Informatics synergies between PaTH and ACT networks

Waqas Amin, MD, Charles Borromeo, MS, Melissa Saul, MS, Michael J Becich, MD, PhD, Shyam Visweswaran, MD, PhD

University of Pittsburgh, Pittsburgh, PA

Summary: We present two clinical data research networks (CDRNs) that are being developed to enable sharing and analyses of patient data across multiple healthcare institutions. Both networks are implementing the i2b2 software tool to store patient data at each institution using centrally defined common data elements, and the SHRINE software tool to perform distributed queries across the network.

Background: To support clinical research, large scale cohort discovery, data sharing, and integration across multiple healthcare institutions is critical. Various institutions have established local clinical research data repositories, and extending these to operate across multiple institutions will enable rapid identification of large cohorts for common conditions as well as accrual of sufficient numbers of patients for rarer diseases. However, major challenges arise due to variability in the source electronic health record (EHR) systems, in semantic interoperability and consistency of data elements, and in governance and regulatory policies across institutions.

The Department of Biomedical Informatics (DBMI) at the University of Pittsburgh is centrally involved in two initiatives that will establish multi-institutional CDRNs. The first initiative is PaTH (University of Pittsburgh, Penn State, Temple University and John Hopkins University) [1] that is funded by the Patient Centered Outcomes Research Institute (PCORI) and the second initiative is ACT (Accrual of patients to Clinical Trials) that is funded by the National Clinical and Translational Science Award (NCATS). The goal of PaTH is to identify and share data on three targeted conditions (idiopathic pulmonary fibrosis (IPF), atrial fibrillation and obesity) across four institutions, while the goal of ACT is to enable rapid identification of patients for accrual to the nation's highest priority clinical trials from 20 institutions. Both initiatives will establish CDRNs by taking advantage of the extensive availability of EHR data and well-established informatics and regulatory expertise at DBMI.

Methods: The two initiatives are synergistic and are guided by three principles: (1) the ability to do exploratory cohort search across multiple sites; (2) the ability to use de-identified data to analyze cohorts; and (3) the capability to re-identify a patient for future enrollment in a trial or to obtain additional data. To realize these principles, both networks have developed similar data models and informatics solutions. The common data models use standardized vocabularies (ICD-9, LOINC, and RxNorm) to represent data elements, the open-source i2b2 (Informatics for Integrating Biology and the Bedside) software tool [2] for establishing local data repositories, and the open-source SHRINE (Shared Health Research Information Network) software tool [3] to perform queries across the network to obtain cohort counts.

Progress: Over a period of 6 months, the PaTH initiative has developed and implemented the PCORnet Common Data Model (CDM) and has loaded data for IPF patients into local i2b2 repositories. The loading of the patient data for the remaining two disease cohorts will occur over the next 12 months. The ACT initiative began 6 months later and has established working groups for data harmonization, technical infrastructure, regulatory and governance policies, and the CDRN will be implemented over the next 12 months.

Discussion: Since the ACT initiative began 6 months later, it is informed to a large extent by the lessons learned in PaTH. The ACT CDM is focused on selected data elements (demographics, diagnoses, procedures, visit characteristics, medications, and selected laboratory test results) and has borrowed the data elements from the PCORnet CDM. ACT selected and implemented ontologies to organize the data elements (e.g., the NDF-RT classification system for drugs), and these ontologies, to a large extent, have been adopted by PaTH. Though both PaTH and ACT have established local i2b2 repositories, the PaTH repositories will be populated with de-identified rich patient data on the three targeted conditions while the ACT i2b2 repositories will be populated with a minimum of three years of de-identified patient data on the selected data elements. Finally, in PaTH the sites will share both counts and individual patient data while in ACT the sites will share patient counts for specific queries. However, it is anticipated that both PaTH and ACT CDRNs will evolve in the subsequent phases to include additional data and features.

The documents generated from both initiatives will be made publically accessible online: PaTH will deposit documents at the PCORnet website and ACT will establish a website for the documents. Once implemented, these federated networks will enable efficient, safe and lower cost multi-site clinical and translational research studies at all PaTH and ACT sites. Achieving this goal will substantially reduce and eventually eliminate a major barrier to conducting highest priority clinical and translational research.

References:

1. Amin W, Tsui FR, Borromeo C, et al. PaTH: Towards a learning health system in the Mid-Atlantic region. J Am Med Inform Assoc. 2014 Jul-Aug;21(4):633-6.

2. Murphy SN, Mendis ME, Berkowitz DA, et al. Integration of clinical and genetic data in the i2b2 architecture. AMIA Annual Symposium Proceedings. 2006:1040.

3. 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 Sep-Oct;16(5):624-30.