Comparing the use of OMOP and Sentinel CDMs for Drug Safety – Implications for European Data

Andrew Bate
Senior Director, Epidemiology Group Lead, Analytics

A Common Data Model for Europe? – Why? Which? How?

December 11th 2017

Pfizer
WORLDWIDE SAFETY & REGULATORY
Worldwide Research & Development

Disclosures and potential conflicts of interest

- I am a full time employee of Pfizer and hold stocks and stock options
- OMOP involvement (until 2013)
  - Member of the Scientific Advisory Board of OMOP
  - Conducted research as part of the OMOP Extended consortium with advice from Patrick Ryan and other OMOP PIs (on converting the UK EMR THIN into OMOP CDM)
- Co-PI on the first, now completed, pilot of IMEDS, a program for non-FDA stakeholder access of the US FDA's Sentinel System data and analytic tools
- The views expressed in this presentation are my own and do not necessarily reflect those of Pfizer

RWE from healthcare databases contributes to Safety Assessment Across Lifecycle

Characterize Patient Risk Profile

Evaluate Medication Risk

Standing Cohorts

- EMRs
- Claims
- Registries

Rapid Queries
Estimate expected risks in indicated populations

Approval

Active Surveillance
Monitor and detect signals in defined patient cohorts using innovative analytic methods

Post Approval Safety Studies
Compare medication risks in the real world, as prescribed and taken during routine clinical practice

Risk Minimization
Evaluate the effectiveness of risk minimization measures (e.g., label/education)

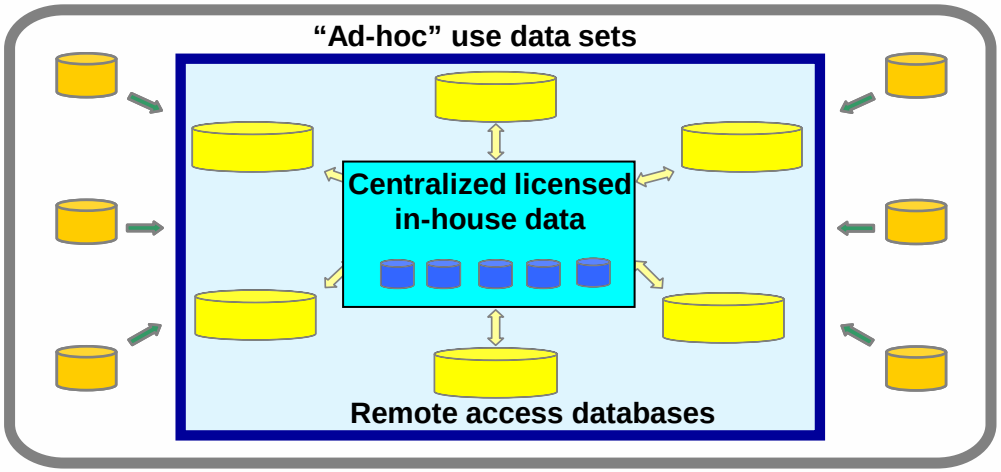
‘Three tiered’ Real World Data Strategy

Suitability of RWD source to address the question of interest
Data capture and its structure
Accessibility
Demonstrability of data and analysis integrity
Recency of data available for analysis
Stakeholder needs

Imperfections in any RWD coupled with huge inter-source heterogeneity result in need for situation specific RWD solutions

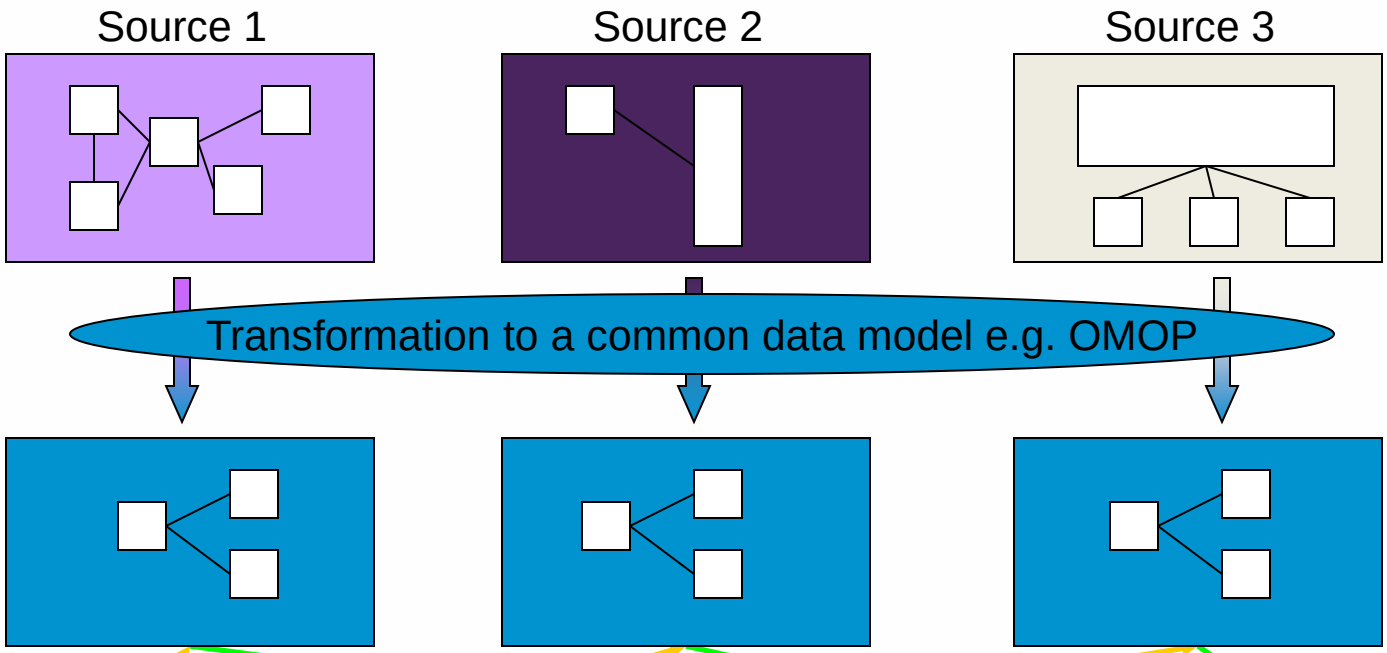
A ‘smorgasbord’ style data strategy

‘Three tiered’ data strategy



Secured appropriate efficient governance

Common data model role in distributed network- the OMOP model



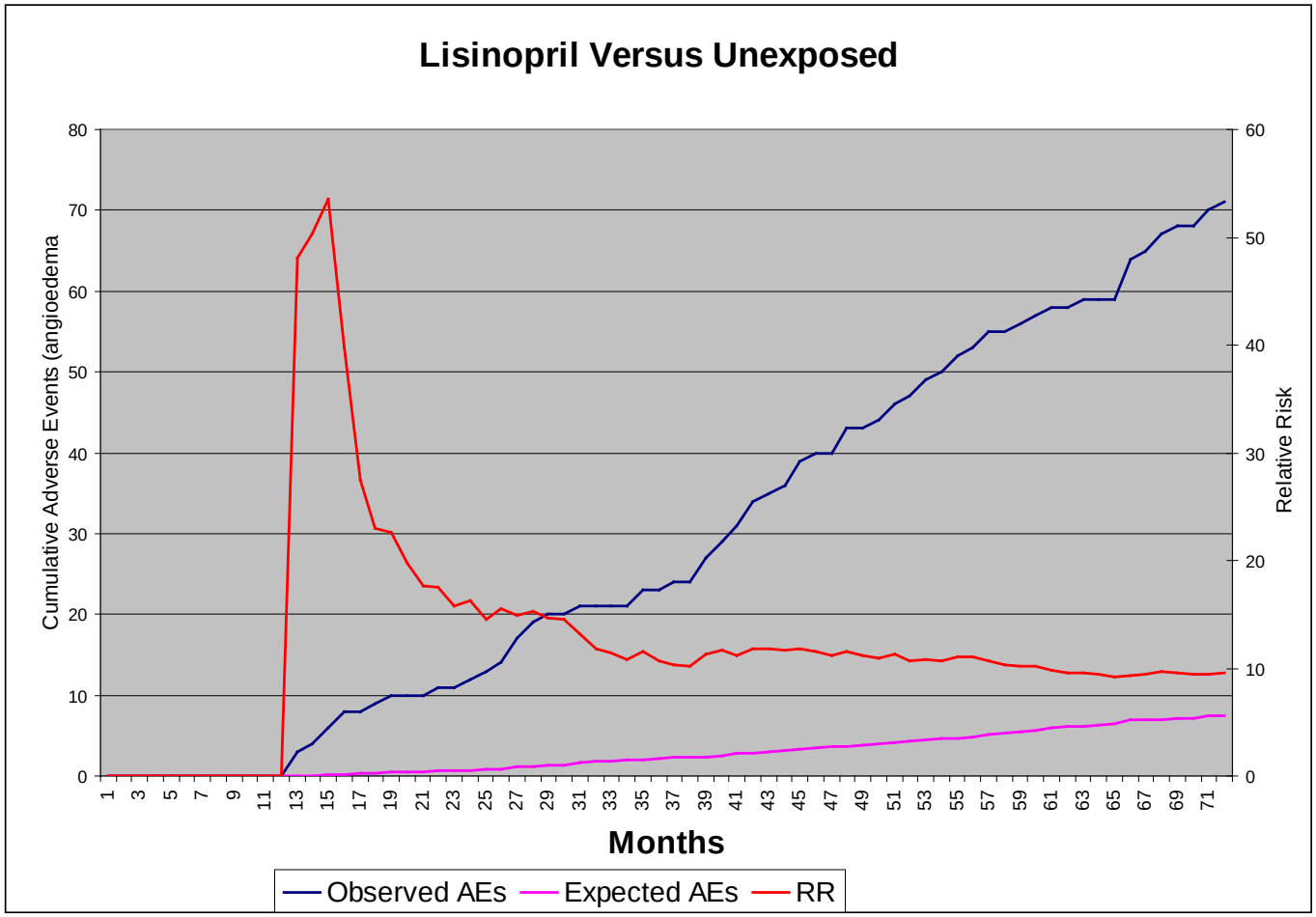
Analysis method

Use of a Common Data Model facilitates fast analysis of multiple databases, and allows analyses across a distributed network. Use of data converted to common denominator can be problematic

OMOP Analysis results

Diagram reference: OMOP

Distributed network analysis: Recording of angioedema for Lisinopril users compared to non-users: 2000-2005



Data from US Health Maintenance Organization research network

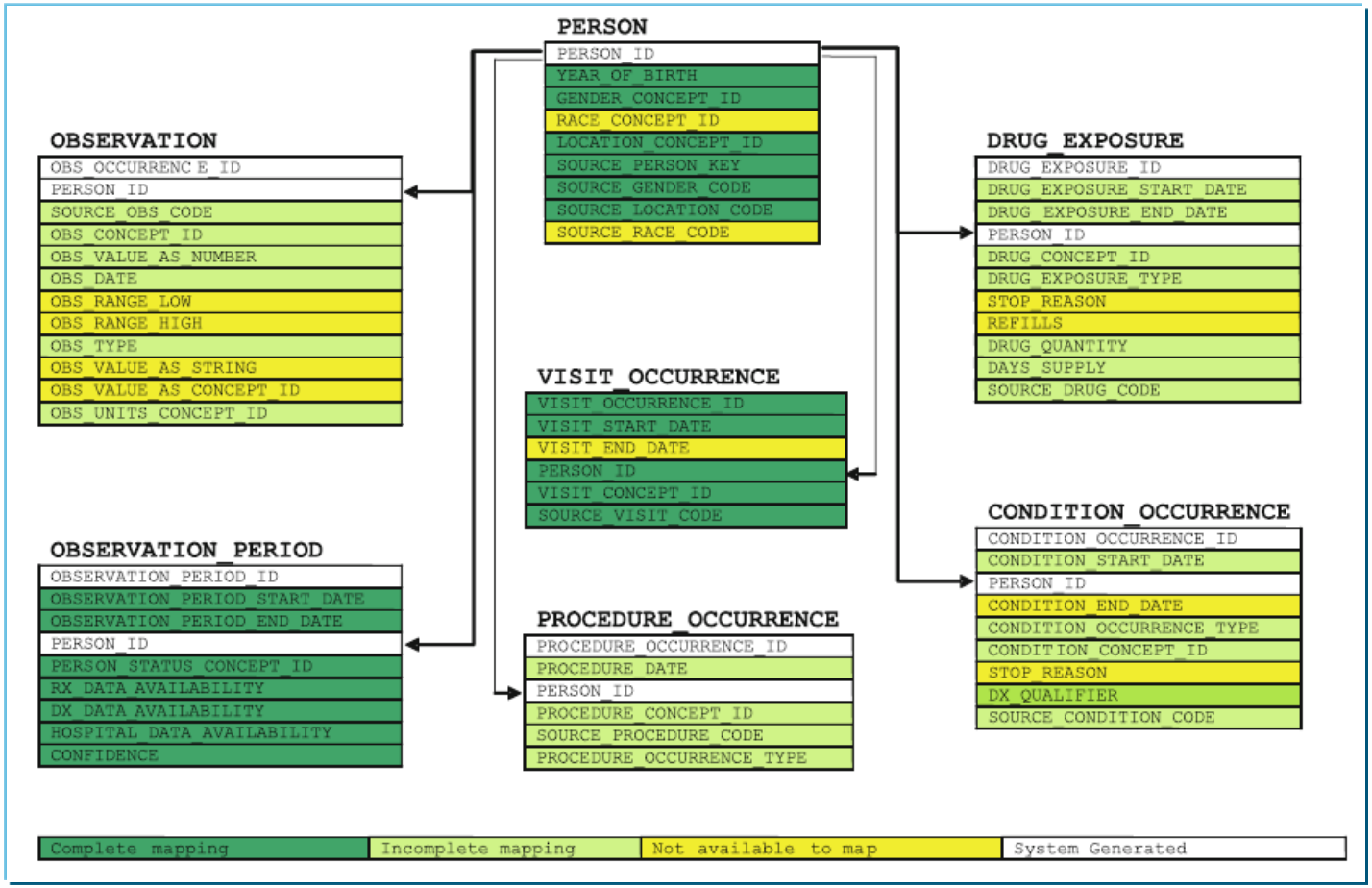
Unpublished data based on work in Brown *et al.*, (2007, 2009) in PDS). Contact: jeff_brown@hphc.org

Signal at month 13; 3 observed and 0.06 expected

Note: Base-case analysis. Outcome: Angioedema. Adjusted for age, sex, and health plan.

Database model heat map – showing goodness of fit of a THIN data conversion into OMOP CDM

Ref Zhou et al 2013



Pfizer WORLDWIDE SAFETY & REGULATORY
Worldwide Research & Development

Shows how well different variables convert into a Common Data Model

Innovation in Medical Development and Surveillance (IMEDS)

REAGAN-UDALL FOUNDATION FOR THE FOOD AND DRUG ADMINISTRATION

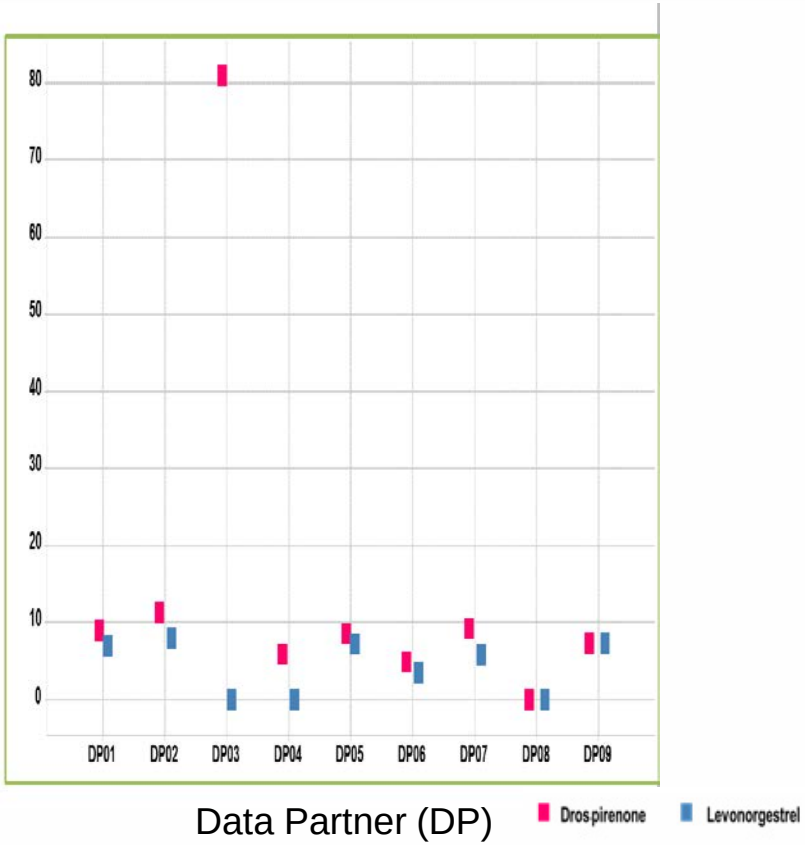
- IMEDS is a program within the Reagan-Udall Foundation for the US FDA and is a public private partnership created to build upon the significance progress made of research methodology by FDA's Sentinel Initiative and the Observational Medicines Outcomes Partnership (OMOP)
- Primary objective is to advance the science and tools necessary to support post-market evidence generation on regulated products, including safety surveillance and evaluations, to facilitate utilization of a robust electronic healthcare data platform for generating better evidence on regulated products in the post-market settings
- See: imeds.reaganudall.org

Pfizer – IMEDS Evaluation Pilot Project
Overview

- Pilot sponsored by Pfizer in two phases:
  - Phase I:Development of Policies and Procedures
  - Phase II: Evaluate IMEDS-Evaluation program by utilizing existing, publicly available Mini-Sentinel summary table programs and modular programs to conduct two demonstration cases
- An Industry First: Pfizer successfully ran two test queries through IMEDS/Sentinel distributed data network
  - Query 1: Evaluate drug – AE association (OCs – VTE)
  - Query 2: Assess effectiveness of a label change (PPIs)
- The IMEDS program is now open for other non-FDA use

IMEDS pilot results for OC VTE query – summary results and incidence rate by Data Partner

	4th Generation OCs	2nd Generation OCs
New Users	350,572	317,363
Dispensings	1,899,922	1,460,766
Days Supplied	62,180,487	63,102,751
Years at Risk	184,485.20	183,852.50
New Episodes w/ Events	158	121
Eligible Members	26,697,378	26,697,378
Member- Years	41768751.5	41852933.9
New Users /Eligible Members (Per 1000 members)	13.13	11.89
Days Supplied/ New User	177.37	198.83
Dispensings/ New User	5.42	4.6
Days Supplied/ Dispensing	32.73	43.2
New Episodes w/ Events /Years at Risk (Per 10000 Years)	8.56	6.58



Rates of VTE were greater for 4th generation than 2nd generation OCs, consistent with the literature

Limited variation across data partners, although some DPs had few events

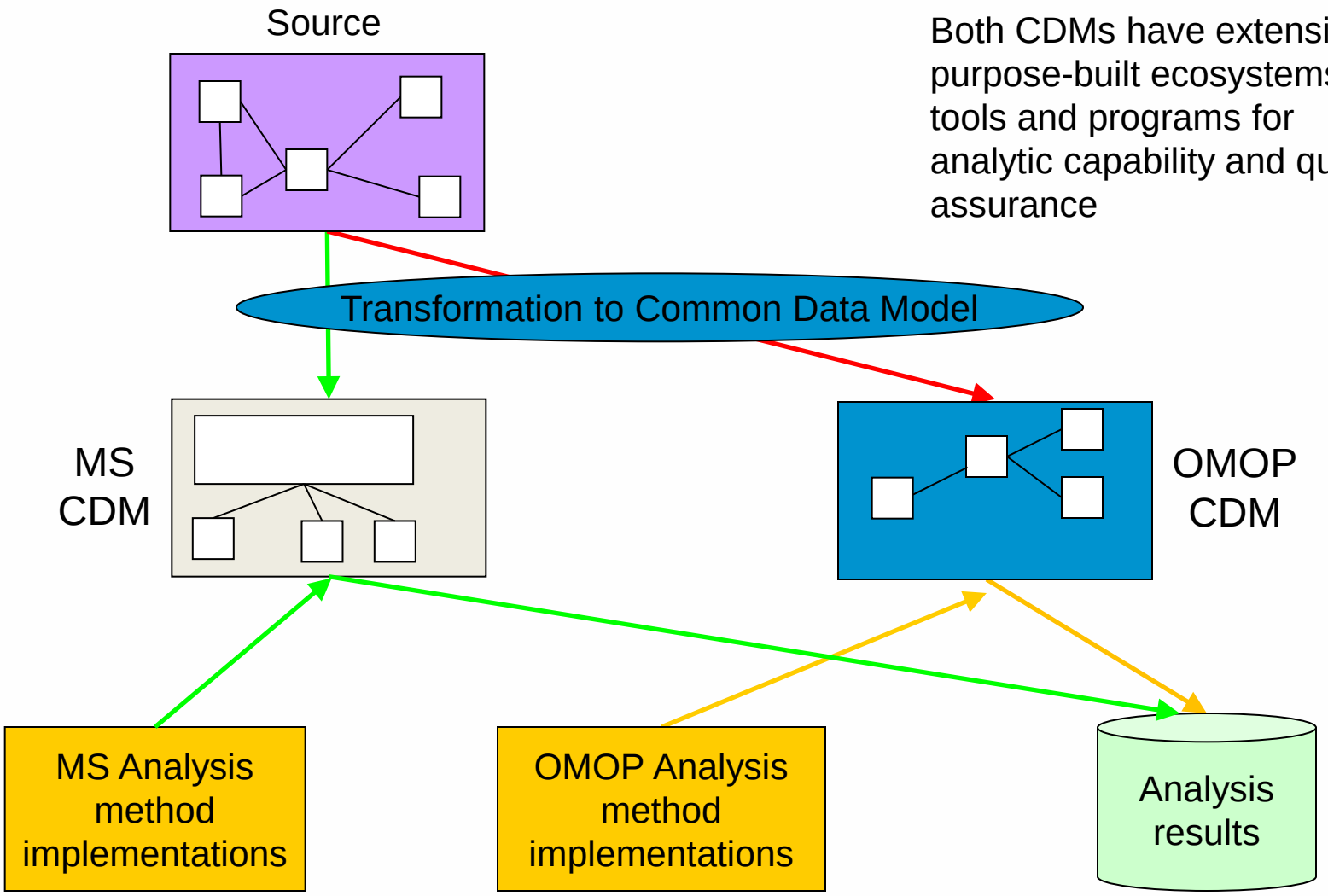
Limitations include: lack of confounding control, simple descriptive analysis techniques, outcomes were defined only by diagnosis codes

Conceivable Complications across CDMs

- A single validated CDM conversion per database release version is valuable for multiple database Pharmacoepidemiological analyses
- However in practice
  - The same database release version can be converted into
    - different Common Data Models
    - different versions of the same Common Data Model
    - the same version of the Common Data Model by different groups
  - Different analytic tools, or versions of Analytic tools may be used for analyses against the same CDM

- The above adds (sometimes unnecessary) complexity to reconciling discordance and concordance across different Pharmacoepidemiological studies and does not help support the credibility of the field
  - Harmonisation efforts and guiding principles in the use of CDM should look to enhance credibility and reproducibility of healthcare database analyses

Testing Two Common Data Models on the Same Data Source



Both CDMs have extensive purpose-built ecosystems of tools and programs for analytic capability and quality assurance

Objective

The Humana-Pfizer CDM project looked to evaluate OMOP and Mini-Sentinel CDMs from an ecosystem perspective to better understand how differences in CDMs and analytic tools affect usability and interpretation of results

- Both CDMs have extensive purpose-built ecosystems of tools and programs for analytics capability and quality assurance

Project Team and Expert Panel

Joint Research Project Team

Pfizer Team

- Xiaofeng Zhou, Kathy Liu, Jim Harnett, Andrew Bate

Humana Team

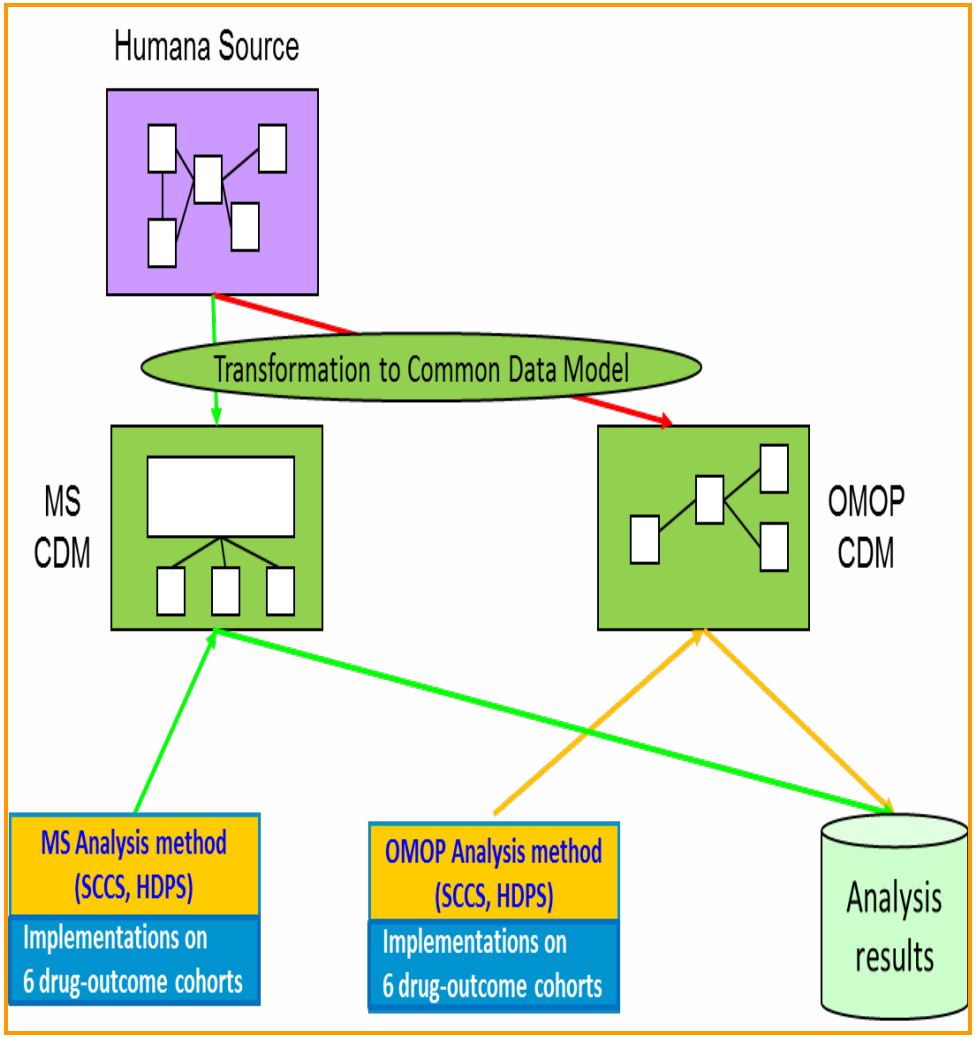
- Brandon Suehs, Yihua Xu, Keran Moll, Margaret Pasquale, Vinit Nair

Expert Panel Members

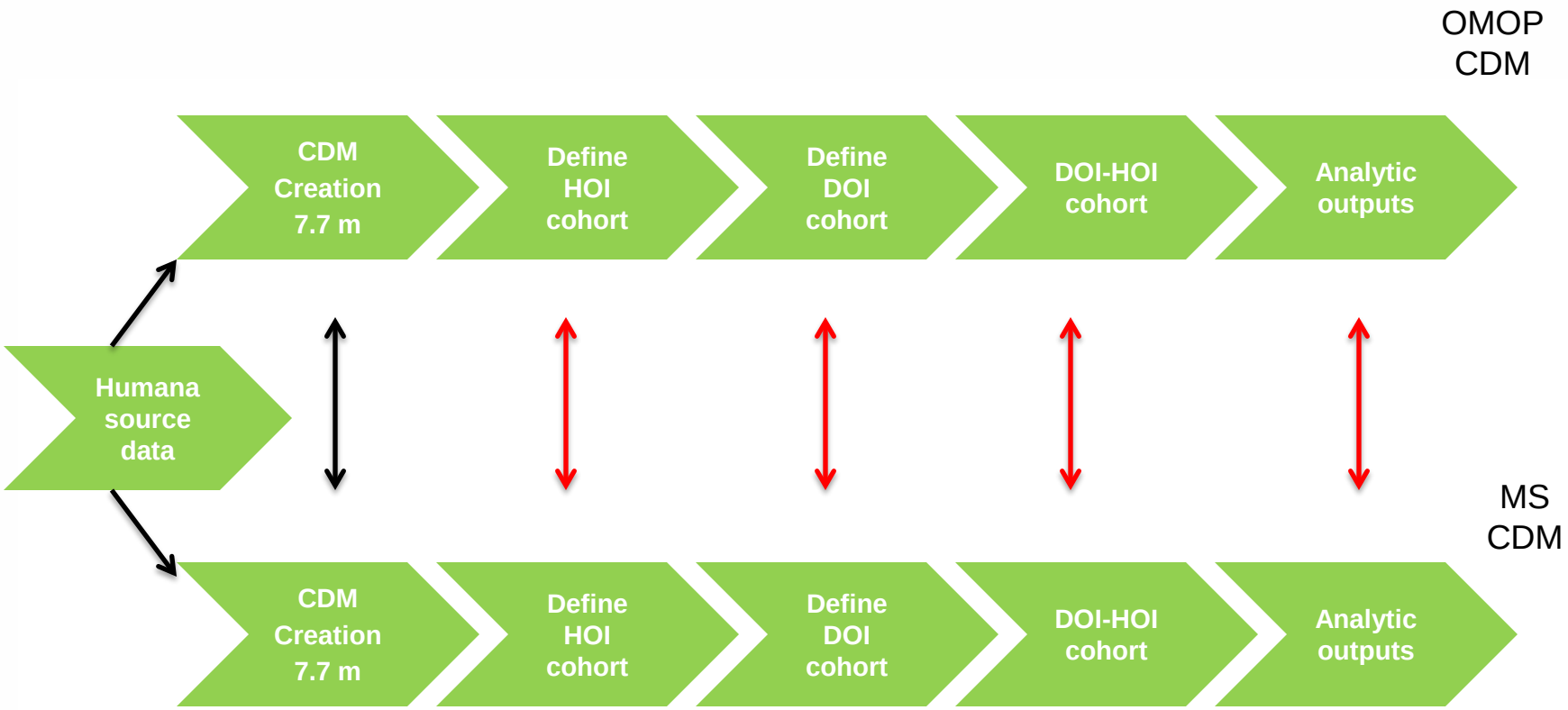
- **Abraham Hartzema**, PharmD, PhD, Professor and Eminent Scholar, University Florida
- **Michael Kahn,** MD, PhD, Professor of Pediatrics, University of Colorado
- **Yola Moride**, PhD, MSc, Professor, Faculty of Pharmacy, University of De Montreal
- **Brian Sauer** PhD, MS, Research Assistant Professor, University of Utah

Method

- Data Source: Humana claims data (2007 -2012)
- Data Mapping: Humana data to OMOP and MS CDMs
- Exposure and Outcome: six established positive drug-outcome pairs
- Analytic Methods:
  - High-dimensional propensity score (HDPS) based analytic procedure
  - Univariate self-controlled case series (SCCS) method
- Comparison:
  - Data at the patient level by source code and mapped concepts
  - Study cohort construction and effect estimates using two analytic methods



Results: Differences in the Key Steps of the Dissection



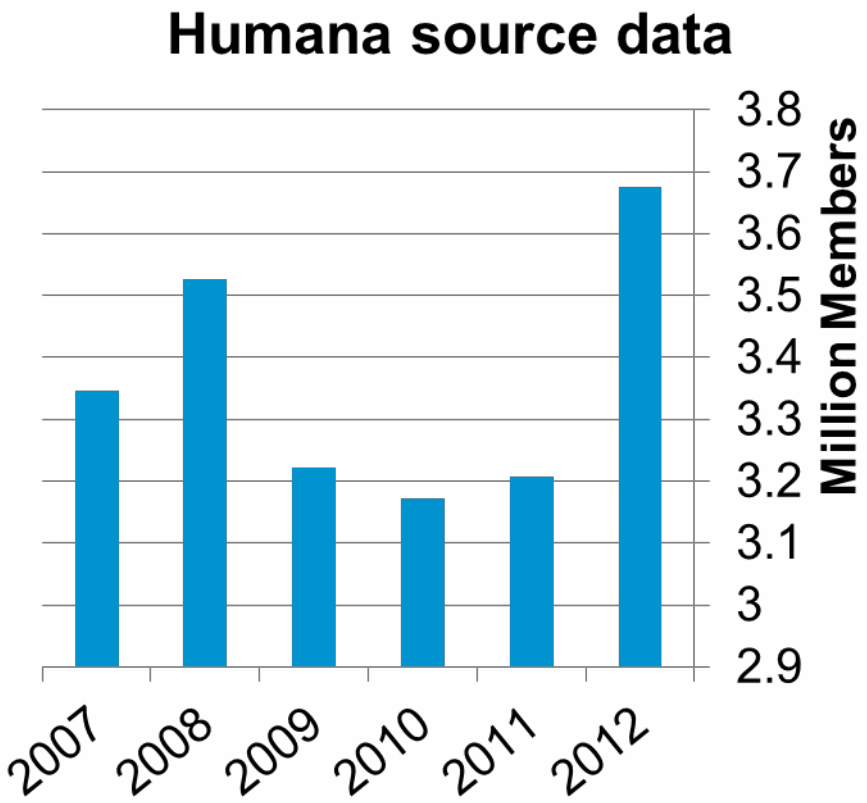
DOI – Drug of Interest
HOI – Health Outcome of Interest
CDM –Common Data Model

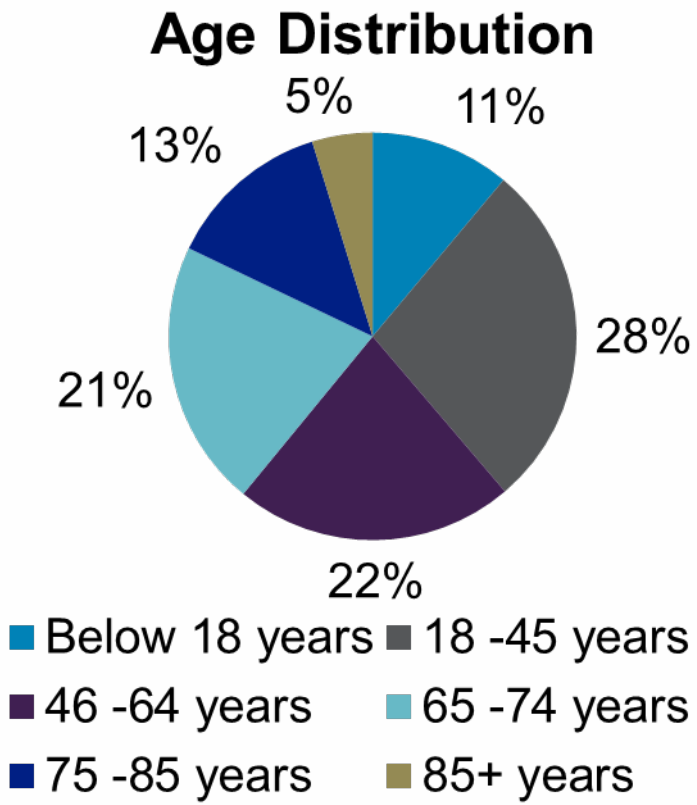
Steps where further discordance was introduced

Step with no or minimal discordance

CDM Construction

Extract, Transform, Load (ETL) of Humana data for 7.6M unique individuals into both OMOP and Mini-Sentinel CDMs

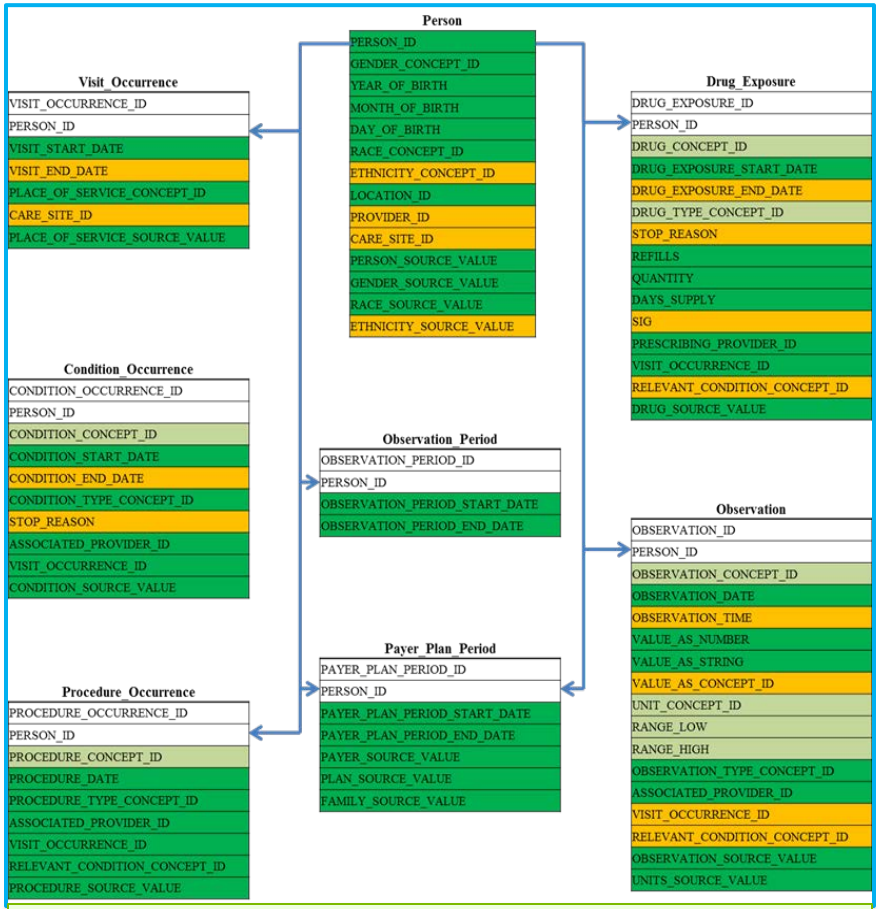




Results: Conceptual Differences in Mapping

- No information loss when mapping source codes into MS CDM
- There was minimal information loss when source data were transformed into OMOP standard vocabulary
- Most unmapped codes in this study had no or minimal impact on the active surveillance method testing.

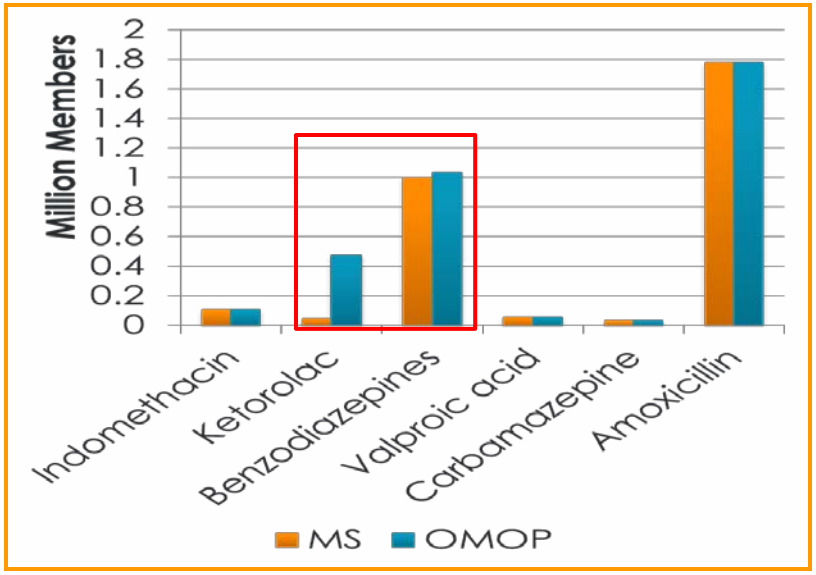
Database heat map: overall mapping quality of the Humana database in OMOP CDM

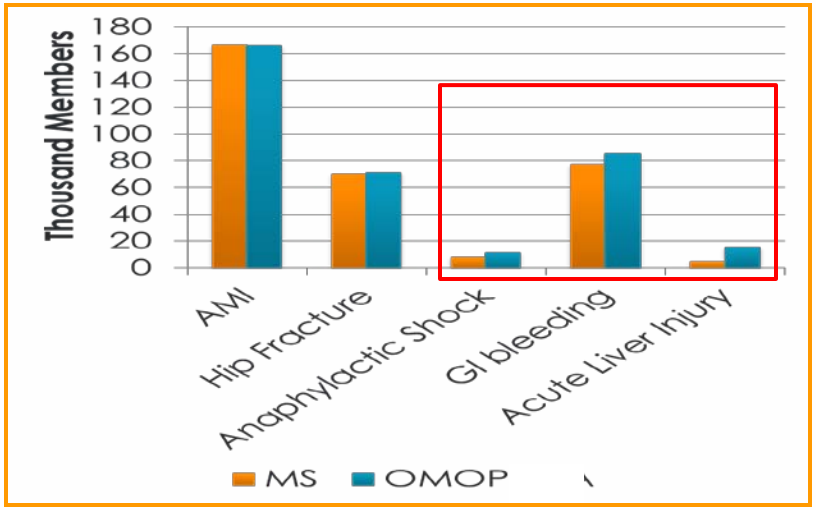


Dark green, complete mapping; light green, incomplete mapping; yellow, not available to map; white, system generated.

Results: Conceptual Differences in Cohort Creation

- Large differences in two DOI and three HOI cohorts extracted from each CDM
- Drug exposure table structure and method to identify cohorts differ across two CDMs



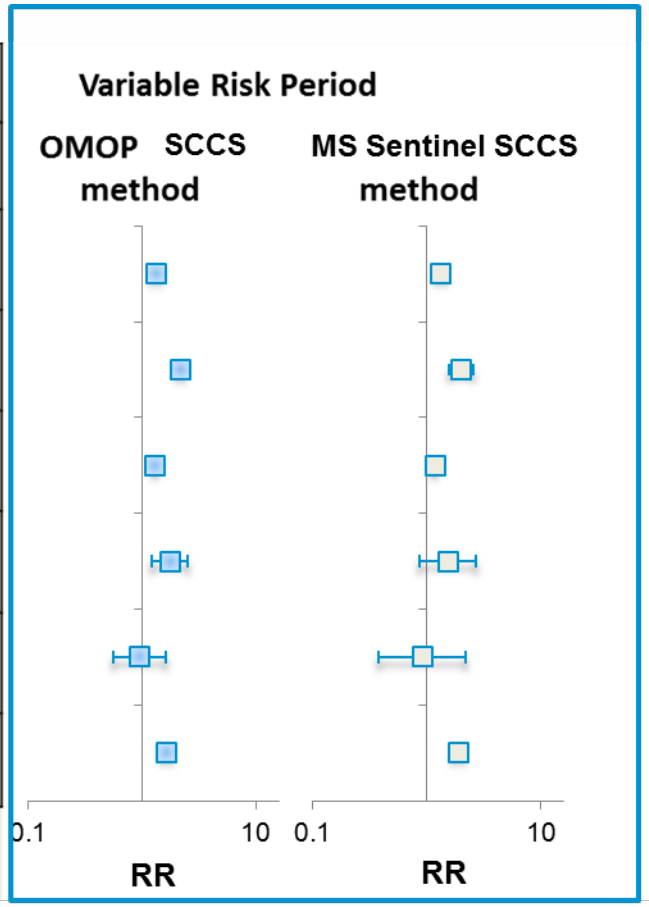


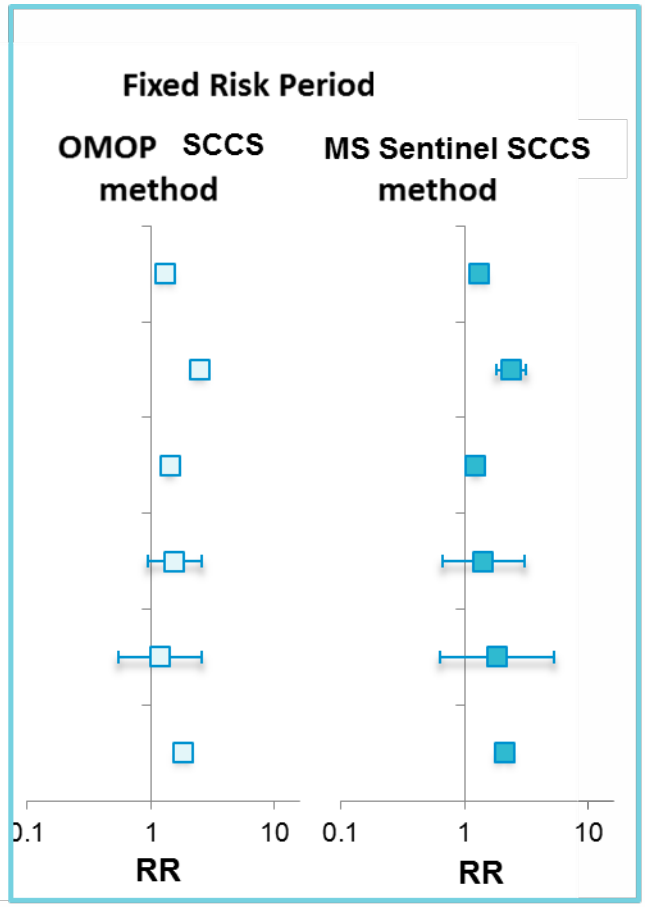
Xu Y, Zhou X, Suehs BT, Hartzema AG, Kahn MG, Moride Y, Sauer BC, Liu Q, Moll K, Pasquale, MK, Nair VP, Bate A, "A comparative assessment of Observational Medical Outcomes Partnership and Mini-Sentinel common data models and analytics: implications for active drug safety surveillance", ***Drug Saf* 2015** 38(8), 749-765.

Results: Self Controlled Case Series (SCCS) Method Testing

Key Finding: Conceptual differences at data model level had slight but not significant Impact on identifying the known safety associations

DOI	**HOI**
Indomethacin	AMI
Ketorolac	GI bleeding
Benzodiazepines	Hip Fracture
Valproic acid	Acute Liver Injury
Carbamazepine	Acute Liver Injury
Amoxicillin	Anaphylactic Shock

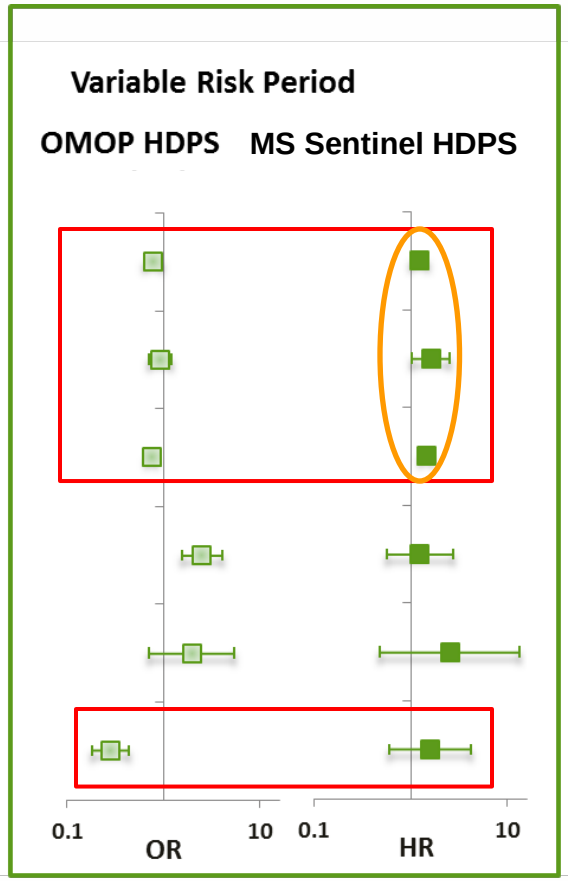


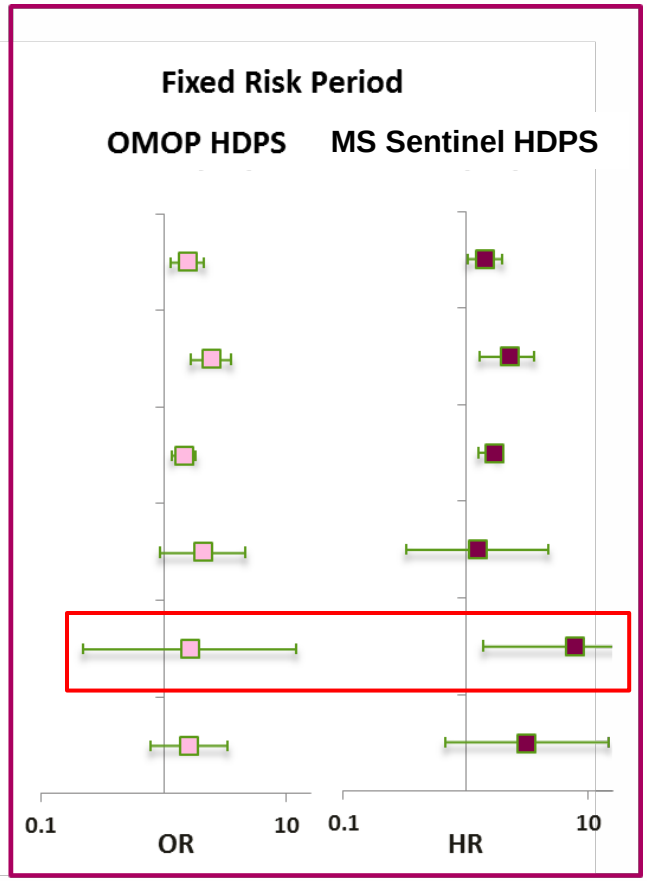


Results: High Dimensional Propensity Score (HDPS) Based Analytic Procedure testing

Key Finding: Differences at ecosystem level can lead to strikingly different risk estimation (primarily due to choice of analytic approach and its implementation)

DOI	HOI
Indomethacin	AMI
Ketorolac	GI bleeding
Benzodiazepines	Hip Fracture
Valproic acid	Acute Liver Injury
Carbamazepine	Acute Liver Injury
Amoxicillin	Anaphylactic Shock





Results: Contrast across Two CDM Ecosystems

	MS CDM	OMOP CDM
CDM conversion	Simple (No standard vocabulary)	Complex (using standard vocabulary)
Unmapped codes	No	Yes (minimal impact in this study)
Data aggregation table	No (embedded in analytic program)	Yes
Drug exposure table including procedure drug codes	No (only medication collected from outpatient pharmacy claims)	Yes
Validation tools	Complex	Complex
HOI/DOI identification	Simple (using source codes)	Complex (using concept ID and relational hierarchy data table)
Analytic procedure (HDPS)	Cox proportional regression model (Account for time to event)	Logistic regression model (Not account for time to event)
Computational Efficiency (HDPS)	More computational intensive	Less computational intensive

*SCCS was not listed as we applied an identical SCCS across both CDMs, thus the performance was comparable across both CDMs.

Some Study Limitations

- The source data is administrative (billing) data from one health plan
- Only six drug-outcome pairs were tested to assess the performance of the two active surveillance methods
- Some drug-outcome pairs were underpowered in the Humana database
- Comparator drugs for this study were chosen from established negative control references.
- We applied published health outcome definitions that only used diagnosis codes.
- A custom SCCS method was applied on both OMOP and Mini-Sentinel CDMs

Summary of CDM comparison

- Strikingly different risk estimation can occur at an ecosystem level, and in our CDM comparison study primarily attributed to the choices of analytic approach and their implementation in the community developed analytic tools.
- The clear conceptual differences between OMOP and Mini-Sentinel CDMs had limited impact on identifying known safety associations in Humana data at the data model level.

Some recommendations from our study

- Transparency needs to be excellent both intra and extra CDM based networks
- No ‘one size fits all’ solutions for CDM based analyses- will always need data outside CDM on occasion
- Cannot consider a CDM in isolation need to consider also accompanying tools and versioning over time
- Need a trusted CDM based infrastructure for ongoing use of value and credibility in evidence generation
  - Include ready replicability to the maximum extent possible outside the network
- Data vendors need to support and understand the importance of doing and validating a CDM conversion, and conduct continual improvement to ensure sustainable routine use as healthcare and database systems change.
- There should be one single instance of a data vendor validated CDM version per database cut

The Hitchhikers guide to the galaxy's 'answer' to healthcare database analysis is…

42

The carefully and appropriately constructed question is what can be difficult to determine… We need to make sure that CDM based analyses help us to get the right answers to the right questions

References

- Bate A, Sobel RE, Marshall J, Daniel G, McCall T, Reynolds RF and Brown J. S531 Oral Contraceptives and VTE Across the Sentinel Data Network – An IMEDS Evaluation Pilot Assessment. Pharmacoepidemiology and Drug Safety 25: S531 915 Suppl 3 AUG 2016
- Bourke A, Bate A, Sauer BC, Brown JS, Hall GC. Evidence generation from healthcare databases: recommendations for managing change. Pharmacoepidemiology and Drug Safety. 2016 25(7): 749-754
- Brown JS et al. 2009. Early adverse drug event signal detection within population-based health networks using sequential methods: key methodologic considerations. PDS18(3): 226-234.
- Sobel RE, Bate A, Marshall J, Daniel G, McCall T, Brown J, and Reynolds RF. Risk Minimization Evaluation in a Distributed Data Network – An IMEDS Evaluation Pilot Assessment of the 2010 Class Label Change for Proton Pump Inhibitors Pharmacoepidemiology and Drug Safety 25: S6 7 Suppl 3 AUG 2016
- Xu Y et al. A Comparative Assessment of Observational Medical Outcomes Partnership and Mini- Sentinel Common Data Models and Analytics: Implications for Active Drug Safety Surveillance. Drug Safety. 2015 38(8), 749-765.
- Zhou X et al 2013 An Evaluation of the THIN Database in OMOP Common Data Model for Active Drug Safety Surveillance. Drug Safety. 36(2): 119-134

Conclusions

- Facilitating data access and ability to conduct a multiple RWD sources in Europe is very important
  - CDMs can play an important role
- Trust in CDM outputs for drug safety experts who do not have interest/expertise in CDMs is critical
- There is a need for continual efforts in ensuring sustainable, reliable and transparent platforms for maintaining for using and further develop CDMs and their associated tools for effective safety surveillance.
  - Sustainability would seem to be even more challenge than a one-off conversion
- In a world of multiple database networks, linkage between the networks is valuable and can facilitate credible healthcare database analyses.