Building Data Capacity for Patient-Centered Outcomes Research: Perspectives from the OHDSI community

Patrick Ryan, PhD
Vice President, Observational Health Data Analytics, Janssen Research and Development
Assistant Professor, Adjunct, Department of Biomedical Informatics, Columbia University Medical Center
OHDSI to advance the science of observational research

• OHDSI is an open science community, founded by scientists at J&J, Columbia University, Stanford, UCLA, Regenstrief Institute, and Iqvia

• OHDSI is an open multi-stakeholder, interdisciplinary collaborative with a mission to improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care

• OHDSI is driving development and adoption of open community data standards, open source analysis software, and open science best practices within regulators, academia, industry, payors, and health systems

GLOBAL OHDSI NETWORK - Industry, Academia, Government

>2,100 Researchers in epidemiology, statistics, informatics, health policy, and clinical sciences

>150 Databases Included

>2B Patient records Represented

GLOBAL OHDSI NETWORK - Industry, Academia, Government

OMOP DRIVING RWE

OMOP COMMON DATA MODEL

+ OPEN SOURCE ANALYSIS CODE

A Clinical Characterization

B Population-level effect estimation

C Patient-level prediction
OHDSI Community

OHDSI Data partner 1

Source data in local structure and vocabularies

Standardized patient-level database (OMOP CDM)

ETL

Standardized analytics (OHDSI tools)

OHDSI tools

OHDSI Data partner 2

Source data

OMOP CDM

ETL

OHDSI tools

OHDSI Data partner 3

Source data

OMOP CDM

ETL

OHDSI tools

OHDSI Data partner n

Source data

OMOP CDM

ETL

OHDSI tools

OHDSI collaborations

Open community data standards (OMOP CDM)

Open source development (OHDSI tools)

Methodological research

Clinical evidence generation

OHDSI Network studies

Pre-specified protocol with analysis specification

Standardized summary statistics results repository

OHDSI collaborations

Evidence dissemination

OHDSI data network

Data standards are a means to an end, not an end in itself

- What *data standards* could make the PCOR data infrastructure more useful for research and other data needs?
- What *data standards* are likely to become more relevant looking forward?
- What needs to be prioritized?
- What *evidence* would be useful to improve health policy and health care, which could be reliably generated by the PCOR data infrastructure?
- How can *data standards* enable real-world *analytics* to meet the relevant *evidence* needs moving forward?
- What needs to be prioritized?
Desired attributes for reliable evidence

<table>
<thead>
<tr>
<th>Desired attribute</th>
<th>Question</th>
<th>Researcher</th>
<th>Data</th>
<th>Analysis</th>
<th>Result</th>
<th>=</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeatable</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
</tr>
<tr>
<td>Reproducible</td>
<td>Identical</td>
<td>Different</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
</tr>
<tr>
<td>Replicable</td>
<td>Identical</td>
<td>Same or different</td>
<td>Similar</td>
<td>Identical</td>
<td>Similar</td>
<td>Similar</td>
</tr>
<tr>
<td>Generalizable</td>
<td>Identical</td>
<td>Same or different</td>
<td>Different</td>
<td>Identical</td>
<td>Similar</td>
<td>Similar</td>
</tr>
<tr>
<td>Robust</td>
<td>Identical</td>
<td>Same or different</td>
<td>Same or different</td>
<td>Different</td>
<td>Similar</td>
<td>Similar</td>
</tr>
<tr>
<td>Calibrated</td>
<td>Similar (controls)</td>
<td>Identical</td>
<td>Identical</td>
<td>Identical</td>
<td>Statistically consistent</td>
<td></td>
</tr>
</tbody>
</table>

- Data standards are necessary to enable replicability, generalizability, and robustness
- Data standards without standardized analytics are not sufficient to ensure reliable evidence
- Standards limited to US data are harmful for US population; public health questions require global data to generate global evidence
Delineating the roles of standards to enable evidence

How can data standards enable real-world analytics to meet the relevant evidence needs moving forward?

- Data standards to enable data exchange
  - HL7 FHIR

- Data standards to harmonize data structure and enable analytics
  - OMOP CDM

- Analytics standards to generate and disseminate evidence
  - HADES

Vocabulary standards to harmonize data content and enable analytics
  - Athena

What evidence would be useful to improve health policy and health care?

Characterization:
- Who are the patients exposed to medical interventions?
- How often do outcomes occur amongst those patients?

Estimation:
- Is the risk of outcome causally related to exposure?
- How does the risk compare with alternative interventions?

Prediction:
- Which risks can be actionably predicted with available data?
- Which patients are at highest risk of adverse events?
Open community data standard to enable analytics: OMOP Common Data Model

- **Concept**
- **Concept_relationship**
- **Concept_ancestor**
- **Vocabulary**
- **Source_to_concept_map**
- **Relationship**
- **Concept_synonym**
- **Drug_strength**
- **Domain**
- **Concept_class**
- **Dose_era**
- **Condition_era**
- **Drug_era**
- **Results Schema**
- **Cohort_definition**
- **Cohort**
- **Standardized derived elements**
- **Standardized health system data**
- **Location**
- **Location_history**
- **Care_site**
- **Provider**
- **Standardized metadata**
- **CDM_source**
- **Metadata**
- **Standardized vocabularies**
- **Concept**
- **Vocabulary**
- **Domain**
- **Concept_class**
- **Concept_relationship**
- **Relationship**
- **Concept_synonym**
- **Concept_ancestor**
- **Source_to_concept_map**
- **Drug_strength**
- **Standardized clinical data**
- **Person**
- **Observation_period**
- **Visit_occurrence**
- **Visit_detail**
- **Condition_occurrence**
- **Drug_exposure**
- **Procedure_occurrence**
- **Device_exposure**
- **Measurement**
- **Note**
- **Note_NLP**
- **Survey_conduct**
- **Observation**
- **Specimen**
- **Fact_relationship**
- **Standardized health economics**
- **Cost**
- **Payer_plan_period**
- **Care_site**
- **Provider**
OHDSI’s standardized vocabularies
to harmonize data content and enable analytics

- 164 Vocabularies across 41 domains
  - MU3 standards: SNOMED, RxNorm, LOINC
  - Disparate sources: ICD9CM, ICD10(CM), Read, NDC, Gemscript, CPT4, HCPCS...
- >9.5 million concepts
  - >3.4 million standard concepts
  - >5.4 million source codes
  - >674,000 classification concepts
- >57 million concept relationships
- >86 million ancestral relationships

Publicly available for download at: [http://athena.ohdsi.org/](http://athena.ohdsi.org/)

As of v5.0 26-FEB-21
Common data model and standardized vocabularies to enable standardized analytics

Transformation to OMOP common data model

Source 1 raw data
- Electronic health records

Source 2 raw data
- Administrative claims

Source 3 raw data
- Clinical data

Source 1 CDM

Source 2 CDM

Source 3 CDM

Open-source analysis code

Open evidence
Evidence needed around the safety of HCQ alone and in combination with azithromycin (AZ)

Standardized analytics: comparative cohort design with large-scale propensity adjustment

Standardized data: 14 databases (claims and EHR) from 6 countries, all using OMOP CDM

Findings: In RA population, HCQ alone is generally safe but in combination with AZ it shows a doubling of risk of 30-day cardiovascular mortality.
Standards enabling evidence for policy: COVID-19 treatment utilization patterns

Use of repurposed and adjuvant drugs in hospital patients with covid-19: multinational network cohort study

Albert Prats-Uribe,1 Anthony G Sena,2,3 Lana Yin Hui Lai, Heba Algholi,7 Osaíd A. Alsher,8 Thamir M Alishammar,9 Cai Paula Casajust,12 Dalia Dawoud,13,14 Asieh Golozar,15,16 Paras P Mehta,18 Mengchun Gong,9 Daniel R Morales,21 Martina Recalde,24,25 Elena Roel,24,25 Karishma Shah,5 N Vignesh Subbian,26 David Vizcaya,27 Lin Zhang,28,29 Ying Jiahe Yang,30 Kristine E Lynch,32 Michael E Matheny, Peter R Rijnbeek,2 George Hripcsak,36 Jennifer CE Lane,3 Marc A Suchard,38 Talita Duarte-Salles,4 Kristin Kostka,4 Daniel Prete-Alhambra4

ABSTRACT
OBJECTIVE
To investigate the use of repurposed and adjuvant drugs in patients admitted to hospital with covid-19 across three continents.

DESIGN
Multinational network cohort study.

SETTING
Hospital electronic health records from the United States, Spain, and China, and nationwide claims data from South Korea.

PARTICIPANTS
303 264 patients admitted to hospital with covid-19 from January 2020 to December 2020.

MAIN OUTCOME MEASURES
Prevalence, or discrepancies of any drug on or 40 days after hospital admission.

For numbered affiliations see end of the article.
Correspondence to: P B Ryan ryan@hdsi.org (ORCID 0000-0002-9727-2139)
Additional material is published online only. To view please visit the journal online.
Gt this: bmj.com/doi/10.1136/bmj.m5138
http://dx.doi.org/10.1136/bmj.m5138
Accepted: 16 April 2021

![Graph showing prevalence of drug use over time with data points and lines]
Data standards are part of methodological standards for reliable evidence

- Patients with cardiovascular diseases and hypertension treated with angiotensin converting enzyme inhibitors (ACEs) angiotensin-II receptor blockers (ARBs) may influence susceptibility to COVID-19 and worsen its severity.

As stated by Watson et al., in relation to one of the published studies, lack of transparency and uncertainties about research standards applied raise doubts about published results. Morales et al. supported the reproducibility of their study by publishing the study protocol in the EU PAS Register ahead of time, providing a start-to-finish executable code, facilitating the sharing and exploration of the complete result set with an interactive web application and asking clinicians and epidemiologists to perform a blinded evaluation of propensity score diagnostics for the treatment comparisons.
Take-home messages

• Data standards should be driven by evidence needs
• Complementary standards required across the evidence generation lifecycle
  – Data standards for exchange
  – Data standards for structure for analytics
  – Vocabulary standards for data content and analytics
  – Analytic standards for evidence generation and dissemination
• Reliable evidence is strengthened by disparate data and analytics standardized within a common shared framework
  – Data network should focus on depth and breadth as required for evidence needs
  – Open science enables adoption of data standards and analytic best practices
  – International collaboration is possible with international standards