Welcome to the journey: OHDSI Symposium 2015

Wifi:
Network: HHONORS-MEETING
Passcode: OHDSI15
Welcome to the journey: Overview of OHDSI: past, present, future

Patrick Ryan, PhD
Janssen Research and Development
Columbia University Medical Center
20 October 2015
Odyssey (*noun*): \oh-d-si\ 

1. A long journey full of adventures
2. A series of experiences that give knowledge or understanding to someone
The Journey of Odysseus

1. Troy
2. Dardanelles
3. Lotus-Eaters
4. Thrinacia (island of the sun god)
5. Ogygia Isle (Calypso)
6. Laestrygonians (cannibals)
7. Aeaea (Circe)
8. Land of the Dead (Tiresias)
9. Aeolus (god of the winds)
10. Sirens
11. Charybdis
12. Scylla
13. Thrinacia
14. Ogygia Isle (Calypso)
15. Phaeacia
16. Ithaca

GRECE

ITALY

SICILY

Cyclops

Olympus

Land of the Dead

Laestrygonians

Troy

Dardanelles

Lotus-Eaters

Ogygia Isle (Calypso)

Olympus

Phaeacia

Ithaca

Thrinacia (island of the sun god)

Aeolus (god of the winds)

Sirens

Scylla

Charybdis

Laestrygonians (cannibals)

Aeaea (Circe)

Land of the Dead (Tiresias)
A journey to OHDSI
Greetings to you, the lucky finder of this GOLDEN TICKET from the OMOP Research Team

Present this ticket at the next OMOP Symposium in the morning and do not be late. You may bring with you one member of your own family...and only one....but no one else...

In your wildest dreams, you could not imagine the marvelous SURPRISES that await YOU!
Thanks to our sponsors

Thanks for the Eugene Washington Engagement Award
Lesson 1: Database heterogeneity:
Holding analysis constant, different data may yield different estimates

“Evaluating the Impact of Database Heterogeneity on Observational Study Results”
Lesson 2: Parameter sensitivity:
Holding data constant, different analytic design choices may yield different estimates

Madigan D, Ryan PB, Scheumie MJ, Therapeutic Advances in Drug Safety, 2013: “Does design matter? Systematic evaluation of the impact of analytical choices on effect estimates in observational studies”
Lesson 3: Empirical performance: Most observational methods do not have nominal statistical operating characteristics.

Lesson 4: Empirical calibration can help restore interpretation of study findings

Lesson 5: Reliable evidence generation isn’t (just) a data/analysis/technology problem

• Understanding the problems requires input and perspective from multiple stakeholders: government, industry, academia, health systems

• Research and development of novel solutions require multi-disciplinary approach: informatics, epidemiology, statistics, clinical sciences

• Adoption and application requires active participation and buy-in from all interested parties (both evidence producers and evidence consumers)

• Major outstanding need: to establish a community of individuals based on shared attitudes, interests and goals where everyone has equal opportunity to participate and contribute
Introducing OHDSI

• The Observational Health Data Sciences and Informatics (OHDSI) program is a multi-stakeholder, interdisciplinary collaborative to create open-source solutions that bring out the value of observational health data through large-scale analytics

• OHDSI has established an international network of researchers and observational health databases with a central coordinating center housed at Columbia University

http://ohdsi.org
Thanks for all of the supporters of the OHDSI community

Full list of acknowledgements: http://www.ohdsi.org/who-we-are/support-for-ohdsi/
OHDSI’s vision

OHDSI collaborators access a network of 1 billion patients to generate evidence about all aspects of healthcare. Patients and clinicians and other decision-makers around the world use OHDSI tools and evidence every day.

http://ohdsi.org
OHDSI: a global community

OHDSI Collaborators:
• >100 researchers in academia, industry, government, health systems
• >10 countries
• Multi-disciplinary expertise: epidemiology, statistics, medical informatics, computer science, machine learning, clinical sciences

http://www.ohdsi.org/who-we-are/collaborators/
Global reach of ohdsi.org

>16,800 distinct viewers from 120 countries in 2015
OMOP CDM now Version 5, following multiple iterations of implementation, testing, modifications, and expansion based on the experiences of the community who bring on a growing landscape of research use cases.
One model, multiple use cases

Standardized health system data
- Location
- Care_site
- Provider
- Payer_plan_period
- Visit_cost
- Procedure_cost
- Drug_cost
- Device_cost
- Cohort
- Cohort_attribute
- Condition_era
- Condition_occurrence
- Procedure_occurrence
- Visit_occurrence
- Observation
- Fact_relationship

Standardized derived elements
- Drug_exposure
- Device_exposure
- Drug_era
- Dose_era
- Domain
- Concept
- Vocabulary
- Concept_class
- Concept_relationship
- Relationship
- Concept_synonym
- Concept_ancestor
- Source_to_concept_map
- Drug_strength
- Cohort_definition
- Attribute_definition

Standardized meta-data
- CDM_source
Databases converted to OMOP CDM within OHDSI Community:
• 53 databases
• 660 million patients
• 12 countries

http://www.ohdsi.org/web/wiki/doku.php?id=resources:data_network
The odyssey to evidence generation

Patient-level data in source system/schema

Mount Olympus
7,929ft • 2,424m
Olympic NP, Washington

TRAVENSE OF THE GODDESSES, MOUNT OLYMPUS
Preparing your data for analysis

WhiteRabbit: profile your source data
RabbitInAHat: map your source structure to CDM tables and fields
ATHENA: standardized vocabularies for all CDM domains
Usagi: map your source codes to CDM vocabulary
CDM: DDL, index, constraints for Oracle, SQL Server, PostgreSQL; Vocabulary tables with loading scripts
ACHILLES: profile your CDM data; review data quality assessment; explore population-level summaries

OHDSI Forums:
Public discussions for OMOP CDM Implementers/developers

http://github.com/OHDSI
Data Evidence sharing paradigms

Single study
- Write Protocol
- Develop code
- Execute analysis
- Compile result

Real-time query
- Develop app
- Design query
- Submit job
- Review result

Large-scale analytics
- Develop app
- Execute script
- Explore results

Patient-level data in OMOP CDM

One-time
Repeated
What evidence does OHDSI seek to generate from observational data?

• Clinical characterization
  – Natural history: Who are the patients who have diabetes? Among those patients, who takes metformin?
  – Quality improvement: What proportion of patients with diabetes experience disease-related complications?

• Population-level estimation
  – Safety surveillance: Does metformin cause lactic acidosis?
  – Comparative effectiveness: Does metformin cause lactic acidosis more than glyburide?

• Patient-level prediction
  – Precision medicine: Given everything you know about me and my medical history, if I start taking metformin, what is the chance that I am going to have lactic acidosis in the next year?
  – Disease interception: Given everything you know about me, what is the chance I will develop diabetes?
What is OHDSI’s strategy to generate evidence?

• **Methodological research**
  – Develop new approaches to observational data analysis
  – Evaluate the performance of new and existing methods
  – Establish empirically-based scientific best practices

• **Open-source analytics development**
  – Design tools for data transformation and standardization
  – Implement statistical methods for large-scale analytics
  – Build interactive visualization for evidence exploration

• **Clinical applications**
  – Identify clinically-relevant questions that require real-world evidence
  – Execute research studies by applying scientific best practices through open-source tools across the OHDSI international data network
  – Promote open-science strategies for transparent study design and evidence dissemination
OHDSI ongoing collaborative activities

**Methodological research**
- Data quality assessment
- Common Data Model evaluation
- ATHENA for standardized vocabularies
- Phenotype evaluation
- Empirical calibration
- LAERTES for evidence synthesis
- Evaluation framework and benchmarking

**Open-source analytics development**
- WhiteRabbit for CDM ETL
- Usagi for vocabulary mapping
- HERMES for vocabulary exploration
- ACHILLES for database profiling
- CIRCE for cohort definition
- CALYPSO for feasibility assessment
- HERACLES for cohort characterization
- CohortMethod
- SelfControlledCaseSeries
- SelfControlledCohort
- TemporalPatternDiscovery
- PatientLevelPrediction
- APHRODITE for predictive phenotyping

**Clinical applications**
- Chronic disease therapy pathways
- HOMER for causality assessment
- PENEOPE for patient-centered product labeling
Poster: Establishing Interoperability Standards between OMOP CDM v4, v5, and PCORnet CDM v1

PCORNet CDRNs:
- NYC CDRN
- PEDSNet
- pSCANNER

Common standard reference model:
OMOP CDM

Common ETL for project destination:
PCORNet CDM
OHDSI ongoing collaborative activities

Poster: Transforming the National Department of Veterans Affairs Data Warehouse to the OMOP Common Data Model
OHDSI ongoing collaborative activities

- Methodological research
- Open-source analytics development
- Clinical application development
- Observational data management

OHDSI Community Booth: ETL 101

Image of a computer screen showing a data management interface with tables for source, CDM, and data fields.
OHDSI ongoing collaborative activities

Observational data management

Open-source analytics development

OHDSI Community Booth:
ATHENA for standardized vocabularies

ATHENA Download Page
Standardized Vocabularies for OMOP CDM

<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>71</td>
<td>ABMS</td>
<td>Provider Specialty (American Board of Medical Specialties)</td>
<td>EULA required</td>
<td>31-JAN-15</td>
</tr>
<tr>
<td>1</td>
<td>SNOMED</td>
<td>Systematic Nomenclature of Medicine - Clinical Terms (SNOMED)</td>
<td>01-OCT-14</td>
<td>01-OCT-14</td>
</tr>
<tr>
<td>2</td>
<td>ICD9CM</td>
<td>International Classification of Diseases, Ninth Revision, Clinical Modification, Volume 1 and 2 (ICD9CM)</td>
<td>01-OCT-14</td>
<td>01-OCT-14</td>
</tr>
<tr>
<td>3</td>
<td>ICD10CM</td>
<td>International Classification of Diseases, Tenth Revision, Clinical Modification, Volume 3 (ICD10CM)</td>
<td>10-OCT-14</td>
<td>12-NOV-14</td>
</tr>
<tr>
<td>5</td>
<td>HCPCS</td>
<td>Healthcare Common Procedure Coding System (HCPCS)</td>
<td>EULA required</td>
<td>20-JUN-15</td>
</tr>
<tr>
<td>6</td>
<td>LoINC</td>
<td>Logical Observation Identifiers Names and Codes (LOINC)</td>
<td>EULA required</td>
<td>20-JUN-15</td>
</tr>
<tr>
<td>7</td>
<td>NDFRT</td>
<td>National Drug File - Reference Terminology (NDFRT)</td>
<td>EULA required</td>
<td>20-JUN-15</td>
</tr>
</tbody>
</table>

Fill out the form, pick the required vocabularies and select the right version.
OHDSI ongoing collaborative activities

Open-source analytic demo: HERMES for vocabulary exploration

Observational data management

Open-source analytics development
OHDSI ongoing collaborative activities

Observational data management

Open-source analytics development

Open-source analytic demo: ACHILLES for database profiling
OHDSI ongoing collaborative activities

Poster:
Determination of Pregnancy Episodes and Outcomes within a Distributed Network of Observational Databases
OHDSI ongoing collaborative activities

Poster:
Size comparison of 17 CDM datasets using IRIS tool

<table>
<thead>
<tr>
<th>MEASURE</th>
<th>RESULT</th>
<th>EXPLANATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>141,805,491</td>
<td>count of patients</td>
</tr>
<tr>
<td>G2</td>
<td>20,328,289,501</td>
<td>count of events</td>
</tr>
<tr>
<td>D2</td>
<td>90,024,522</td>
<td>count of patients with at least 1 Dx and 1 Rx</td>
</tr>
<tr>
<td>D3</td>
<td>112,148,500</td>
<td>count of patients with at least 1 Dx and 1 Proc</td>
</tr>
<tr>
<td>D4</td>
<td>5,939,621</td>
<td>count of patients with at least 1 Obs, 1 Dx and 1 Rx</td>
</tr>
<tr>
<td>D5</td>
<td>277,975</td>
<td>count of deceased patients</td>
</tr>
</tbody>
</table>
OHDSI ongoing collaborative activities

Clinical characterization

Methodological research

Poster:
Lessons from CIRCE implementation of eMERGE phenotype definitions into actionable CDM v5 SQL queries
OHDSI ongoing collaborative activities

Open-source analytics demos:
CIRCE for cohort definition
CALYPSO for feasibility assessment
OHDSI ongoing collaborative activities

Clinical characterization

Open-source analytics development

Open-source analytics demos:
HERACLES for cohort characterization

PENELLOPE exposure: celecoxib
OHDSI ongoing collaborative activities

Poster: 
Exploration of the Epidemiology of Endometriosis

Endometriosis

Very common
More than 3 million US cases per year

- Treatable by a medical professional
- Requires a medical diagnosis
- Lab tests or imaging often required
- Chronic: can last for years or be lifelong

With endometriosis, the tissue can be found on the ovaries, fallopian tubes or the intestines.
The most common symptoms are pain and menstrual irregularities.
Effective treatments, such as hormones and excision surgery, are available.
OHDSI ongoing collaborative activities

Presentation:
Treatment pathways in chronic disease
OHDSI ongoing collaborative activities

Population-level estimation

Methodological research

Open-source analytic demo:
LAERTES for evidence integration
OHDSI ongoing collaborative activities

**Poster:**
**Accuracy of an Automated Knowledgebase for Identifying Adverse Drug Reactions**

Figure 1. Histograms of predicted probabilities with AUCs for positive/negative controls in the various reference sets, using the model trained on both OMOP and EU-ADR ref. set.

- OMOP: AUC = 0.87 (0.83-0.92)
- EUADR: AUC = 0.95 (0.90-0.99)
- AZCERT: AUC = 0.86 (0.79-0.93)

OMOP: Observational Medical Outcomes Partnership, EU-ADR: Exploring and Understanding Adverse Drug Reactions, AUC: area under the curve
OHDSI ongoing collaborative activities

Poster:
How high can we go? Evaluating massively high-dimensional propensity score and outcome models in large-scale observational studies
OHDSI ongoing collaborative activities

Population-level estimation

Open-source analytics development

Open-source analytics demos:

- Cohort Method
- Self-controlled case series
- Empirical calibration

![Graph showing density distribution for two groups, treated and comparator, with preference score on the x-axis and density on the y-axis.](image)
OHDSI ongoing collaborative activities

Poster:
A Climate-Wide Journey to Explore Mechanisms Underlying Birth Month Disease Risk Associations
OHDSI ongoing collaborative activities

Poster:
Discovering the hidden risk factors: An empirical evaluation of incorporating feature-learning methods into a risk model framework using the OMOP CDM
OHDSI ongoing collaborative activities

Open-source analytic demo:
APHRODITE for predictive phenotyping

Package ‘Aphrodite’
October 15, 2015

Type Package
Title Automated PHenotype Routine for Observational Definition Identification Training and Evaluation (APHRODITE) - Phenotype building tool using Fuzzy labels
Version 1.2
Date 2015-09-21
Author Juan M. Banda [aut, cre],
Kate Niehaus [aut],
Marc A. Suchard [aut],
Martijn J. Schuemie [aut]
Maintainer Juan M. Banda <jmbanda@stanford.edu>
Description Aphrodite uses noisy class labels to create silver standard training corpora to construct phenotype models in conjunction with expert knowledge codified in existing ontologies and a comprehensive representation of the patient clinical record to learn phenotype models.
License Apache License 2.0 file LICENSE
Depends R (>= 3.1.0)
data.table
OHDSI ongoing collaborative activities

Poster: Lift your Anchors and Begin the OHDSI with APHRODITE

Figure 1: (Left) Phenotypes currently being identified in real-time at BIDMC. (Right) Display screen where one of the phenotypes has been used to recommend a pathway of care.
OHDSI ongoing collaborative activities

Methodological research

Open-source analytics development

Clinical applications

Poster: OHDSI Cloud Architecture
OHDSI commercial ecosystem
Journey through the OHDSI collaborator showcase

- Data modeling and standardization
- Clinical applications
- Methods research
- Analytics development
- Ecosystem
- Registration
- Tech
Panel Discussion – Experiences from the OHDSI international data network
Panel Discussion – The Value and Challenges of Evidence from Observational Data: A Multi-Stakeholder Perspective

• Moderator: David Madigan, PhD, Executive Vice President and Dean of the Faculty of Arts and Sciences at Columbia University

• Robert Ball, MD, MPH, ScM, Deputy Director – Office of Surveillance and Epidemiology, CDER, US Food and Drug Administration

• Invited: Robert Califf, MD, Deputy Commissioner of Medical Products and Tobacco, US Food and Drug Administration

• Nareesa Mohammed-Rajput, MD, Medical Director of Clinical Informatics, Suburban Hospital part of Johns Hopkins Medicine

• Maryan Zirkle MD, MS, MA, Program Officer – CER Methods and Infrastructure Program, PCORI

• Lesley Wise, Vice President of PV Risk Management and Pharmacoepidemiology, Takeda Pharmaceuticals
Future of OHDSI

This is your journey....

....where do we go from here?
I asked you to participate...

https://www.surveymonkey.com/r/59GTY6X

Let’s generate some evidence...
Standardized large-scale analytics tools under development within OHDSI

ACHILLES: Database profiling

CIRCE: Cohort definition

HERMES: Vocabulary exploration

HERACLES: Cohort characterization

CALYPSO: Feasibility assessment

OHDSI Methods Library: CYCLOPS
   - CohortMethod
   - SelfControlledCaseSeries
   - SelfControlledCohort
   - TemporalPatternDiscovery
   - Empirical Calibration

LAERTES: Drug-AE evidence base

PLATO: Patient-level predictive modeling

PENELlope: Patient-centered product labeling

HOMER: Population-level causality assessment
Thank you OHDSI Symposium Organizing Committee

David Sontag
NYU

Chunhua Weng
Columbia University

Jon Duke
Regenstrief

Ana Szarfman
FDA

Charlie Bailey
CHOP

Gregory Fusco
Takeda
Thank you Maura Beaton
Join the journey

Interested in OHDSI?
Questions or comments?

Contact:
ryan@ohdsi.org