Facilitating Health Data Sharing Across Diverse Practices and Communities

Ching-Ping Lin[1], Robert A. Black MA[1], Jay LaPlante[1], Gina A. Keppel MPH[1],
Leah Tuzzio MPH[2], Alfred O. Berg MD[1], Ron J. Whitener JD, Dedra S. Buchwald MD[1],
Laura-Mae Baldwin MD[1], Paul A. Fishman PhD[2], Sarah M. Greene MPH[2],
John H. Gennari PhD[1], Peter Tarczy-Hornoch MD[1], Kari A. Stephens PhD[1]
[1]University of Washington, Seattle, WA; [2]Group Health Research Institute, Seattle, WA

Abstract

Health data sharing with and among practices is a method for engaging rural and underserved populations, often with strong histories of marginalization, in health research. The Institute of Translational Health Sciences, funded by a National Institutes of Health Clinical and Translational Science Award, is engaged in the LC Data QUEST project to build practice and community based research networks with the ability to share semantically aligned electronic health data. We visited ten practices and communities to assess the feasibility of and barriers to developing data sharing networks. We found that these sites had very different approaches and expectations for data sharing. In order to support practices and communities and foster the acceptance of data sharing in these settings, informaticists must take these diverse views into account. Based on these findings, we discuss system design implications and the need for flexibility in the development of community-based data sharing networks.

Introduction and Background

A key aim of the National Institutes of Health (NIH) Roadmap is to broaden the participation of communities and practice-based care settings in medical and health services research, both to increase the capability to mount large-scale clinical studies with a diversity of participants and to accelerate the integration of new findings into care practices.[1] Without bi-directional translational pathways between scientific discoveries and primary care, neither individual patient care nor population health will change.[2] Practice-based research networks have been valuable environments for describing health disparities, framing care guidelines for primary care settings and increasing the external validity of research.[3, 4] A major challenge facing researchers when working with practices, especially practices in rural areas, is the complexity of creating valid study designs that take into account the high cost of travel, recruitment and data collection at multiple sites as well as the availability of statistically significant sample sizes.[4, 5] We propose to address this challenge through the creation of locally-situated and controlled clinical data repositories capable of sharing data with outside researchers.

Querying across heterogeneous data sets that appear integrated as single data source is known as *federated querying*.[6] Our approach will semantically align repositories within a research network to support federated queries across multiple community-based practices.

A recipient of an NIH Clinical and Translational Science Award, The Institute of Translational Health Sciences (ITHS) has developed the LC Data QUEST (Locally Controlled Data Query, Extraction, Standardization and Translation) pilot project to create research networks that can perform federated health data queries across network members. LC Data QUEST is a partnership between ITHS and two key communities in the Washington, Wyoming, Alaska, Montana, Idaho (WWAMI) region: American Indian and Alaska Natives (AI/AN) and community-based primary care practices, initially represented by those in the UW Family Medicine Residency Network (FMRN). Our goal for these partnerships is to link health researchers with the large, geographically dispersed communities that represent extraordinary diversity across race/ethnicity, culture, rural/urban location, geography, health service delivery and financing systems and the health status of their members. Simultaneously, participating communities and practices will determine their own research priorities and offer opportunities to bring research expertise to bear on pressing and often unaddressed health issues.

As a first step towards building trust, pioneering federated query projects such as the electronic Primary Care Research Network (ePCRN) and the Shared Health Research Information Network (SHRINE) have focused on anonymized aggregate count data and granting access to a limited set of trusted researchers.[7, 8] Providing accurate aggregate count data as an introductory step towards broader data sharing can help partners can realize the benefits of relevant outcome measures or increase the efficiency of identifying eligible study participants from electronic medical records (EMRs) while limiting the risk to their data, institution or patients. Because of the challenges in building trust, developing governance, operational processes, shared data elements and ontological mappings, we have chosen to follow this in-

cremental model for the implementation of LC Data QUEST.

We are developing data sharing capabilities at three FMRN and three AI/AN practices. This initial pilot project serves as a proof-of-concept for the utility and feasibility of a federated query network among practices across the WWAMI region. The first phase of the pilot is limited to aggregate count data with the ultimate goal of sharing broader, patient-level data sets in subsequent phases. We will implement a technical infrastructure to create data repositories at partner sites by locating, extracting and aligning EMR data. We will also implement a federated query tool to access the data repositories and return aggregate counts. As these technical foundations are put into place, governance, training, and research support will increase research capacity at community practice sites.

In order to select our six pilot sites, we visited ten candidate practices and communities to determine their technical and institutional readiness for LC Data QUEST. We also sought to understand what our partners hoped to gain from sharing and combining their data with other practices. These evaluations helped us assess feasibility and identify barriers and challenges to implementing a federated query project. From these conversations, a picture emerged that the FMRN and AI/AN practices envisioned different data sharing models and definitions of research networks that reflect their distinct goals and research priorities. These differing models significantly influenced the system requirements we developed to serve both types of communities concurrently. While commercial database products supporting distributed data systems exist readily on the market, we recognize that to support a distributed clinical data sharing system, we need to satisfy local priorities, values and governance requirements.[8]

Setting

We evaluated five FMRN and five AI/AN practices distributed across the WWAMI region (Washington (5), Montana (1), Idaho (2), Alaska (2)). The family medicine practices are independent members of the University of Washington Family Medicine Residency Network training program. On average, each practice supports 30-35 clinical providers, including both faculty and residents, each working a variable number of half days in the clinic. The AI/AN practices included two general types. The first type were

practices for which the tribe receives funding from the United States to operate, but is managed completely by the tribe itself. The second type were clinics funded and managed by the Indian Health Service, a division of the United States Health and Human Services. Substantially smaller than FMRN practices, they support 3-5 providers. A few practices in both AI/AN and FMRN settings have quality improvement resources and procedures that include EMR data analysis, but none have developed their own data repositories for the secondary use of health data.

In identifying stakeholders and leaders necessary to support and authorize a data sharing project, we spoke to a diverse set of practice leaders including providers, technical staff, and administrators. AI/AN communities also required forming partnerships with tribal leaders to gain trust and ensure protection of tribal sovereignty in relation to health data.

Data Sharing in Local Context

Both FMRN and AI/AN practices share a goal for improving their patients' health and increasing their patients' accessibility to clinical trial participation. The FMRN practices, as training programs, envision a research network in which clinicians and clinical researchers (either local to the practice or at a remote academic center) partner to develop and evaluate study questions and feasibilities.

A key motivation for FMRN practices is the opportunity to collaborate with colleagues to further medical knowledge and develop novel medical research. Also important to these practices is leveraging the data in the EMR at both the individual and population level for quality improvement. Potential projects may come from remote academic researchers, local practitioners/researchers or a resident working on a quality improvement or clinical research requirement at his/her residency training program. As seen in Figure 1a, academic researchers partner with FMRN practices within the research network. As projects are proposed, the research network determines whether there is interest and willingness among practices to participate.

AI/AN communities are also interested in the investigation of relevant research questions. However, in contrast to the FMRN sites, they also envision combining health data to garner funding and services for common health goals. AI/AN sites in the WWAMI region represent small populations often of
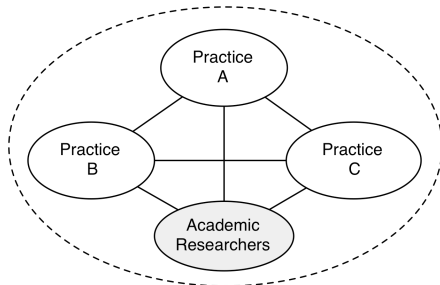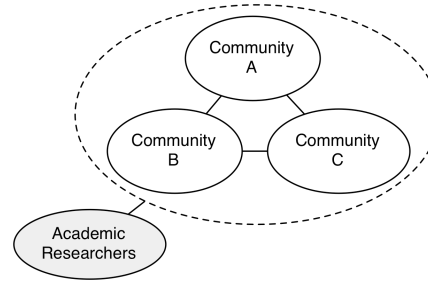
Figure 1a. Family Medicine Residency Network

Figure 1b. American Indian/Alaska Native Communities

Figure 1: Contrasting views of research networks. 1a shows that for FMRN members, practices collaborate collegially with academic researchers. In contrast, 1b shows that for AI/AN communities, academic researchers are viewed as outside the network, but communities may choose to partner with them if projects are locally beneficial.

fewer than 10,000, 5,000, or even 1,000 members, The ability to leverage semantically aligned data and expedient data extractions across tribal practices affords the development of programs, grant, and research opportunities that otherwise would be difficult, if not impossible. The combined populations of two or more tribal communities represent a larger sampling population for clinical research, inferring higher statistical power.

AI/AN communities have had a contentious relationship with outside researchers, especially in the area of data sharing and publication of results without community oversight.[9, 10] From our discussions with these communities, it is clear that the research process is viewed with suspicion and mistrust given harm endured by their people from historical exploitation from the research community.[11] The nature of scientific research, namely the priority given to benefiting the researcher rather than the participants, is often at odds culturally with the priorities of the tribal community. Figure 1b shows academic researchers are viewed as being outside the AI/AN research network.

Despite this tension, AI/AN communities were willing to engage in this project as it contains a core value of community control over the research process, data and results. Therefore, this health improvement project was developed in partnership with tribal communities and reflects community interests. Any future research project using LC Data QUEST will be vetted through a tribal authorization process to determine whether it is acceptable, non-detrimental, and of benefit to the health and well being the community. In addition, regular updates and reporting to tribal leadership is required through tribal resolutions and data sharing agreements. Figure 1b illustrates that with community control and appropriate governance, AI/AN communities are willing to partner with academic researchers to benefit from resources and

expertise that address community-defined health priorities.

Systems Requirement Implications

From our site visits and discussions, we developed a set of system requirements for a query tool supporting the initial aggregate count phase of LC Data QUEST. While there will be a common technical infrastructure to support both FMRN and AI/AN practices, the operational processes and rules may vary site to site. We recognize the diverse priorities of our partner sites and we were additionally sensitive to the following concerns:

- *Preventing researchers fishing for research questions*: Researchers with no specific research question should not use the data network to haphazardly query for ideas.

- *Sensitive or stigmatized diseases*: Queries relating to mental illness, substance abuse, sexually transmitted infections and other sensitive health areas may need to be more carefully governed.

Local Control

To gain the support and trust of partners, our system requirements support the philosophy of local control. As a result, we have developed the following system requirements:

- All data repositories will reside locally. This approach differs from a central repository solution that stores copies of all sharable data from multiple practices for aggregate analysis.
- Each specific query must be vetted and sanctioned by communities or practices through appropriate Institutional Review Boards (IRBs). At the AI/AN tribes, this process will include a tribal review process in addition to practice review. This differs from other federated query projects such as SHRINE whose institutions have ap-

proved all queries on the available data by named researchers.[8]

- Practices can review the query and results before returning the data to the requester, including previously authorized requesters.
- Practices must be able to withdraw their data repository from the data sharing network at any time.
- All queries must be logged and audited locally.
- Practices must be able to query their own repositories. This may be an unnecessary requirement for academic institutions or large hospitals with existing data repositories, but for many community practices, LC Data QUEST will represent their first transformed and aligned data source.

Our support of local control is both due to practices' and tribes' wishes and also because we believe it will better facilitate the expansion of LC Data QUEST to additional sites. Practices and communities will be more receptive to participating in LC Data QUEST if they are assured that all data reside locally except for those they choose to share. The ability for practices to detach and re-attach their practice from the network at-will similarly limits the risk and exposure from joining.

System requirements such as auditing and logging of queries translate more directly to technical system specifications while others could be implemented using operations processes such as restricting users to a single, trusted operator. Careful consideration of the available human and technical resources as well as site preferences will determine how these requirements are met. Certainly the balance between technical and human roles in the operational processes may change in future redesigns.

De-identification of Data

In order to protect the identity of practices, tribes, and patients, the aggregate count results will not be attributed to or broken down by practice. If further authorization is obtained, practices can allow themselves to be identified for further contact with researchers. Supporting a conservative and flexible model allows for variable security and boundary paradigms. For instance, practices may develop different policies depending if researchers come from within the community or practice, within the research network or from the outside.

Discussion

The development of LC Data QUEST's federated query system, governance and operational processes will provide new avenues for clinical researchers as well as empower local communities to address their own health concerns and facilitate practices' ability to improve quality and implement practice innovations. We are committed to partnering with practices to serve local needs and honor individual autonomy, local sensitivities and values to fully ensure that sites are invested in the goals of the networks so that they are not exploited for research experimentation purposes. This is not merely an ethical position or method to garner support from practices, but key to the long-term financial and institutional sustainability of the research networks.[12]

From an informatics perspective, practice and community support are critical for maintaining data quality and by extension, the efficacy of the data sharing system. In this paper we did not discuss the substantial task of locating, extracting, cleaning and performing ontological mapping of the raw EMR data to build the foundational data repositories. Several of the pilot practices use the same EMR technology, but as other projects have reported, this does not eliminate the effort required to create meaningfully comparable data due to differences in workflow or coding nor does it address the issue of inaccurate or incomplete medical data.[8, 13]

Projects such as Distributed Ambulatory Research in Therapeutics Network (DARTNet), a data sharing network of health practices located in Colorado, have learned that an important strategy for creating quality data and reducing data gaps is to include providers and practices in a quality assurance process.[13] This can be done through a point-of-care decision support tool or periodic data quality checking. Regardless of the method, maintaining data quality requires ongoing commitment from practices.

An additional challenge we face is validating the data sharing network. We have been working with practices to develop a set of clinically relevant, prototypical queries that will allow us to test several types of data sets across the sites. Examples include cohort discovery for grant proposals and compiling medical health data relevant to defined cohorts across multiple sites for research and quality assurance processes.

We have reported the general reflections of our conversations with practices and communities. We describe how data sharing research networks are perceived given current work practices, social relationships and in the cases of tribes, political and legal requirements. In reality, the introduction of LC Data QUEST represents a novel information technology intervention that may lead to as yet unexplored partnerships and uses. As we have outlined, new technical capabilities and resulting quality improvement measures may influence clinical workflow. New

partnerships may form between AI/AN and FMRN practices based on geographic area or common research interests. Extracting and mapping the same data elements for both communities in the pilot and collaborating closely with sites will support these new possibilities.

We began with descriptive requirements gathering to develop a system design. The implementation of a new system would immediately produce possibilities that lead to new work practices and new requirements. In short, we recognize that our design necessitates the flexibility to support an evolving landscape. Our system requirements support tighter control in the initial stages to build a trusted foundation. Yet we leave the door open for more streamlined processes in the future as the LC Data QUEST matures and includes more practices, broader data types and patient-level data.

Conclusion

We have reported on our initial site evaluations and system requirements for building research networks across the WWAMI region with the capacity to share semantically aligned clinical data. Our partner sites include family medicine residency and American Indian/Alaskan Native practices and communities representing geographically dispersed, often underserved, diverse and rural populations. Research networks in these settings have enormous potential such as increasing recruitment of study participants and introducing new programs and therapies into communities. Motivations for network participants include greater access to clinical trials, the furthering of medical knowledge, increased access to data that can improve practice quality and function, and the opportunity for improving community health. Our two types of communities have diverse perceptions of research networks and the role of outside academic researchers. As a result, we have developed our system requirements to respect local values, regulations, sensitivities and objectives. Finally, we recognize that LC Data QUEST affords new possibilities for collaboration and work practice and so we anticipate our requirements and design will evolve as these research networks mature.

Acknowledgements

References

1. Zerhouni EA. US Biomedical Research: Basic, Translational, and Clinical Sciences. JAMA. 2005;294(11):1352-8.

2. Westfall JM, Mold J, Fagnan L. Practice-Based Research--"Blue Highways" on the NIH Roadmap. JAMA. 2007;297(4):403-6.

3. Nutting PA, Beasley JW, Werner JJ. Practice-Based Research Networks Answer Primary Care Questions. JAMA. 1999;281(8):686-8.

4. Lindbloom EJ, Ewigman BG, Hickner JM. Practice-Based Research Networks: The Laboratories of Primary Care Research. [Article]. Medical Care. 2004;2(4 suppl):III-45-III-9.

5. American Academy of Family Physicians. Methods for Practice-Based Research Networks: Challenges and Opportunities. Practice-Based Research Networks Methods Conference; 2001; San Antonio, TX; 2001.

6. Schatz B, Mischo WH, Cole TW, Hardin JB, Bishop AP, Hsinchun C. Federating diverse collections of scientific literature. Computer. 1996;29(5):28-36.

7. Peterson KA, Fontaine P, Speedie S. The Electronic Primary Care Research Network (ePCRN): A New Era in Practice-based Research. J Am Board Fam Med. 2006;19(1):93-7.

8. Weber GM, Murphy SN, McMurry AJ, et al. The Shared Health Research Information Network (SHRINE): A Prototype Federated Query Tool for Clinical Data Repositories. J Am Med Inform Assoc. 2009;16(5):624-30.

9. Dalton R. When two tribes go to war. Nature. 2004;430(6999):500-2.

10. Weiss KM. Richard H. Ward, Ph.D. (June 7, 1943-February 14, 2003): wild ride of the Valkyries. Am J Hum Genet. 2003;72(5):1079-83.

11. Hodge FS, Weinmann S, Roubideaux Y. Recruitment of American Indians and Alaska Natives Into Clinical Trials. Annals of Epidemiology. 2000;10(8, Supplement 1):S41-S8.

12. Mold JW, Peterson KA. Primary Care Practice-Based Research Networks: Working at the Interface Between Research and Quality Improvement. Ann Fam Med. 2005;3(suppl_1):S12-20.

13. 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. 2009;151(5):338-40.