Welcome to OHDSI 2019: This is our community

George Hripcsak, MD, MS, Chair of the Department of Biomedical Informatics at Columbia University Medical Center

Harlan M. Krumholz, MD, SM, Harold H. Hines, Jr. Professor of Medicine at Yale University School of Medicine; Director, Yale New Haven Hospital Center for Outcomes Research and Evaluation (CORE)

@hmkyale

2019 Theme: Continuous evaluation

How do we know we are making progress on our journey?

Thank you to our sponsors!

janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson

BAYER

CASE WESTERN RESERVE UNIVERSITY EST. 1826 SCHOOL OF MEDICINE

CLEVELAND INSTITUTE FOR COMPUTATIONAL BIOLOGY

IMS Health & Quintiles are now IQVIA™

And contributions from viewers like you!

Google Cloud

ODYSSEUS DATA SERVICES INC

We thank the FDA for their generous support of the 2019 OHDSI symposium through the FDA SCIENTIFIC CONFERENCE GRANT PROGRAM (R13)

OHDSI is

an open science community

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 community

We're all in this journey together…

256 collaborators in 27 different countries over six continents

OHDSI's community engagement

- Active community online discussion: forums.ohdsi.org
  - **>2,770** distinct users have made **>18,700** posts on **>3,250** topics
  - Implementers, Developers, Researchers, CDM Builders, Vocabulary users, OHDSI in Korea, OHDSI in China, OHDSI in Europe
- Weekly community web conferences for all collaborators to share their research ideas and progress
- >25 workgroups for solving shared problems of interest
  - ex: Common Data Model, Population-level Estimation, Patient-level Prediction, Phenotype, NLP, GIS, Oncology, Women of OHDSI
- Quarterly tutorials in OHDSI tools and best practices, taught by OHDSI collaborators for OHDSI collaborators
- OHDSI Symposiums held annually in North America, Europe and Asia to provide the community face-to-face opportunities to showcase research collaborations
- Follow us on Twitter @OHDSI and LinkedIn

OHDSI Home | Forums | Wiki | Github

OHDSI

all categories | all tags | Categories | Latest | Top | + New Topic

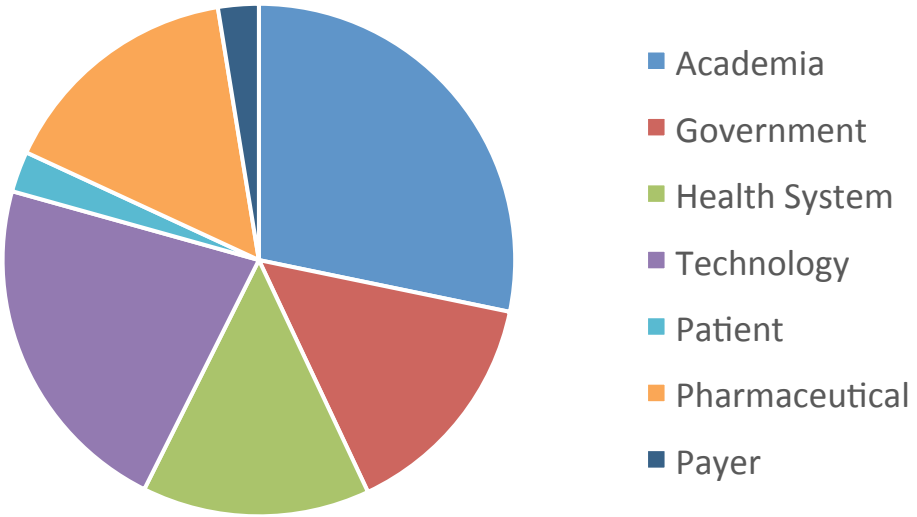
Year SEP 13, 2018 - SEP 13, 2019

Topic	Category	Users	Replies	Views	Activity
EHR data to OMOP CDM Work Group themis, workinggroup	Implementers	M S W	91	2.0k	4h
New WG - Book of OHDSI 32	General	D A	82	1.6k	
New Comprehensive Hierarchy for Providers, Visits (and Place of Service, Specialty, Care Site) 52	Vocabulary Users	R	50	1.2k	
각 cdm의 차이점에 대한 질문입니다	OHDSI in Korea		7	1.1k	
Atlas Setup Failing atlas	Implementers	S A	39	919	
Loading Profiles and Cohort Generation in Atlas	Developers	L P	27	840	
How to determine whether a drug is brand or generics?	General	R F	9	826	Jan 28
Some errors using Atlas after installation	Developers	A P S	27	813	Apr 22
ACEI ARB and Lung Ca 12	Researchers	A H	20	735	Nov '18
What is a phenotype in the context of observational research?	Researchers		42	720	May 24
Mapping OMOP CDM to FHIR cdm	General	N	2	710	Dec '18

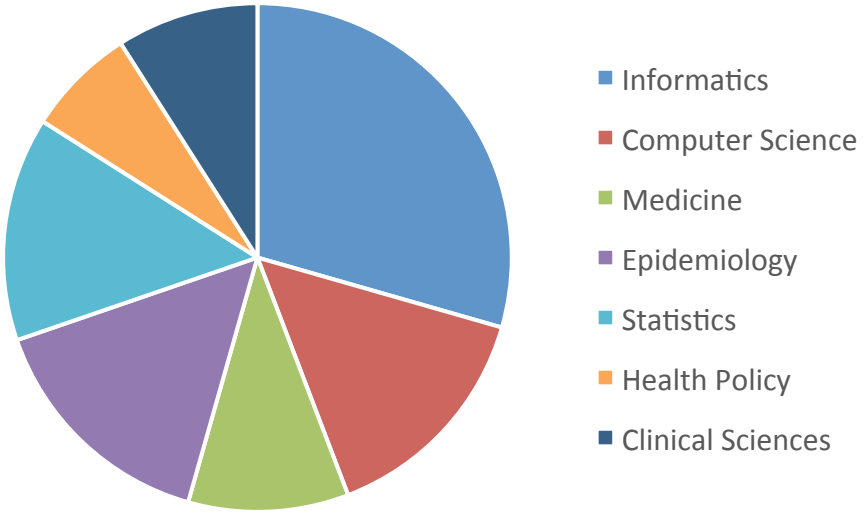
In the last year, we've seen tremendous activity and interest across a wide range of topics in multiple categories (Implementers, Vocabulary Users, Developers, Researchers)

Diversity of the OHDSI community represented today at the OHDSI Symposium

Stakeholder group



Disciplinary perspective



Relationship with OHDSI community	Persons
I am new to OHDSI and curious to learn more	240
I actively participate in OHDSI meetings and work groups	177
I use OHDSI tools and methods to support my research	176
I have an OMOP CDM instance	125
I am in the process of converting my data into the OMOP CDM	95
I actively participate in discussions on the OHDSI forum	74
I am participating in an OHDSI network research study	55
I contribute code to the OHDSI GitHub	48

OHDSI is

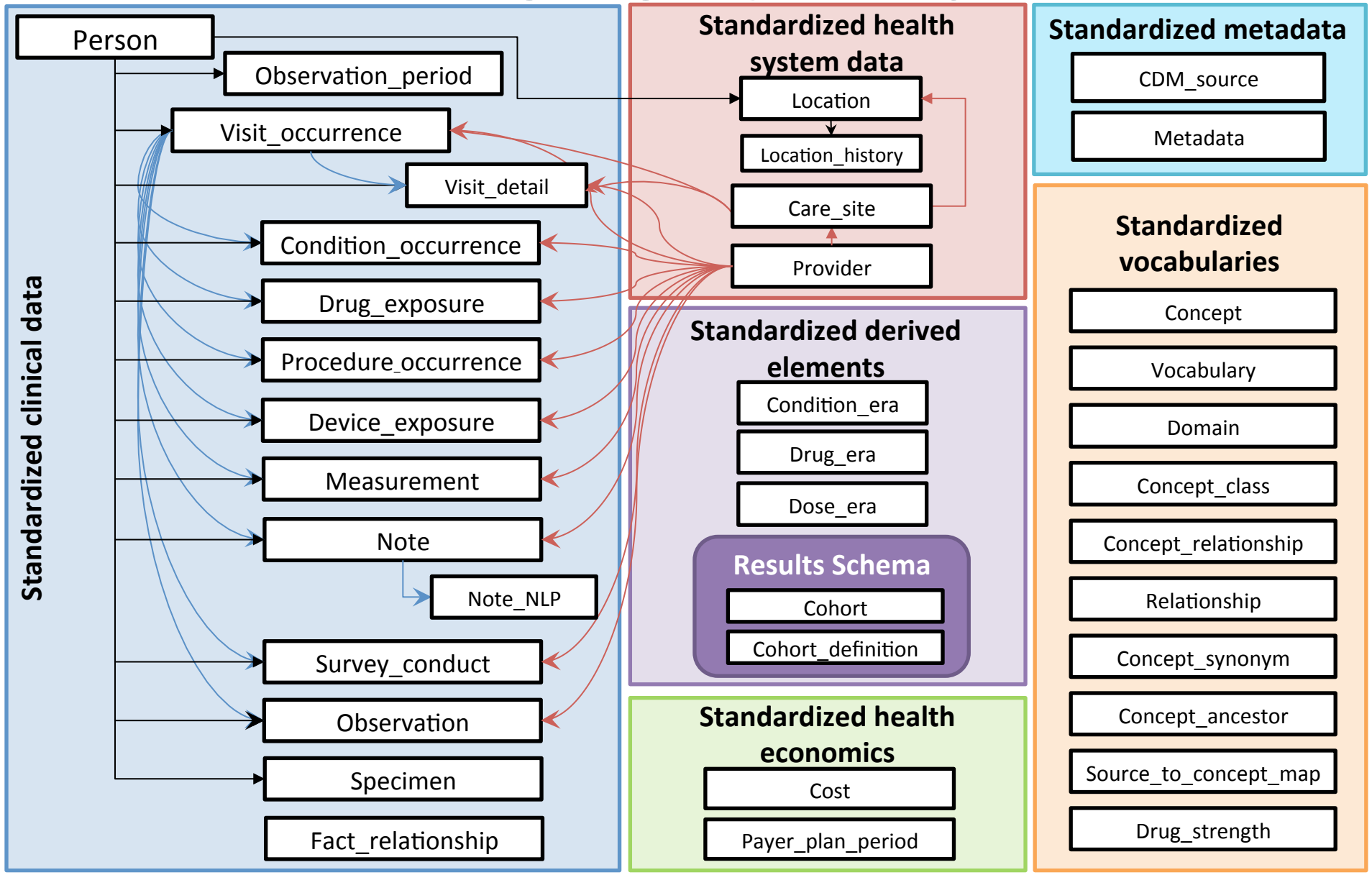
an international data network

Data across the OHDSI community

- 152 entries on 2019 OHDSI data network inventory
- 133 different databases with patient-level data from various perspectives:
  - Electronic health records, administrative claims, hospital systems, clinical registries, health surveys, biobanks
- Data in 18 different countries, with >369 million patient records from outside US

All using one open community data standard:
OMOP Common Data Model

Open community data standard: OMOP CDM v6



OHDSI's standardized vocabularies

- >130 Vocabularies across 40 domains
  - MU3 standards: SNOMED, RxNorm, LOINC
  - Disparate sources: ICD9CM, ICD10(CM), Read, NDC, Gemscript, CPT4, HCPCS…
- >7.4 million concepts
  - >3.0 million standard concepts
  - >3.8 million source codes
  - >511,000 classification concepts
- >45 million concept relationships
- >74 million ancestral relationships

OHDSI is

advancing science

What is OHDSI's strategy to deliver reliable 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 evidence generation**
  - 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

Highlights of progress from the community: Data standards

- Increased adoption of OMOP CDM
- Evaluation of vocabulary
- Expanded vocabulary
- Community collaboration around conventions (THEMIS)
- Added rigor around data quality (see Clair and Andrew)

Journal of the American Medical Informatics Association, 25(12), 2018, 1618–1625
doi: 10.1093/jamia/ocy124
Advance Access Publication Date: 3 November 2018
Research and Applications

Research and Applications

Effect of vocabulary mapping for conditions on phenotype cohorts

George Hripcsak,[1,2,3] Matthew E Levine,[1,2] Ning Shang,[1,2] and Patrick B Ryan[1,2,4]

[1]Department of Biomedical Informatics, Columbia University, New York, New York, USA, [2]Observational Health Data Sciences and Informatics (OHDSI), New York, New York, USA, [3]Medical Informatics Services, NewYork-Presbyterian Hospital, New York, New York, USA, and [4]Epidemiology Analytics, Janssen Research and Development, Titusville, New Jersey, USA

Corresponding Author: George Hripcsak, MD, MS, Department of Biomedical Informatics, Columbia University Irving Medical Center, 622 W 168th St, PH20, New York, NY 10032, USA (hripcsak@columbia.edu)

Received 27 April 2018; Revised 13 August 2018; Editorial Decision 22 August 2018; Accepted 3 September 2018

ABSTRACT

Objective: To study the effect on patient cohorts of mapping condition (diagnosis) codes from source billing vocabularies to a clinical vocabulary.

Materials and Methods: Nine International Classification of Diseases, Ninth Revision, Clinical Modification (ICD9-CM) concept sets were extracted from eMERGE network phenotypes, translated to Systematized Nomenclature of Medicine - Clinical Terms concept sets, and applied to patient data that were mapped from source ICD9-CM and ICD10-CM codes to Systematized Nomenclature of Medicine - Clinical Terms codes using Observational Health Data Sciences and Informatics (OHDSI) Observational Medical Outcomes Partnership (OMOP) vocabulary mappings. The original ICD9-CM concept set and a concept set extended to ICD10-CM were used to create patient cohorts that served as gold standards.

Results: Four phenotype concept sets were able to be translated to Systematized Nomenclature of Medicine - Clinical Terms without ambiguities and were able to perform perfectly with respect to the gold standards. The other 5 lost performance when 2 or more ICD9-CM or ICD10-CM codes mapped to the same Systematized Nomenclature of Medicine - Clinical Terms code. The patient cohorts had a total error (false positive and false negative) of up to 0.15% compared to querying ICD9-CM source data and up to 0.26% compared to querying ICD9-CM and ICD10-CM data. Knowledge engineering was required to produce that performance; simple automated methods to generate concept sets had errors up to 10% (one outlier at 250%).

Discussion: The translation of data from source vocabularies to Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) resulted in very small error rates that were an order of magnitude smaller than other error sources.

Conclusion: It appears possible to map diagnoses from disparate vocabularies to a single clinical vocabulary and carry out research using a single set of definitions, thus improving efficiency and transportability of research.

Journal of Biomedical Informatics 96 (2019) 103239

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

HemOnc: A new standard vocabulary for chemotherapy regimen representation in the OMOP common data model

Jeremy L. Warner[a,b,*], Dmitry Dymshyts[c], Christian G. Reich[d], Michael J. Gurley[e], Harry Hochheiser[f], Zachary H. Moldwin[g], Rimma Belenkaya[h], Andrew E. Williams[i], Peter C. Yang[b,j]

[a] Vanderbilt University Medical Center, Nashville, TN, United States
[b] HemOnc.org, LLC, Lexington, MA, United States
[c] Odysseus Data Services, Inc., Cambridge, MA, United States
[d] IQVIA, Cambridge, MA, United States
[e] Northwestern University, Chicago, IL, United States
[f] University of Pittsburgh, Pittsburgh, PA, United States
[g] University of Illinois at Chicago College of Pharmacy, Chicago, IL, United States
[h] Memorial Sloan Kettering Cancer Center, New York, NY, United States
[i] Tufts University, Medford, MA, United States
[j] Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States

Highlights of progress from the community: Methods research

- Phenotype definition
- Phenotype evaluation
- Study design evaluation

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Facilitating phenotype transfer using a common data model

George Hripcsak[a,b,*], Ning Shang[a], Peggy L. Peissig[c], Luke V. Rasmussen[d], Cong Liu[a], Barbara Benoit[e], Robert J. Carroll[f], David S. Carrell[g], Joshua C. Denny[f,h], Ozan Dikilitas[i], Vivian S. Gainer[e], Kayla Marie Howell[j], Jeffrey G. Klann[e], Iftikhar J. Kullo[i], Todd Lingren[k], Frank D. Mentch[l], Shawn N. Murphy[e], Karthik Natarajan[a,b], Jennifer A. Pacheco[d], Wei-Qi Wei[f], Ken Wiley[m], Chunhua Weng[a]

[a] Department of Biomedical Informatics, Columbia University, New York, NY, United States
[b] Medical Informatics Services, NewYork-Presbyterian Hospital, New York, NY, United States
[c] Center for Precision Medicine Research, Marshfield Clinic Research Institute, Marshfield, WI, United States
[d] Northwestern University Feinberg School of Medicine, Chicago, IL, United States
[e] Research Information Science and Computing, Partners Healthcare, Boston, MA, United States
[f] Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States
[g] Kaiser Permanente Washington Health Research Institute, Seattle, WA, United States
[h] Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, United States
[i] Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN, United States
[j] Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN, United States
[k] Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States
[l] Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, United States
[m] National Human Genome Research Institute, NIH, Bethesda, MD, United States

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

PheValuator: Development and evaluation of a phenotype algorithm evaluator

Joel N. Swerdel[a,b,*], George Hripcsak[b,c], Patrick B. Ryan[a,b,c]

[a] Janssen Research & Development, 920 Route 202, Raritan, NJ 08869, USA
[b] OHDSI Collaborators, Observational Health Data Sciences and Informatics (OHDSI), 622 West 168th Street, PH-20, New York, NY 10032, USA
[c] Columbia University, 622 West 168th Street, PH20, New York, NY 10032, USA

ARTICLE INFO

Keywords:
Phenotype algorithms
Validation
Diagnostic predictive modeling

ABSTRACT

Background: The primary approach for defining disease in observational healthcare databases is to construct phenotype algorithms (PAs), rule-based heuristics predicated on the presence, absence, and temporal logic of clinical observations. However, a complete evaluation of PAs, i.e., determining sensitivity, specificity, and positive predictive value (PPV), is rarely performed. In this study, we propose a tool (PheValuator) to efficiently estimate a complete PA evaluation.

Methods: We used 4 administrative claims datasets: OptumInsight's de-identified Clinformatics™ Datamart (Eden Prairie,MN); IBM MarketScan Multi-State Medicaid); IBM MarketScan Medicare Supplemental Beneficiaries; and IBM MarketScan Commercial Claims and Encounters from 2000 to 2017. Using PheValuator involves (1) creating a diagnostic predictive model for the phenotype, (2) applying the model to a large set of randomly selected subjects, and (3) comparing each subject's predicted probability for the phenotype to inclusion/exclusion in PAs. We used the predictions as a 'probabilistic gold standard' measure to classify positive/negative cases. We examined 4 phenotypes: myocardial infarction, cerebral infarction, chronic kidney disease, and atrial fibrillation. We examined several PAs for each phenotype including 1-time (1X) occurrence of the diagnosis code in the subject's record and 1-time occurrence of the diagnosis in an inpatient setting with the diagnosis code as the primary reason for admission (1X-IP-1stPos).

Results: Across phenotypes, the 1X PA showed the highest sensitivity/lowest PPV among all PAs. 1X-IP-1stPos yielded the highest PPV/lowest sensitivity. Specificity was very high across algorithms. We found similar results between algorithms across datasets.

Conclusion: PheValuator appears to show promise as a tool to estimate PA performance characteristics.

Received: 2 August 2018 | Revised: 29 April 2019 | Accepted: 5 May 2019

DOI: 10.1002/sim.8215

RESEARCH ARTICLE

WILEY Statistics in Medicine

A plea to stop using the case-control design in retrospective database studies

Martijn J. Schuemie[1,2,3] | Patrick B. Ryan[1,2,4] | Kenneth K.C. Man[5,6,7,8] | Ian C.K. Wong[5,6] | Marc A. Suchard[1,3,9,10] | George Hripcsak[1,4,11]

[1]Observational Health Data Sciences and Informatics, New York, New York

[2]Epidemiology Analytics, Janssen Research and Development, Titusville, New Jersey

[3]Department of Biostatistics, University of California, Los Angeles, California

[4]Department of Biomedical Informatics, Columbia University Medical Center, New York, New York

[5]Centre for Safe Medication Practice and Research, Department of Pharmacology and Pharmacy, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong

[6]Research Department of Practice and Policy, UCL School of Pharmacy, London, UK

[7]Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands

[8]Department of Social Work and Social Administration, Faculty of Social Science, The University of Hong Kong, Pokfulam, Hong Kong

[9]Department of Biomathematics, University of California, Los Angeles, California

The case-control design is widely used in retrospective database studies, often leading to spectacular findings. However, results of these studies often cannot be replicated, and the advantage of this design over others is questionable. To demonstrate the shortcomings of applications of this design, we replicate two published case-control studies. The first investigates isotretinoin and ulcerative colitis using a simple case-control design. The second focuses on dipeptidyl peptidase-4 inhibitors and acute pancreatitis, using a nested case-control design. We include large sets of negative control exposures (where the true odds ratio is believed to be 1) in both studies. Both replication studies produce effect size estimates consistent with the original studies, but also generate estimates for the negative control exposures showing substantial residual bias. In contrast, applying a self-controlled design to answer the same questions using the same data reveals far less bias. Although the case-control design in general is not at fault, its application in retrospective database studies, where all exposure and covariate data for the entire cohort are available, is unnecessary, as other alternatives such as cohort and self-controlled designs are available. Moreover, by focusing on cases and controls it opens the door to inappropriate comparisons between exposure groups, leading to confounding for which the design has few options to adjust for. We argue that this design should no longer be used in these types of data. At the very least, negative control exposures should be used to prove that the concerns raised here do not apply.

Highlights of progress from the community: Open source development

- ATLAS 2.7.3 released
- Criteria2Query published
- Community contributions for multiple OMOP CDM utilities

atlas.ohdsi.org/#/home

ATLAS

- Home
- Data Sources
- Search
- Concept Sets
- Cohort Definitions
- Characterizations
- Cohort Pathways
- Incidence Rates
- Profiles
- Estimation
- Prediction
- Jobs
- Configuration
- Feedback

Apache 2.0
open source software

provided by
OHDSI
join the journey

ryan@ohdsi.org

Home

Welcome to ATLAS.

ATLAS is an open source application developed as a part of OHDSI intended to provide a unified interface to patient level data and analytics.

Documentation

The ATLAS user guide can be found here.

Getting Started

Define a New Cohort — Begin performing research by defining the group of people you intend to study

Search the Vocabulary — Search the different ontologies used to describe patient level data around the world

Release Notes

ATLAS Version 2.7.3 Release Notes
WebAPI Version 2.7.3 Release Notes

This latest release contains **7** feature enhancements and issue resolutions:

- Cohort definitions creation date is 4 hours greater than actual while being on EST timezone
- Do not call user/refresh endpoint case of IAP authentication
- Characterization pop-up shows wrong percentage
- Role import / export works incorrectly
- Title Consistency
- Active Directory groups mapping issue
- Cannot save concept set modification in cohort definition

Journal of the American Medical Informatics Association, 26(4), 2019, 294–305
doi: 10.1093/jamia/ocy178
Advance Access Publication Date: 7 February 2019
Research and Applications

Research and Applications

Criteria2Query: a natural language interface to clinical databases for cohort definition

Chi Yuan,[1,2] Patrick B. Ryan,[1,3] Casey Ta,[1] Yixuan Guo,[1] Ziran Li,[1] Jill Hardin,[3] Rupa Makadia,[3] Peng Jin,[1] Ning Shang,[1] Tian Kang,[1] and Chunhua Weng[1]

[1]Department of Biomedical Informatics, Columbia University, New York, New York, USA, [2]Department of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing, Jiangsu Province, P.R. China, and [3]Epidemiology Analytics, Janssen Research and Development, Titusville, New Jersey, USA

Corresponding Author: Chunhua Weng, PhD, Department of Biomedical Informatics, Columbia University, 622 West 168th Street, PH-20, Room 407, New York, NY 10032, USA (chunhua@columbia.edu)

Received 7 September 2018; Revised 16 November 2018; Editorial Decision 27 November 2018; Accepted 29 November 2018

ABSTRACT

Objective: Cohort definition is a bottleneck for conducting clinical research and depends on subjective decisions by domain experts. Data-driven cohort definition is appealing but requires substantial knowledge of terminologies and clinical data models. Criteria2Query is a natural language interface that facilitates human-computer collaboration for cohort definition and execution using clinical databases.

Materials and Methods: Criteria2Query uses a hybrid information extraction pipeline combining machine learning and rule-based methods to systematically parse eligibility criteria text, transforms it first into a structured criteria representation and next into sharable and executable clinical data queries represented as SQL queries conforming to the OMOP Common Data Model. Users can interactively review, refine, and execute queries in the ATLAS web application. To test effectiveness, we evaluated 125 criteria across different disease domains from ClinicalTrials.gov and 52 user-entered criteria. We evaluated F1 score and accuracy against 2 domain experts and calculated the average computation time for fully automated query formulation. We conducted an anonymous survey evaluating usability.

Results: Criteria2Query achieved 0.795 and 0.805 F1 score for entity recognition and relation extraction, respectively. Accuracies for negation detection, logic detection, entity normalization, and attribute normalization were 0.984, 0.864, 0.514 and 0.793, respectively. Fully automatic query formulation took 1.22 seconds/criterion. More than 80% (11+ of 13) of users would use Criteria2Query in their future cohort definition tasks.

Conclusions: We contribute a novel natural language interface to clinical databases. It is open source and supports fully automated and interactive modes for autonomous data-driven cohort definition by researchers with minimal human effort. We demonstrate its promising user friendliness and usability.

Bioinformatics, 2019, 1–4
doi: 10.1093/bioinformatics/btz409
Advance Access Publication Date: 19 June 2019
Application Note

Data and text mining

PatientExploreR: an extensible application for dynamic visualization of patient clinical history from electronic health records in the OMOP common data model

Benjamin S. Glicksberg [1], Boris Oskotsky[1], Phyllis M. Thangaraj[2,3,4,†], Nicholas Giangreco [2,3,4,†], Marcus A. Badgeley [5,†], Kipp W. Johnson[5,†], Debajyoti Datta[1], Vivek A. Rudrapatna[1,6], Nadav Rappoport[1], Mark M. Shervey[5], Riccardo Miotto[5], Theodore C. Goldstein[1], Eugenia Rutenberg[1], Remi Frazier[7], Nelson Lee[7], Sharat Israni[1], Rick Larsen[7], Bethany Percha[5], Li Li[5], Joel T. Dudley[5], Nicholas P. Tatonetti[2,3,4] and Atul J. Butte[1,8,*]

[1]Bakar Computational Health Sciences Institute, University of California, San Francisco, San Francisco, CA 94158, USA, [2]Department of Biomedical Informatics, [3]Department of Systems Biology, [4]Department of Medicine, Columbia University, New York, NY 10032, USA, [5]Departments of Genomics and Data Science, Icahn Institute for Genomic Sciences and Multiscale Biology, Icahn School of Medicine at Mount Sinai, Institute of Next Generation Healthcare, New York, NY 10029, USA, [6]Division of Gastroenterology, Department of Medicine, University of California, San Francisco, CA 94158, USA, [7]Enterprise Information and Analytics, University of California, San Francisco, San Francisco, CA 94158, USA and [8]Center for Data-Driven Insights and Innovation, University of California Health, Oakland, CA 94607, USA

JAMIA Open, 2(1), 2019, 10–14
doi: 10.1093/jamiaopen/ooy059
Advance Access Publication Date: 4 January 2019
Application Notes

Application Notes

ROMOP: a light-weight R package for interfacing with OMOP-formatted electronic health record data

Benjamin S. Glicksberg,[1] Boris Oskotsky,[1] Nicholas Giangreco,[2,†] Phyllis M. Thangaraj,[2,†] Vivek Rudrapatna,[1] Debajyoti Datta,[1] Remi Frazier,[3] Nelson Lee,[3] Rick Larsen,[3] Nicholas P. Tatonetti[2] and Atul J. Butte[1]

[1]Department of Pediatrics Bakar Computational Health Sciences Institute, University of California San Francisco, San Francisco, California, USA, [2]Departments of Biomedical Informatics, Systems Biology, and Medicine, Columbia University, New York, New York, USA and [3]Academic Research Systems, Department of Enterprise Data Warehouse University of California San Francisco, San Francisco, California, USA

[†]These two authors contributed equally to the study.

Corresponding Author: Atul J. Butte, MD, PhD, Bakar Computational Health Sciences Institute, University of California San Francisco, San Francisco, CA 94158, USA (Atul.Butte@ucsf.edu)

Received 3 July 2018; Revised 26 October 2018; Editorial Decision 29 November 2018; Accepted 2 December 2018

Highlights of progress from the community: Clinical applications

LEGEND basic viewer

About | Specific research questions

Indication
Hypertension

Exposure group
Drug class

☐ Include combination exposures

Target
ACE inhibitors

Comparator
Thiazide or thiazide-like diuretics

Outcome
Abdominal pain

Data source
- ☑ CCAE
- ☑ CUMC
- ☑ IMSG
- ☑ JMDC
- ☑ MDCD
- ☑ MDCR
- ☑ NHIS_NSC
- ☑ Optum
- ☑ Panther
- ☑ Meta-analysis

Analysis
- ☐ PS stratification, on-treatment
- ☐ PS stratification, intent-to-treat

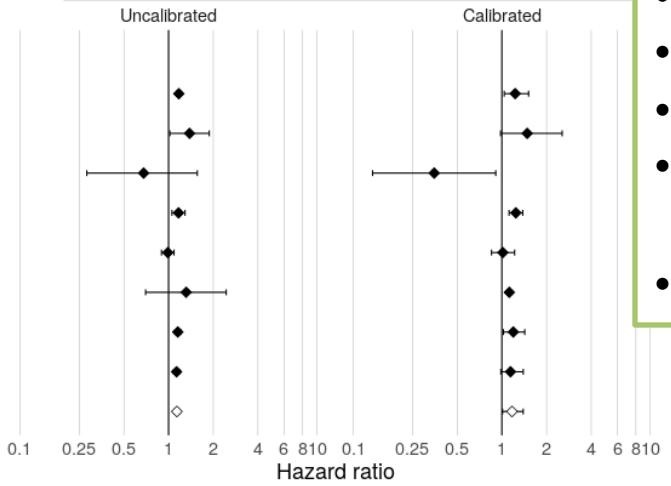
Show 15 entries

Analysis	Data source	HR	LB	UB	P	Cal.HR	Cal.LB	Cal.UB	Cal.P
PS matching, on-treatment	CCAE	1.17	1.13	1.22	0.00	1.23	1.04	1.51	0.02
PS matching, on-treatment	CUMC	1.39	1.02	1.88	0.04	1.48	0.98	2.55	0.01
PS matching, on-treatment	IMSG	0.68	0.28	1.56	0.38	0.35	0.13	0.91	0.03
PS matching, on-treatment	MDCD	1.17	1.05	1.29	0.00	1.24	1.12	1.38	0.00
PS matching, on-treatment	MDCR	0.99	0.90	1.09	0.80	1.01	0.85	1.22	0.79
PS matching, on-treatment	Meta-analysis	1.14	1.09	1.18	0.00	1.17	1.01	1.39	0.01
PS matching, on-treatment	NHIS_NSC	1.31	0.70	2.44	0.39	1.12	NA	2.24	0.75
PS matching, on-treatment	Optum	1.15	1.11	1.21	0.00	1.19	1.02	1.43	0.02
PS matching, on-treatment	Panther	1.13	1.08	1.18	0.00	1.14	0.98	1.39	0.00

Showing 1 to 9 of 9 entries Previous 1 Ne

Power | Propensity scores | Covariate balance | Systematic error | Forest plot

Source	HR (95% CI)	Calibrated HR (95% CI)
CCAE	1.17 (1.13-1.22)	1.23 (1.04-1.51)
CUMC	1.39 (1.02-1.88)	1.48 (0.98-2.55)
IMSG	0.68 (0.28-1.56)	0.35 (0.13-0.91)
MDCD	1.17 (1.05-1.29)	1.24 (1.12-1.38)
MDCR	0.99 (0.90-1.09)	1.01 (0.85-1.22)
NHIS_NSC	1.31 (0.70-2.44)	1.12 (NA-2.24)
Optum	1.15 (1.11-1.21)	1.19 (1.02-1.43)
Panther	1.13 (1.08-1.18)	1.14 (0.98-1.39)
Summary (I^2 = 0.50)	1.14 (1.09-1.18)	1.17 (1.01-1.39)



LEGEND workgroup focus on clinically important hypotheses:

- ACE vs. THZ
- ACE vs. ARB
- Beta blocker vs. first-line
- Chlorthalidone vs. hydrochlorothiazide
- Mono vs. combo therapy

http://data.ohdsi.org/LegendBasicViewer/

OHDSI is

building collaborations

FDA Biologics Effectiveness and Safety (BEST) Initiative

Biologics Effectiveness and Safety (BEST) Initiative: Incorporating ISBT-128 Codes into OHDSI's OMOP Common Data Model to Build a National Hemovigilance System to Monitor Transfusion-Related Adverse Events

Joyce Obidi[1], Kinnera Chada[1], Joann Gruber[1], Graça Dores[1], Alan Williams[1], Emily Storch[1], Juan M Banda[2] Saurabh Gombar[2], Deepa Balraj[2], Ross Hayden[3], Paul Biondich[3], Shaun Grannis[3], George Hripcsak[4], Thomas Falconer[4], Karthik Natarajan[4], Dmitry Dymshyts[6], Sara Dempster[7], Christian Reich[7], Nandini Selvam[7], Nerissa Williams[7], Steven Anderson[1], Azadeh Shoaibi[1]

[1]Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, MD, USA; [2]Stanford University, Stanford, CA, USA; [3]Regenstrief Institute, Indianapolis, Indiana, USA; [4]Columbia University, New York, NY, USA; [5]Observational Health Data Sciences and Informatics, New York, NY, USA; [6]Odysseus Data Services Inc., Cambridge, MA, USA; [7]IQVIA, Cambridge, MA, USA

INTRODUCTION

The U.S. FDA Center for Biologics Evaluation and Research (CBER) regulates collection of whole blood and blood components utilized in transfusion[1].

CBER's Role in Blood Safety

To protect recipients of blood and blood components and to monitor transfusion-related adverse events (AEs)

BEST Initiative

Biologics Effectiveness and Safety (BEST) Initiative is a component of the CBER Sentinel Program. The BEST Initiative is made up of a distributed network of data providers that use claims and electronic health record (EHR) data sources transformed into a common data model (CDM).

Infrastructure for Hemovigilance

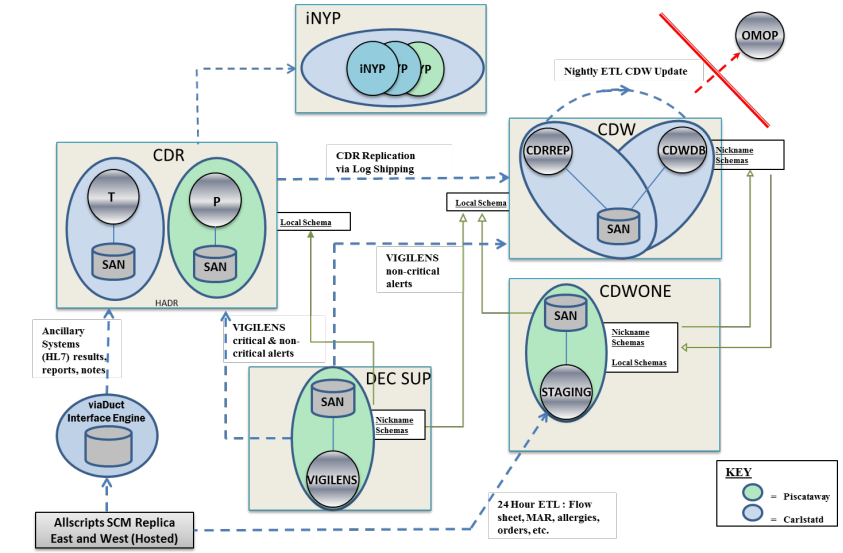
The most detailed blood and blood components data are included in the Information Standard for Blood and Transplant (ISBT)-128 coding system[2]. In laying the infrastructure for a hemovigilance system, we incorporated the ISBT-128 coding system into the CDM used by the BEST Initiative.

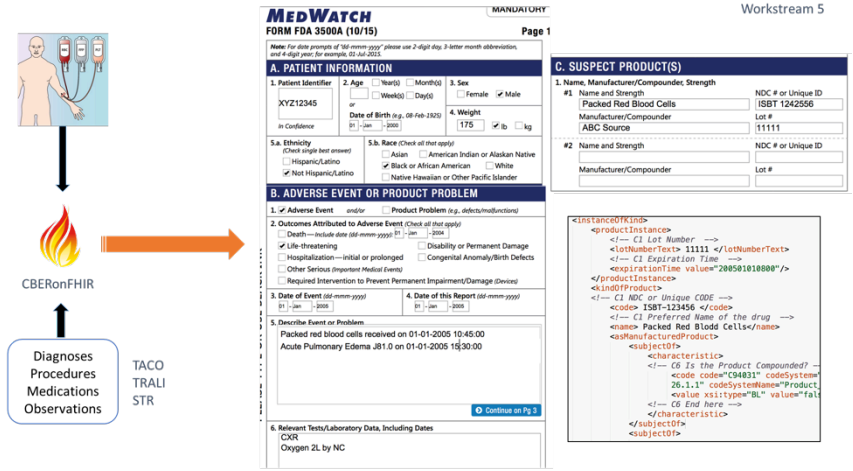
OBJECTIVE

The aim of this study was to build a component of the infrastructure for a national hemovigilance system using EHR data sources to monitor transfusion-related AEs by incorporating the ISBT-128 coding system into the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) of the Observational Health Data Sciences and Informatics (OHDSI) consortium[3].

METHODS

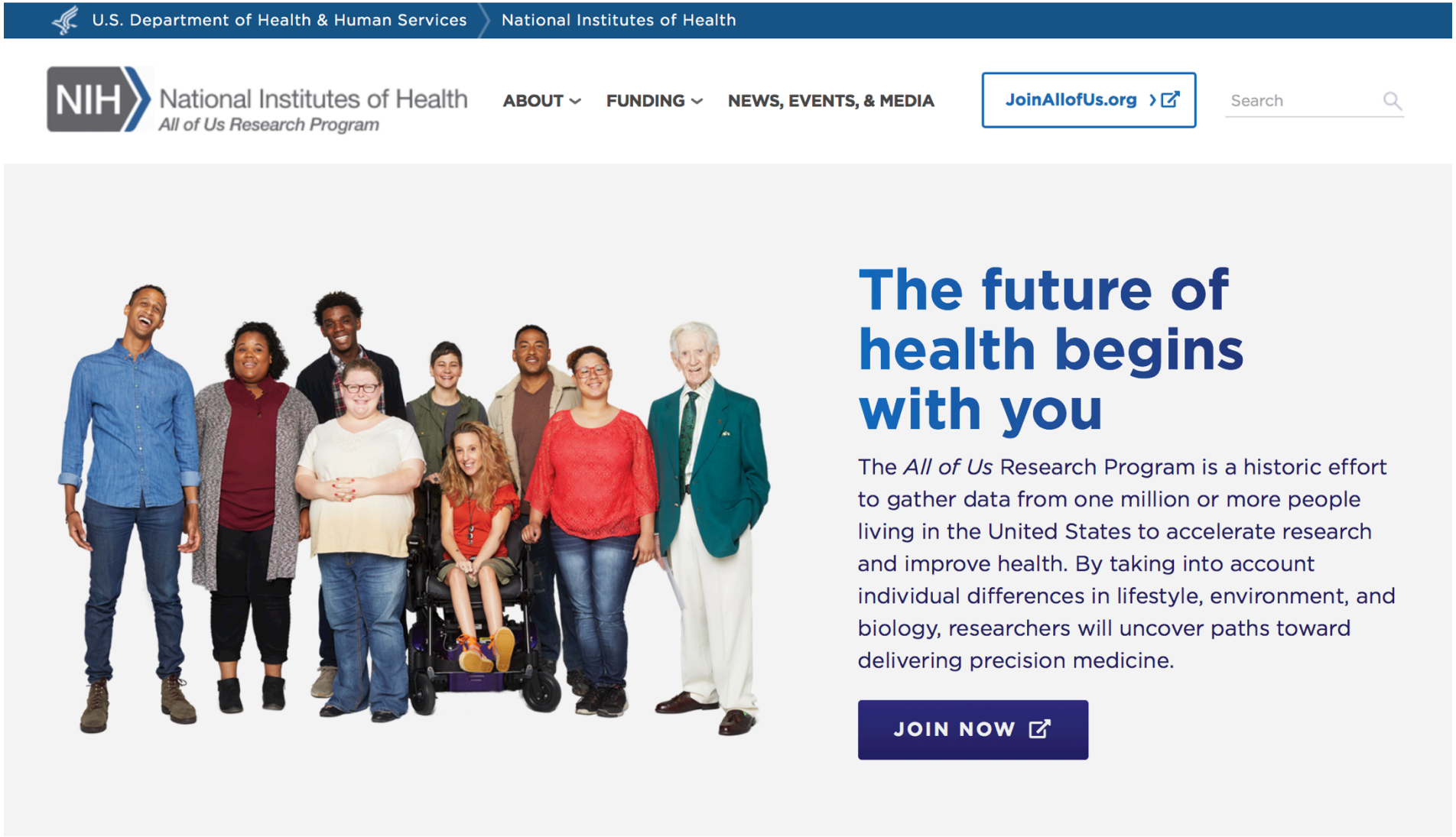
The CBER BEST Initiative is a collaboration with IQVIA, OHDSI Consortium, Columbia University, Stanford University, Indiana University, Regenstrief Institute, Georgia Institute of Technology, and University of California Los Angeles. Within the BEST Initiative, we used three EHR databases that cover approximately 24 million patient records from geographically diverse areas of the U.S. We added a library of 14,543 ISBT-128 codes to the OMOP CDM. Each EHR data source requested access to its corresponding blood bank data and transformed its data into the OMOP CDM containing the newly added ISBT-128 codes. By querying the databases, we determined the type and frequency of ISBT-128 codes used in patient records from 2010-2017 within the blood banks of EHR data providers participating in the BEST Initiative.





NIH *All of Us Research Program*

- 1,000,000 diverse participants
- Clinical data in OMOP CDM



NIH *All Of Us Research Program*

PLOS ONE

RESEARCH ARTICLE

Data model harmonization for the All Of Us Research Program: Transforming i2b2 data into the OMOP common data model

Jeffrey G. Klann[1,2,3]*, **Matthew A. H. Joss**[1], **Kevin Embree**[4], **Shawn N. Murphy**[1,2,3]

1 Research Information Science and Computing, Partners Healthcare, Boston, Massachusetts, United States of America, **2** Harvard Medical School, Boston, Massachusetts, United States of America, **3** Laboratory of Computer Science, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, United States of America, **4** Personalized Medicine, Partners Healthcare, Boston, Massachusetts, United States of America

* jeff.klann@mgh.harvard.edu

Check for updates

OPEN ACCESS

Citation: Klann JG, Joss MAH, Embree K, Murphy SN (2019) Data model harmonization for the All Of Us Research Program: Transforming i2b2 data into the OMOP common data model. PLoS ONE 14(2): e0212463. https://doi.org/10.1371/journal.pone.0212463

Editor: Christian Lovis, Hopitaux Universitaires de Geneve, SWITZERLAND

Received: December 4, 2018

Accepted: February 2, 2019

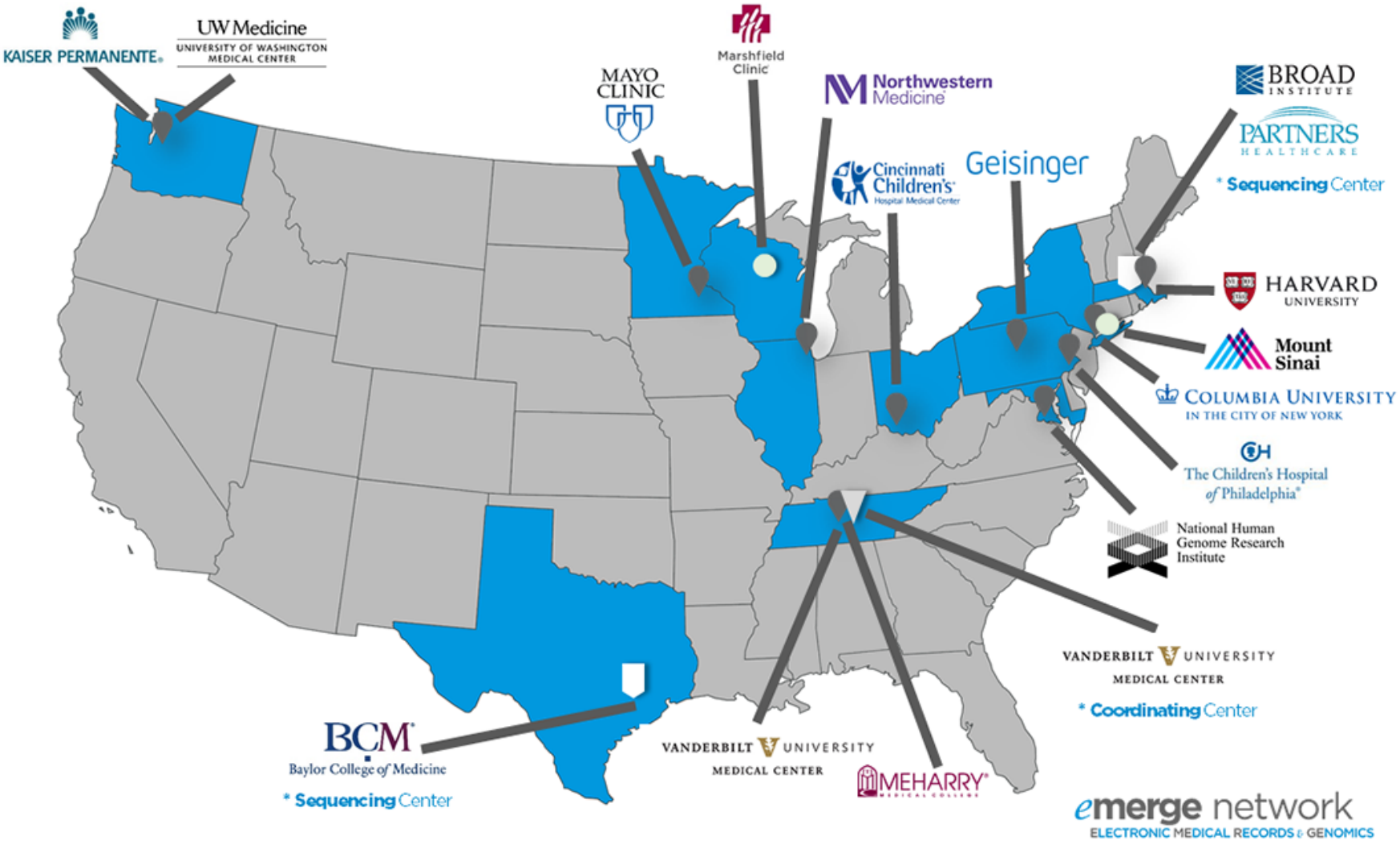
Published: February 19, 2019

Abstract

Background

The All Of Us Research Program (AOU) is building a nationwide cohort of one million patients' EHR and genomic data. Data interoperability is paramount to the program's success. AOU is standardizing its EHR data around the Observational Medical Outcomes Partnership (OMOP) data model. OMOP is one of several standard data models presently used in national-scale initiatives. Each model is unique enough to make interoperability difficult. The i2b2 data warehousing and analytics platform is used at over 200 sites worldwide, which uses a flexible ontology-driven approach for data storage. We previously demonstrated this ontology system can drive data reconfiguration, to transform data into new formats without site-specific programming. We previously implemented this on our 12-site Accessible Research Commons for Health (ARCH) network to transform i2b2 into the Patient Centered Outcomes Research Network model.

Electronic Medical Records and Genomics (eMERGE) Network



Electronic Medical Records and Genomics (eMERGE) Network

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Facilitating phenotype transfer using a common data model

George Hripcsak[a,b,*], Ning Shang[a], Peggy L. Peis
Barbara Benoit[e], Robert J. Carroll[f], David S. Car
Vivian S. Gainer[e], Kayla Marie Howell[j], Jeffrey
Frank D. Mentch[l], Shawn N. Murphy[e], Karthik N
Ken Wiley[m], Chunhua Weng[a]

[a] Department of Biomedical Informatics, Columbia University, New York, NY, Unite
[b] Medical Informatics Services, NewYork-Presbyterian Hospital, New York, NY, Uni
[c] Center for Precision Medicine Research, Marshfield Clinic Research Institute, Mars
[d] Northwestern University Feinberg School of Medicine, Chicago, IL, United States
[e] Research Information Science and Computing, Partners Healthcare, Boston, MA, U
[f] Department of Biomedical Informatics, Vanderbilt University Medical Center, Nash
[g] Kaiser Permanente Washington Health Research Institute, Seattle, WA, United Stat
[h] Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, Uni
[i] Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN, United State
[j] Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University
[k] Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States
[l] Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA,
[m] National Human Genome Research Institute, NIH, Bethesda, MD, United States

Background: Implementing clinical phenotypes across a network is labor intensive and potentially error prone. Use of a common data model may facilitate the process.

Methods: Electronic Medical Records and Genomics (eMERGE) sites implemented the Observational Health Data Sciences and Informatics (OHDSI) Observational Medical Outcomes Partnership (OMOP) Common Data Model across their electronic health record (EHR)-linked DNA biobanks. Two previously implemented eMERGE phenotypes were converted to OMOP and implemented across the network.

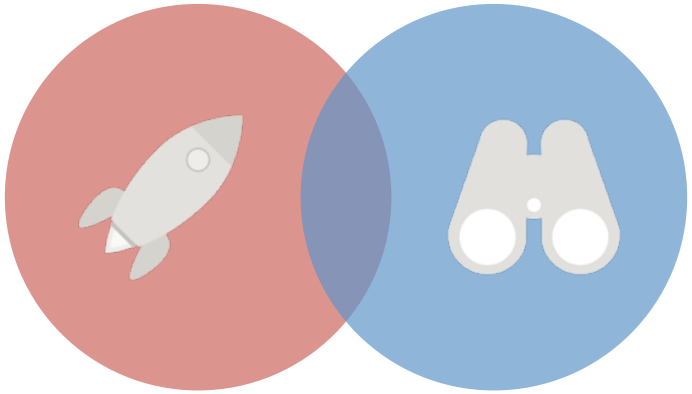
Results: It was feasible to implement the common data model across sites, with laboratory data producing the greatest challenge due to local encoding. Sites were then able to execute the OMOP phenotype in less than one day, as opposed to weeks of effort to manually implement an eMERGE phenotype in their bespoke research EHR databases. Of the sites that could compare the current OMOP phenotype implementation with the original eMERGE phenotype implementation, specific agreement ranged from 100% to 43%, with disagreements due to the original phenotype, the OMOP phenotype, changes in data, and issues in the databases. Using the OMOP query as a standard comparison revealed differences in the original implementations despite starting from the same definitions, code lists, flowcharts, and pseudocode.

Conclusion: Using a common data model can dramatically speed phenotype implementation at the cost of having to populate that data model, though this will produce a net benefit as the number of phenotype implementations increases. Inconsistencies among the implementations of the original queries point to a potential benefit of using a common data model so that actual phenotype code and logic can be shared, mitigating human error in re-interpretation of a narrative phenotype definition.

The European Health Data and Evidence Network

Mission

Our mission is to provide a new paradigm for the discovery and analysis of health data in Europe, by building a large-scale, federated network of data sources standardised to a common data model



Vision

The European Health Data & Evidence Network (EHDEN) aspires to be the trusted observational research ecosystem to enable better health decisions, outcomes and care

Objectives

EHDEN
EUROPEAN HEALTH DATA & EVIDENCE NETWORK

Harmonisation

Harmonise in excess of **100 million** anonymised **health records** to the OMOP common data model, supported by an ecosystem of certified SMEs, and technical architecture for a federated network

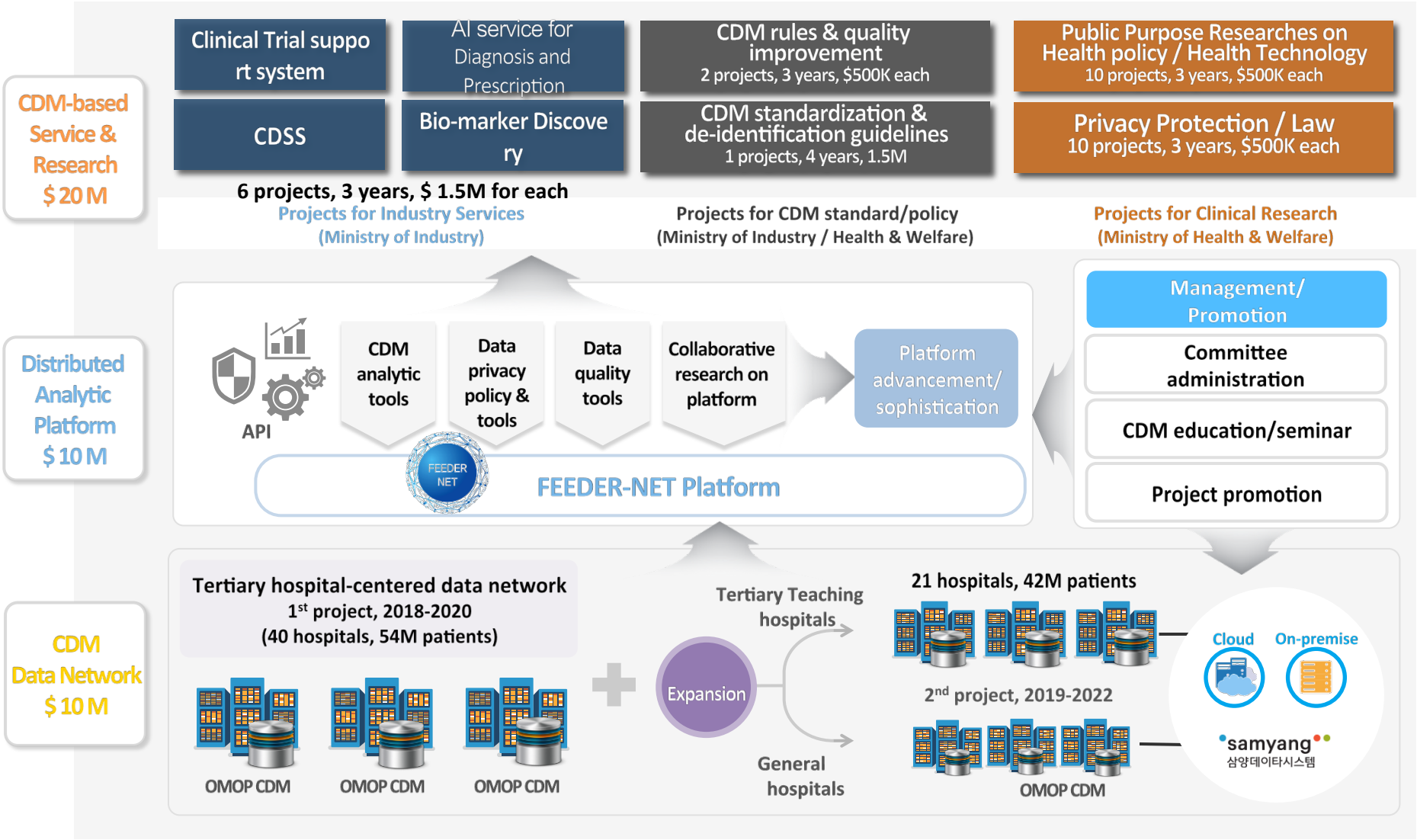
Evidence

Impact our understanding of, and improvement of, clinical **outcomes for patients** within diverse healthcare systems in the EU

Community

Establish a self-sustaining **open science collaboration** in Europe, supporting academia, industry, regulators, payers, government, NGOs and others

National CDM Projects in Korea
2018-2022



FEEDER-NET: FEderated E-health big Data for Evidence Renovation NETwork

FEEDER-NET Data Network in Korea

Data Network of 60+ Hospitals, 98M Patients

70% of Tertiary Teaching Hospitals

서울 Seoul

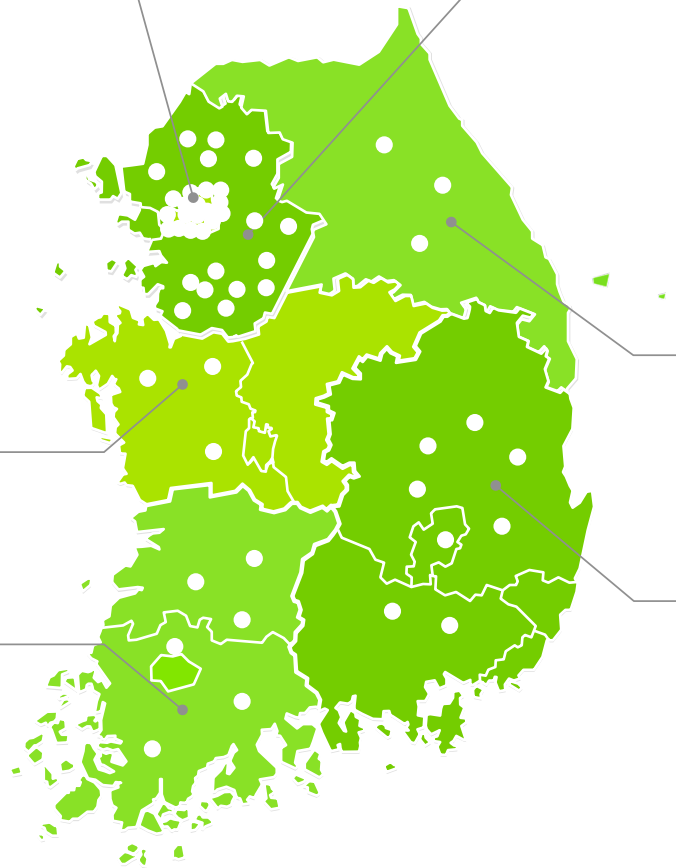
- 서울아산병원 Asan Medical Center
- SNUH 서울대학교병원
- 명지병원 MYONGJI HOSPITAL
- 가톨릭대학교 서울성모병원
- 인제대학교서울백병원
- 가톨릭대학교 여의도성모병원
- 한국원자력의학원
- 가톨릭대학교 성바오로병원
- 세브란스병원 SEVERANCE HOSPITAL
- 고려대학교안암병원
- 고려대학교구로병원
- 삼성서울병원
- 강동경희대학교병원
- 경희대학교병원
- 건국대학교병원
- 원자력병원
- 한양대학교의료원
- 중앙보훈병원
- 목동병원
- 강북삼성병원
- 한림대학교강동성심병원
- 순천향대학교 서울병원
- 연세대학교의과대학 강남세브란스병원

충청도 Chungcheongdo

- 건양대학교병원 KONYANG UNIV. HOSPITAL
- CNU 충남대학교병원
- 순천향대학교 천안병원

전라도 Jeonlado

- 예수병원
- 원광대학교병원 WONKWANG UNIVERSITY HOSPITAL
- 전북대학교병원
- 전라북도 군산의료원
- 진안군의료원 JINANGUN MEDICAL CENTER
- 전라북도 남원의료원



Incheon / Kyeonggi 인천/경기

- 가톨릭대학교 부천성모병원
- 세종병원
- SNUH 분당서울대학교병원
- 가톨릭대학교 의정부성모병원
- 메디플렉스 세종병원
- 국립암센터 NATIONAL CANCER CENTER
- 인천사랑병원
- 순천향대학교 부천병원
- 인하대병원
- 아주대학교병원 Ajou University Hospital
- 고려대학교안산병원
- 가천대길병원 Gachon University Gil Hospital
- 용인세브란스병원 YONGIN SEVERANCE HOSPITAL
- 동국대학교일산병원
- CHA 의과학대학교 차병원

Gangwondo 강원도

- 가톨릭관동대학교 국제성모병원
- 원주세브란스기독병원
- 강원대학교병원

Gyeongsangdo 경상도

- 울산대학교병원 ULSAN UNIVERSITY HOSPITAL
- 명지병원 MYONGJI HOSPITAL
- 순천향대학교 구미병원
- KNUH 경북대학교병원
- PNUH 부산대학교병원
- 동국대학교경주병원
- 칠곡경북대학교병원
- 대구가톨릭대학교병원

OHDSI is

a community of collaborators

OHDSI-EHDEN STUDY-ATHON

External Validation package available at:
https://github.com/OHDSI/StudyProtocolSandbox/tree/master/mortalityValidation

- 5 day event
- ~30 researchers
- Result: 2 papers

Results of the study available at:
http://data.ohdsi.org/oxfordMortalityExternalValidation/

Case Western Reserve University : OHDSI face-to-face documentation-a-thon

OHDSI China Symposium 2019

The Journey From Data to Evidence OHDSI Europe 2019

- A platform to stimulate community building: 250 participants from **27** countries
- OHDSI Europe in action: 35 posters, 8 software demos
- Educate and train the community: 5 full day tutorials

www.ohdsi-europe.org

The Journey From Data to Evidence OHDSI Europe 2019

Fudan University – OHDSI tutorials

OHDSI Korea – Study design datathon

Upcoming symposia

OHDSI Korea

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

OHDSI Korea Symposium

12-14th December 2019,

KONJIAM Resort, Gwangju, Gyeonggi-Do, Republic of Korea

OHDSI EUROPE
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

European OHDSI Symposium 2020

27-29 March 2020 – Oxford, UK

Mathematical Institute,
University of Oxford

NDORMS
NUFFIELD DEPARTMENT OF ORTHOPAEDICS, RHEUMATOLOGY AND MUSCULOSKELETAL SCIENCES

2019 Theme: Continuous evaluation

How do we know we are making progress on our journey?

OHDSI evaluates itself and publishes the results

- OMOP CDM vocabulary evaluation
  - Automated translation of database works
  - Best not to automated the translation of cohort definitions
- eMERGE phenotype implementation
  - Without CDM, narrative+flowchart+pseudocode+code list -> inconsistent
  - With CDM, can improve consistency and efficiency but caveats
- PheValuator phenotype evaluation
  - Can estimate performance without manually curating gold standard
  - Estimates are imperfect

Today’s focus

How do we know we are making progress toward using real-world evidence for regulatory decision-making?

Symposium • Day 2

Time	Description
7:30 - 8:00 am	Registration with light breakfast **(Grand Ballroom F-H Foyer)**
8:00 - 9:00 am	**Welcome to OHDSI 2019: This is our community (Grand Ballroom DE)** • **George Hripcsak, MD, MS**, Vivian Beaumont Allen Professor and Chair of Biomedical Informatics at Columbia University Irving Medical Center; Director of Medical Informatics Services at NewYork-Presbyterian Hospital/Columbia Campus • **Harlan Krumholz, MD**, Harold H. Hines, Jr. Professor of Medicine, Epidemiology and Public Health at Yale University
9:00 - 11:00 am	**Plenary Session: A journey toward real-world evidence for regulatory decision-making** Building confidence in **real-world data**: Data quality reporting • **Clair Blacketer, MPH, PMP**, Associate Director of Epidemiology Analytics at Janssen Research & Development; PhD Student at Erasmus Medical Center Rotterdam • **Andrew Williams, PhD**, Senior Informatics Advisor at Tufts Medical Center ——— Establishing scientific best practices for **real-world analysis**: Book Of OHDSI • **Martijn Schuemie, PhD**, Senior Director and Research Fellow of Epidemiology Analytics at Janssen Research & Development; Visiting Scholar of Biostatistics at the University of California, Los Angeles • **David Madigan, PhD**, Professor of Statistics at Columbia University ——— Proving reliable **real-world evidence**: Replicating RCTs using LEGEND • **Patrick Ryan, PhD**, Vice President of Observational Health Data Analytics at Janssen Research & Development; Adjunct Assistant Professor of Biomedical Informatics at Columbia University • **George Hripcsak, MD, MS**, Vivian Beaumont Allen Professor and Chair of Biomedical Informatics at Columbia University Irving Medical Center; Director of Medical Informatics Services at NewYork-Presbyterian Hospital/Columbia Campus
11:00 - 11:30 am	**Stakeholder panel: What has been done? Where should we go? How do we get there?** **Moderator: Harlan M. Krumholz, MD, SM**, Harold H. Hines, Jr. Professor of Medicine, Epidemiology and Public Health at Yale University **Panelists: Joseph Ross, MD, MHS**, Professor of Internal Medicine at Yale University **Azadeh Shoaibi, PhD, MHS**, Lead of CBER Sentinel Program at US Food and Drug Administration **Fatemah Alnofal, MIT**, Research Specialist at the Saudi Food and Drug Authority
11:30 - 1:00 pm	**OHDSI Collaborator Showcase: Part 1 (Grand Ballroom F-H)** Software demonstrations and poster presentations highlighting the scientific progress throughout the OHDSI community **Buffet Lunch served in the Foyer at 12:30pm*
1:00 - 2:00 pm	**OHDSI Collaborator Showcase: Part 2, Lightning Talks (Grand Ballroom DE)** **Moderator:** James Weaver, MPH, MS, Associate Director of Epidemiology Analytics at Janssen Research and Development **Speakers: Mui Van Zandt**, Director of OMOP Data Networks at IQVIA **Rimma Belenkaya, MA, MS**, Data Modeler/Knowledge Manager at Memorial Sloan Kettering Cancer Center; **Juan M. Banda, PhD**, Assistant Professor of Computer Science at Georgia State Univ.; **Rupa Makadia, PhD, MS**, Associate Director of Epidemiology Analytics at Janssen Research and Development; **Anastasiya Nestsiarovich, MD, PhD**, Postdoctoral Fellow at the Univ. of New Mexico; **Seng Chan You, MD, MS**, Medical Doctor at the Department of Biomedical Informatics at Ajou University; and **Alison Callahan, PhD**, Research Scientist at the Center for Biomedical Informatics Research at Stanford University
2:00 - 3:00 pm	**OHDSI Collaborator Showcase: Part 3 (Grand Ballroom F-H)** Software demonstrations and poster presentations highlighting the scientific progress throughout the OHDSI community
3:00 - 4:30 pm	**Community Evidence in Action (Grand Ballroom DE)** **Introduction by: Mui Van Zandt**, Director of OMOP Data Networks at IQVIA **European Health Data & Evidence Network (EHDEN) / Oxford study-a-thon – Exploring knee arthroplasty: Peter Rijnbeek, PhD**, Associate Professor of Health Data Science at Erasmus Medical Center Rotterdam and **Dani Prieto-Alhambra, MD, PhD**, Professor of Pharmaco- and Device Epidemiology at University of Oxford ——— **Center for Surgical Sciences – Personalizing surgery for colorectal cancer: Ismail Gögenur, MD, DMSc**, Professor and Director of Center for Surgical Science (CSS) at the Zealand University Hospital, Denmark and **Gregory Klebanov, MS**, Chief Technology Officer at Odysseus Data Services, Inc. ——— **Women of OHDSI – Predicting breast cancer to improve screening: Maura Beaton, MS**, Project Manager of OHDSI at Columbia University; **Kristin Kostka, MPH**, Associate Director, OMOP Data Networks – Americas at IQVIA; **Jenna Reps, PhD**, Associate Director of Epidemiology Analytics at Janssen Research and Development; and **Anna Ostropolets, MD**, PhD Student at Columbia University
4:30 - 5:30 pm	**Closing session: Growing up on a journey** **Patrick Ryan, PhD**, Vice President of Observational Health Data Analytics at Janssen Research & Development; Adjunct Assistant Professor of Biomedical Informatics at Columbia University **Christian Reich, MD, PhD**, Vice President of Real World Solutions at IQVIA *Best Contribution and Titan award winners will be announced during this time*
5:30 - 7:30 pm	**Networking reception (Grand Ballroom F-H)** *Light refreshments will be served*

A journey toward real-world evidence for real-world decision-making

The current medical research enterprise cannot keep pace with the information needs of patients, clinicians, administrators, and policy makers. The flow of new knowledge is too slow, and its scope is too narrow.

By Harlan M. Krumholz

Big Data And New Knowledge In Medicine: The Thinking, Training, And Tools Needed For A Learning Health System

ABSTRACT Big data in medicine—massive quantities of health care data accumulating from patients and populations and the advanced analytics that can give those data meaning—hold the prospect of becoming an engine for the knowledge generation that is necessary to address the extensive unmet information needs of patients, clinicians, administrators, researchers, and health policy makers. This article explores the ways in which big data can be harnessed to advance prediction, performance, discovery, and comparative effectiveness research to address the complexity of patients, populations, and organizations. Incorporating big data and next-generation analytics into clinical and population health research and practice will require not only new data sources but also new thinking, training, and tools. Adequately utilized, these reservoirs of data can be a practically inexhaustible source of knowledge to fuel a learning health care system.

Health Aff (Millwood). 2014 Jul; 33(7): 1163–1170

The medical research community's delay in adopting big data approaches has left it particularly ill prepared for a precision medicine future that is designed to provide personalized information and individualized care.

By Harlan M. Krumholz

Big Data And New Knowledge In Medicine: The Thinking, Training, And Tools Needed For A Learning Health System

ABSTRACT Big data in medicine—massive quantities of health care data accumulating from patients and populations and the advanced analytics that can give those data meaning—hold the prospect of becoming an engine for the knowledge generation that is necessary to address the extensive unmet information needs of patients, clinicians, administrators, researchers, and health policy makers. This article explores the ways in which big data can be harnessed to advance prediction, performance, discovery, and comparative effectiveness research to address the complexity of patients, populations, and organizations. Incorporating big data and next-generation analytics into clinical and population health research and practice will require not only new data sources but also new thinking, training, and tools. Adequately utilized, these reservoirs of data can be a practically inexhaustible source of knowledge to fuel a learning health care system.

Health Aff (Millwood). 2014 Jul; 33(7): 1163–1170

Medicine aspires to a learning health care system, but is failing to rapidly learn from the data being generated through the course of clinical care.

By Harlan M. Krumholz

Big Data And New Knowledge In Medicine: The Thinking, Training, And Tools Needed For A Learning Health System

ABSTRACT Big data in medicine—massive quantities of health care data accumulating from patients and populations and the advanced analytics that can give those data meaning—hold the prospect of becoming an engine for the knowledge generation that is necessary to address the extensive unmet information needs of patients, clinicians, administrators, researchers, and health policy makers. This article explores the ways in which big data can be harnessed to advance prediction, performance, discovery, and comparative effectiveness research to address the complexity of patients, populations, and organizations. Incorporating big data and next-generation analytics into clinical and population health research and practice will require not only new data sources but also new thinking, training, and tools. Adequately utilized, these reservoirs of data can be a practically inexhaustible source of knowledge to fuel a learning health care system.

Health Aff (Millwood). 2014 Jul; 33(7): 1163–1170

COMMENT

Growing access to diverse 'real-world' data sources is enabling new approaches to close persistent evidence gaps about the optimal use of medical products in real-world practice.

Accelerating development of scientific evidence for medical products within the existing US regulatory framework

Rachel E. Sherman[1], Kathleen M. Davies[1], Melissa A. Robb[1], Nina L. Hunter[1] and Robert M. Califf[1,2]

Growing access to diverse 'real-world' data sources is enabling new approaches to close persistent evidence gaps about the optimal use of medical products in real-world practice. Here, we argue that contrary to widespread impressions, existing FDA regulations embody sufficient flexibility to accommodate the emerging tools and methods needed to achieve this goal.

Nat Rev Drug Discov. 2017 May;16(5):297-298

COMMENT

Here, we argue that contrary to widespread impressions, existing FDA regulations embody sufficient flexibility to accommodate the emerging tools and methods needed to achieve this goal.

Accelerating development of scientific evidence for medical products within the existing US regulatory framework

Rachel E. Sherman[1], Kathleen M. Davies[1], Melissa A. Robb[1], Nina L. Hunter[1] and Robert M. Califf[1,2]

Growing access to diverse 'real-world' data sources is enabling new approaches to close persistent evidence gaps about the optimal use of medical products in real-world practice. Here, we argue that contrary to widespread impressions, existing FDA regulations embody sufficient flexibility to accommodate the emerging tools and methods needed to achieve this goal.

Nat Rev Drug Discov. 2017 May;16(5):297-298

…Congressional mandate requires "the coordination of relevant Federal health programs to build data capacity for comparative clinical effectiveness research . . . from multiple sources, including electronic health records"

The NEW ENGLAND JOURNAL of MEDICINE

SOUNDING BOARD

Transforming Evidence Generation to Support Health and Health Care Decisions

Robert M. Califf, M.D., Melissa A. Robb, M.S.(Reg.Sci.), B.S.N., Andrew B. Bindman, M.D., Josephine P. Briggs, M.D., Francis S. Collins, M.D., Ph.D., Patrick H. Conway, M.D., Trinka S. Coster, M.D., Francesca E. Cunningham, Pharm.D., Nancy De Lew, M.A., Karen B. DeSalvo, M.D., M.P.H., Christine Dymek, Ed.D., Victor J. Dzau, M.D., Rachael L. Fleurence, Ph.D., Richard G. Frank, Ph.D., J. Michael Gaziano, M.D., M.P.H., Petra Kaufmann, M.D., Michael Lauer, M.D., Peter W. Marks, M.D., Ph.D., J. Michael McGinnis, M.D., M.P.P., Chesley Richards, M.D., M.P.H., Joe V. Selby, M.D., M.P.H., David J. Shulkin, M.D., Jeffrey Shuren, M.D., J.D., Andrew M. Slavitt, M.B.A., Scott R. Smith, Ph.D., B. Vindell Washington, M.D., M.H.C.M., P. Jon White, M.D., Janet Woodcock, M.D., Jonathan Woodson, M.D., and Rachel E. Sherman, M.D., M.P.H.

N Engl J Med. 2016 Dec 15;375(24):2395-2400

… governmental agencies and partners in the private sector, including those that fund research, are now collaborating on the focused development of infrastructure for the generation of evidence that can support a learning health system.

The NEW ENGLAND JOURNAL of MEDICINE

SOUNDING BOARD

Transforming Evidence Generation to Support Health and Health Care Decisions

Robert M. Califf, M.D., Melissa A. Robb, M.S.(Reg.Sci.), B.S.N., Andrew B. Bindman, M.D., Josephine P. Briggs, M.D., Francis S. Collins, M.D., Ph.D., Patrick H. Conway, M.D., Trinka S. Coster, M.D., Francesca E. Cunningham, Pharm.D., Nancy De Lew, M.A., Karen B. DeSalvo, M.D., M.P.H., Christine Dymek, Ed.D., Victor J. Dzau, M.D., Rachael L. Fleurence, Ph.D., Richard G. Frank, Ph.D., J. Michael Gaziano, M.D., M.P.H., Petra Kaufmann, M.D., Michael Lauer, M.D., Peter W. Marks, M.D., Ph.D., J. Michael McGinnis, M.D., M.P.P., Chesley Richards, M.D., M.P.H., Joe V. Selby, M.D., M.P.H., David J. Shulkin, M.D., Jeffrey Shuren, M.D., J.D., Andrew M. Slavitt, M.B.A., Scott R. Smith, Ph.D., B. Vindell Washington, M.D., M.H.C.M., P. Jon White, M.D., Janet Woodcock, M.D., Jonathan Woodson, M.D., and Rachel E. Sherman, M.D., M.P.H.

N Engl J Med. 2016 Dec 15;375(24):2395-2400

If RWD and RWE are to be effectively leveraged for public health purposes, there will need to be shared learning and collaboration across clinicians, patients, health care systems, pharmaceutical companies, and regulators.

VIEWPOINT

Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness

Jacqueline Corrigan-Curay, JD, MD
Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland.

Leonard Sacks, MD
Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland.

Janet Woodcock, MD
Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland.

For hundreds of years, the development of new medical treatments relied on "real-world" experience. Discoveries such as citrus fruit curing scurvy described in the 1700s or insulin as a treatment for diabetes in the 1920s long preceded the advent of the modern randomized clinical trial. What these diseases had in common was a reliable method of diagnosis, a predictable clinical course, and a large and obvious effect of the treatment.

In the late 1940s, the medical community began to adopt the use of randomized clinical designs for drug trials.[1] The recognition that anecdotal reports based on clinical practice observations were often misleading led to the nearly complete replacement of this "real-world evidence" (RWE) approach to evidence generated using the modern clinical trial model. Although moving medical science toward greater scientific rigor, this transformation simultaneously diminished the use (and minimized the value) of evidence generated from practice-based observations. Randomization and blinding became the gold standard for determining the effect of treatment. With

records (EHRs), together with rising costs and recognized limitations of traditional trials, has renewed interest in the use of real-world data (RWD) to enhance the efficiency of research and bridge the evidentiary gap between clinical research and practice. RWD can be defined as data relating to patient health status or the delivery of health care routinely collected from a variety of sources, such as the EHR and administrative data.

Under the 21st Century Cures Act, the Food and Drug Administration is tasked with developing a program to evaluate the use of RWE to support approval of new indications for approved drugs or to satisfy postapproval study requirements.[2] RWE can be defined as the clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of RWD. A framework for this program will be published by the end of 2018.

The FDA routinely uses RWD to provide evidence about drug safety, drawing on claims and pharmacy data from more than 100 million individuals in its Sentinel System.[3] In addition, FDA regulations have long recog-

JAMA 2018 Sep 4;320(9):867-868

What do I love about OHDSI?

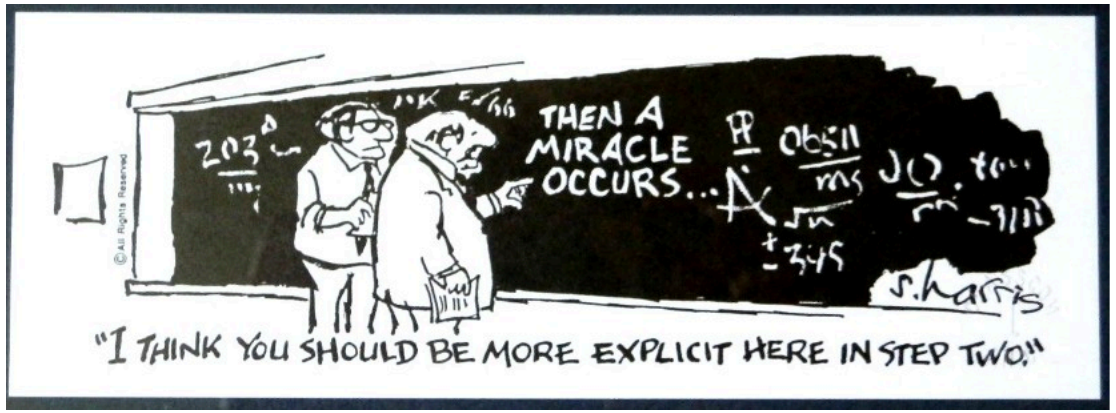
- Spirit of collaboration, kindness, generosity

What do I love about OHDSI?

- Principles of transparency, open science, integrity

What do I love about OHDSI?

- Scientific rigor; reproducibility, validity



Key Challenges

- Data quality
- Data spectrum
- Causal inference
- Communication/Education
- Application

Next Session

- Data quality
  - Plenary #1: Blacketer and Williams
- Analytic standards
  - Plenary #2: Schuemie and Madigan
- Evidence quality
  - Plenary #3: Ryan and Hripcsak