Establishing Interoperability Standards between OMOP CDM v4, v5, and PCORnet CDM

Rimma Belenkaya, MS, MA[1], Parsa Mirhaji MD, PhD [1], Mark Khayter [2], Don Torok, MS [2], Ritu Khare, PhD[3], Toan Ong, PhD[4], Lisa Schilling, MD, MSPH[4]

[1]Montefiore Medical Center, Bronx, NY; [2]Ephir, Inc., Boston, MA (pSCANNER CDRN); [3]The Children's Hospital of Philadelphia, Philadelphia, PA; [4]School of Medicine, University of Colorado, Anschutz Medical Campus, Aurora, CO

Introduction

PCORnet, the National Patient-Centered Clinical Research Network, funded by the Patient-Centered Outcomes Research Institute (PCORI), integrates data from 11 heterogeneous networks to enable large-scale comparative effectiveness research. While the PCORnet Common Data Model (CDM) has been evolving, all 11 networks choose to first integrate their source data into more established CDMs, such as i2b2 and OMOP, and then port these data into PCORnet CDM. Crosswalking from healthcare source systems to OMOP CDM and then to PCORnet CDM poses a substantial challenge. To ensure data harmonization with minimal loss of source granularity, comprehensive CDM interoperability standards are required.

Aim

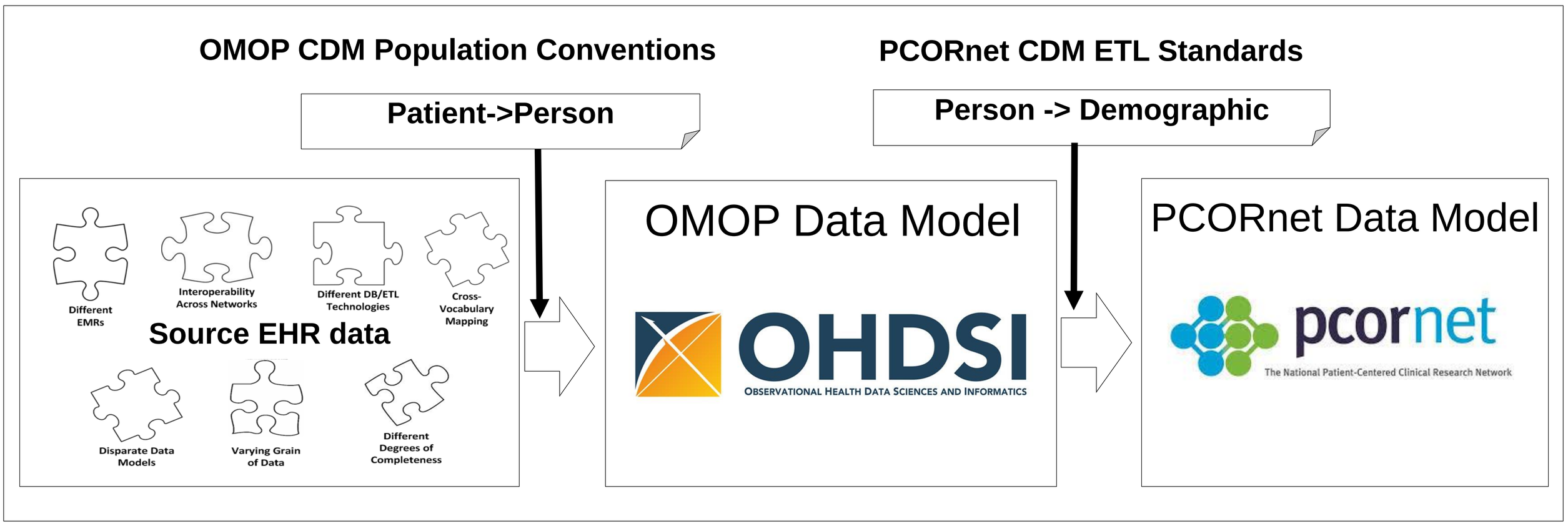
To create interoperability standards between OMOP and PCORnet CDM to support data integration for comparative effectiveness research.

Methods

The data extraction, transformation and loading (ETL) from the source electronic health data to the PCORnet CDM is a two-step process: (1) Populate the OMOP CDM while enforcing, to the extent possible, the PCORnet requirement. (2) Convert data in the OMOP CDM to PCORnet CDM via a set of mechanistic transformation rules.

The collaborative process encouraged greater scrutiny of decisions which required that decisions have solid justification. For the most part, consensus was always reached. When consensus was not reached it was typically due to the requirement of additional knowledge or understanding of either of the two CDMs, and despite a lack of consensus, a transformation convention was established. Once the OMOP CDM population conventions have been established, the process of creating OMOP-to-PCORnet ETL standards is reduced to describing simple mappings and transformations.

Crosswalk from source – to – OMOP CDM – to – PCORNet CDM



Challenges and solutions of the ETL process

Challenge	Solution
Differences in data structure and domain between the OMOP CDM and PCORnet CDM	Conventions to perform schema mappings between the two CDMs.
Interpretation of unknown values (i.e. 'refused to answer', NULL, unknown and unmapped values).	Extend utilization of PCORnet source concepts as standard concepts in the OMOP vocabulary
Non-existence of the interoperability standards between OMOP and PCORnet CDMs	1. Identify matching domains, attributes and vocabularies between the two CDMs 2. Propose a solution to account for data elements that are missing in the OMOP CDM 3. Add data representation conventions in OMOP CDM that provide closer alignment between the two models.

Results

- Interoperability standards from source (i.e. electronic health records) to OMOP CDM v4 (Specific to PCORI CDM) and from source to OMOP CDM v5 (specific to PCORI CDM).
- A conventions document for populating the OMOP CDM, and an extract-transform-load specifications document for transforming to the PCORnet CDM v1, for both OMOP CDM v4 and OMOP CDM v5. The documents will be publicly available for the OHDSI community.

Conclusions

With rare exceptions, the OMOP CDM supports greater granularity of data representation in both the CDM and vocabulary than PCORnet CDM. These features allow for adequate preservation of source data granularity, and transformation from the more granular OMOP to less granular PCORnet representation is straightforward.

It is possible to use the OMOP CDM as an intermediary data representation when converting various healthcare datasets to PCORnet CDM. However it was necessary to use vocabulary concepts that violated the domain rule established in the OMOP CDM v5 standard. In addition attributes required in PCORnet that do not exist in OMOP can be represented without altering the standard table schema, but require a set of documented conventions to be understood and extractable from the OMOP CDM.

References

1. Kaushal R, et al. Changing the research landscape: the New York City Clinical Data Research Network. J Am Med Inform Assoc: 21 (4). 2014 Jul: p. 587-590.
2. Ohno-Machado L, et al. pSCANNER: patient-centered Scalable National Network for Effectiveness Research. Journal of the American Medical Informatics Association 21.4. 2014: p. 621-626.
3. Forrest CB, et al. PEDSnet: a National Pediatric Learning Health System. J Am Med Inform Assoc: 21 (4). 2014 Jul 1: p. 602-606.