Welcome to OHDSI F2F 2017!
March 17-18
Atlanta, GA
Not the Glitz of the OHDSI Annual Symposium

There’s a different path forward that we can now take. One outlined by Schuemie and Peter Rijnbeek.

For reliable Evidence Generation to work and succeed, openness and transparency must be our common creed. Sharing your protocol, posting your code, reproducible in every step of the workflow.
F2F is the OHDSI conference that rolls up its sleeves, spits in its hands, and goes to work.

- Carl Sandburg
Who’s Here?

- Amazon.com
- AstraZeneca
- BlueCross BlueShield of South Carolina
- Case Western Reserve University
- Children's Hospital of Philadelphia, PEDSnet
- China Medical University, Taiwan
- Cloudera
- Columbia University
- ConvergeHealth by Deloitte
- Department of Veterans Affairs
- Ephir, Inc.
- Erasmus MC
- Georgia Tech
- Google
- IBM Research
- Janssen R&D
- McKinsey & Company
- Medical University of South Carolina
- Northwestern University
- Odysseus Data Services
- QuintilesIMS
- SHYFT Analytics
- Stanford University
- U of Texas Health Science Center at Houston
- U of Colorado, pSCANNER
- UCLA
- University of Massachusetts Medical School
- University of New Mexico Health Sciences
- University of Pittsburgh
- University of Utah, VA, VINCI
- USC
- Weill Cornell Medicine
Thank you for your sponsorship!

couldera

Georgia Tech Research Institute

CHAI
It’s not too latte.
A little context...
OHDSI’s Mission

To improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.
Along the way...
# Vocabulary Mappings

<table>
<thead>
<tr>
<th>Vocabulary ID (CDM V4.5)</th>
<th>Vocabulary Code (CDM V5)</th>
<th>Vocabulary Name</th>
<th>Available</th>
<th>Latest Update</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SNOMED</td>
<td>Systematic Nomenclature of Medicine - Clinical Terms (IHTSDO)</td>
<td></td>
<td>05-OCT-16</td>
</tr>
<tr>
<td>2</td>
<td>ICD9CM</td>
<td>International Classification of Diseases, Ninth Revision, Clinical Modification, Volume 1 and 2 (NCHS)</td>
<td></td>
<td>01-OCT-14</td>
</tr>
<tr>
<td>3</td>
<td>ICD9Proc</td>
<td>International Classification of Diseases, Ninth Revision, Clinical Modification, Volume 3 (NCHS)</td>
<td></td>
<td>01-OCT-14</td>
</tr>
<tr>
<td>5</td>
<td>HCPCS</td>
<td>Healthcare Common Procedure Coding System (CMS)</td>
<td></td>
<td>28-OCT-15</td>
</tr>
<tr>
<td>6</td>
<td>LOINC</td>
<td>Logical Observation Identifiers Names and Codes (Regenstrief Institute)</td>
<td></td>
<td>24-JUN-16</td>
</tr>
<tr>
<td>7</td>
<td>NDFRT</td>
<td>National Drug File - Reference Terminology (VA)</td>
<td></td>
<td>03-OCT-16</td>
</tr>
<tr>
<td>8</td>
<td>RxNorm</td>
<td>RxNorm (NLM)</td>
<td></td>
<td>03-OCT-16</td>
</tr>
<tr>
<td>9</td>
<td>NDC</td>
<td>National Drug Code (FDA and manufacturers)</td>
<td></td>
<td>16-OCT-15</td>
</tr>
<tr>
<td>10</td>
<td>GPI</td>
<td>Medi-Span Generic Product Identifier (Wolters Kluwer Health)</td>
<td>License required</td>
<td>06-MAY-15</td>
</tr>
<tr>
<td>11</td>
<td>UCUM</td>
<td>Unified Code for Units of Measure (Regenstrief Institute)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Gender</td>
<td>OMOP Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Race</td>
<td>Race and Ethnicity Code Set (USBC)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Place of Service</td>
<td>Place of Service Codes for Professional Claims (CMS)</td>
<td>EULA required</td>
<td>01-SEP-16</td>
</tr>
<tr>
<td>15</td>
<td>MedDRA</td>
<td>Medical Dictionary for Regulatory Activities (MSSO)</td>
<td>EULA required</td>
<td>01-SEP-16</td>
</tr>
<tr>
<td>16</td>
<td>Multum</td>
<td>Corner Multum (Corner)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Read</td>
<td>NHS UK Read Codes Version 2 (HSCIC)</td>
<td></td>
<td>18-MAR-16</td>
</tr>
<tr>
<td>18</td>
<td>OXMiS</td>
<td>Oxford Medical Information System (OCHIP)</td>
<td></td>
<td>27-APR-15</td>
</tr>
<tr>
<td>19</td>
<td>Indication</td>
<td>Indications and Contraindications (FDB)</td>
<td>License required</td>
<td>19-NOV-15</td>
</tr>
<tr>
<td>20</td>
<td>ETC</td>
<td>Enhanced Therapeutic Classification (FDB)</td>
<td>License required</td>
<td>19-NOV-15</td>
</tr>
<tr>
<td>21</td>
<td>ATC</td>
<td>WHO Anatomic Therapeutic Chemical Classification</td>
<td></td>
<td>03-OCT-16</td>
</tr>
<tr>
<td>22</td>
<td>Multix</td>
<td>Multix (FDB)</td>
<td>License required</td>
<td>03-OCT-16</td>
</tr>
<tr>
<td>28</td>
<td>VA Product</td>
<td>VA National Drug File Product (VA)</td>
<td></td>
<td>03-OCT-16</td>
</tr>
<tr>
<td>31</td>
<td>SMQ</td>
<td>Standardised MedDRA Queries (MSSO)</td>
<td></td>
<td>03-OCT-16</td>
</tr>
<tr>
<td>32</td>
<td>VA Class</td>
<td>VA National Drug File Class (VA)</td>
<td></td>
<td>03-OCT-16</td>
</tr>
<tr>
<td>33</td>
<td>Cohort</td>
<td>Legacy OMOP HOI or DOI cohort</td>
<td></td>
<td>03-OCT-16</td>
</tr>
<tr>
<td>34</td>
<td>ICD10</td>
<td>International Classification of Diseases, Tenth Revision (WHO)</td>
<td></td>
<td>01-DEC-16</td>
</tr>
<tr>
<td>35</td>
<td>ICD10PCS</td>
<td>ICD-10 Procedure Coding System (CMS)</td>
<td></td>
<td>18-MAY-16</td>
</tr>
<tr>
<td>40</td>
<td>DRG</td>
<td>Diagnosis-related group (CMS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>MDC</td>
<td>Major Diagnostic Categories (CMS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>APC</td>
<td>Ambulatory Payment Classification (CMS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>Revenue Code</td>
<td>UB04/CMS1450 Revenue Codes (CMS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>Ethnicity</td>
<td>OMOP Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>MeSH</td>
<td>Medical Subject Headings (NLM)</td>
<td></td>
<td>09-MAY-16</td>
</tr>
<tr>
<td>47</td>
<td>NUCC</td>
<td>National Uniform Claim Committee Health Care Provider Taxonomy Code Set (NUCC)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data Source Exploration

- **Person Summary**
  - Source name: Demo data - 1K synthetic patients
  - Number of persons: 1K

- **Year of Birth**
  - Bar chart showing distribution of people by year of birth from 1910 to 1980.

- **Population by Gender**
  - Pie chart showing distribution of male and female population.

- **Population by Race**
  - Pie chart showing distribution by race categories.

- **Population by Ethnicity**
  - Pie chart showing distribution by ethnicity categories.

- **Data Density**
  - Line chart showing total rows by category over years 2006 to 2010.
# Data Quality Checking

<table>
<thead>
<tr>
<th>Message Type</th>
<th>Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERROR</td>
<td>101-Number of persons by age, with age at first observation period; should not have age &lt; 0, (n=848)</td>
</tr>
<tr>
<td>ERROR</td>
<td>103 - Distribution of age at first observation period (count = 1); min value should not be negative</td>
</tr>
<tr>
<td>ERROR</td>
<td>114-Number of persons with observation period before year-of-birth; count (n=851) should not be &gt; 0</td>
</tr>
<tr>
<td>ERROR</td>
<td>206 - Distribution of age by visit_concept_id (count = 7); min value should not be negative</td>
</tr>
<tr>
<td>ERROR</td>
<td>209-Number of visit records with end date &lt; start date; count (n=168074) should not be &gt; 0</td>
</tr>
<tr>
<td>ERROR</td>
<td>301-Number of providers by specialty concept_id; 224 concepts in data are not in correct vocabulary (Specialty)</td>
</tr>
<tr>
<td>ERROR</td>
<td>400-Number of persons with at least one condition occurrence, by condition_concept_id; 115 concepts in data are correct vocabulary (SNOMED)</td>
</tr>
<tr>
<td>ERROR</td>
<td>406 - Distribution of age by condition_concept_id (count = 753); min value should not be negative</td>
</tr>
<tr>
<td>ERROR</td>
<td>411-Number of condition occurrence records with end date &lt; start date; count (n=182349) should not be &gt; 0</td>
</tr>
<tr>
<td>ERROR</td>
<td>506 - Distribution of age at death by gender (count = 2); min value should not be negative</td>
</tr>
</tbody>
</table>
Health Analytics Workflows
Computable Phenotypes

Cohort definition: A cohort is defined as the set of persons satisfying one or more inclusion criteria for a duration of time. One person may qualify for one cohort multiple times during non-overlapping time intervals. Cohorts are constructed in ATLAS by specifying cohort entry criteria and cohort exit criteria. Cohort entry criteria involve selecting one or more initial events, which determine the start date for cohort entry, and optionally specifying additional inclusion criteria which filter to the qualifying events. Cohort exit criteria are applied to each cohort entry record to determine the end date when the person's episode no longer qualifies for the cohort.

As detailed at: https://phekb.org/phenotype/type-2-diabetes-mellitus

Initial event cohort: Events are recorded time-stamped observations for the persons, such as drug exposures, conditions, procedures, measurements and visits. All events have a start date and end date, though some events may have a start date and end date with the same value (such as procedures or measurements). The event index date is set to be equal to the event start date.

- People having any of the following: Add Initial Event...
  - a condition occurrence of T2DM Diagnosis
  - occurrence start is: Before YYYY-MM-DD

- a drug exposure of T2DM Prescriptions

with continuous observation of at least 0 days before and 0 days after event index date

Limit initial events to: all events per person.

Initial event inclusion criteria: From among the initial events, include:

- having all of the following criteria: Add New Criteria...
  - with at most 0 using all occurrences of:
    - a condition occurrence of T2DM Diagnosis

starting between All days Before and All days After event index date and ending any time.
"ConceptSets": [
  {
    "id": 0,
    "name": "[PHEKB] T2DM",
    "expression": [
      "items": [
        {
          "concept": {
            "CONCEPT_CLASS_ID": "Clinical Finding",
            "CONCEPT_CODE": "422014003",
            "CONCEPT_ID": 443732,
            "CONCEPT_NAME": "Disorder due to type 2 diabetes mellitus",
            "DOMAIN_ID": "Condition",
            "INVALID_REASON": "V",
            "INVALID_REASON.Caption": "Valid",
            "STANDARD_CONCEPT": "S",
            "STANDARD_CONCEPT.Caption": "Standard",
            "VOCABULARY_ID": "SNOMED"
          }
        },
        "includeDescendants": true
      }
    ]
  },
  {
    "concept": {
      "CONCEPT_CLASS_ID": "Clinical Finding",
      "CONCEPT_CODE": "421750000",
      "CONCEPT_ID": 443734,
      "CONCEPT_NAME": "Ketoacidosis in type 2 diabetes mellitus",
      "DOMAIN_ID": "Condition",
      "INVALID_REASON": "V",
      "INVALID_REASON.Caption": "Valid",
      "STANDARD_CONCEPT": "S",
      "STANDARD_CONCEPT.Caption": "Standard",
      "VOCABULARY_ID": "SNOMED"
    },
    "isExcluded": true,
    "includeDescendants": true
  },
  {
    "concept": {
      "CONCEPT_CLASS_ID": "Clinical Finding",
      "CONCEPT_CODE": ""
Standardized Cohort Diagnostics

Propensity score distribution

Covariate balance

https://github.com/OHDSI/CohortMethod
Predictive Modeling

PatientLevelPrediction

Features

- Takes a cohort and outcome of interest as input.
- Extracts the necessary data from a database in OMOP Common Data Model format.
- Uses a large set of covariates including for example all drugs, diagnoses, procedures, as well as age, comorbidity indexes, etc.
- Various machine learning algorithms can be used to develop predictive models.
- Includes function for evaluating predictive models
- Includes functions to plot and explore model performance (ROC + Calibration)
- Supported outcome models are l1 logistic regression, Random forest, Gradient boosting machines, Naive Bayes, KNN and MLP.

Screenshots
And we’ve generated evidence...
Supporting Medication Safety: Risk of Angioedema with Levetiracetam
OHDSI in the Literature

Articles

[HTML] Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers
Abstract The vision of creating accessible, reliable clinical evidence by accessing the clinical experience of hundreds of millions of patients across the globe is a reality. The Observational Health Data Sciences and Informatics (OHDSI) has built on learnings from the
Cited by 31 Related articles All 6 versions Cite Save

Characterizing treatment pathways at scale using the OHDSI network
G Hripcsak, PB Ryan, JD Duke... - Proceedings of the ..., 2016 - National Acad Sciences
Abstract Observational research promises to complement experimental research by providing large, diverse populations that would be infeasible for an experiment. Observational research can test its own clinical hypotheses, and observational studies also
Cited by 8 Related articles All 6 versions Web of Science: 2 Cite Save

[HTML] Improving detection of arrhythmia drug-drug interactions in pharmacovigilance data through the implementation of similarity-based modeling
S Vilar, T Lorberbaum, G Hripcsak, NP Tatonetti - PloS one, 2015 - journals.plos.org
... Department of Biomedical Informatics, Columbia University, New York, NY, United States of America, Department of Systems Biology, Columbia University, New York, NY, United States of America, Observational Health Data Sciences and Informatics (OHDSI), New York, NY ...
Cited by 4 Related articles All 14 versions Web of Science: 1 Cite Save More
Characterizing treatment pathways at scale using the OHDSI network
G Hripcsak, PB Ryan, JD Duke... - Proceedings of the ... , 2016 - National Acad Sciences
Abstract Observational research promises to complement experimental research by providing large, diverse populations that would be infeasible for an experiment. Observational research can test its own clinical hypotheses, and observational studies also
Cited by 8   Related articles   All 6 versions   Web of Science: 2   Cite   Save

Robust empirical calibration of p-values using observational data
MJ Schuemie, G Hripcsak, PB Ryan... - Statistics in ..., 2016 - Wiley Online Library
... Martijn J. Schuemie,. Corresponding author: ORCID:orcid.org/0000-0002-0817-5361. Janssen Research and Development LLC, Titusville, NJ, USA; Observational Health Data Sciences and Informatics (OHDSI). ... E-mail: schuemie@ohdsi.org. Search for more papers by this author ...
Related articles   All 5 versions   Cite   Save

Sharing Clinical Big Data While Protecting Confidentiality and Security: Observational Health Data Sciences and Informatics
RW Park - Healthcare Informatics Research, 2017 - synapse.koreamed.org
... Recently, distributed research networks (DRNs), such as Observational Health Data and Informatics (OHDSI, pronounced "Odyssey"), the National Patient Centered Clinical Research Network (PICORNET), or Sentinel Initiatives have gained popularity among clinical data ...
All 4 versions   Cite   Save   More
OHDSI across the globe
안녕하세요.
저는 삼성연합의과학원 디지털헬스학과의 우현기 입니다.

이번 2017 OHDSI Symposium 에 참여하여 매우 유익한 강의들을 들을 수 있어서 뜻깊은 시간이었고, 이런 자리를 마련해주신 관계자 분들의 노고에 깊은 감사를 드립니다.

몇가지 질문 및 요청사항이 있어서 토픽을 남깁니다.

1. 지난 수요일(3/8) 심포지엄 중 두번째 연사였던 Martijn Schuemie 의 강의 내용 중, R shiny 를 통하여 여러 분석데이터들을 도표화 하여 웹에 띄워 주는 아래와 같은 화면을 볼 수 있었고 매우 인상적이었습니다.
   (1) R shiny를 활용한 위 화면의 URL 주소를 좀 알 수 있을까요? (사진을 찍었는데 잘 보이지가 않네요 😞)
   (2) 위 프로그램의 R shiny Code도 혹시 함께 볼 수 있는 방법이 있을까요?

2. R shiny 로 작업할때 다루는 데이터인 csv 파일내의 행(row)이 약 55만개가 넘어가면 아래와 같은 에러가 나오는 경우가 있었습니다.
   **Error: 'from' cannot be NA, NaN or infinite**
   (1) 위와 같은 에러의 해결방법과 혹시 R shiny에서 다룰 수 있는 데이터의 양(용량 혹은 행, 열의 갯수)에 대한 Limitation이 있는지 궁금합니다. R shiny 써보신 분들은 위와 같은 에러를 접하신 분이 계시다면 의견을 부탁드립니다.
안녕하세요.
저는 삼성융합의과학원 디지털헬스학과의 우현기입니다.

이번 2017 OHDSI Symposium에 참여하여 매우 유익한 강의들을 들을 수 있어서 뜻깊은 시간이었고, 이런 자리를 마련해주신 관계자 분들의 노고에 깊은 감사를 드립니다.

몇가지 질문 및 요청사항이 있어서 토픽을 남깁니다.

1. 지난 수요일(3/8) 심포지엄 중 두번째 연사였던 Martijn Schuemie의 강의 내용 중, R shiny를 통하여 여러 분석데이터들을 도표화 하여 웹에 띄워 주는 아래와 같은 화면을 볼 수 있었고 매우 인상적이었습니다.

@whk Schuemie의 강의에 나온 내용에 대해서는 Schuemie에게 직접 문의하시면 될 듯 합니다. 관심 가져 주셔서 감사합니다. ^^
## Friendly Forums

<table>
<thead>
<tr>
<th>Metric</th>
<th>Today</th>
<th>Yesterday</th>
<th>Last 7 Days</th>
<th>Last 30 Days</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Visits</td>
<td>56</td>
<td>48</td>
<td>252</td>
<td>958</td>
<td>15863</td>
</tr>
<tr>
<td>New Users</td>
<td>4</td>
<td>5</td>
<td>21</td>
<td>58</td>
<td>892</td>
</tr>
<tr>
<td>Topics</td>
<td>5</td>
<td>3</td>
<td>15</td>
<td>47</td>
<td>1225</td>
</tr>
<tr>
<td>Posts</td>
<td>30</td>
<td>28</td>
<td>128</td>
<td>329</td>
<td>6601</td>
</tr>
<tr>
<td>Time to first response</td>
<td>1.87</td>
<td>2.22</td>
<td>17.01</td>
<td>28.8</td>
<td>167.9</td>
</tr>
</tbody>
</table>
Too Responsive?

Christian_Reich

@Gowtham_Rao, @hripcsa:

Just got scolded by the Forum website:

Let others join the conversation
This topic is clearly important to you – you've posted more than 28% of the replies here. Are you sure you're providing adequate time for other people to share their points of view, too?

So, this is the last one before I shut up. 😄
Breadth of Topics Evolving with Community

What -omics vocabularies are you using?

Philosophical question on oncology ETL

Creating HEDIS measures as cohort definitions in ATLAS

Using NLP sentiment to determine inclusion

OMOP use in Global Epi for CDC? Gates Foundation?

OMOP in Hadoop?

How do we define what is a ‘good study’?
Greetings,

I am Markeese Robinson and I work at a Federally Qualified Community Health Center located in Jackson, Mississippi.

I need help with the OMOP conversion tools that are listed on the OHDSI website. We would also like to understand the process to convert the EHR data to the OMOP format.

I look forward to hearing from you. Thank you in advance for any assistance that you can offer.
JACKSON-HINDS COMPREHENSIVE HEALTH CENTER (JHCHC)

As the largest provider of primary health care services to the poor, uninsured and undeserved population in Mississippi, the prime objective of JHCHC is to eliminate disparities in health care access for these groups, which often fall between the gaps of private insurance and personal income. Payments for services rendered are based on a number of factors including the patient’s ability to pay as well as grant fund availability for various services.
Time to roll up our sleeves, spit in our hands, and go to work.
# Today’s Agenda

<table>
<thead>
<tr>
<th>Time</th>
<th>Session Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:30 – 10:30am</td>
<td>Working Group Breakout Session – Part I  &lt;br&gt;  - Common data model and vocabulary – Proposal review (Room A)  &lt;br&gt;  - Population-level estimation (Room B)  &lt;br&gt;  - Hadoop (Room B)  &lt;br&gt;  - Vocabulary Visualization (Room B)  &lt;br&gt;  - Orientation for newcomers (Room C)</td>
</tr>
<tr>
<td>10:30 – 11:00am</td>
<td>Break</td>
</tr>
<tr>
<td>11:00 – 12:00pm</td>
<td>Working Group Breakout Session – Part II  &lt;br&gt;  - Common data model and vocabulary – Proposal review (Room A)  &lt;br&gt;  - Architecture (Room B)  &lt;br&gt;  - Patient-level prediction  &lt;br&gt;  - Natural language processing  &lt;br&gt;  - Orientation for newcomers</td>
</tr>
<tr>
<td>12:00 – 1:00pm</td>
<td>Reconvene the community  &lt;br&gt;  - Summary of key points from each work group  &lt;br&gt;  - Hack-a-thon presentation: Framing of target problems</td>
</tr>
<tr>
<td>1:00 – 1:30pm</td>
<td>Lunch</td>
</tr>
<tr>
<td>1:30 – End of day</td>
<td>Hack-a-thon – Three possible tracks  &lt;br&gt;  - Phenotyping and cohort building  &lt;br&gt;  - Large scale statistical computing  &lt;br&gt;  - Design session: UI experience and information dissemination</td>
</tr>
</tbody>
</table>
### Tomorrow’s Agenda

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:00 – 12:30pm</td>
<td>Continue hack-a-thon activities across three tracks:</td>
</tr>
<tr>
<td></td>
<td>• Phenotyping and cohort building</td>
</tr>
<tr>
<td></td>
<td>• Large scale statistical computing</td>
</tr>
<tr>
<td></td>
<td>• Design session: UI experience and information dissemination</td>
</tr>
<tr>
<td>12:30 – 1:30pm</td>
<td>Lunch</td>
</tr>
<tr>
<td>1:30 – 3:00pm</td>
<td>Reconvene the community</td>
</tr>
<tr>
<td></td>
<td>• Review outcomes from each hack-a-thon track</td>
</tr>
<tr>
<td>3:00 – 3:30pm</td>
<td>Break</td>
</tr>
<tr>
<td>3:30 – 5:30pm</td>
<td>Open community discussion</td>
</tr>
<tr>
<td></td>
<td>• Next steps for following through on group projects</td>
</tr>
<tr>
<td></td>
<td>• Other priorities for collaborative projects</td>
</tr>
<tr>
<td></td>
<td>• Other ways to engage the community and make contributions</td>
</tr>
<tr>
<td>5:30pm</td>
<td>Wrap-up</td>
</tr>
</tbody>
</table>