Assessment of Factors and Approaches to Mapping Laboratory Results in PCORnet

Michelle M. Smerek[1], Elisa Priest, Dr.PH[2], S. Trent Rosenbloom, MD, MPH, FACMI[3], Jon E. Puro, MPA:HA[4], Pedro Rivera, MSCS[4], Shelley A. Rusincovitch[5], Rahul Jain, MPH, CPHIMS[6], and Keith Marsolo, PhD[7]

[1]Duke Clinical Research Institute, Durham, NC; [2]Center for Clinical Effectiveness, Baylor Scott & White Health, Dallas, TX; [3]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN; [4]OCHIN, Inc., Portland, OR; [5]Duke Translational Research Institute, Durham, NC; [6]Health Services Research Portfolio, Louisiana Public Health Institute, New Orleans, LA; [7]Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH

Introduction

The Patient-Centered Outcomes Research Institute (PCORI) launched the National Patient-Centered Clinical Research Network (PCORnet), which seeks to create a "network of networks," to serve as a highly representative national infrastructure for research. PCORnet operates in a distributed fashion, with queries and analytic programs distributed to data partners, who maintain possession and control of their data. Key to PCORnet is the Common Data Model (CDM), which enables interoperability across network partners through standardized representations of multiple data domains. Version 2.0 of the PCORnet CDM saw the introduction of laboratory results, which have not been harmonized across institutions as frequently as more traditional domains such as demographics and diagnoses. In alignment with Meaningful Use and other initiatives[1], the PCORnet CDM includes LOINC (The Logical Observation Identifiers Names and Codes) as an underlying standard for representing lab results. LOINC was created to serve as a common set of identifiers, names and codes for clinical and laboratory observations[2]. While the LOINC standard has been around for over 20 years[3], prior to Meaningful Use, there was no mandate that health systems map their laboratory tests to LOINC, leading most systems to rely on local codes. Regulations will eventually spur more adoption of LOINC, but until then, those wishing to perform research using laboratory results across data partners will need to map their data. There are several tools created by the LOINC team at Regenstrief to assist with the mapping process, such as the Regenstrief LOINC Mapping Assistant (RELMA®) and several guidance documents[4]. Despite the existence of these resources, there remain challenges in mapping local laboratory data to LOINC. To ensure that PCORnet includes analyzable categories of laboratory data for research, it is vital to understand the factors that make this process challenging, and to provide pragmatic approaches to accomplishing this goal.

Methods

All of the data partners that are members of a PCORnet Clinical Data Research Network (CDRN), were asked to map local lab results to eleven common laboratory tests (Table 1). Initial guidance was provided with the PCORnet CDM Specifications, but networks were encouraged to develop strategies best suited to their unique situations. Partners were encouraged to provide feedback on challenges or promising approaches to the Data Standards, Security and Network Infrastructure (DSSNI) Task Force, and the issues were also discussed at several CDM Implementation Forums, which were open to all network participants. These findings are summarized below.

Table 1. Laboratory result common measures that are part of the PCORnet Common Data Model, version 3.0

A1c	Creatinine	International normalized ratio
Creatine kinase total	Hemoglobin	Troponin I cardiac
Creatine kinase MB	Low-density lipoprotein	Troponin T cardiac (qualitative)
Creatine kinase MB/creatine kinase total		Troponin T cardiac (quantitative)

Results

A few of the challenges reported by the data partners are listed below.

Standardization varies: The level of standardization varies by the lab providing the data to the health care provider. Although data from large, outside labs is often more standardized than data from local hospitals and point of care labs, the standardization can vary by region or branch. For example, some systems allow free text entry for lab descriptions as well as results, adding to the complexity of the mapping task.

Incomplete or nonsensical local names: The semantic quality of local lab result names can vary widely, and often the person performing the mapping is not a laboratory data domain expert, so cannot discern the meaning of the names.

Multiple possible mappings: Many systems use internally developed codes for labs instead of LOINC. In this case, there is not always a one-to-one relationship between a local laboratory test and a corresponding LOINC and it is often not apparent which common measure is the best to map to.

Level of granularity: It can be a challenge to determine the right level of granularity for a specific test[5]. For example, it may be possible to map a Hemoglobin test to either, 30313-1: Hemoglobin [Mass/volume] in Arterial blood, or 718-7: Hemoglobin [Mass/volume] in Blood. The former code is more specific, but the latter is the one more likely to be used across a network.

CPT to LOINC mapping resource incomplete: One data partner investigated the National Library of Medicine's CPT to LOINC cross walk, but found that many of the target codes were not included in this resource.

These findings are consistent with those of Raebel et al[6] who highlight that "Clinical laboratory test results obtained during routine healthcare delivery are not uniformly coded nor documented in a standardized manner."

Despite these challenges, networks also reported a number of successful strategies that enabled them to complete the mapping task.

Pareto principle: Many groups have found that a small number of local tests account for the majority of total test volume[5]. One data partner employed an 80/20 rule to target high frequency labs for mapping.

Map, review, repeat: One group worked in an iterative fashion - grouping subsets of lab results in the source data that seemed to correspond to the PCORnet lab result test names, reviewing with lab domain experts, then mapping again.

Cross-functional mapping team: Several groups related the importance of technical analysts consulting with a clinical lab subject matter expert who understands the lab work flow and the lab tests to help with decision making during mapping. In addition, engaging the clinical team allows analysts to provide feedback that can be used to increase standardization and data quality.

Align quality criteria with use case: One integrated delivery system with experience with large datasets found[7] that "perfect is the enemy of the good". They defined "appropriateness for use" criteria and set a maximum level for acceptable missing values. In this case, it is valid to map the majority of the lab codes in the data, and also acceptable for some to be missing. This pragmatic approach is often counterintuitive to data analysts focused on producing 'perfect' data. However, many papers have described the challenges of using EHR data for research[8]. Considering both the data quality issues inherent in EHR data and the goals of the research project, it may be appropriate to take this pragmatic approach to mapping lab data.

Discussion

This initial exercise has been tremendously helpful in allowing PCORnet to understand the state of laboratory result data across the network. Given the heterogeneous combination of organizations contributing data, it is vital to understand the spectrum of factors and challenges the data partners are facing in order to develop strategies that will enable all to accomplish this important task. Sharing information via the CDM Implementation Forums, and a spin-off 'Lab Mapping Interest Group' have proved very beneficial. To gather additional feedback, we intend to send out a survey to all data partners, which will provide additional insight. We hope that the results will allow us to create tailored mapping strategies that can be customized to each data partner's unique situation.

References

1. Blumenthal D, Tavenner RN. The "Meaningful Use" regulation for electronic health records. N Engl J Med 2010; 363:501-504
2. McDonald CJ, Huff SM, Suico JG, et al. LOINC, a universal standard for identifying laboratory observations: a 5- year update. Clin Chem 2003;49;624-33.
3. Dixon BE, Hook J, Vreeman DJ. Learning from the crowd in terminology mapping: the LOINC experience. Lab Med 2015 Spring;46(2):168-174.
4. Logical Object Identifiers Names and Codes (LOINC)® Available at: https://loinc.org/. Accessed Sept 21, 2015.
5. Lin MC, Vreeman DJ, McDonald CJ, Huff SM. A Characterization of Local LOINC Mapping for Laboratory Tests in Three Large Institutions. Methods Inf Med. 2011; 50(2):105-14.
6. Raebel MA, Haynes K, Woodworth TS, et al. Electronic clinical laboratory test results data tables: lessons from Mini-Sentinel. Pharmacoepidemiol Drug Saf. 2014 Jun; 23(6):609-18.
7. Priest EL, Klekar C, Cantu G, Berryman C, Garinger G, Hall L, Kouznetsova M, Kudyakov R, Masica A. Developing electronic data methods infrastructure to participate in collaborative research networks. EGEMS (Wash DC). 2014 Dec 2;2(1):1126.8.
8. Bayley KB, Belnap T, Savitz L, Masica AL, Shah N, Fleming NS. Challenges in using electronic health record data for CER: experience of 4 learning organizations and solutions applied. Med Care. 2013 Aug;51(8 Suppl 3):S80-6.
9. Khan AN, Griffith SP, Moore C, et al. Standardizing laboratory data by mapping to LOINC. J Am Med Inform Assoc. 2006 May-Jun;13(3):353-5.