Standardizing phenotypes for systematic observational analysis: Lessons from OHDSI

Patrick Ryan, PhD
Janssen Research and Development
15 November 2014
A critical need for our community: 
Open-source phenotype library

OMOP established an open-source library of Health Outcomes of Interest (HOI) definitions for use in observational studies. These are a subset of all conditions that are of importance due to their historical associations with drug toxicities, their medical significance, and/or public health implications.

We have found that there is very little consensus in the literature about outcome definitions. We therefore decided to create a library of HOI definitions that cover a wide representation of the prior work for testing in the OMOP experiment: broad and narrow selections of diagnosis codes, combination of diagnosis codes with diagnostic or therapeutic procedures and lab values.

Library of HOI Definitions

- Angioedema
- Aplastic Anemia
- Acute Liver Injury
- Acute Kidney Injury
- Bleeding
- Hip Fracture
- Hospitalization
- Acute Myocardial Infarction
- Mortality After Myocardial Infarction
- GI Ulcer Hospitalization

omop.org/HOI
Alternative definitions for Acute liver injury

**D**: Occurrence of at least one diagnosis code

**D+P**: Occurrence of at least one diagnosis code and
*diagnostic* procedure <=30d before
*OR* *treatment* procedure >=60d after

**D+P+L**: Occurrence of at least one diagnosis code
and *diagnostic* procedure <=30d before
*OR* *treatment* procedure >=60d after
and laboratory results indicative of Hy's law:
(ALT >= 3xULN OR AST >= 3xULN) AND Bilirubin >= 2xULN
within 7 days

**L**: Laboratory results indicative of Hy's law:
(ALT >= 3xULN OR AST >= 3xULN) AND Bilirubin >= 2xULN
within 7 days
Assessment of Case Definitions for Identifying Acute Liver Injury in Large Observational Databases

Aaron J. Katz · Patrick B. Ryan · Judith A. Racoosin · Paul E. Stang

Table 2 The prevalence and demographic characteristics of patients satisfying four distinct definitions of acute liver injury, by database

<table>
<thead>
<tr>
<th>Database</th>
<th>Total population [n (000s)]</th>
<th>Patients satisfying acute liver injury definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>MarketScan® Lab Database</td>
<td>1,229</td>
<td>D [n (%)] 154,357 (12.6) L [n (%)] 191 (0.02)</td>
</tr>
<tr>
<td>Regenstrief Institute</td>
<td>2,002</td>
<td>D [n (%)] 215,930 (10.8) L [n (%)] 7,349 (0.4)</td>
</tr>
<tr>
<td>Partners HealthCare System</td>
<td>2,942</td>
<td>D [n (%)] 264,083 (9.0) L [n (%)] 36,267 (1.2)</td>
</tr>
<tr>
<td>National Patient Care Database of the Veterans Health Administration</td>
<td>3,202</td>
<td>D [n (%)] 345,519 (10.8) L [n (%)] NA</td>
</tr>
<tr>
<td>Humana Inc.</td>
<td>5,197</td>
<td>D [n (%)] 447,886 (8.6) L [n (%)] NA</td>
</tr>
<tr>
<td>GE Healthcare</td>
<td>11,216</td>
<td>D [n (%)] 514,118 (4.6) L [n (%)] 6,461 (0.1)</td>
</tr>
<tr>
<td>SDI Health (now IMS Health, Inc.)</td>
<td>90,485</td>
<td>D [n (%)] 6,491,416 (7.2) L [n (%)] NA</td>
</tr>
</tbody>
</table>

D = definition A, L = definition B, D+P = definition C, D+P+L = definition D.
Lab value / ULN

Days from index date

when criteria of (diagnosis + procedure + lab) was satisfied

--- 2*ULN
EDDIE result: little consistency in outcome definition selection

Stang et al, Drug Safety 2013
Alternative Outcome Definitions and Their Effect on the Performance of Methods for Observational Outcome Studies

Christian G. Reich · Patrick B. Ryan · Martijn J. Schuemie
OMOP finding: Restricting outcome definition did not impact predictive accuracy

- Narrowing definition does not ensure proper trade-off of lower sensitivity vs. higher specificity
- PPV is irrelevant to assessment of precision and accuracy in effect estimation

Fig. 5 AUC values of alternative definitions for each of the three HOI across the four databases. Definitions with only few test cases are labeled. Upper left diagram red markers are equivalent to Fig. 4. HOI health outcome of interest, MDCD MarketScan Multi-state Medicaid, MDCR MarketScan Medicare Supplemental Beneficiaries, CCAE MarketScan Commercial Claims and Encounters, GE GE Centricity
Phenotyping is the biggest barrier on the critical path to large-scale real-world evidence generation.

Standardize data structure and content: OMOP CDM and Vocabularies

Standardize phenotypes

Standardize cohort definition syntax and execution: CIRCE

Standardize patient-level prediction: PLATO

Standardize population-level estimation for safety surveillance and comparative effectiveness: HOMER

Standardize clinical characterization for disease natural history and quality improvement: HERACLES

Standardize database profiling: ACHILLES

http://ohdsi.org