Masys D. Never too old for anonymity: a statistical standard for demographic data sharing via the HIPAA Privacy Rule. *J Am Med Inform Assoc* 2011;18:3–10.
29 Dwork C, Pottenger R. Toward practicing privacy. *J Am Med Inform Assoc* 2013;20:102–8.