Bridging the divide: Leveraging the PCORnet CDM to make EHR data generally accessible for patient-centered outcome research

Yaxing Liu, PhD¹,², Benjamin A. Goldstein¹,³, & the Clinical Research Datamart (CRDM) Working Group
1. Children’s Health and Discovery Initiative, Duke University
2. Duke Health Technology Solutions, Duke University Health System
3. Department of Biostatistics and Bioinformatics, Duke University

**Background:** Many academic health systems struggle with how best to make EHR data accessible to clinical researchers. Many of these centers – including our own – participate in research networks like PCORnet, which spur the creation of curated EHR datamarts. As a general data model, the PCORnet common data model (CDM) does not necessarily meet all of research needs for a given research community. Moreover, the CDM contains protected health information, which needs to be properly governed. In this poster we describe our efforts to build off of the PCORnet CDM to make EHR data accessible to our local research community within a protected analytics computing environment. This work was partially funded by a Clinical and Translational Sciences Award.

**Methods:** We envisioned our target users being statistical analysts supporting research projects as well as computationally oriented clinicians. We started with our institution’s implementation of the PCORnet CDM v4.1. After having focus sessions with different research groups, we identified additional “data sidecars” that enhance the PCORnet CDM to provide the additional data elements required for different use cases and medical specialties. Our PCORnet model currently sits on an Oracle Server within our data warehouse. We created views to be accessed via analytic software (i.e. R, SAS, Python) within Duke’s Protected Analytics Computing Environment (PACE).

**Results:** Figure 1 shows a schematic of how we built off the PCORnet CDM. We obtained a single IRB to allow the data to exist. Each project then requires project-specific IRBs. The database is also meant to be linkable to other data assets such as specialty-specific datamarts (e.g. tumor registry, cath lab data etc.) as well as external data assets (e.g. environmental data, neighborhood data). Table 1 lists the data sidecars that we developed to extend the PCORnet CDM to meet our research needs. We note that some of these data sidecars can fit into the PCORnet v4.1 CDM, but were not implemented at our institution. We performed user tests across different analytic teams to assess the extent the environment met user needs and led to consistent analytic queries.

**Table 1: Additional Sidecars**

<table>
<thead>
<tr>
<th>Time Varying Addresses with Geo-location</th>
<th>Provider Specialties</th>
<th>Detailed Encounter Info</th>
<th>Immunizations</th>
<th>Psychiatric Data</th>
<th>External Death Data</th>
</tr>
</thead>
</table>

**Conclusions:**
By leveraging and extending the existing PCORnet CDM we have been able to efficiently build out an EHR data resource that can support retrospective analyses for population health and health services research. In future, we will develop a GitLab-based code repository to extend best practices for working with EHR data. We will monitor user engagement and assess changes in research efficiency.