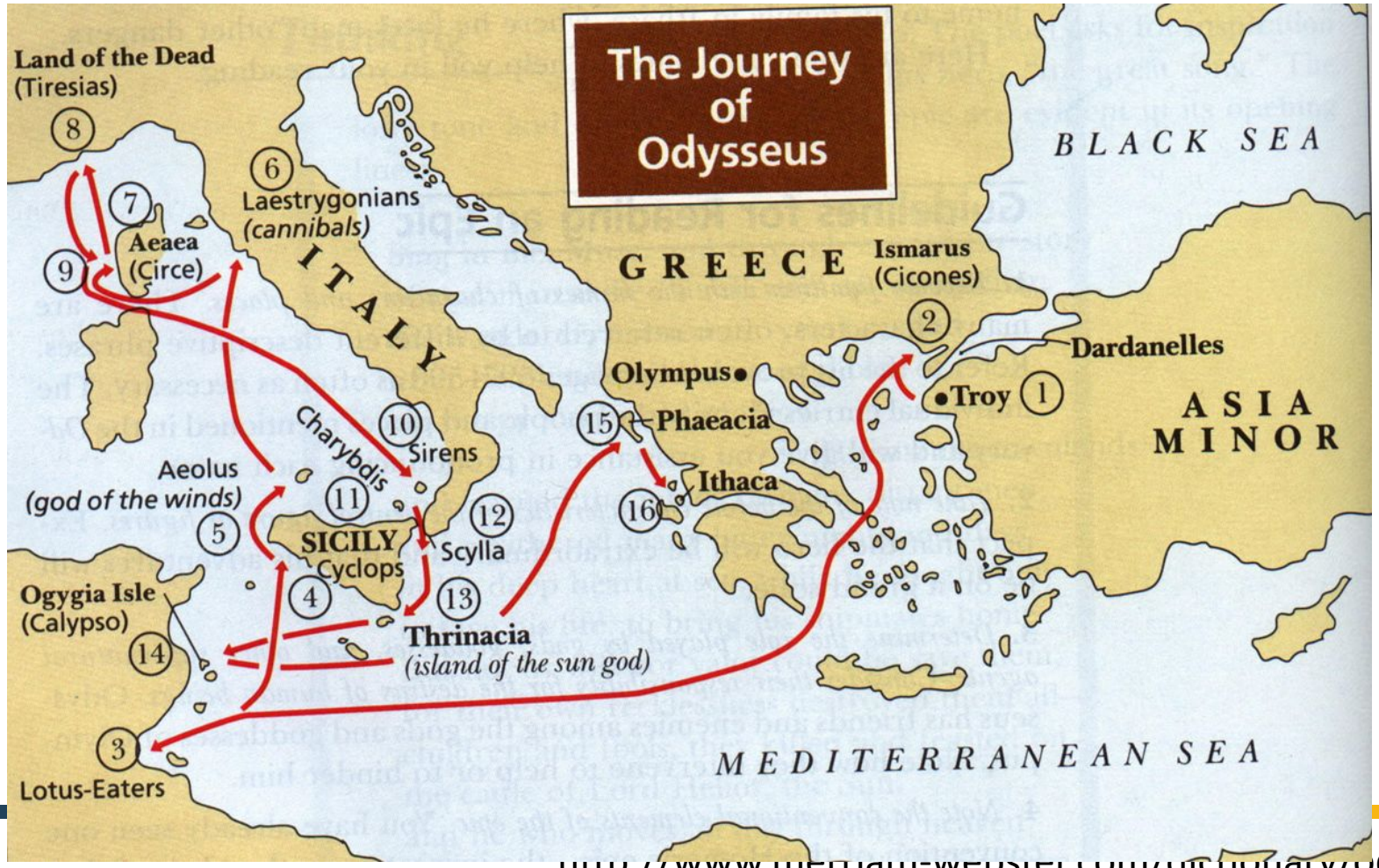
OHDSI's Journey: Where we've been, where we can go

Patrick Ryan, PhD
Janssen Research and Development
Columbia University Medical Center
10 January 2017

Odyssey (*noun*): \oh-d-si\

1. A long journey full of adventures



http://www.merriam-webster.com/dictionary/odyssey

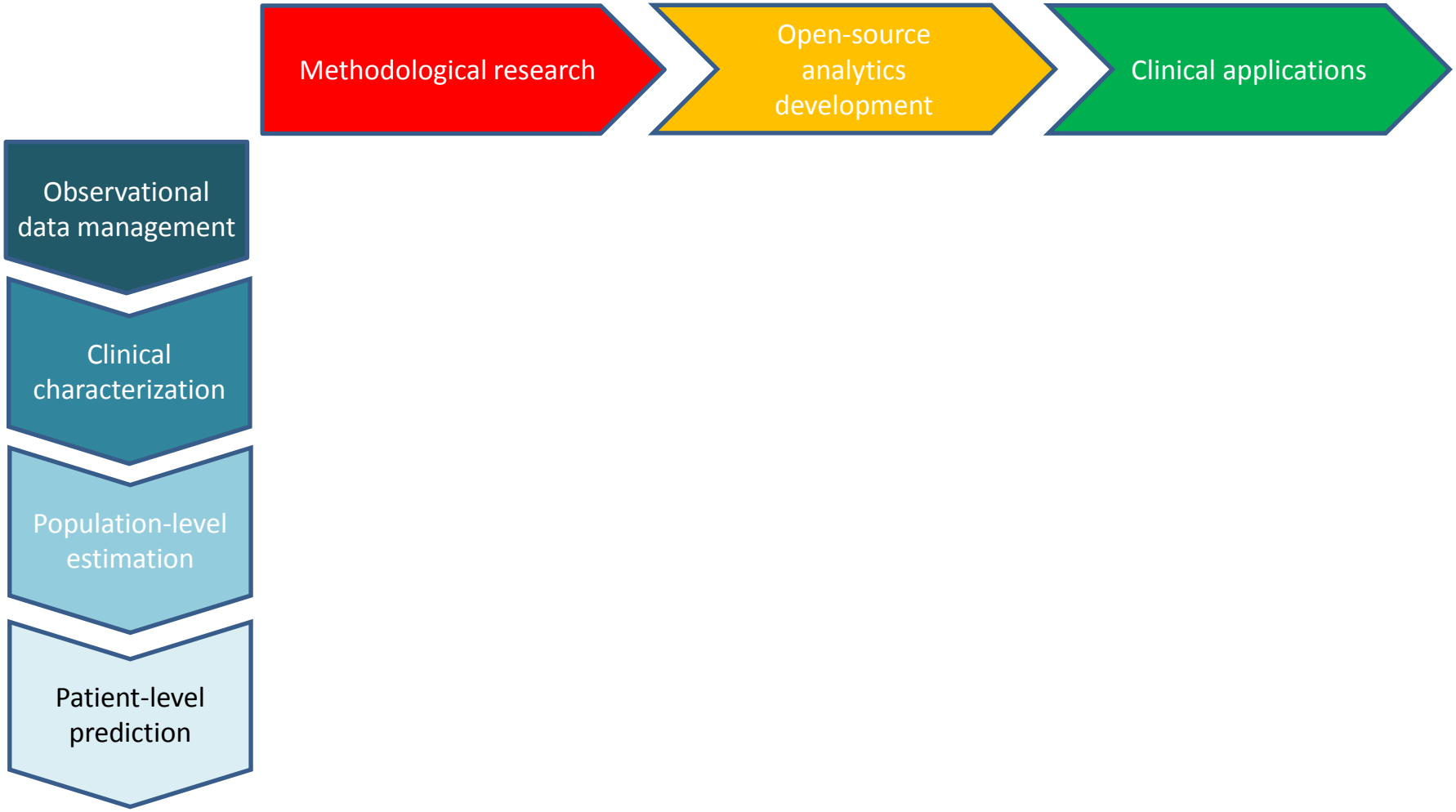
OHDSI's mission

To improve health, by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.

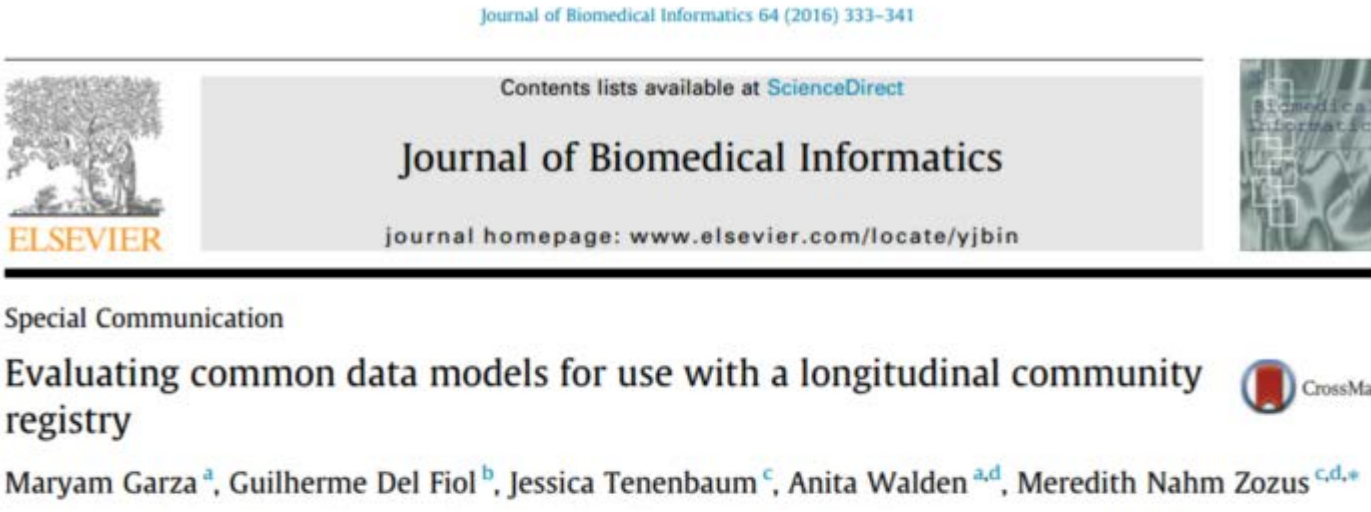
OHDSI's values

- **Innovation**: Observational research is a field which will benefit greatly from disruptive thinking. We actively seek and encourage fresh methodological approaches in our work.
- **Reproducibility**: Accurate, reproducible, and well-calibrated evidence is necessary for health improvement.
- **Community**: Everyone is welcome to actively participate in OHDSI, whether you are a patient, a health professional, a researcher, or someone who simply believes in our cause.
- **Collaboration**: We work collectively to prioritize and address the real world needs of our community's participants.
- **Openness**: We strive to make all our community's proceeds open and publicly accessible, including the methods, tools and the evidence that we generate.
- **Beneficence**: We seek to protect the rights of individuals and organizations within our community at all times.

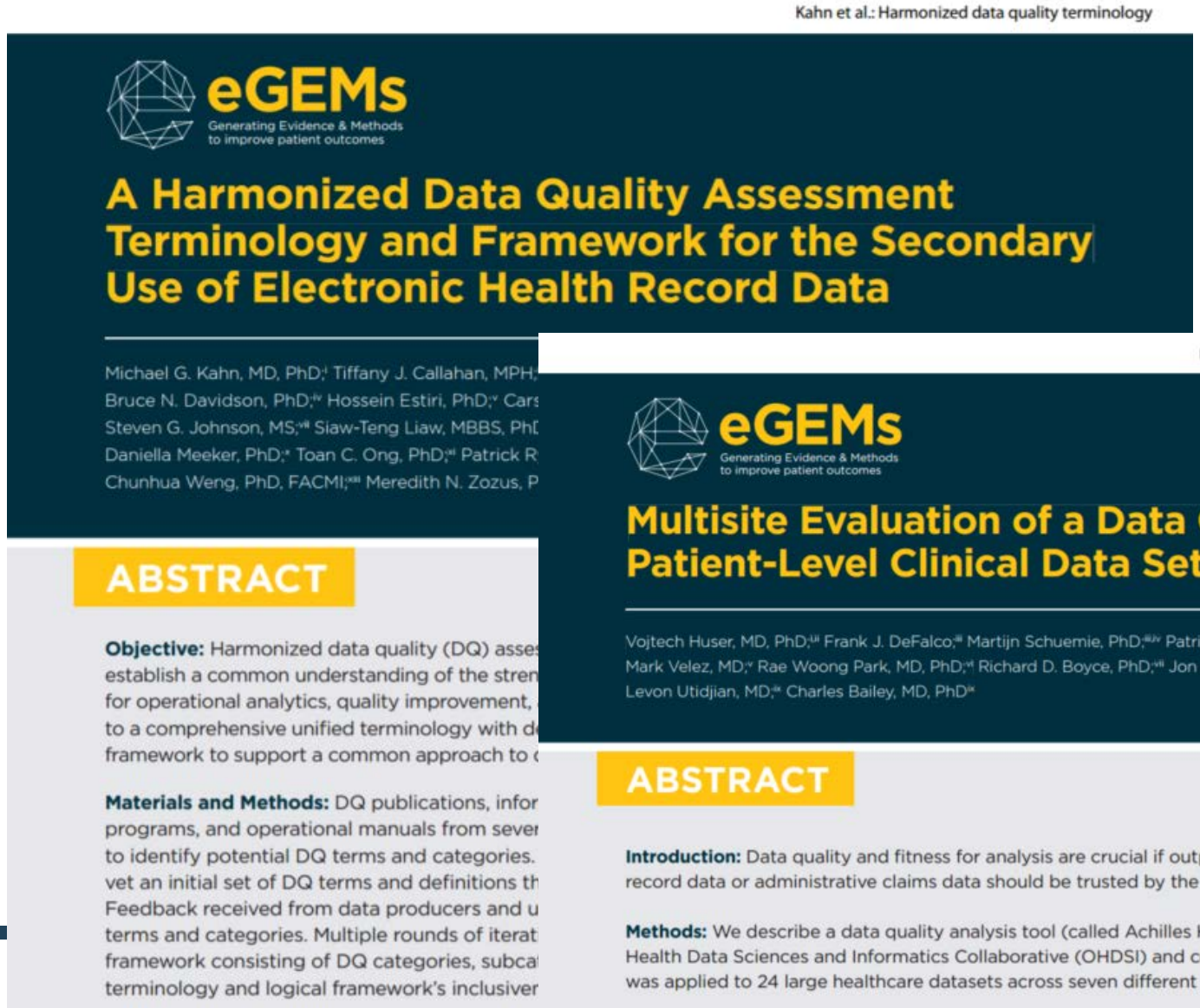
OHDSI's areas of focus



Journey toward best practices in data management



Journey toward best practices in data quality assessment



Kahn et al.: Harmonized data quality terminology

eGEMs
Generating Evidence & Methods to improve patient outcomes

A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data

Michael G. Kahn, MD, PhD;[i] Tiffany J. Callahan, MPH;
Bruce N. Davidson, PhD;[iv] Hossein Estiri, PhD;[v] Cars
Steven G. Johnson, MS;[vii] Siaw-Teng Liaw, MBBS, PhD
Daniella Meeker, PhD;[x] Toan C. Ong, PhD;[xi] Patrick R
Chunhua Weng, PhD, FACMI;[xiii] Meredith N. Zozus, P

ABSTRACT

Objective: Harmonized data quality (DQ) asses
establish a common understanding of the stren
for operational analytics, quality improvement,
to a comprehensive unified terminology with d
framework to support a common approach to

Materials and Methods: DQ publications, infor
programs, and operational manuals from sever
to identify potential DQ terms and categories.
vet an initial set of DQ terms and definitions th
Feedback received from data producers and u
terms and categories. Multiple rounds of iterat
framework consisting of DQ categories, subca
terminology and logical framework's inclusiver

Huser et al.: Evaluation of Achilles Heel (Data Quality Tool)

eGEMs
Generating Evidence & Methods to improve patient outcomes

Multisite Evaluation of a Data Quality Tool for Patient-Level Clinical Data Sets

Vojtech Huser, MD, PhD;[iii] Frank J. DeFalco;[iii] Martijn Schuemie, PhD;[iii,v] Patrick B. Ryan, PhD;[iii] Ning Shang, PhD;[v]
Mark Velez, MD;[v] Rae Woong Park, MD, PhD;[vi] Richard D. Boyce, PhD;[vii] Jon Duke, MD, MS;[viii] Ritu Khare, PhD;[ix]
Levon Utidjian, MD;[ix] Charles Bailey, MD, PhD[ix]

ABSTRACT

Introduction: Data quality and fitness for analysis are crucial if outputs of analyses of electronic health record data or administrative claims data should be trusted by the public and the research community.

Methods: We describe a data quality analysis tool (called Achilles Heel) developed by the Observational Health Data Sciences and Informatics Collaborative (OHDSI) and compare outputs from this tool as it was applied to 24 large healthcare datasets across seven different organizations.

Journey toward open data standardization

www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES

» Translational research

» Outcomes research

» Adverse effects

» Drug safety

Data Descriptor: A curated and standardized adverse drug event resource to accelerate drug safety research

Juan M. Banda[1], Lee Evans[2], Rami S. Vanguri[3], Nicholas P. Tatonetti[3], Patrick B. Ryan[4] & Nigam H. Shah[1]

Received: 17 December 2015
Accepted: 24 March 2016
Published: 10 May 2016

Identification of adverse drug reactions (ADRs) during the post-marketing phase is one of the most important goals of drug safety surveillance. Spontaneous reporting systems (SRS) data, which are the mainstay of traditional drug safety surveillance, are used for hypothesis generation and to validate the newer approaches. The publicly available US Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS) data requires substantial curation before they can be used appropriately, and applying different strategies for data cleaning and normalization can have material impact on analysis results. We provide a curated and standardized version of FAERS removing duplicate case records, applying standardized vocabularies with drug names mapped to RxNorm concepts and outcomes mapped to SNOMED-CT concepts, and pre-computed summary statistics about drug-outcome relationships for general consumption. This publicly available resource, along with the source code, will accelerate drug safety research by reducing the amount of time spent performing data management on the source FAERS reports, improving the quality of the

Journey toward leveraging open data to inform new analyses

Journal of Biomedical Informatics 66 (2017) 72–81

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Accuracy of an automated knowledge base for identifying drug adverse reactions

E.A. Voss [a,b,c,*], R.D. Boyce [d,c], P.B. Ryan [a,e,c], J. van der Lei [b,c], P.R. Rijnbeek [b,c], M.J. Schuemie [a,c]

[a] Epidemiology Analytics, Janssen Research & Development, LLC, Raritan, NJ, United States
[b] Erasmus University Medical Center, Rotterdam, Netherlands
[c] Observational Health Data Sciences and Informatics (OHDSI), New York, NY, United States
[d] University of Pittsburgh, Pittsburgh, PA, United States
[e] Columbia University, New York, NY, United States

ARTICLE INFO

Article history:
Received 21 July 2016
Revised 8 December 2016
Accepted 10 December 2016
Available online 16 December 2016

Keywords:
Pharmacovigilance
Adverse drug reaction
Machine-learning experiment
Knowledge base
Health outcome

ABSTRACT

Introduction: Drug safety researchers seek to know the degree of certainty with which a particular drug is associated with an adverse drug reaction. There are different sources of information used in pharmacovigilance to identify, evaluate, and disseminate medical product safety evidence including spontaneous reports, published peer-reviewed literature, and product labels. Automated data processing and classification using these evidence sources can greatly reduce the manual curation currently required to develop reference sets of positive and negative controls (i.e. drugs that cause adverse drug events and those that do not) to be used in drug safety research.

Methods: In this paper we explore a method for automatically aggregating disparate sources of information together into a single repository, developing a predictive model to classify drug-adverse event relationships, and applying those predictions to a real world problem of identifying negative controls for statistical method calibration.

Results: Our results showed high predictive accuracy for the models combining all available evidence, with an area under the receiver-operator curve of $\geqslant 0.92$ when tested on three manually generated lists of drugs and conditions that are known to either have or not have an association with an adverse drug event.

Conclusions: Results from a pilot implementation of the method suggests that it is feasible to develop a scalable alternative to the time-and-resource-intensive, manual curation exercise previously applied to develop reference sets of positive and negative controls to be used in drug safety research.

Journey toward large-scale open science

PNAS

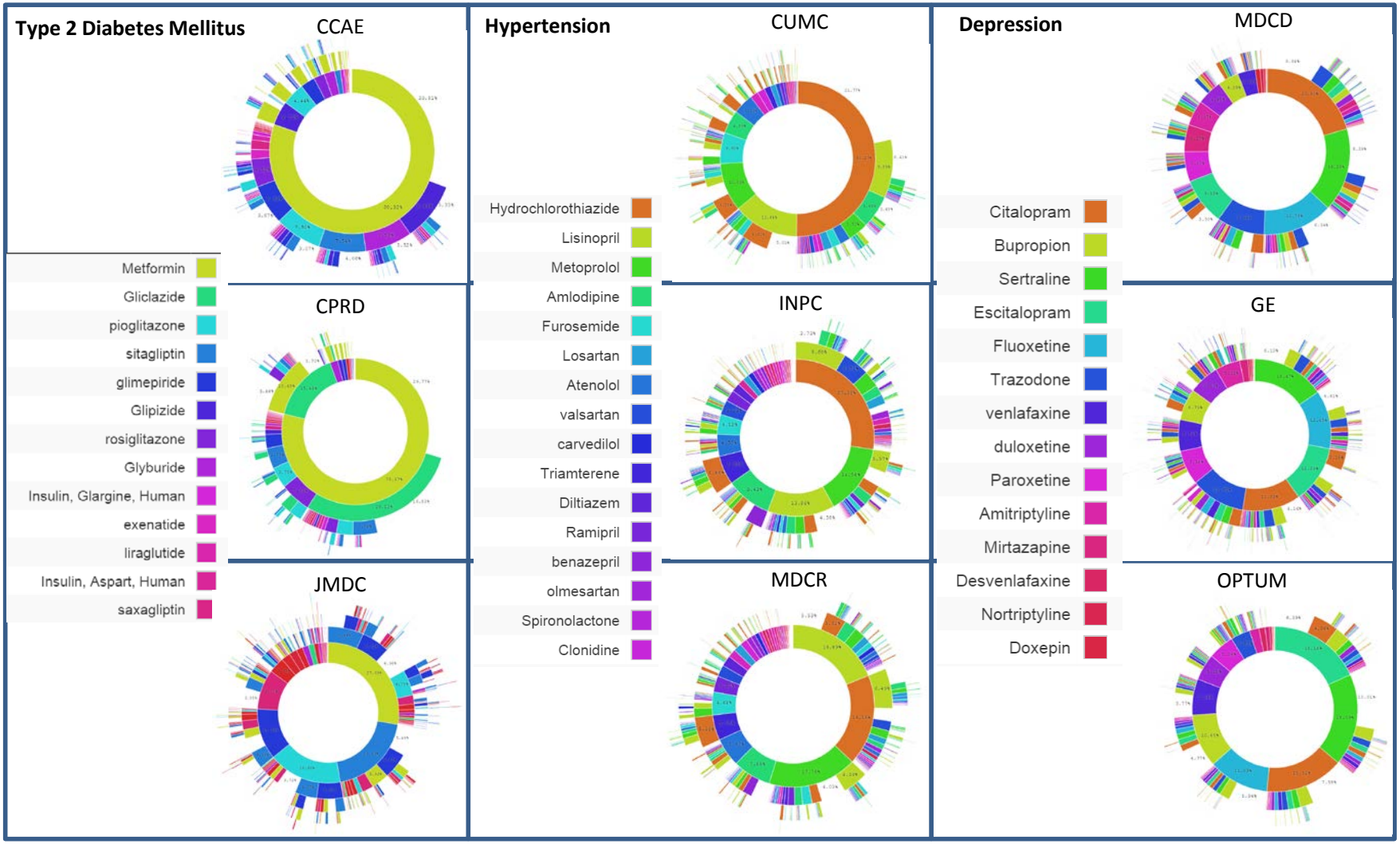
COLLOQUIUM PAPER

Characterizing treatment pathways at scale using the OHDSI network

George Hripcsak[a,b,c,1], Patrick B. Ryan[c,d], Jon D. Duke[c,e], Nigam H. Shah[c,f], Rae Woong Park[c,g], Vojtech Huser[c,h], Marc A. Suchard[c,i,j,k], Martijn J. Schuemie[c,d], Frank J. DeFalco[c,d], Adler Perotte[a,c], Juan M. Banda[c,f], Christian G. Reich[c,l], Lisa M. Schilling[c,m], Michael E. Matheny[c,n,o], Daniella Meeker[c,p,q], Nicole Pratt[c,r], and David Madigan[c,s]

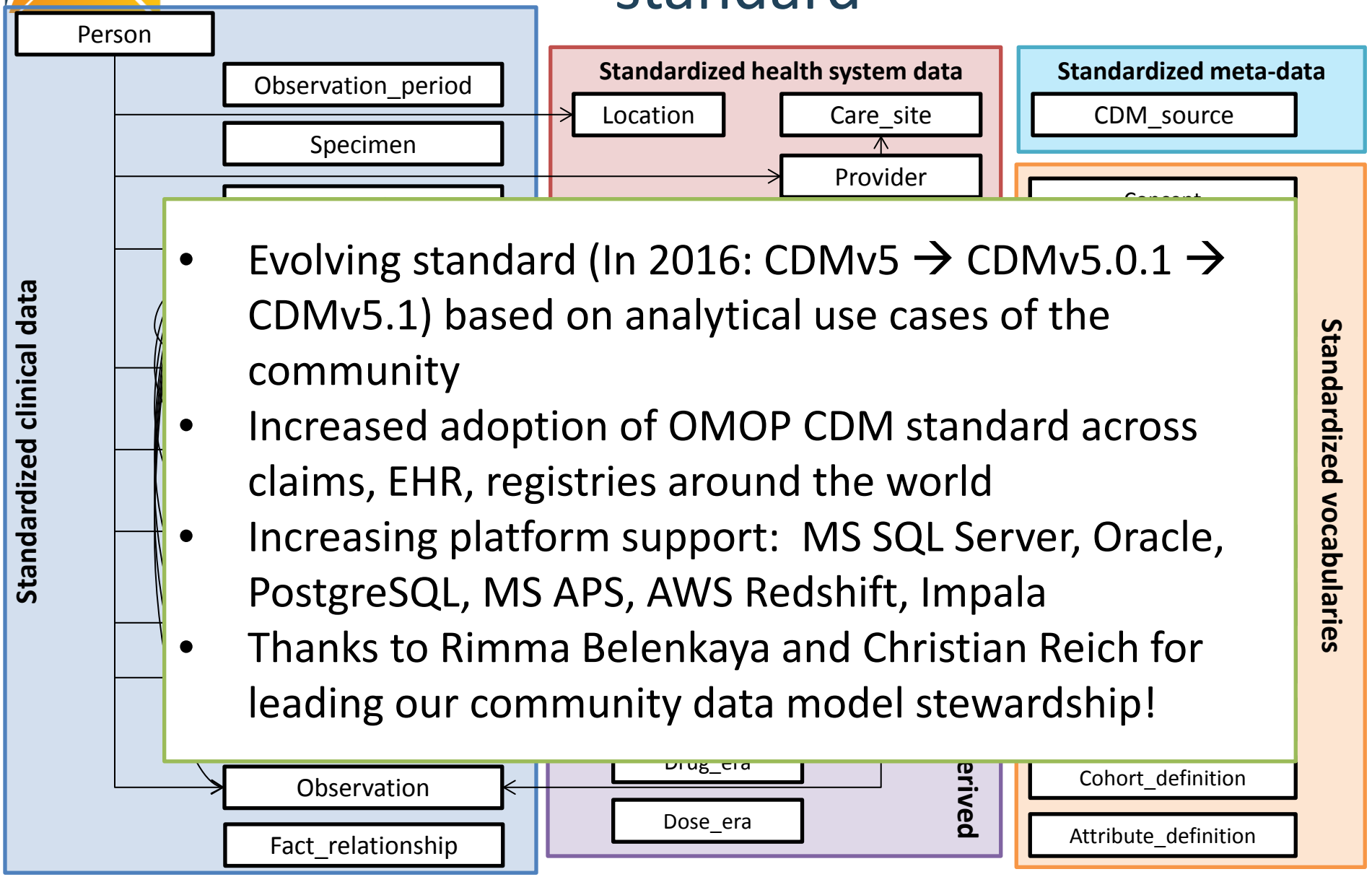
[a]Department of Biomedical Informatics, Columbia University Medical Center, New York, NY 10032; [b]Medical Informatics Services, NewYork-Presbyterian Hospital, New York, NY 10032; [c]Observational Health Data Sciences and Informatics, New York, NY 10032; [d]Epidemiology Analytics, Janssen Research and Development, Titusville, NJ 08560; [e]Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN 46205; [f]Center for Biomedical Informatics Research, Stanford University, CA 94305; [g]Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, South Korea, 443-380; [h]Lister Hill National Center for Biomedical Communications (National Library of Medicine), National Institutes of Health, Bethesda, MD 20894; [i]Department of Biomathematics, University of California, Los Angeles, CA 90095; [j]Department of Biostatistics, University of California, Los Angeles, CA 90095; [k]Department of Human Genetics, University of California, Los Angeles, CA 90095; [l]Real World Evidence Solutions, IMS Health, Burlington, MA 01809; [m]Department of Medicine, University of Colorado School of Medicine, Aurora, CO 80045; [n]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37212; [o]Geriatric Research, Education and Clinical Center, VA Tennessee Valley Healthcare System, Nashville, TN 37212; [p]Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90089; [q]Department of Pediatrics, University of Southern California, Los Angeles, CA 90089; [r]Division of Health Sciences, University of South Australia, Adelaide, SA, Australia 5001; and [s]Department of Statistics, Columbia University, New York, NY 10027

Journey to clinical characterization insights required network research



Hripcsak et al, PNAS, 2016

Journey of an open community data standard



- Evolving standard (In 2016: CDMv5 → CDMv5.0.1 → CDMv5.1) based on analytical use cases of the community
- Increased adoption of OMOP CDM standard across claims, EHR, registries around the world
- Increasing platform support: MS SQL Server, Oracle, PostgreSQL, MS APS, AWS Redshift, Impala
- Thanks to Rimma Belenkaya and Christian Reich for leading our community data model stewardship!

OHDSI's standardized vocabularies

- 69 Vocabularies across 25 domains
  -
  -
- 3,526,752 concepts
  - 1,174,628 standard concepts
  - 1,418,696 source codes
  - 323,159 classification concepts
- 24,064,798 concept relationships

Thank you Christian and the Odysseus team for continue to steward, maintain, and improve this invaluable resource for the entire community!

Our journey as a community of collaborators

OHDSI Collaborators:

- >140 researchers in academia, industry, government, health systems
- >20 countries
- Multi-disciplinary expertise: epidemiology, statistics, medical informatics, computer science, machine learning, clinical sciences

Databases converted to OMOP CDM within OHDSI Community:

- >50 databases
- >660 million patients

Forums.ohdsi.org

- 361 distinct users who have posted
- 5,823 posts on 1,100 topics
- Active discussions across all categories:
  - Implementers, Developers, Researchers, CDM Builders, Vocabulary users

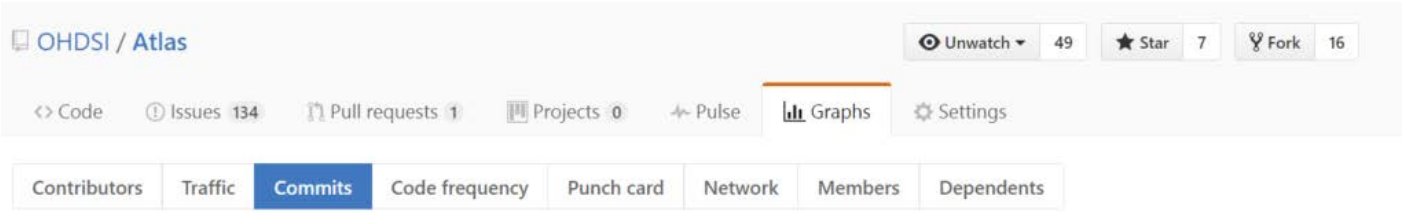
Thank you Christian Reich, Lee Evans, and Chris Knoll for being our most diligent community responders!

Journey toward open-source analytics development

- 61 developers on OHDSI GitHub repositories
- Applications in development or released for:
  - CDM ETL design and implementation
  - Clinical characterization (ACHILLES, ATLAS)
  - Population-level effect estimation (CohortMethod)
  - Patient-level prediction
  - OHDSI network studies (protocol + source code)

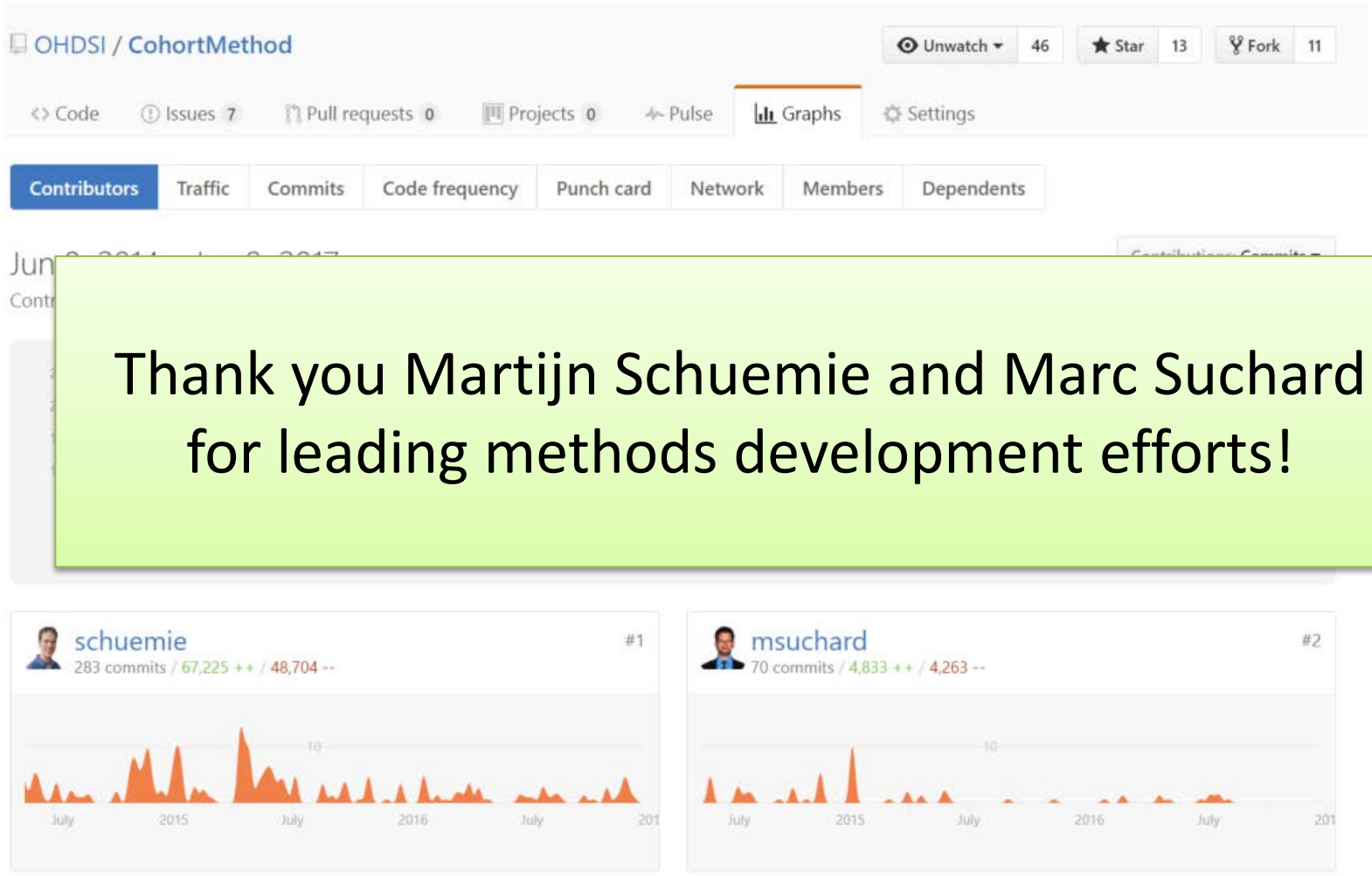
Journey toward open-source analytics development

ATLAS – a single community platform for:
-vocabulary browsing
-database characterization
-cohort definition
-incidence rate
-patient profiles
-population-level effect estimation study design
….



Thank you Frank DeFalco for leading the OHDSI architecture workgroup!

Journey toward open-source analytics development



Thank you Martijn Schuemie and Marc Suchard for leading methods development efforts!

OHDSI Symposium 2016

- 410 registrants from 11 countries, 27 US states

Congratulations to Christophe Lambert, Sigfried Gold, and Rupa Makadia on winning the OHDSI Best Poster awards!

- 48 poster presentations in observational data management, methodological research, analytics technology and infrastructure, and clinical applications

Thank you OHDSI tutorial faculty!




CDM/ETL: Rimma Belenkaya, Karthik Natarajan, Mark Velez, Erica Voss


Technology stack: Taha Abdul-Basser, Lee Evans, Karthik Natarajan, Mark Velez




Cohort definition: Juan Banda, Jon Duke, Chris Knoll, Nigam Shah

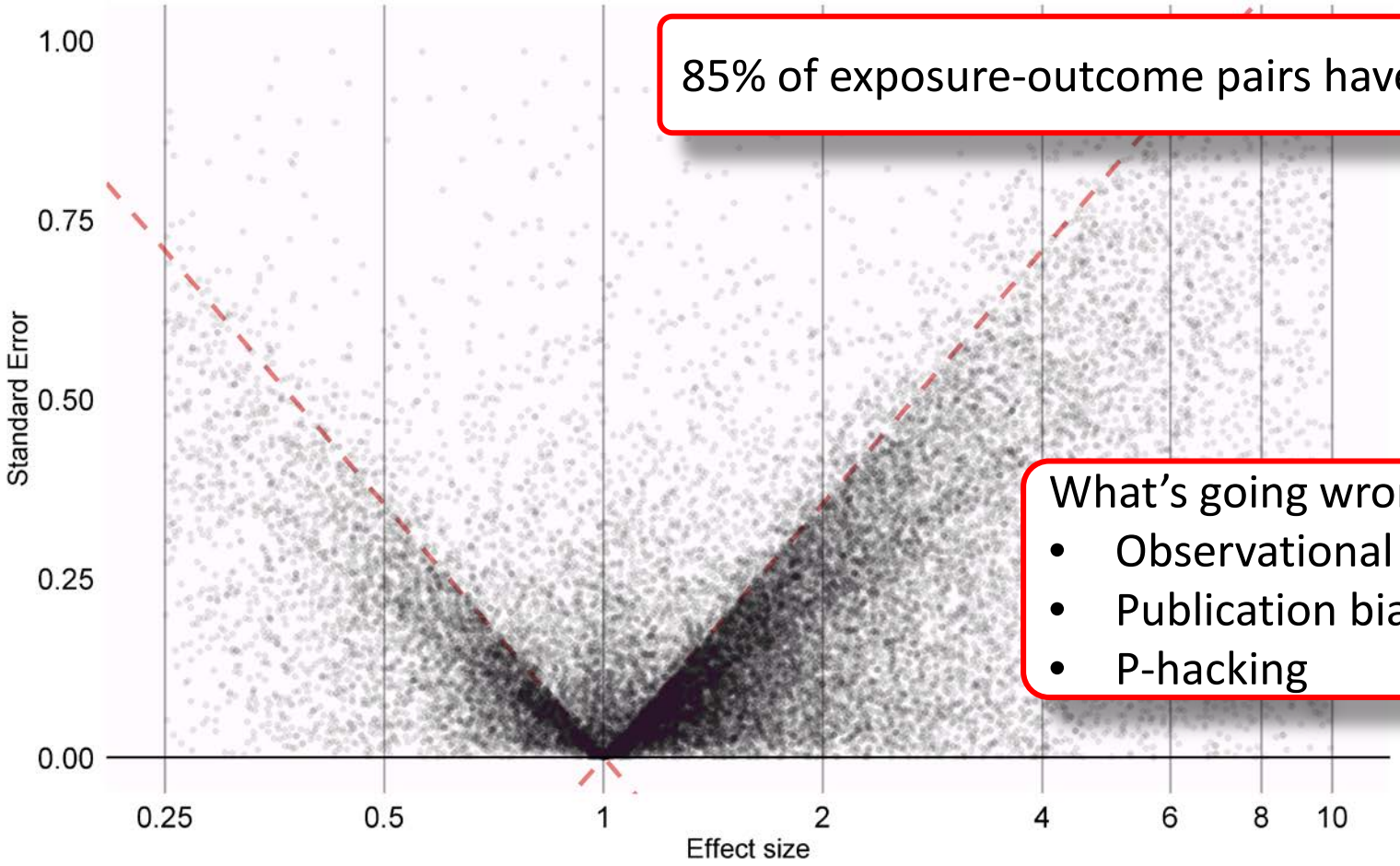
Vocabulary:
Frank DeFalco, George Hripcsak, Christian Reich






Population-level estimation:
David Madigan, Martijn Schuemie, Marc Suchard

Observational research results in literature



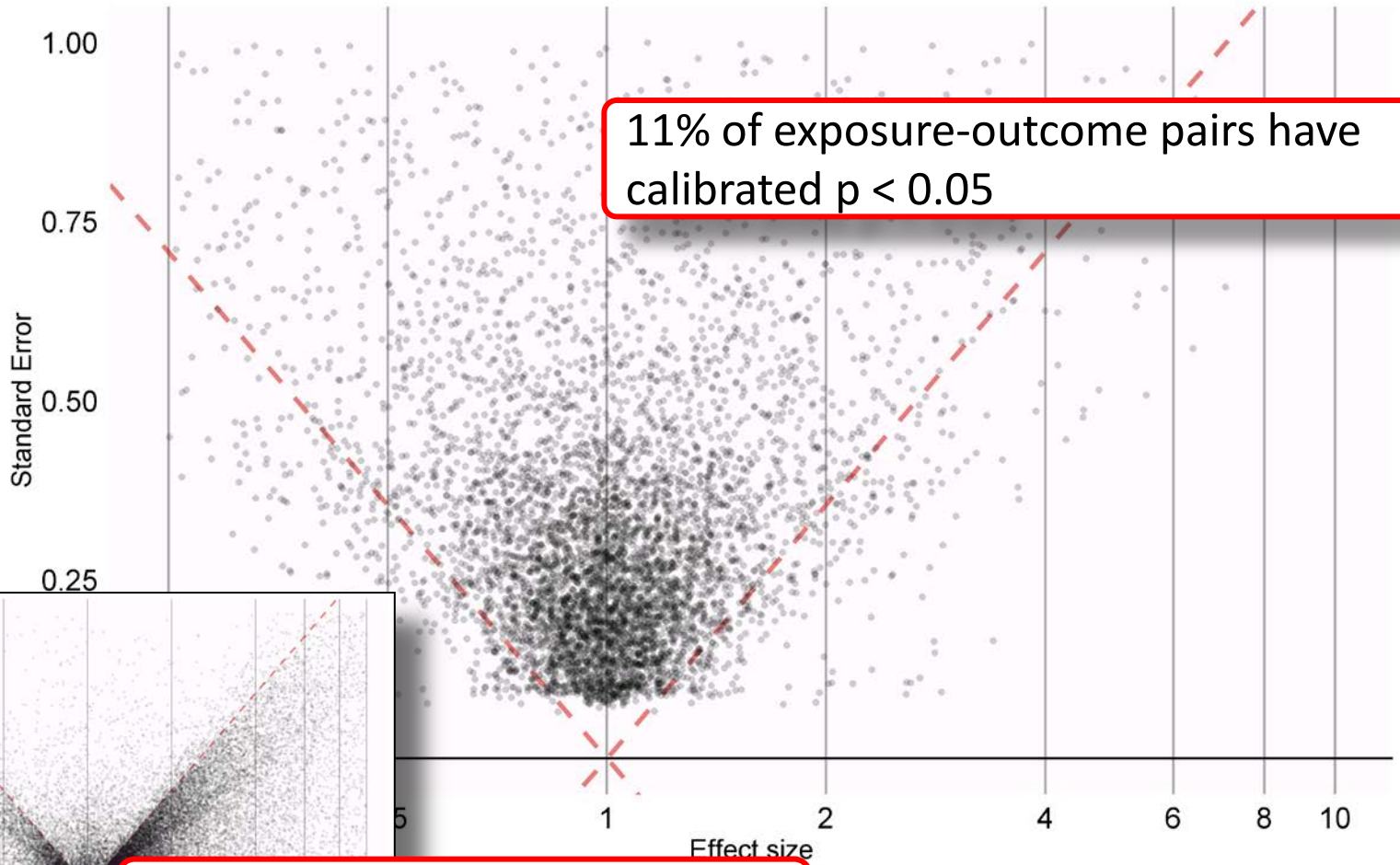
85% of exposure-outcome pairs have p < 0.05

What's going wrong?

- Observational study bias
- Publication bias
- P-hacking

29,982 estimates
11,758 papers

Large-scale analysis can start to produce reliable evidence to enable an honest learning healthcare system



11% of exposure-outcome pairs have calibrated p < 0.05

In literature, 85% have p < 0.05

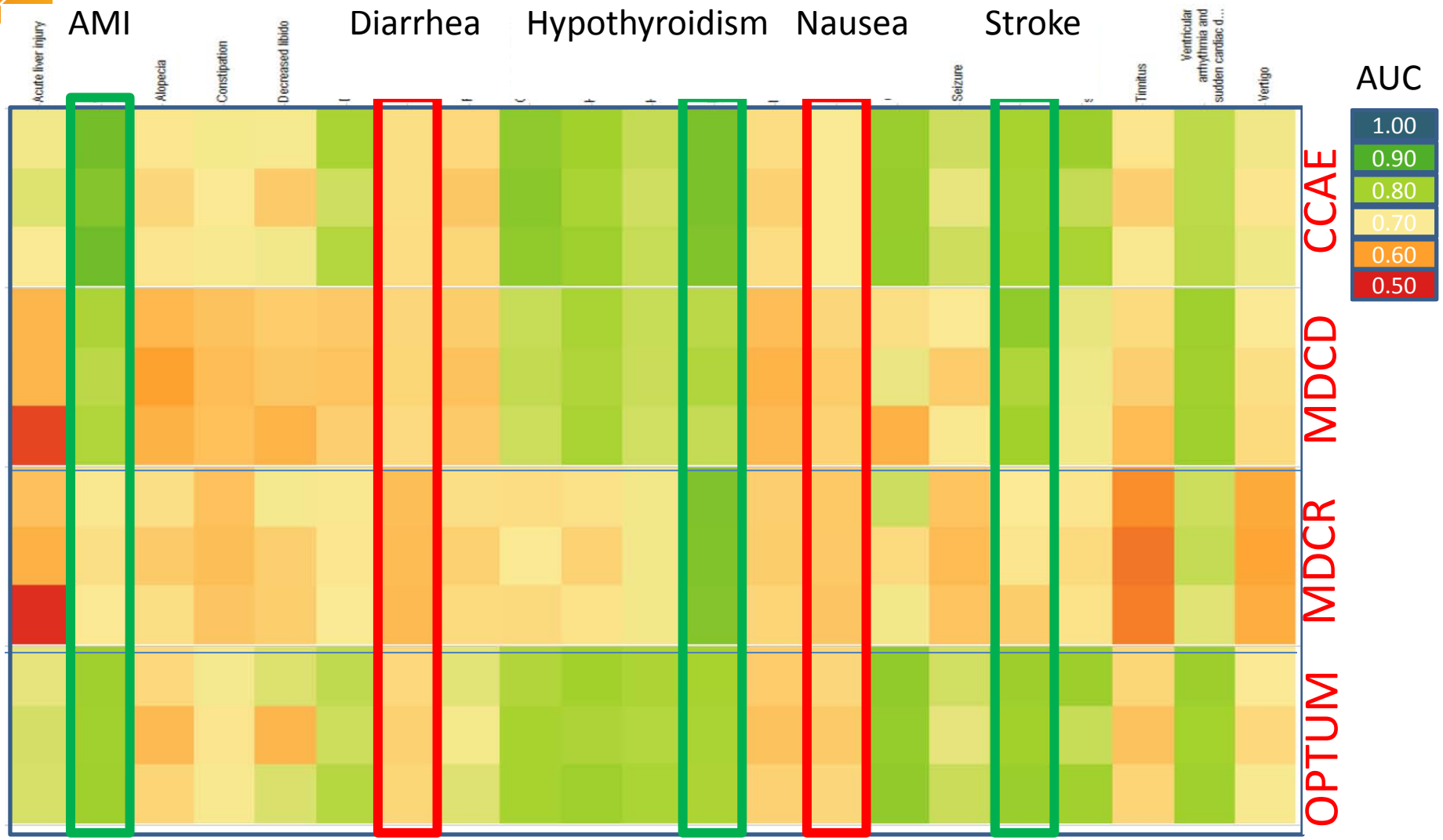
OHDSI's recommended best practices for population-level effect estimation

- Write and share protocol
- Open source study code
- Use validated software
- Replicate across databases

- Produce standard diagnostics
- Include negative controls
- Create positive controls
- Calibrate confidence interval and p-value

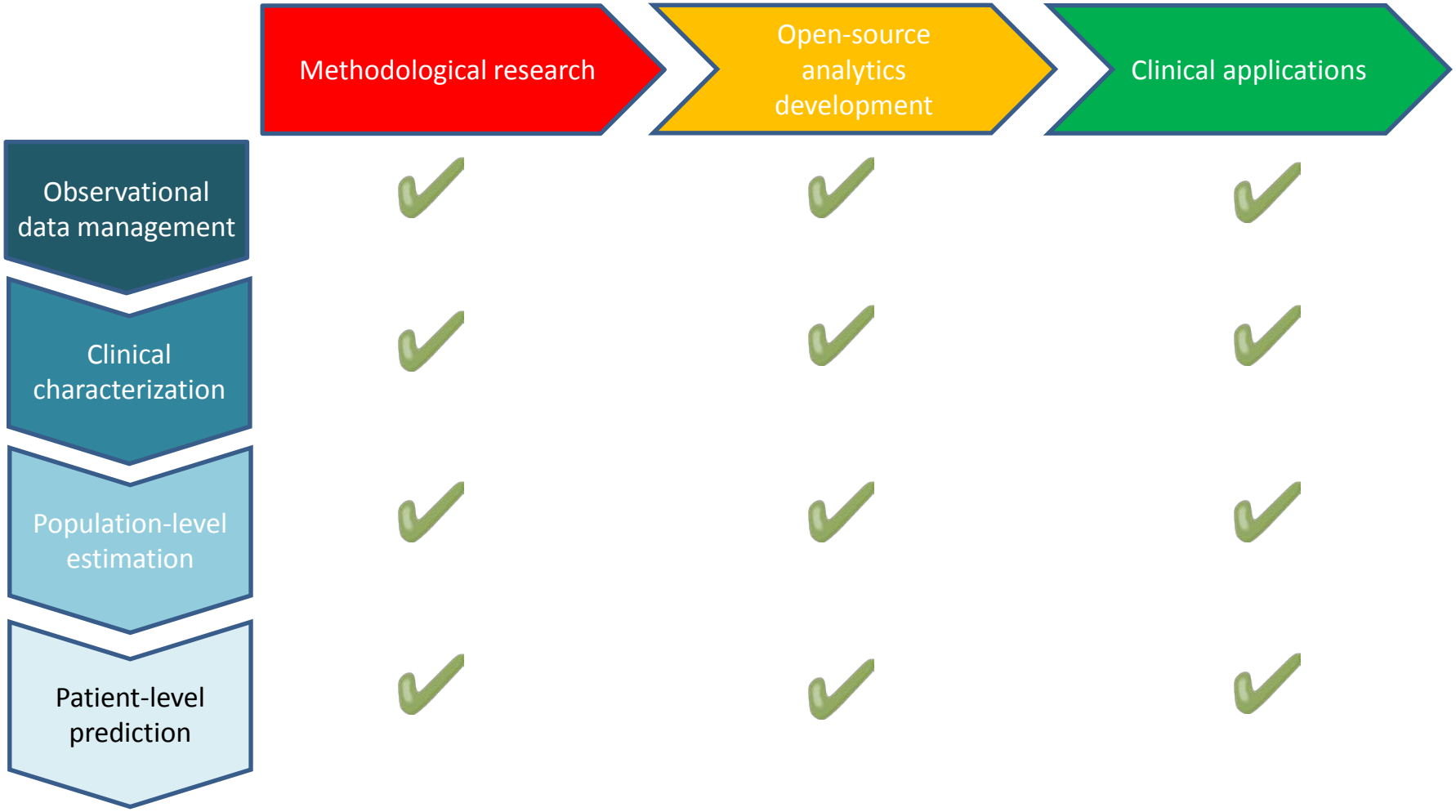
- Don't provide only the effect estimate
- Also share protocol, study code, diagnostics and evaluation
- Produce evidence at scale

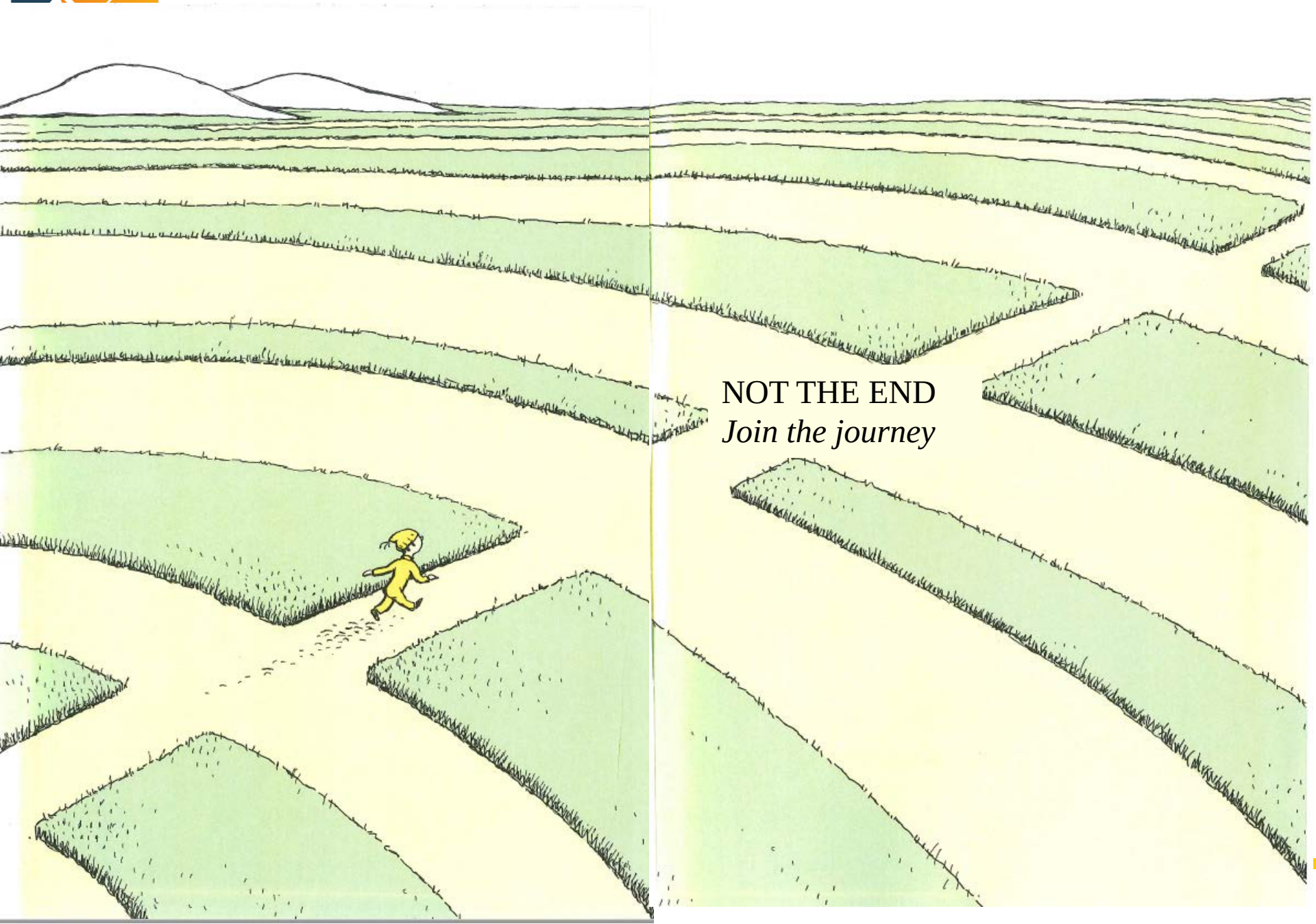
Populations can be used to accurately predict outcomes for individuals



Rijnbeek, OHDSI Symposium, 2016

OHDSI's areas of focus:
A look back on progress from 2016…




NOT THE END
Join the journey

At the start of 2016…

Indication
Synagis® (palivizumab) is a prescription medication that is used to help prevent a serious lung disease caused by respiratory syncytial virus (RSV) in children at high risk for severe lung disease from RSV.

Select Safety Information
Common side effects of Synagis include fever and rash. Other possible side effects include skin reactions around the area where the shot was given (like redness, swelling, warmth, or discomfort).

Please see complete Important Safety Information on pages 22-24 and accompanying full Prescribing Information, including Patient Information.

Doctor X: “This paper says there’s side effects, but I’ve never seen them happen”

SYNAGIS- palivizumab injection, solution
MedImmune, LLC

HIGHLIGHTS OF PRESCRIBING INFORMATION
These highlights do not include all the information needed to use SYNAGIS safely and effectively. See full prescribing information for SYNAGIS.
SYNAGIS® (palivizumab) injection, for intramuscular use
Initial U.S. Approval: 1998

INDICATIONS AND USAGE

Synagis is a respiratory syncytial virus (RSV) F protein inhibitor monoclonal antibody indicated for the prevention of serious lower respiratory tract disease caused by RSV in children at high risk of RSV disease.

- Safety and efficacy were established in children with bronchopulmonary dysplasia (BPD), infants with a history of premature birth (less than or equal to 35 weeks gestational age), and children with hemodynamically significant congenital heart disease (CHD).
- The safety and efficacy of Synagis have not been established for treatment of RSV disease. (1)

DOSAGE AND ADMINISTRATION

15 mg per kg of body weight, administered intramuscularly prior to commencement of the RSV season and remaining doses administered monthly throughout the RSV season. (2.1)
Children undergoing cardio-pulmonary bypass should receive an additional dose of Synagis as soon as possible after the cardio-pulmonary bypass procedure (even if sooner than a month from the previous dose). Thereafter, doses should be administered monthly as scheduled. (2.1, 12.3)

DOSAGE FORMS AND STRENGTHS

Single-dose liquid solution vials: 50 mg per 0.5 mL and 100 mg per 1 mL. (3)

CONTRAINDICATIONS

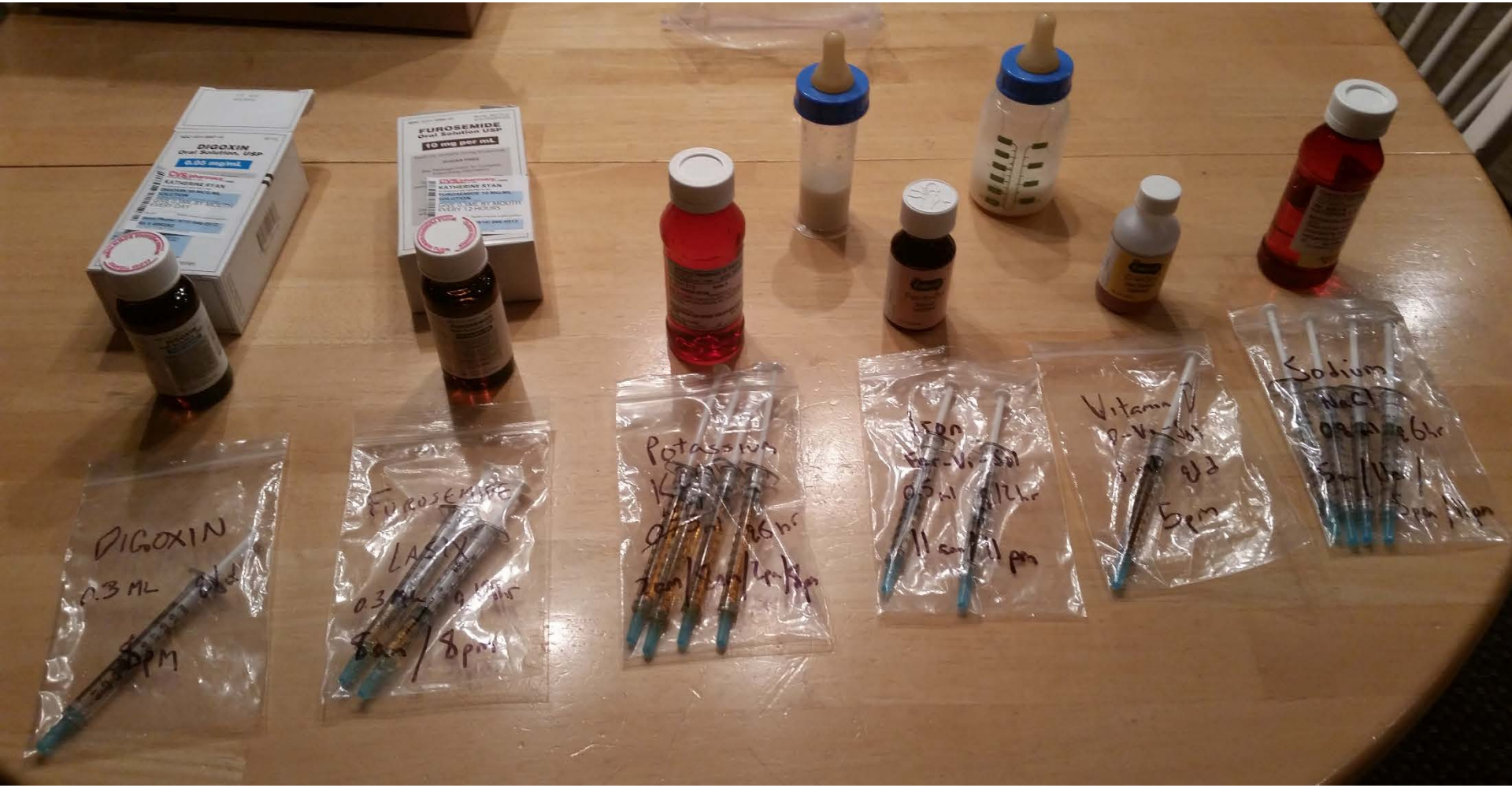
Previous significant hypersensitivity reaction to Synagis. (4)

WARNINGS AND PRECAUTIONS

- Anaphylaxis and anaphylactic shock (including fatal cases), and other severe acute hypersensitivity reactions have been reported. Permanently discontinue Synagis and administer appropriate medications if such reactions occur. (5.1)
- As with any intramuscular injection, Synagis should be given with caution to children with thrombocytopenia or any coagulation disorder. (5.2)
- Palivizumab may interfere with immunological-based RSV diagnostic tests such as some antigen detection-based assays. (5.3, 12.4)

ADVERSE REACTIONS

Adverse reactions occurring greater than or equal to 10% and at least 1% more frequently than placebo are fever and rash. (6.1)


DIGOXIN
Oral Solution, USP
0.05 mg/mL
FUROSEMIDE
Oral Solution USP
10 mg per mL
DIGOXIN
0.3 ML
8PM
Potassium
Sodium

In the middle of 2016…

Scientific Studies: Last Week Tonight with John Oliver (HBO)

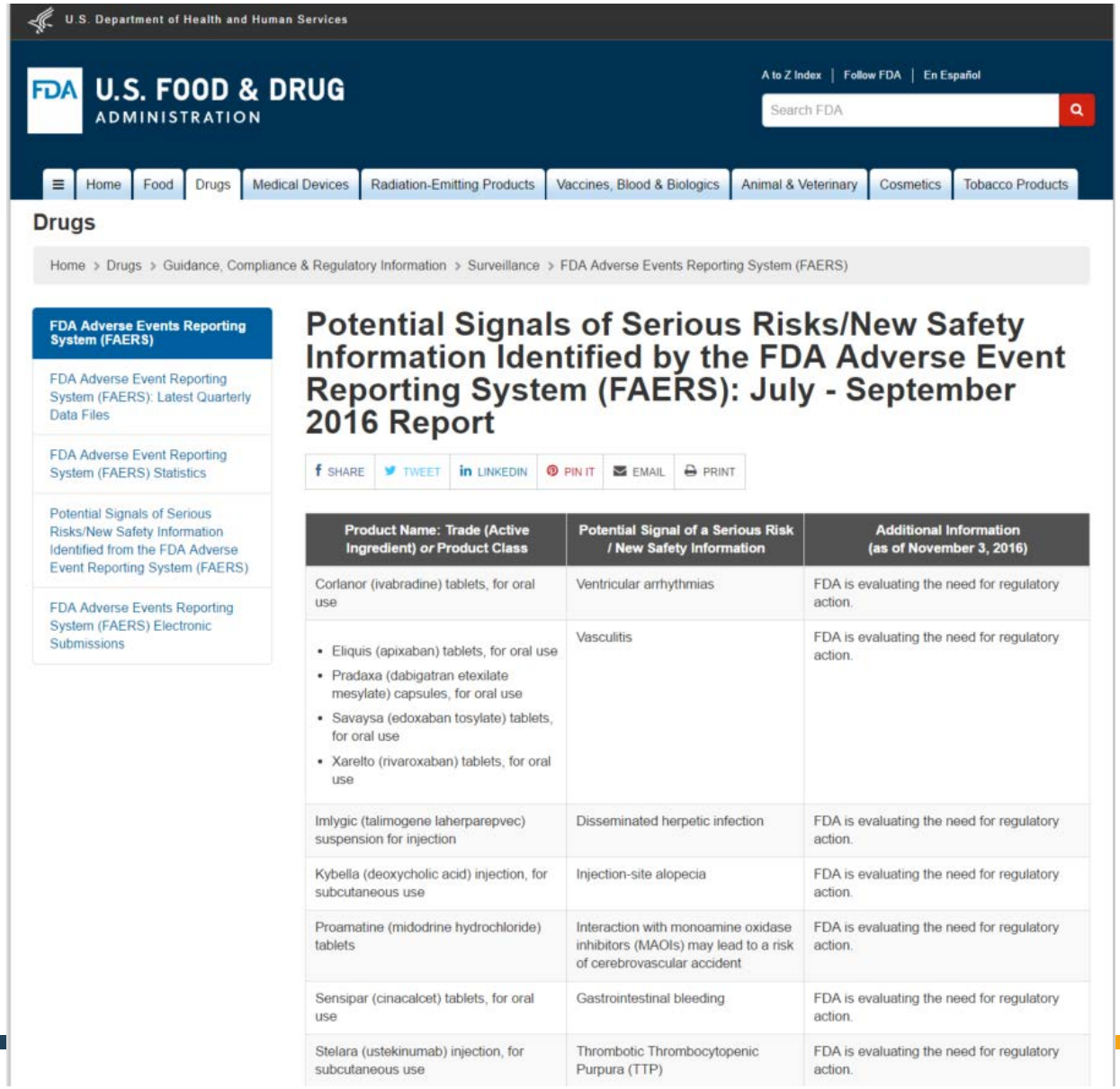
LastWeekTonight

Subscribe 4,427,629

8,638,416 views

Add to Share More 97,246 2,876

At the end of 2016…



U.S. Department of Health and Human Services

FDA U.S. FOOD & DRUG ADMINISTRATION

A to Z Index | Follow FDA | En Español

Search FDA

Home | Food | Drugs | Medical Devices | Radiation-Emitting Products | Vaccines, Blood & Biologics | Animal & Veterinary | Cosmetics | Tobacco Products

Drugs

Home > Drugs > Guidance, Compliance & Regulatory Information > Surveillance > FDA Adverse Events Reporting System (FAERS)

FDA Adverse Events Reporting System (FAERS)

- FDA Adverse Event Reporting System (FAERS): Latest Quarterly Data Files
- FDA Adverse Event Reporting System (FAERS) Statistics
- Potential Signals of Serious Risks/New Safety Information Identified from the FDA Adverse Event Reporting System (FAERS)
- FDA Adverse Events Reporting System (FAERS) Electronic Submissions

Potential Signals of Serious Risks/New Safety Information Identified by the FDA Adverse Event Reporting System (FAERS): July - September 2016 Report

SHARE | TWEET | LINKEDIN | PIN IT | EMAIL | PRINT

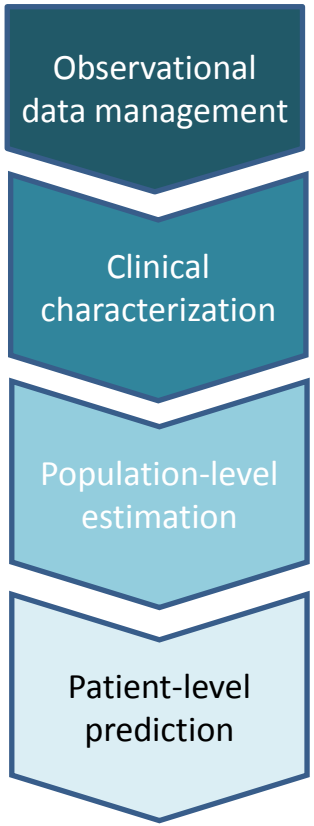
Product Name: Trade (Active Ingredient) *or* Product Class	Potential Signal of a Serious Risk / New Safety Information	Additional Information (as of November 3, 2016)
Corlanor (ivabradine) tablets, for oral use	Ventricular arrhythmias	FDA is evaluating the need for regulatory action.
• Eliquis (apixaban) tablets, for oral use • Pradaxa (dabigatran etexilate mesylate) capsules, for oral use • Savaysa (edoxaban tosylate) tablets, for oral use • Xarelto (rivaroxaban) tablets, for oral use	Vasculitis	FDA is evaluating the need for regulatory action.
Imlygic (talimogene laherparepvec) suspension for injection	Disseminated herpetic infection	FDA is evaluating the need for regulatory action.
Kybella (deoxycholic acid) injection, for subcutaneous use	Injection-site alopecia	FDA is evaluating the need for regulatory action.
Proamatine (midodrine hydrochloride) tablets	Interaction with monoamine oxidase inhibitors (MAOIs) may lead to a risk of cerebrovascular accident	FDA is evaluating the need for regulatory action.
Sensipar (cinacalcet) tablets, for oral use	Gastrointestinal bleeding	FDA is evaluating the need for regulatory action.
Stelara (ustekinumab) injection, for subcutaneous use	Thrombotic Thrombocytopenic Purpura (TTP)	FDA is evaluating the need for regulatory action.

Why our journey isn't finished yet

- No patient should have to wonder "what's the chance that this event might happen to me?"
- All patients deserve to have reliable evidence about the safety and comparative effectiveness of alternative treatments when making medical decisions
- The potential for a 'learning healthcare system' can only be realized if we agree to learn together and collaborate to build a system

Building the LHC of observational research?

OHDSI's areas of focus: A look forward at 2017…



- **Generate and disseminate more clinical evidence**
- Maintain and evolve open community data and vocabulary standards
- Develop and improve tools to enable large-scale analysis
- Establish and promote community best practices
- Strengthen and expand collaborations across OHDSI research network
- Advance scholarship in observational data science through publication, presentations, and education
- **Generate and disseminate more clinical evidence**

Join the journey

- Discussion / questions / comments

ryan@ohdsi.org