Alone we can do so little, together we can map so much

Kristine E Lynch, PhD[1,2], Elizabeth Hanchrow, RN, MSN[3], Elise Gatsby, MPH[1], Benjamin Viernes, MPH[1,2], Daniel Park, BS[3], Scott L DuVall, PhD[1,2], Michael E Matheny, MD[3,4]

[1]VA Informatics and Computing Infrastructure, Salt Lake City, Utah, USA; [2]University of Utah, Salt Lake City, Utah, USA; [3]Vanderbilt University Medical Center, Nashville, TN, USA; [4]Tennessee Valley Healthcare System, Nashville, TN, USA

Abstract

The Department of Veterans Affairs (VA) Informatics and Computing Infrastructure (VINCI) team has extended considerable effort to transform VA electronic health record data into the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM). While the overall proportion of VA laboratory data mapped to OMOP standard concepts is high (>90%), and continues to improve over time, individual clinical domains may still be greatly affected by concepts that remain unmapped or incorrectly mapped. OHDSI's 'Virtual Study-A-Thon to Aid Response to COVID-19' provided opportunity for VINCI to improve *mapping of clinical concepts that were not previously prioritized for curation because of their relative low frequency in the source data. Influenza related concepts (Influenza A, B, and Parainfluenza 1-4) are one example. As a result of VINCI's participation in the Study-A-Thon, manual annotation efforts were devoted to influenza concepts and mapping significantly improved— 77.2% unmapped to mapped lab tests and 21.6% incorrect to correct mapped lab tests. Participation in events like the Study-A-Thon is a unique, yet practical, way to prioritize VINCI's quality control efforts and ensure continued far-reaching improvements to VA OMOP data.*

Research Category: Observational data management

Introduction

In 2015 VA Informatics and Computing Infrastructure (VINCI) began transforming the VA's complex electronic record data into the OMOP CDM for use by the VA research community.[1] The VA OMOP project supports VINCI's mission to improve the healthcare of Veterans by providing researchers access to integrated national datasets and tools for analysis. The VA OMOP instance is iteratively developed and mapping improvements are integrated into the instance approximately every quarter. While the ultimate goal is to perform quality control on all data elements, data most commonly requested by users and most frequently occurring source concepts are prioritized for review. Consequentially, some data elements, though transformed into the OMOP CDM, may still reflect underlying source data quality issues.

Laboratory data present a particularly complex data mapping problem because proper encoding to LOINC requires significant domain expertise and investment by healthcare systems using EHRs.[2] Laboratory test names are stored as non-standardized string variables in source data and are often, but not always, accompanied by a corresponding LOINC. While the VA has curated laboratory data mappings using LOINC codes for years across all facilities, there are numerous inconsistencies, gaps, and mis-mappings. This variability is attributable to both between and within facility mapping efforts, making quality control challenging.

The ODHSI 'Virtual Study-A-Thon to Aid Response to COVID-19' (March 26-29, 2020) presented an opportunity to improve the quality of influenza laboratory data that otherwise may not have been prioritized within the normal quarterly update cycle. We describe the quality of influenza data pre- and post-Study-A-Thon.

Methods

A broad string search of source lab test names was performed to ascertain all possible permutations of Influenza A, B, and parainfluenza 1-4 lab tests. This included both antibody and antigen tests. Data were manually reviewed by a nurse annotator. The primary tasks were to 1) assign a LOINC code if missing and 2) adjudicate LOINC code if not missing. The annotator used a combination of source component (test name), property (units), system (specimen), and timing data to choose the correct test name-to-LOINC mapping and used the Regenstrief's LOINC website lookup tool (http://loinc.org/) and top 2000 lab

mapping guidance document for a reference.

Results

There were 7,578 distinct source representations of influenza lab data, of which 5,848 (77.2%) were unmapped before manual review owing to the lack of LOINCs in source data. Of the 22.8% (n= 1,730) that were mapped to a standard concept (i.e., had a LOINC code in source data), 95% were mapped to an incorrect LOINC code and had to be manually remapped. Examples of each mapping scenario are shown in Table 1. Common mistakes were interchanging antigen and antibody tests (Table 1, example 2a) and influenza virus tests erroneously labeled as haemophilus influenza (Table 1, example 2b).

Although the majority of influenza LOINCs were unmapped (77.5%), among those test that were mapped, only 6.2% of instances were correctly assigned (Table 2). Figure 2 shows one example (LOINC 44563-5, Influenza virus A Ag [presence] in nose) of the potential impact of mis-mapping over time. Except for 2018, the manually mapped data resulted in more instances of Influenza virus A antigen tests (overall). Similar patterns were observed for many of the individual LOINC codes.

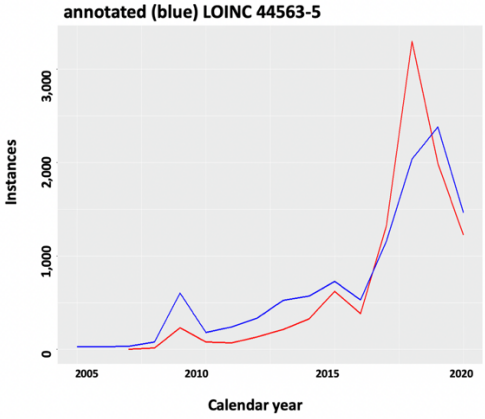
Table 1. Examples of the three Influenza mapping scenarios

Source lab test name (topography)	Original LOINC code	Manually mapped LOINC code
	1. UNMAPPED	
a. Influenza B antigen (Nasopharynx)	NULL	**43895-2-7** (Influenza virus B Ag [presence] in nasopharynx)
b. Influenza A (PCR) (Nasopharynx)	NULL	**43874-7** (Influenza virus A Ag [presence] in nasopharynx)
	2. MAPPED INCORRECTLY	
a. Influenza A Antigen (Nasopharynx)	**22365-1** (Influenza virus A Ab [titer] in Serum)	**43874-4** (Influenza virus A Ag [presence] in nasopharynx)
b. Flu A (Nasopharynx)	**31833-7** (Haemophilus influenza A Ag [presence] in unspecified specimen)	**43874-7** (Influenza virus A Ag [presence] in nasopharynx)
	3. MAPPED CORRECTLY	
a. Influenza B (Pharynx)	**31863-4** (Influenza virus V Ag [presence] in throat)	**31863-4** (Influenza virus V Ag [presence] in throat)
b. PCR Parainfluenza 3 (Nasopharynx)	**29910-7** (Parainfluenza virus 3 RNA [presence] in unspecified specimen)	**29910-7** (Parainfluenza virus 3 RNA [presence] in unspecified specimen)

Table 2. Source representations and instance counts according to the three mapping scenarios

Mapping scenario	Source	%	Instance count	%
Total	7,578	100%	986,7932	100%
1. Unmapped	5,848	77.2%	265,260	26.9%
2. Incorrect	1,640	21.6%	660,454	66.9%
3. Correct	90	1.2%	61,218	6.2%

Figure 1. Instances of original (red) and annotated (blue) LOINC 44563-5



Discussion/Conclusion

Participation in the OHDSI's Study-A-Thon provided opportunity to quickly reprioritize and redirect staff efforts to align with emergent research priorities. In the case of VA OMOP, one of the most useful quality control checks has been real-world research, where we find issues that need to be corrected out of necessity and timeliness. These types of collaborative events help to improve one of the largest instances of OMOP CDM in the world.

References

1. Lynch KE, Deppen SA, DuVall SL, et al. Incrementally transforming electronic medical records into the Observational Medical Outcomes Partnership Common Data Model: A multidimensional quality assurance approach. Appl Clin Inform. 2019;10(5):794-803.
2. Stram M, Gigliotti T, Hartman D, et al. Logical Observation Identifiers Names and Codes for laboratorians. Arch Pathol Lab Med. 2020;144(2):229-239.