Materials we can discuss today

- Common Data Model
- Vocabulary
- Analysis use cases
- ETL tips/tricks

OMOP Common Data Model

Patrick Ryan, PhD
Janssen Research and Development
on behalf of OHDSI team
25 March 2014

Observational Medical Outcomes Partnership

OMOP was a public-private partnership established to inform the appropriate use of observational healthcare databases for studying the effects of medical products:

- Conducting methodological research to empirically evaluate the performance of alternative methods on their ability to identify true associations
- Developing tools and capabilities for transforming, characterizing, and analyzing disparate data sources across the health care delivery spectrum
- Establishing a shared resource so that the broader research community can collaboratively advance the science

http://omop.org

Introducing OHDSI

- The Observational Health Data Sciences and Informatics (OHDSI) program is a multi-stakeholder, interdisciplinary collaborative to create open-source solutions that bring out the value of observational health data through large-scale analytics
- OHDSI has established an international network of researchers and observational health databases with a central coordinating center housed at Columbia University

OHDSI's mission

To transform medical decision making by creating reliable scientific evidence about disease natural history, healthcare delivery, and the effects of medical interventions through large-scale analysis of observational health databases for population-level estimation and patient-level predictions

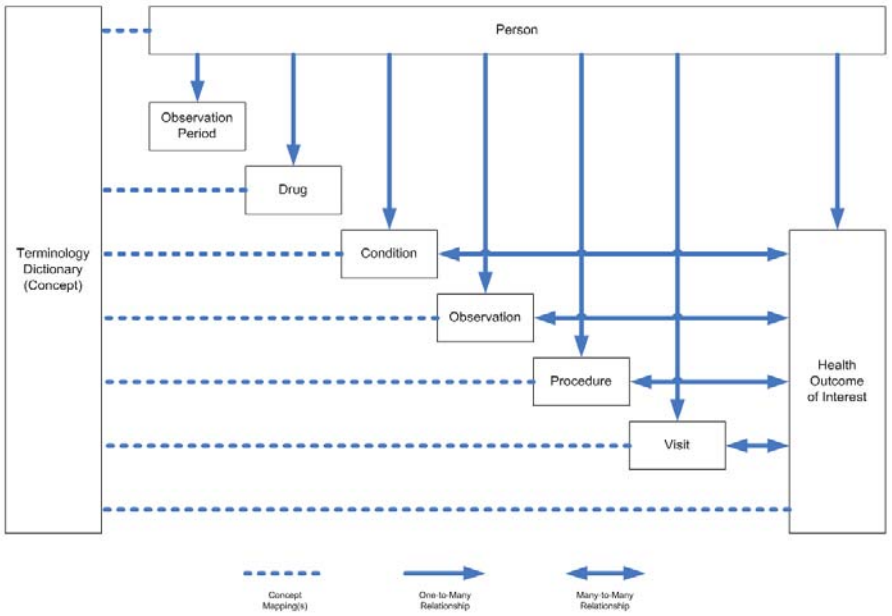
http://ohdsi.org

Objectives in OMOP Common Data Model development

- One model to accommodate both administrative claims and electronic health records
  - Claims from private and public payers, and captured at point-of-care
  - EHRs from both inpatient and outpatient settings
  - Also used to support registries and longitudinal surveys
- One model to support collaborative research across data sources both within and outside of US
- One model that can be manageable for data owners and useful for data users (efficient to put data IN and get data OUT)
- Enable standardization of structure, content, and analytics focused on specific use cases

Evolution of the OMOP Common data model

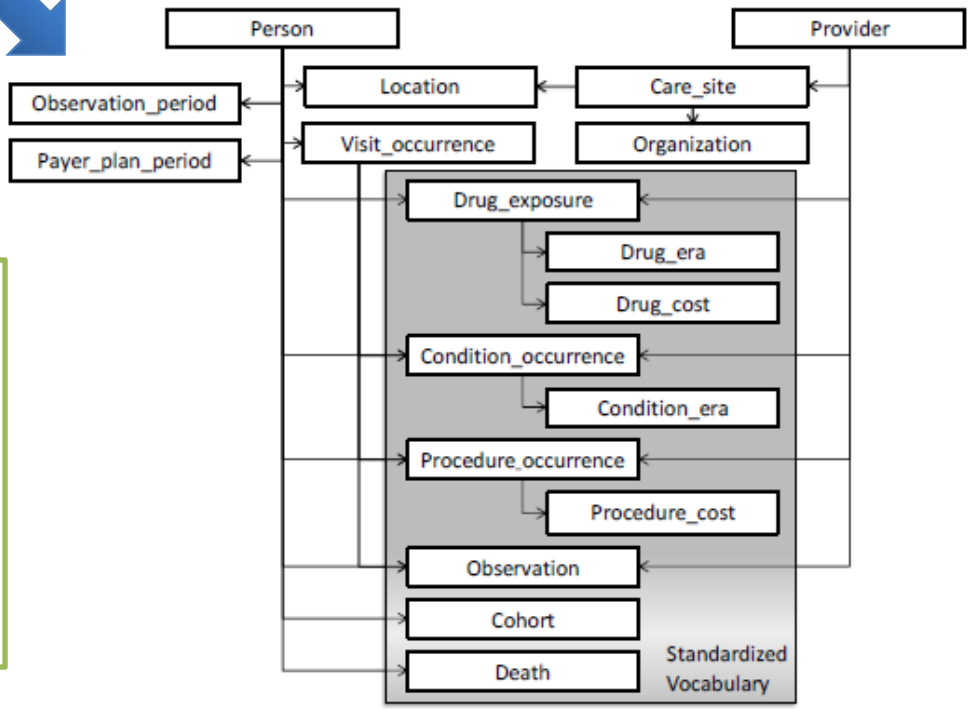
OMOP CDMv2



Drug Safety Surveillance

CER/PCOR

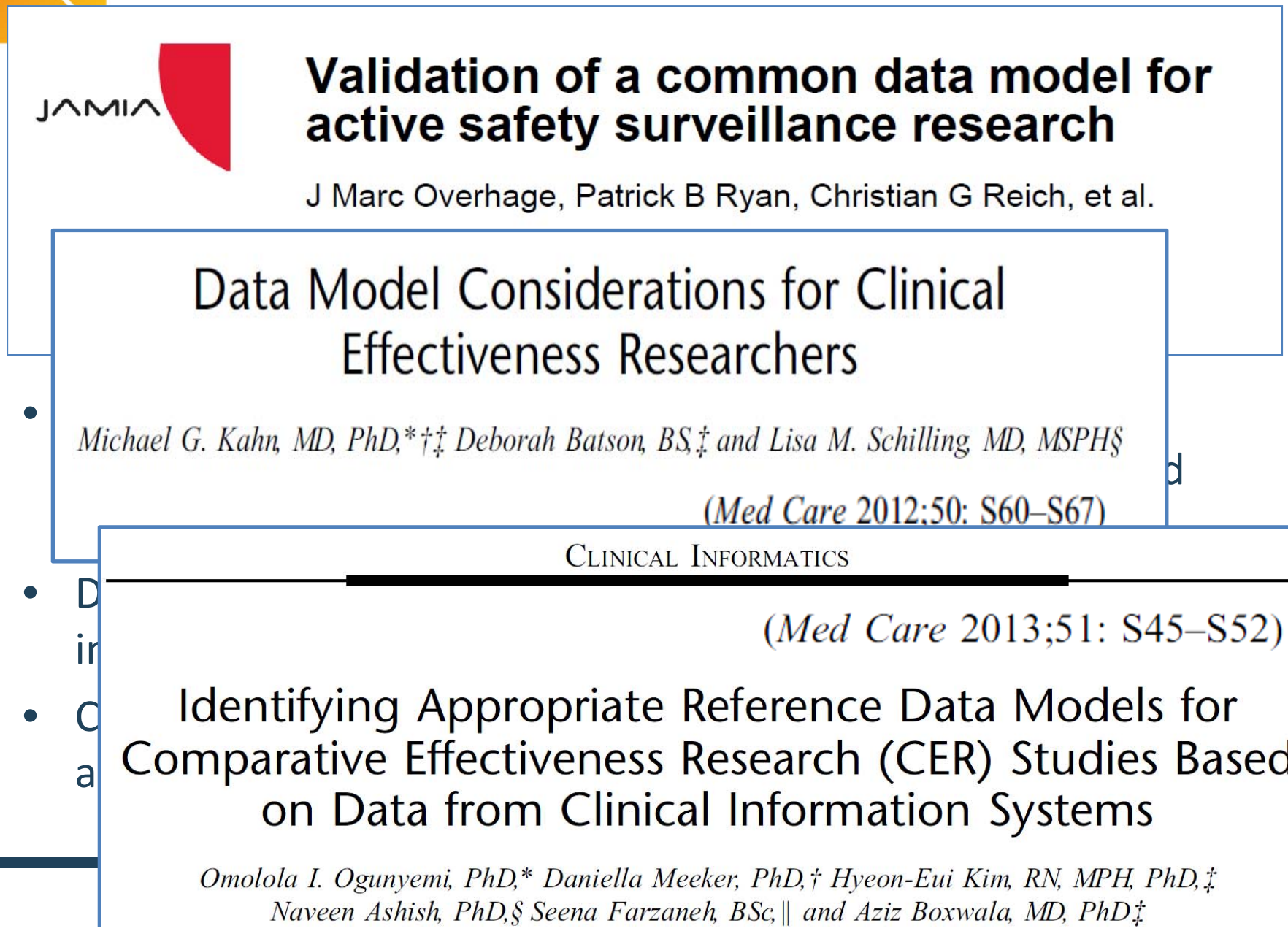
OMOP CDMv4



OMOP CDM now Version 4, following multiple iterations of implementation, testing, modifications, and expansion based on the experiences of the OMOP community who bring on a growing landscape of research use cases.

http://omop.org/CDM

Lessons learned about data model



JAMIA

Validation of a common data model for active safety surveillance research

J Marc Overhage, Patrick B Ryan, Christian G Reich, et al.

Data Model Considerations for Clinical Effectiveness Researchers

Michael G. Kahn, MD, PhD,†‡ Deborah Batson, BS,‡ and Lisa M. Schilling, MD, MSPH§*

(*Med Care* 2012;50: S60–S67)

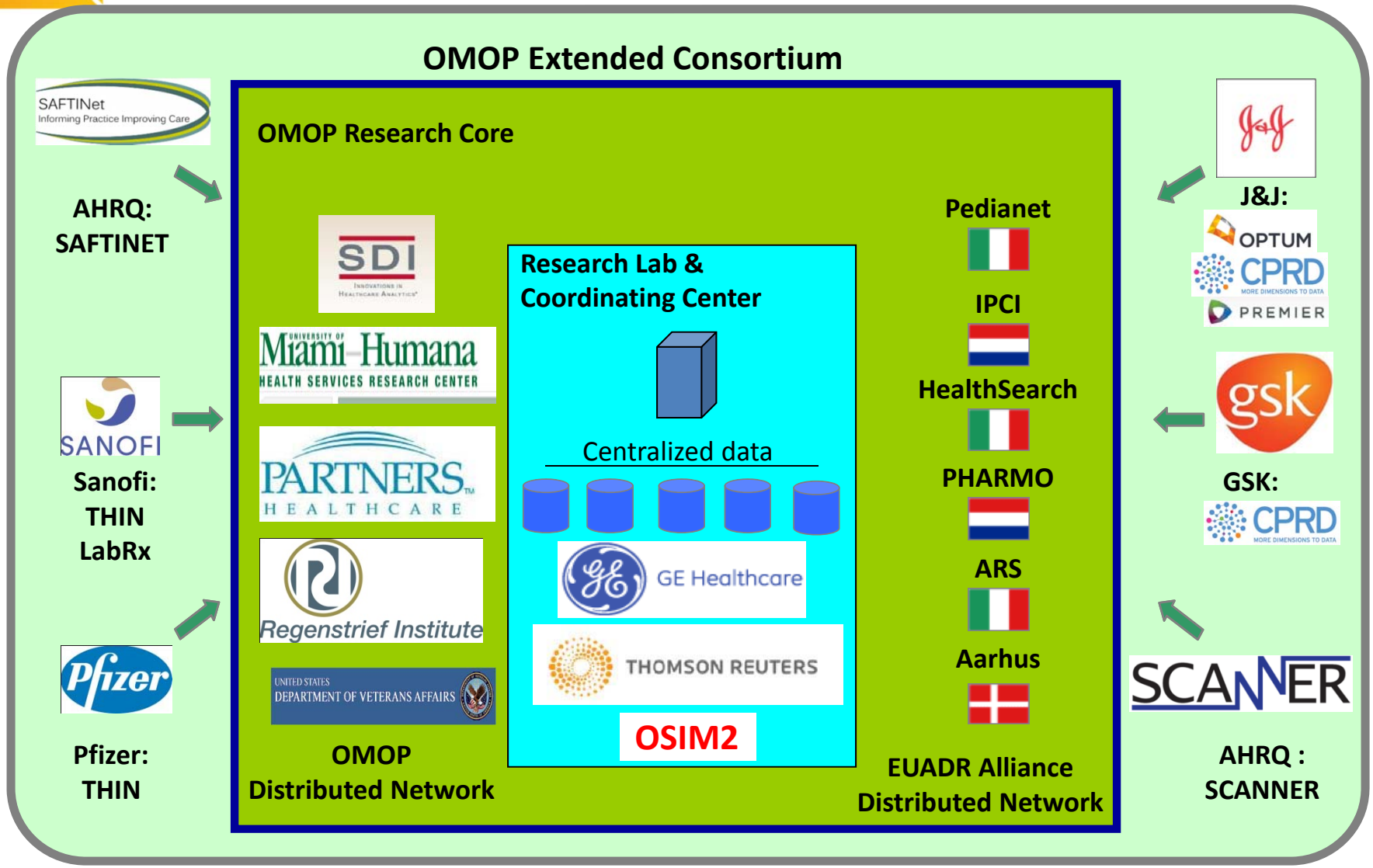
CLINICAL INFORMATICS

(*Med Care* 2013;51: S45–S52)

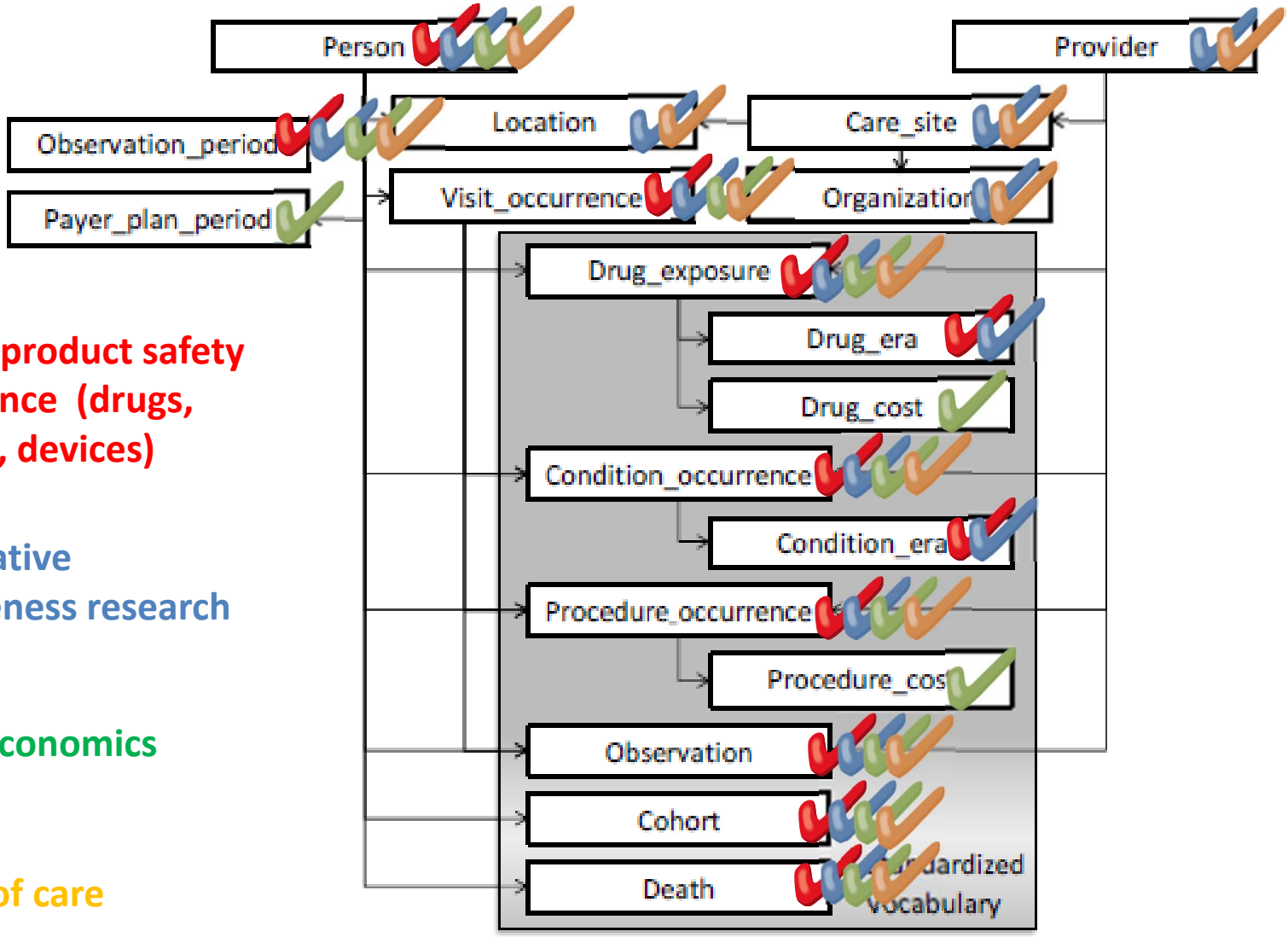
Identifying Appropriate Reference Data Models for Comparative Effectiveness Research (CER) Studies Based on Data from Clinical Information Systems

Omolola I. Ogunyemi, PhD, Daniella Meeker, PhD,† Hyeon-Eui Kim, RN, MPH, PhD,‡ Naveen Ashish, PhD,§ Seena Farzaneh, BSc,‖ and Aziz Boxwala, MD, PhD‡*

OMOP Data Community continues to evolve…



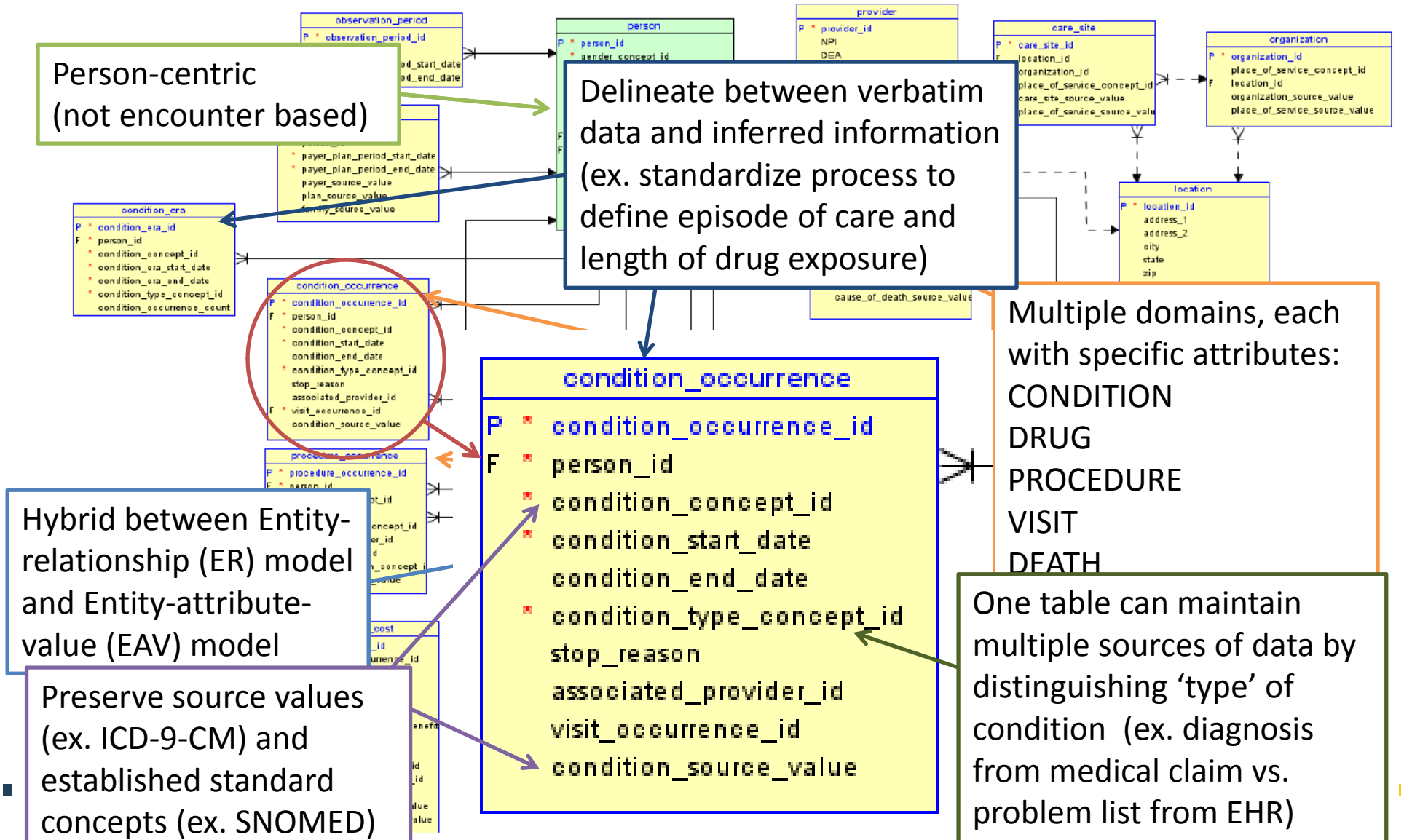
One model, multiple use cases



- **Medical product safety surveillance (drugs, vaccines, devices)**
- **Comparative effectiveness research**
- **Health economics**
- **Quality of care**

http://omop.org/CDM

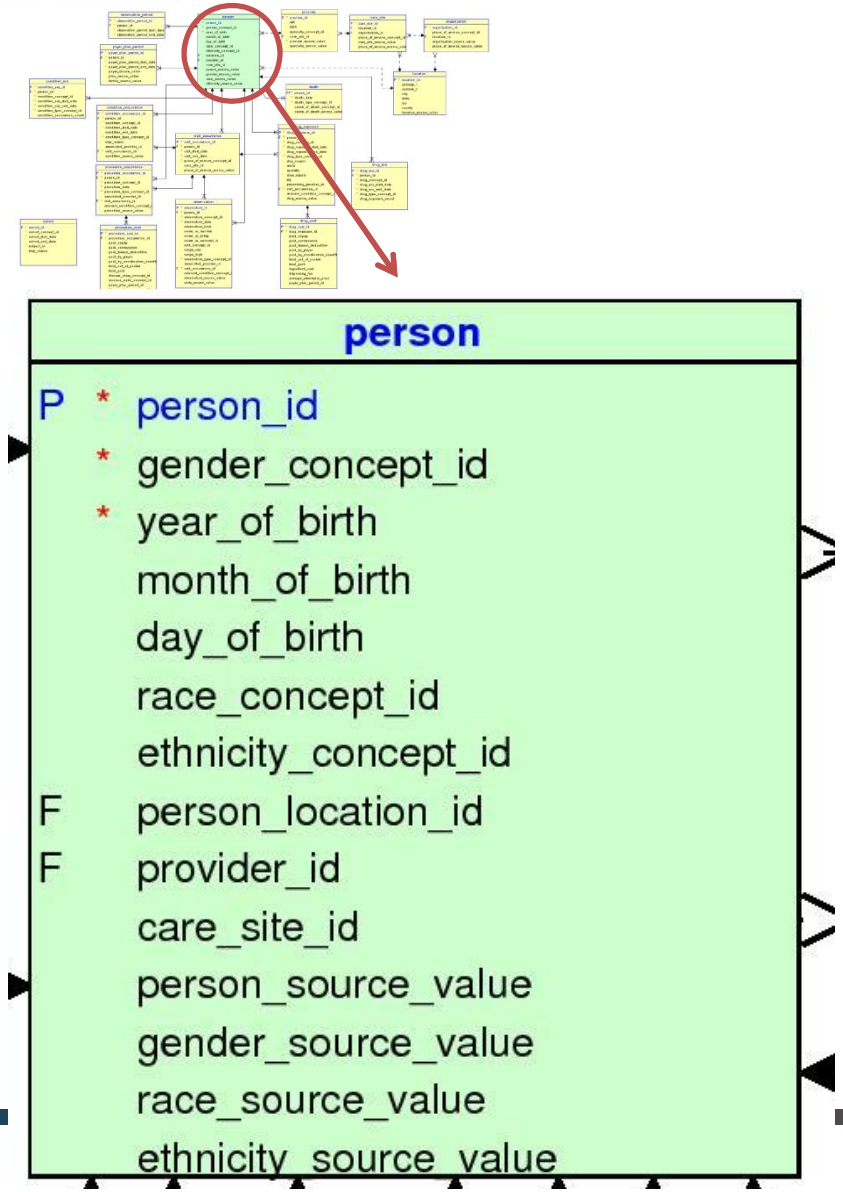
Key concepts within OMOP CDM v4



Standard Variable Name Conventions across CDMv4

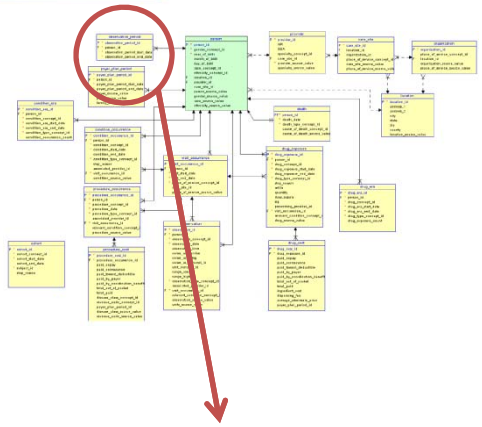
- <entity>_CONCEPT_ID:
  - Foreign key into the Standard Vocabulary, which serves as the primary basis for all standardized analytics
  - Ex: CONDITION_CONCEPT_ID = 31967 contains reference value for SNOMED concept of 'Nausea'
- <entity>_SOURCE_VALUE:
  - Verbatim information from the source data, typically used in ETL to map to CONCEPT_ID, and not to be used by any standard analytics
  - Ex: CONDITION_SOURCE_VALUE = '787.02' was the ICD-9 code captured as a diagnosis from the administrative claim
- <entity>_TYPE_CONCEPT_ID:
  - Delineates the origin of the source information, standardized within the Vocabulary
  - Ex: DRUG_TYPE_CONCEPT_ID can allow analysts to discriminate between 'Pharmacy dispensing' and 'Prescription written'
- <entity>_ID:
  - Unique identifiers for key entities, which can serve as foreign keys to establish relationships across entities
  - Ex: PERSON_ID uniquely identifies each individual. VISIT_OCCURRENCE_ID uniquely identifies a PERSON encounter at a point of care.

Person



- Uniquely identifies all individuals with person-level information in all entities
- Captures standardized demographics
- Allows maintenance of current location and primary provider and care site
- PERSON_ID is a foreign key for all person-level information

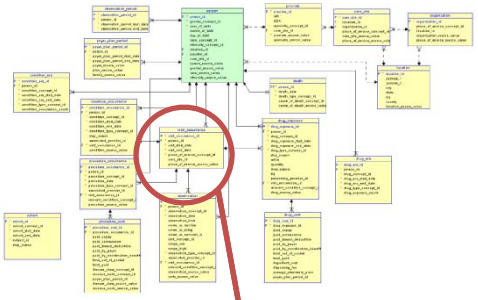
Observation period

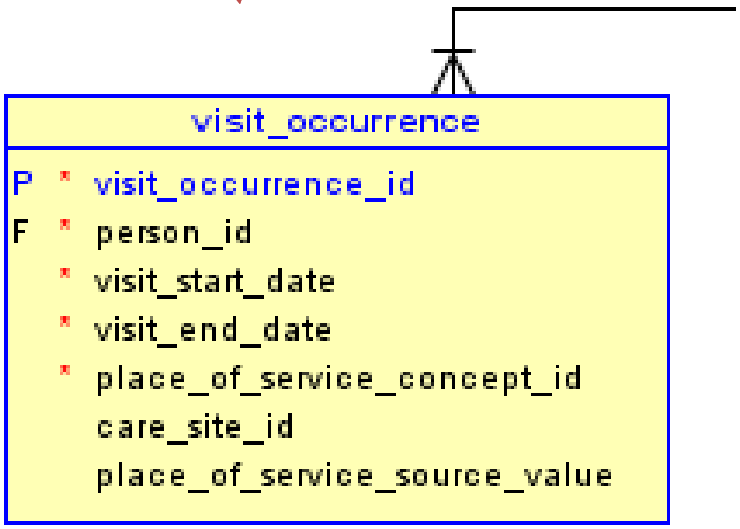


observation_period	
P	observation_period_id
F	person_id
	observation_period_start_date
	observation_period_end_date

- Spans of time where data source has capture of observational data
- One person may have multiple periods if there is interruption in data capture
- For claims databases, this is analogous to ‘continuous enrollment’ during time with both medical and pharmacy benefits
- For EHR databases with no defined capture, rules may vary

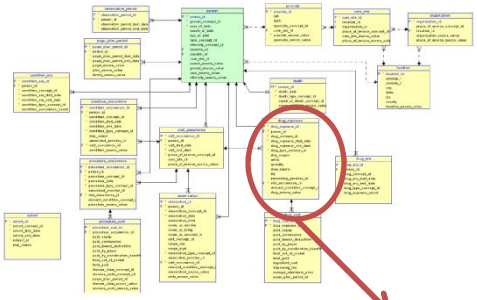
Visit occurrence

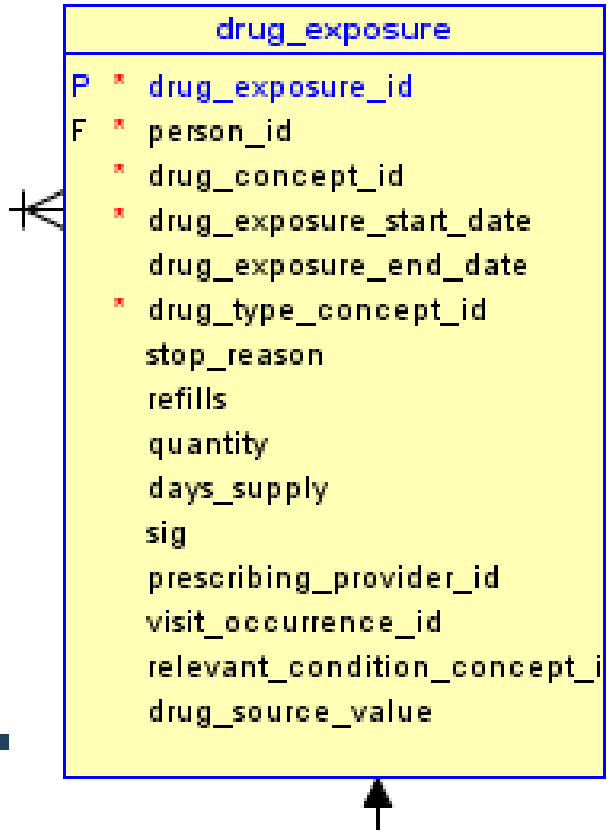




- Unique identifier for each encounter, can be used as foreign key for observations recorded during each visit
- Main classification for place-of-service:
  - Outpatient
  - Inpatient
  - Emergency room
- Outpatient and ER visits are typically completed on the same day (so VISIT_START_DATE = VISIT_END_DATE)
- Inpatient visits are typically admissions/discharges such that VISIT_END_DATE – VISIT_START_DATE = Length of stay

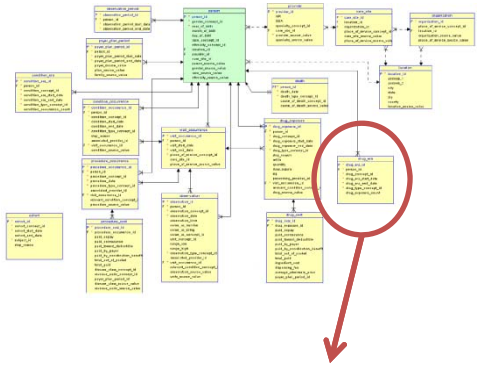
Drug exposure





- Record for each exposure, with associated verbatim information as available
- In claims data, a drug exposure row for each pharmacy dispensing claim AND a row for each procedure on medical claim which is drug administration
- No inferred information in DRUG_EXPOSURE → we will develop standardized approach to derive END_DATE in DRUG_ERA

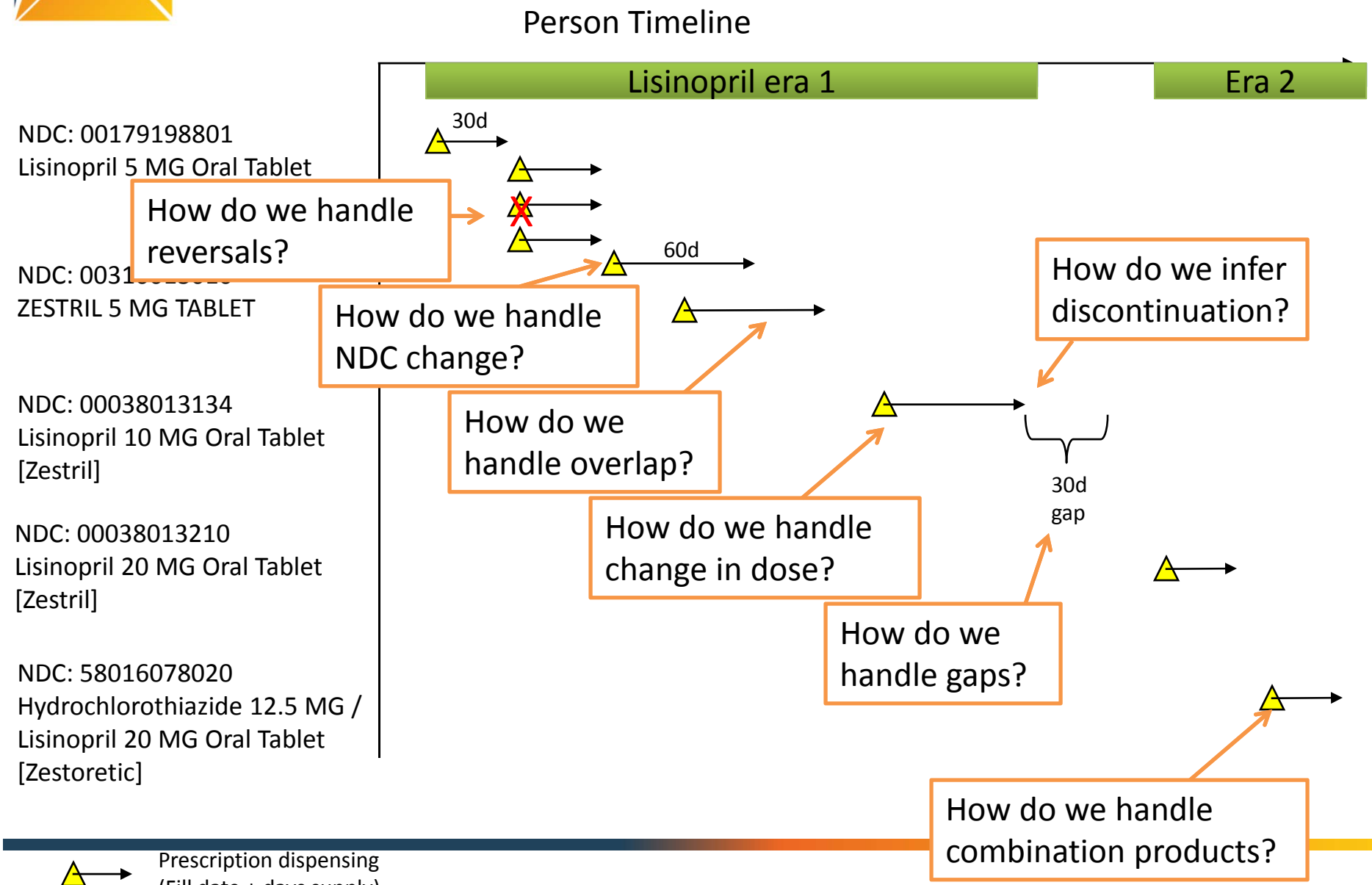
Drug era



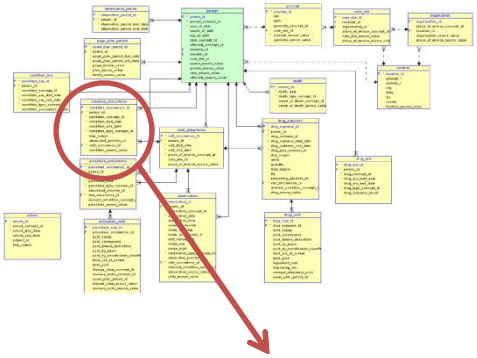
drug_era	
P	drug_era_id
F	person_id
	drug_concept_id
	drug_era_start_date
	drug_era_end_date
	drug_type_concept_id
	drug_exposure_count

- Span of time a person is assumed to be continuously exposed to a specific ingredient
- Fully derived from DRUG_EXPOSURE table, no source data needed
- Drug records aggregated to RxNorm ingredient level
- Standard analytics applied to enforce consistent assumptions around gaps between prescriptions

Illustrating inferences needed within longitudinal pharmacy claims data for one patient



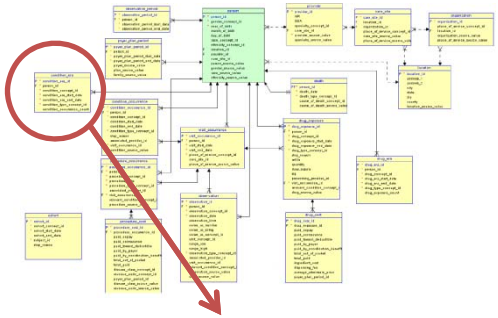
Condition occurrence

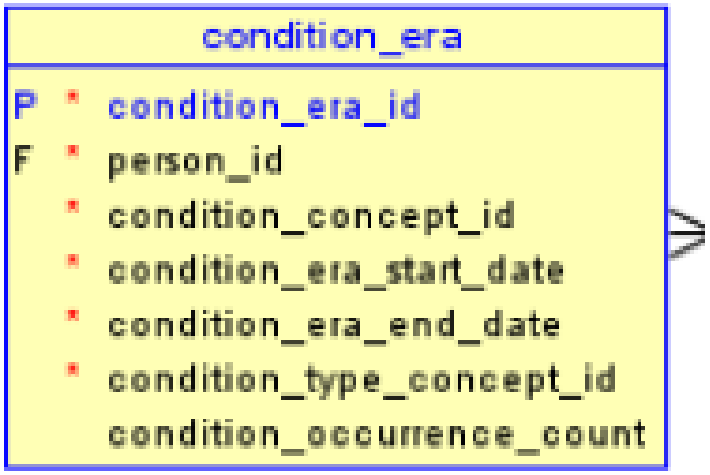


condition_occurrence	
P *	condition_occurrence_id
F *	person_id
*	condition_concept_id
*	condition_start_date
	condition_end_date
*	condition_type_concept_id
	stop_reason
	associated_provider_id
	visit_occurrence_id
	condition_source_value

- Record for each condition observed in the dataset
- From claims data, there would be one condition occurrence record for each diagnosis code recorded from all medical claims
- Optionally allow linkage to a corresponding visit and associated provider
- VISIT_OCCURRENCE_ID allows linkage to specific CARE_SITE through the VISIT_OCCURRENCE table

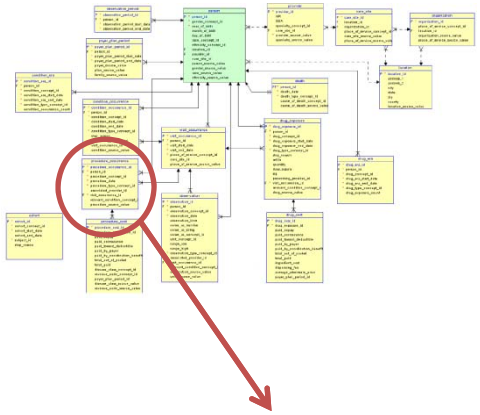
Condition era

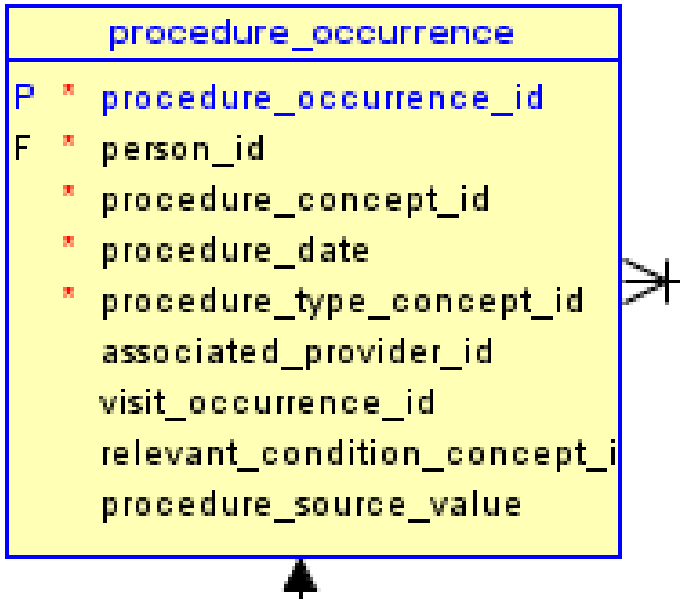




- Analogous to DRUG_ERA, a derived table to consolidate condition occurrence records into meaningful units for analysis
- Combines diagnosis instances into continuous episodes of care for a given condition
- Standardizes some aspects of data cleaning, but does not replace need to make study-specific choices around disease onset and length of illness

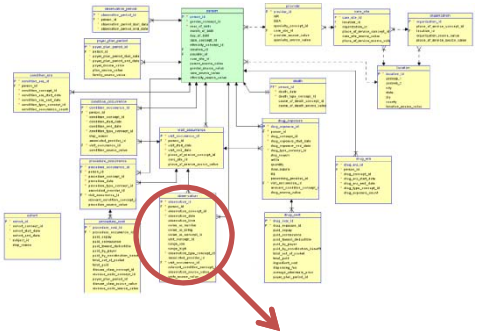
Procedure occurrence





- Record for each medical procedure observed
- In claims data, procedures may be observed on both outpatient claims (CPT, HCPCS) and inpatient claims (ICD9P)

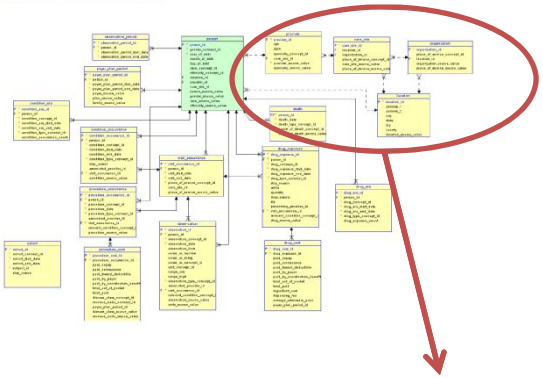
Observation



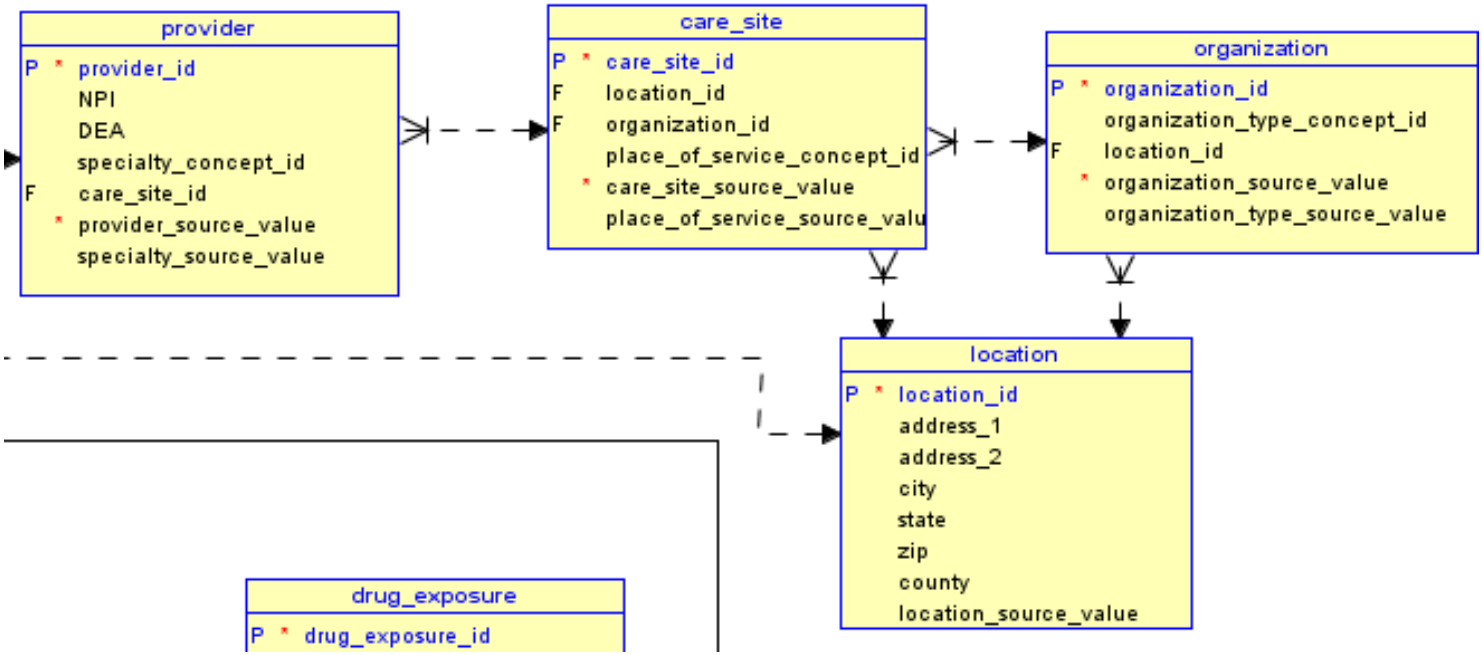
observation	
P *	observation_id
F *	person_id
*	observation_concept_id
*	observation_date
	observation_time
	value_as_number
	value_as_string
	value_as_concept_id
	unit_concept_id
	range_low
	range_high
*	observation_type_concept_id
	associated_provider_id
	visit_occurrence_id
	relevant_condition_concept_i
	observation_source_value
	units_source_value

- "Catch-all" that can accommodate a wide array of other types of clinical observations, such as laboratory values, signs/symptoms, family history, rule-out conditions
- Observation values can be numeric, string, or concept
- Laboratory data requires standardizing CONCEPT (LOINC) and UNIT (UCUM), and values commonly compared with normal range

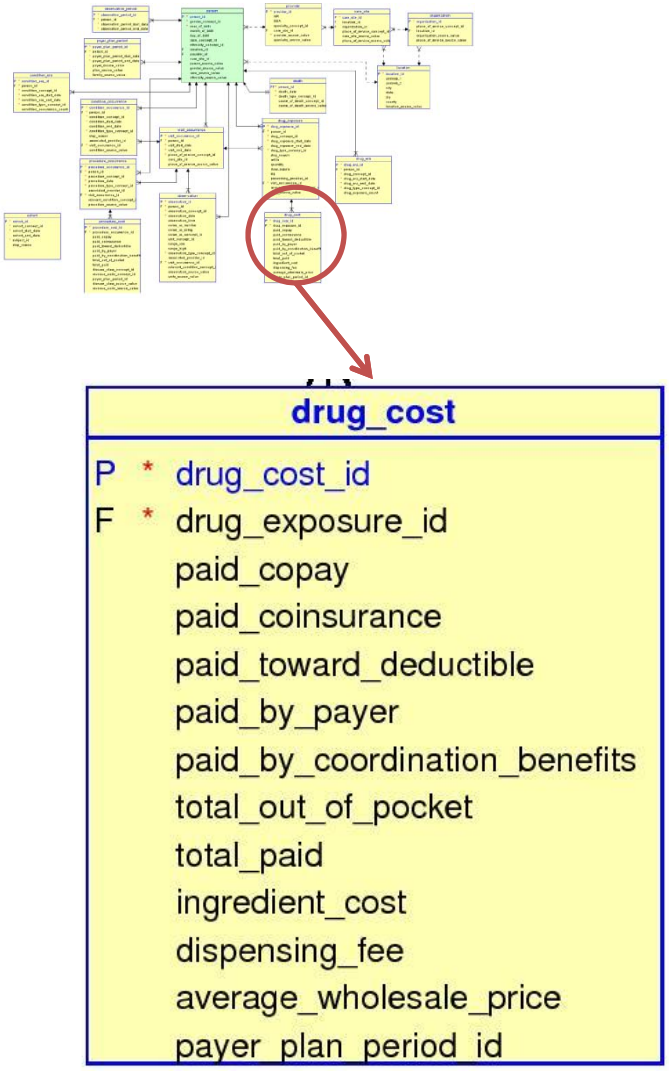
Provider / Care site / Organization



- Tables to record administration of care, used in system/quality analyses
- Data not available in many sources

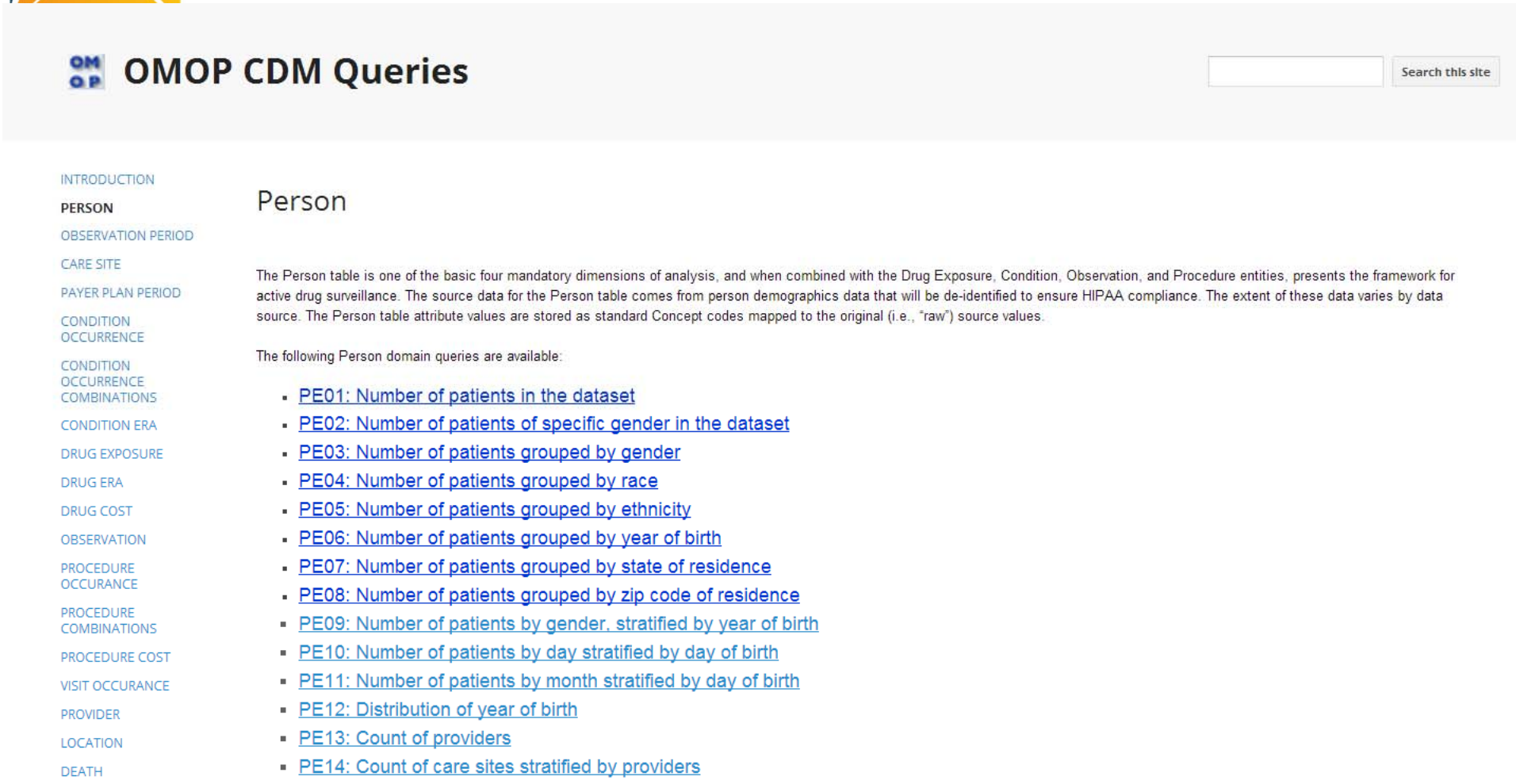


Incorporating financial information



- Each DRUG_EXPOSURE can have one or more DRUG_COST entries
- Each PROCEDURE_OCCURRENCE can have one or more PROCEDURE_COST entries
- Focus on amount paid, and can broken down by typically available categories
- Cost records can optionally link to person's payer/plan information, as available

Example queries for many use cases



http://cdmqueries.omop.org/

Other resources

- Video presentations of tutorials on how to query data and conduct basic analyses: http://omop.org/2013Symposium
- CDM specifications and example ETLs from organizations converting their data to the OMOP CDM: http://omop.org/CDM
- Vocabulary specifications and release notes: http://omop.org/Vocabularies
- Publications using the OMOP CDM: http://omop.org/PublicationList

Concluding thoughts

- OMOP common data model can accommodate different types of observational healthcare data in one standardized, easy-to-use format
- Many different types of observational analyses can be implemented using the same basic building blocks
- Common data model, linked with OMOP vocabulary, can make your analyses standardized and efficient across a network of databases

Join the journey

Interested in OHDSI?

Questions or comments?

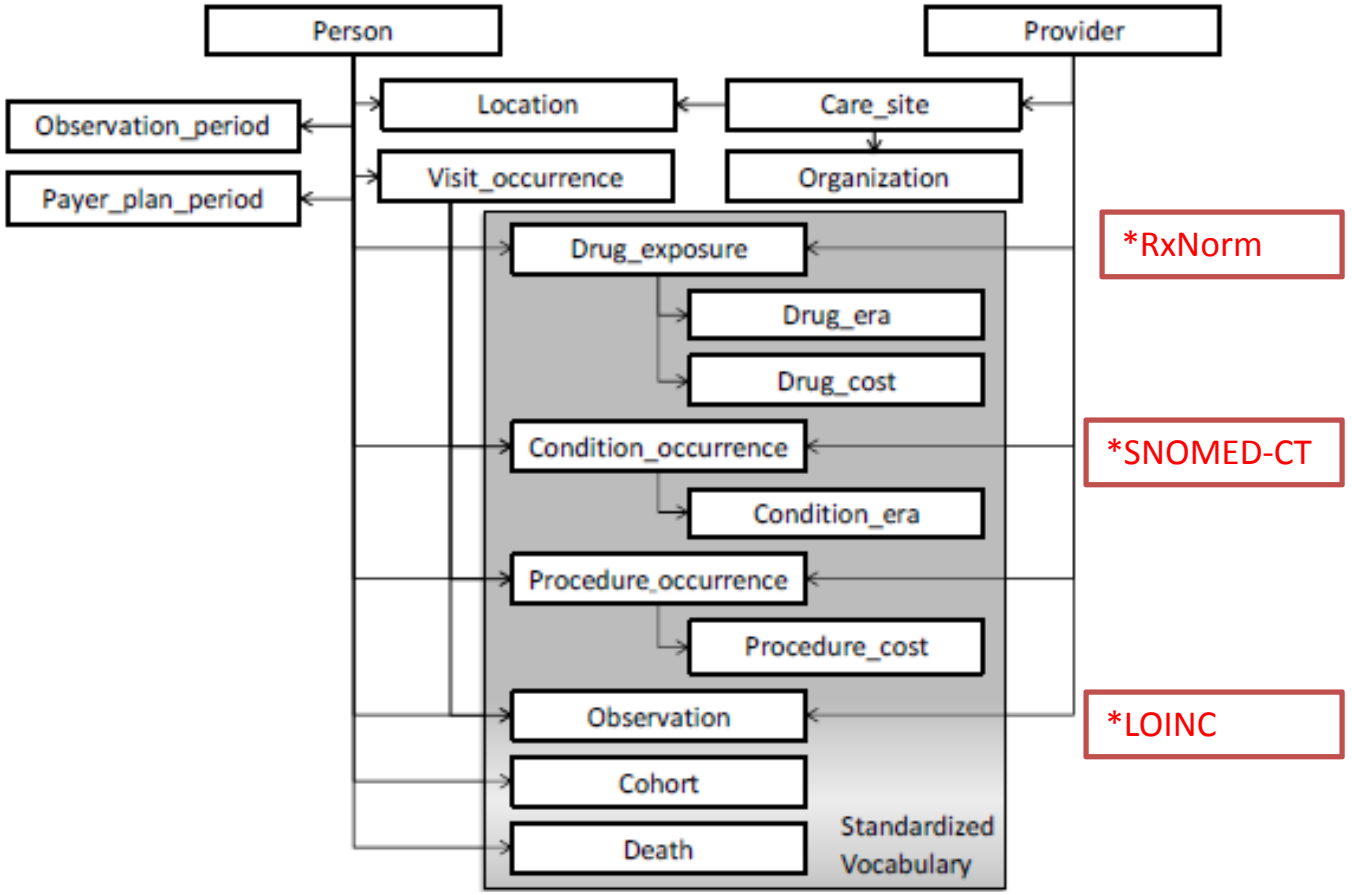
Contact:

ryan@ohdsi.org

OMOP vocabulary

Patrick Ryan, PhD
Janssen Research and Development
on behalf of OHDSI team
25 March 2014

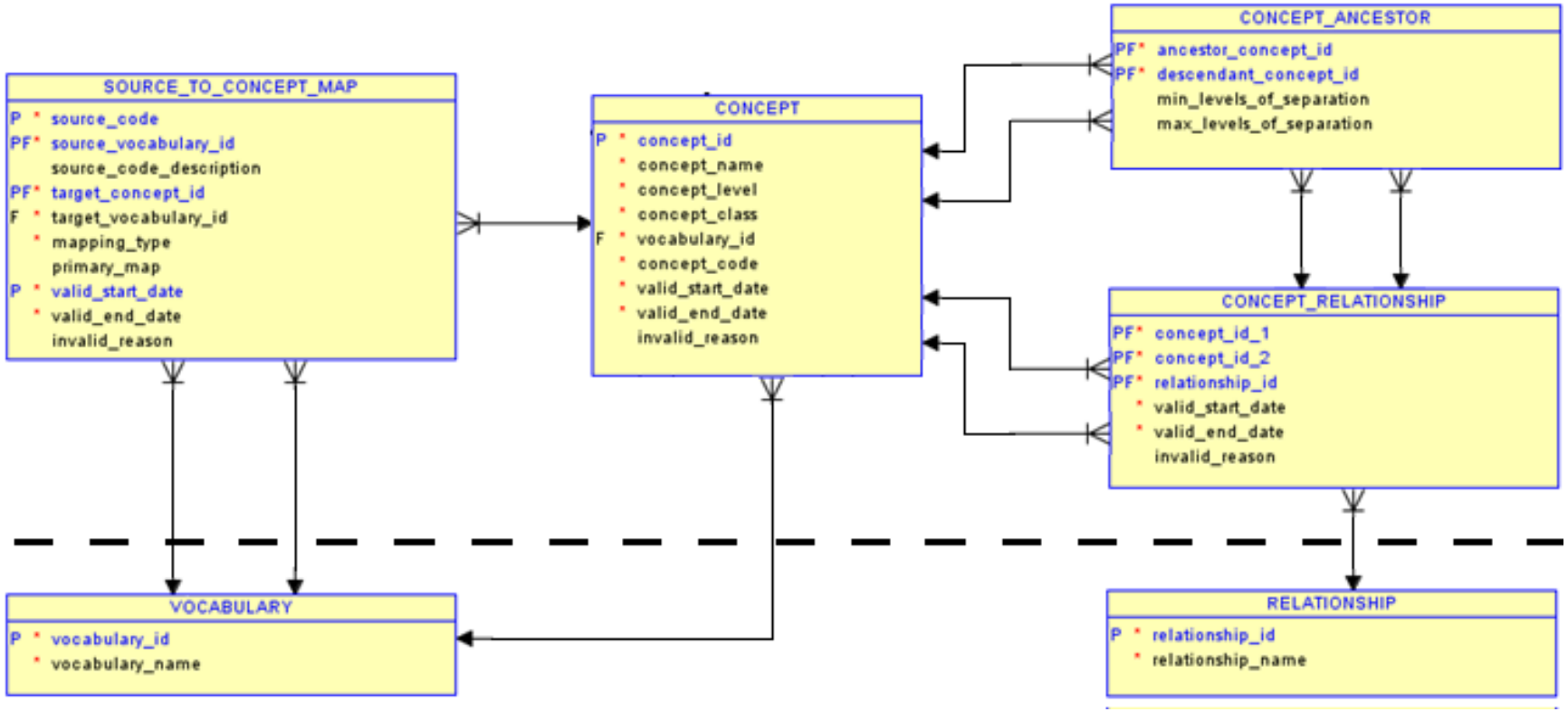
Standardizing to a common data model



- Developed with broad stakeholder input
- Designed to accommodate disparate types of data (claims and EHRs)
- Optimized to use case of standardized large-scale analytics
- Conceived for active medical product surveillance, but extensible for other use cases
- Applied successfully across OMOP data community
- **Standards-based, conforming to ONC Meaningful Use Stage 2 recommendations**

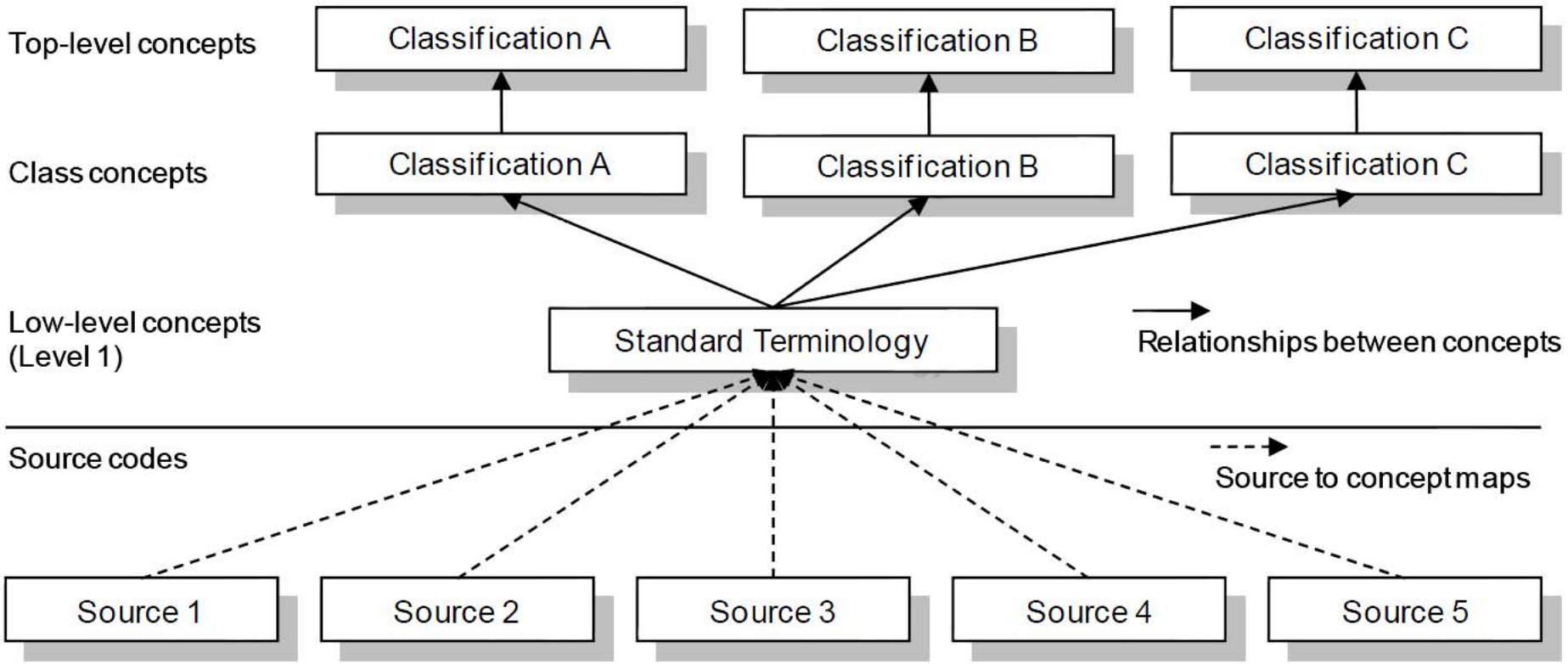
http://omop.org/CDM

OMOP Common Vocabulary Model



1. All content: concepts in **`concept`** table
2. Source codes mapped to concepts in **`source_to_concept_map`**
3. Direct relationships between concepts listed in **`concept_relationship`**
4. Multi-step hierarchical relationships pre-processed in **`concept_ancestor`**

General structure of vocabulary



OMOP Common Vocabulary Model

What it is

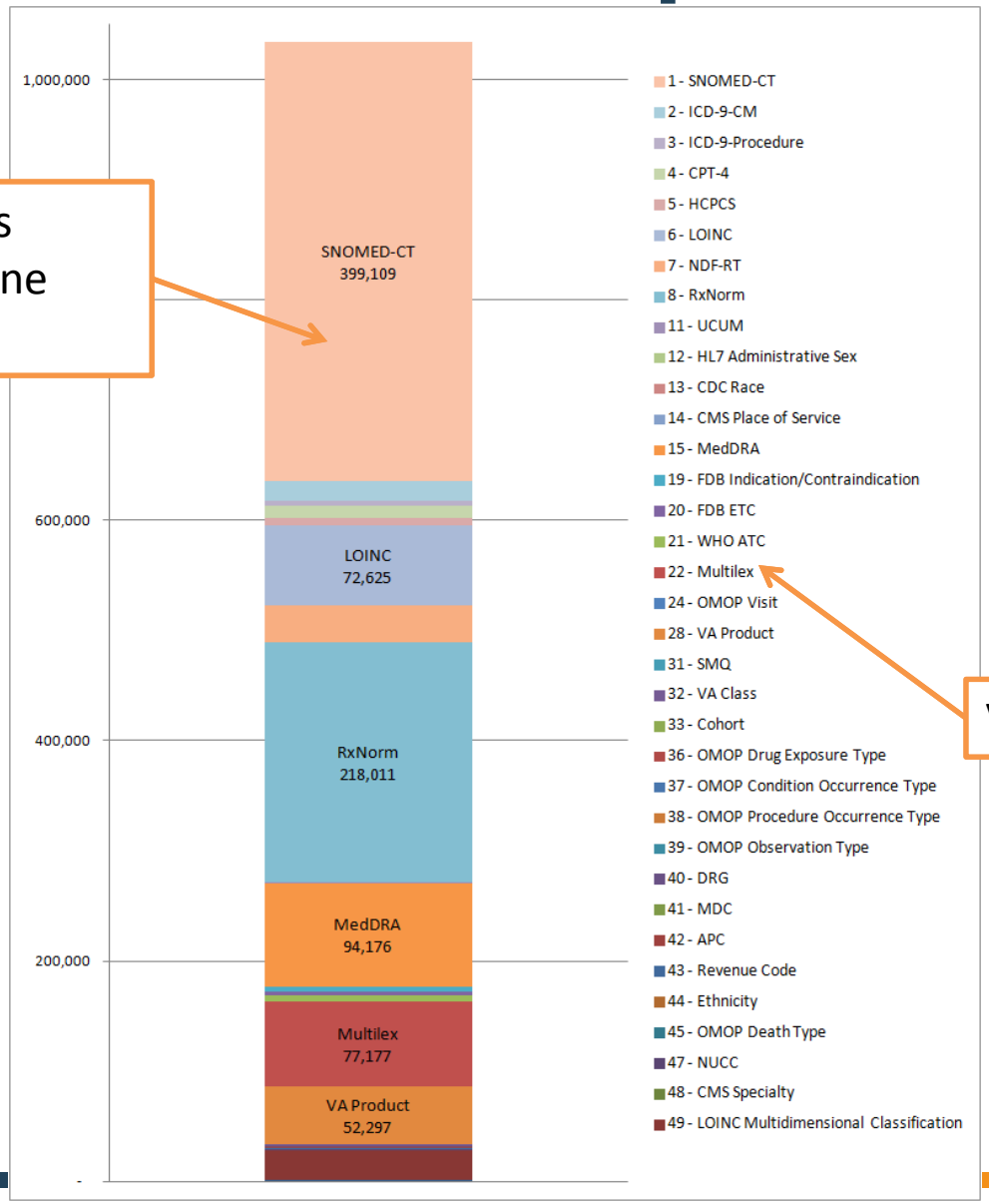
- Standardized structure to house existing vocabularies used in the public domain
- Compiled standards from disparate public and private sources
- Built on the shoulders of National Library of Medicine's Unified Medical Language System (UMLS)
- Directed, hierarchical representation to facilitate efficient queries within analysis use cases

What it's not

- OMOP didn't create 'new' vocabularies
- Use of OMOP vocabulary does not 'replace' the source values in your system
- Static dataset – the vocabulary updates regularly to keep up with the continual evolution of the sources
- Finished product – vocabulary maintenance and improvement is ongoing activity that requires community participation and support

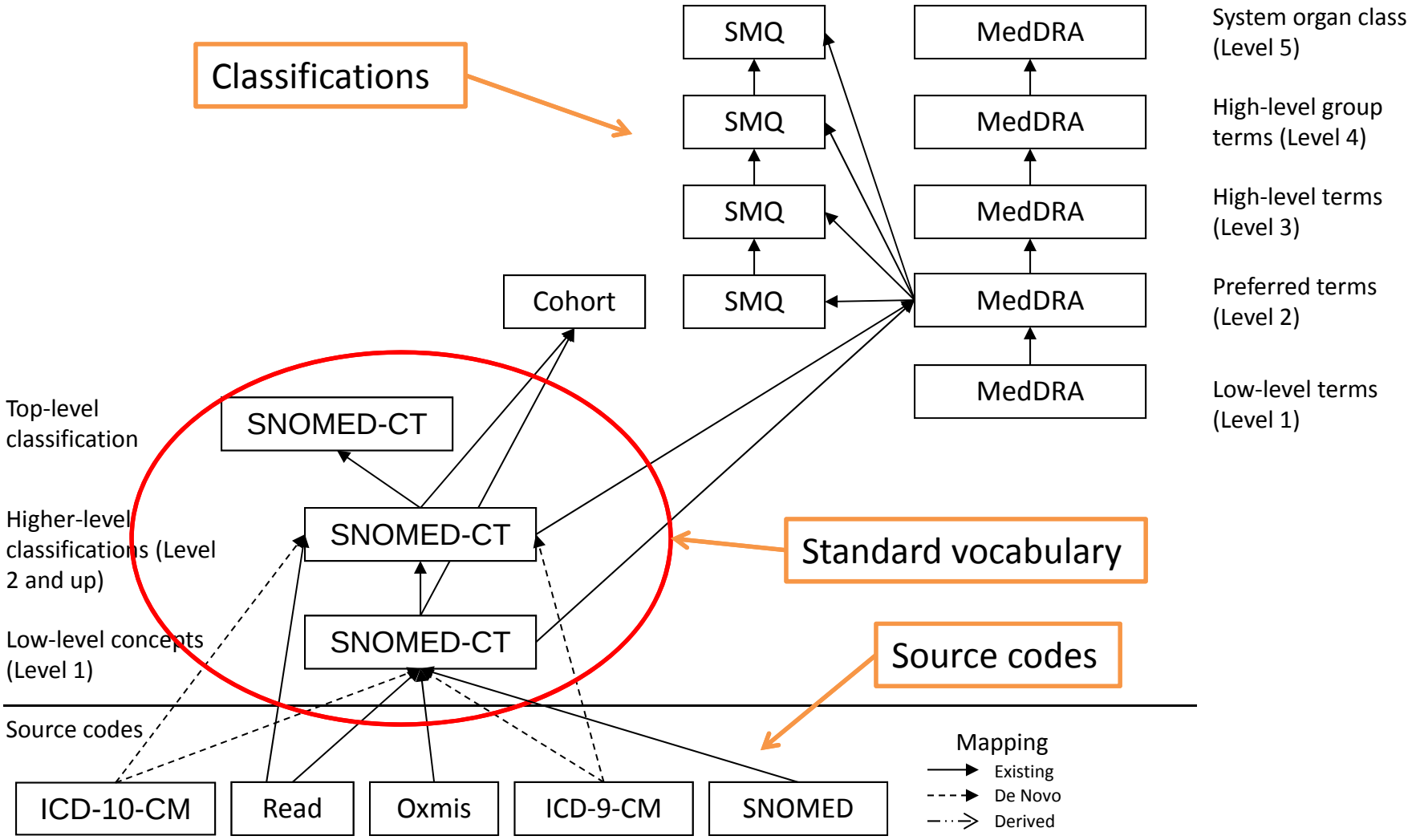
The **concept** Table

All vocabularies stacked up in one table



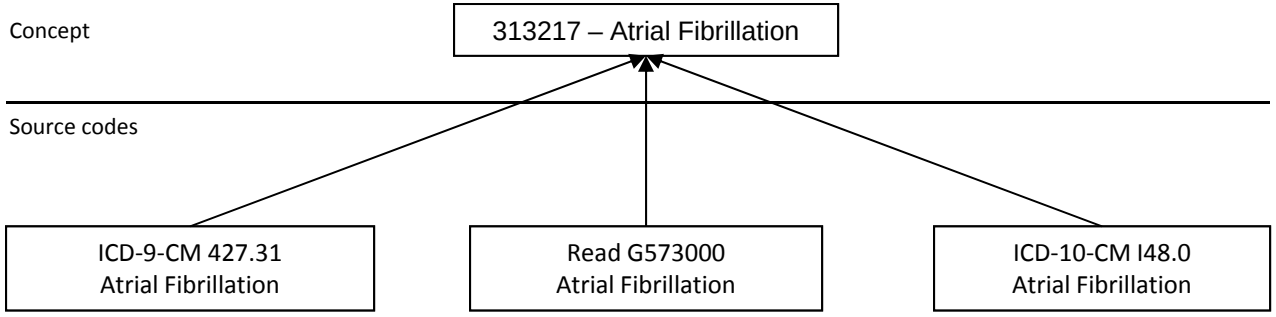
Vocabulary ID/Name

Disease Concepts



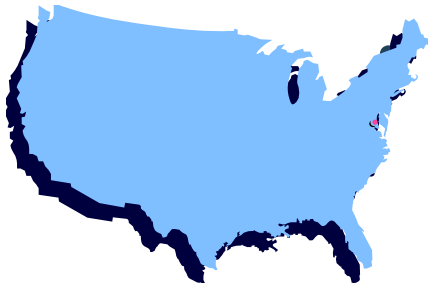
Codes from Different Vocabularies all Map to Concept

```
/* ICD-9-CM */
SELECT * FROM source_to_concept_map WHERE source_code = '427.31';
/* Read */
SELECT * FROM source_to_concept_map WHERE source_code = 'G573000';
/* ICD-10-CM */
SELECT * FROM source_to_concept_map WHERE source_code = 'I48.0';
```



Why not Use Codes?

Reason #1: Different data use different codes



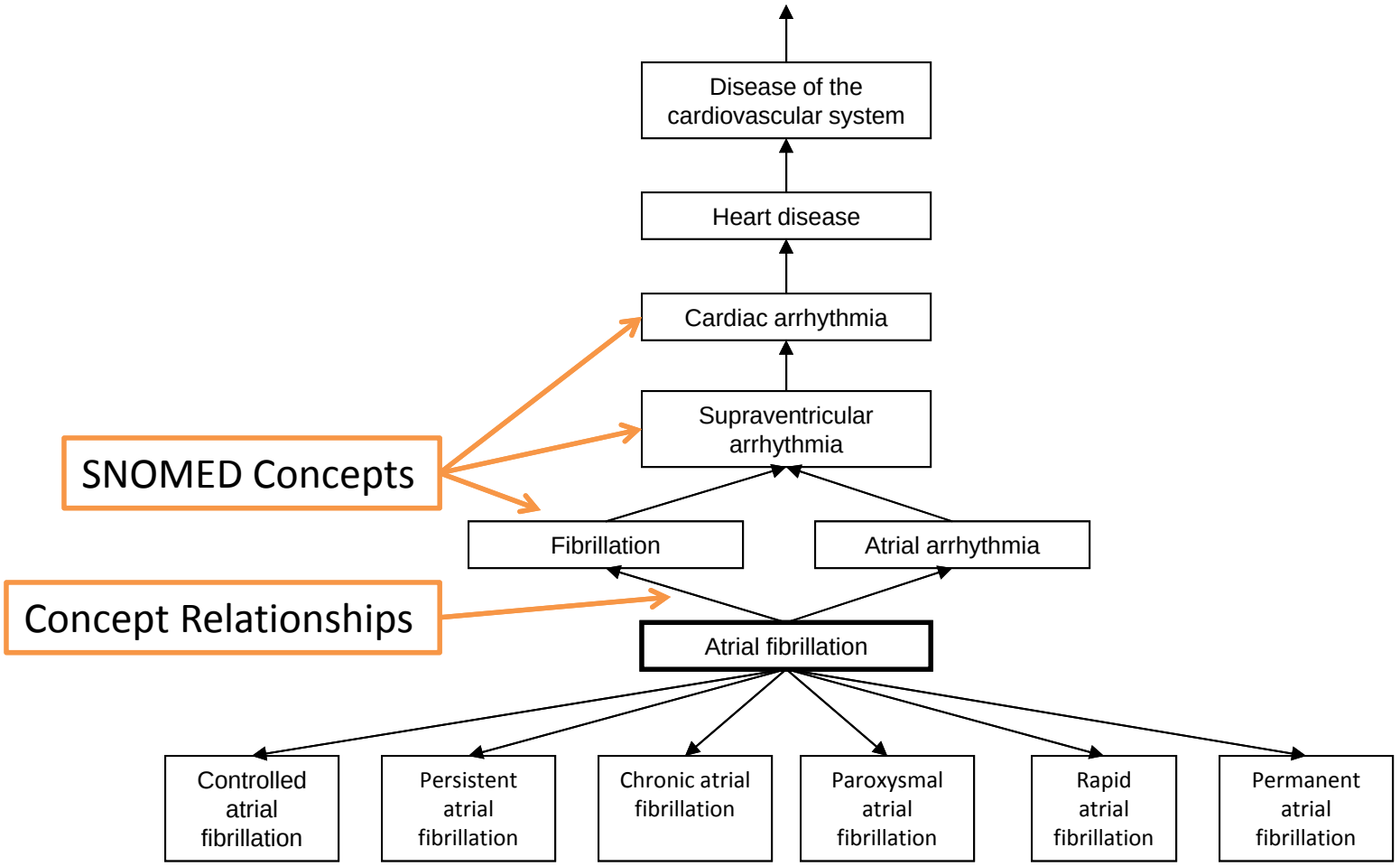
- Administrative Claims
  - Conditions: ICD-9-CM
    - Until 2014, when ICD-10-CM becomes CMS billing standard
  - Drugs: NDC
    - If you only care about pharmacy dispensing
      If the product is administered by physician, then HCPCS, ICD9P, CPT4

- Electronic health records
  - Conditions: ICD-9-CM, SNOMED, ICD-10-CM, free text
  - Drugs: NDC, Multum, GPI, VA Product, ChargeMaster, free text

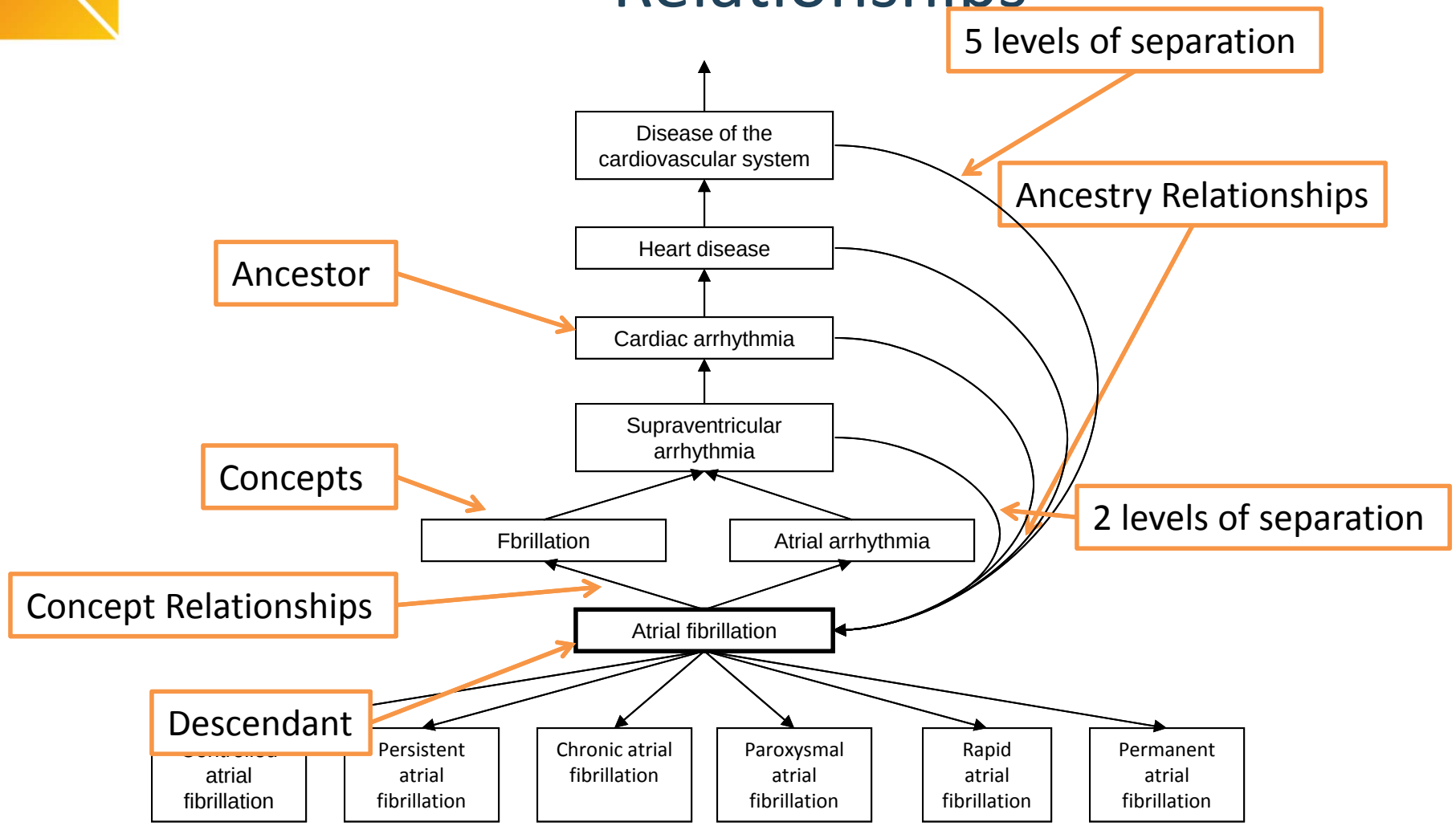
Codes Used in the World

- Conditions
  - READ, OXMIS, ICD-9-CM, ICD-10-CM, ICPC, MedDRA, freetext in different languages
- Drugs
  - Multilex, DM+D, ATC, NPI, NDC* (*for different nations), freetext in different languages

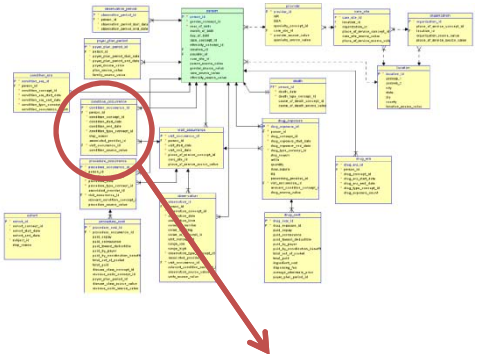
Reason #2: Disease Hierarchy



Ancestry Relationships: Higher-Level Relationships



How vocabulary integrates with CDM: Condition occurrence



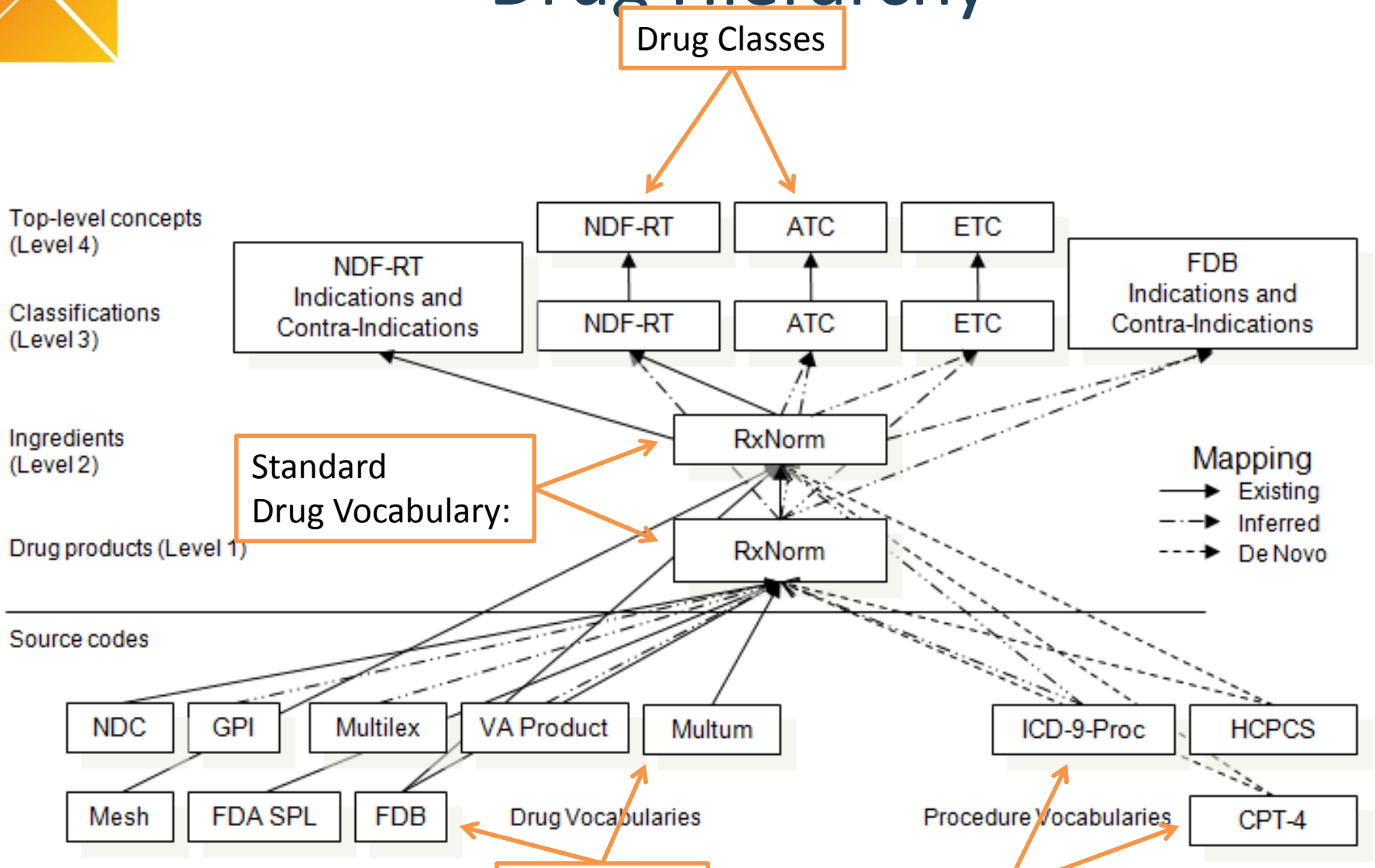
condition_occurrence	
P	condition_occurrence_id
F	person_id
	condition_concept_id
	condition_start_date
	condition_end_date
	condition_type_concept_id
	stop_reason
	associated_provider_id
	visit_occurrence_id
	condition_source_value

- CONDITION_SOURCE_VALUE would contain the code originally found in raw database
  - ex. ICD-9-CM for US administrative claims
- CONDITION_CONCEPT_ID would contain concept from standard vocabulary
  - ex. corresponding SNOMED concept that ICD-9-CM mapped to in SOURCE_TO_CONCEPT_MAP
- CONDITION_TYPE_CONCEPT_ID would contain concept for provenance of the data
  - ex. primary inpatient diagnosis

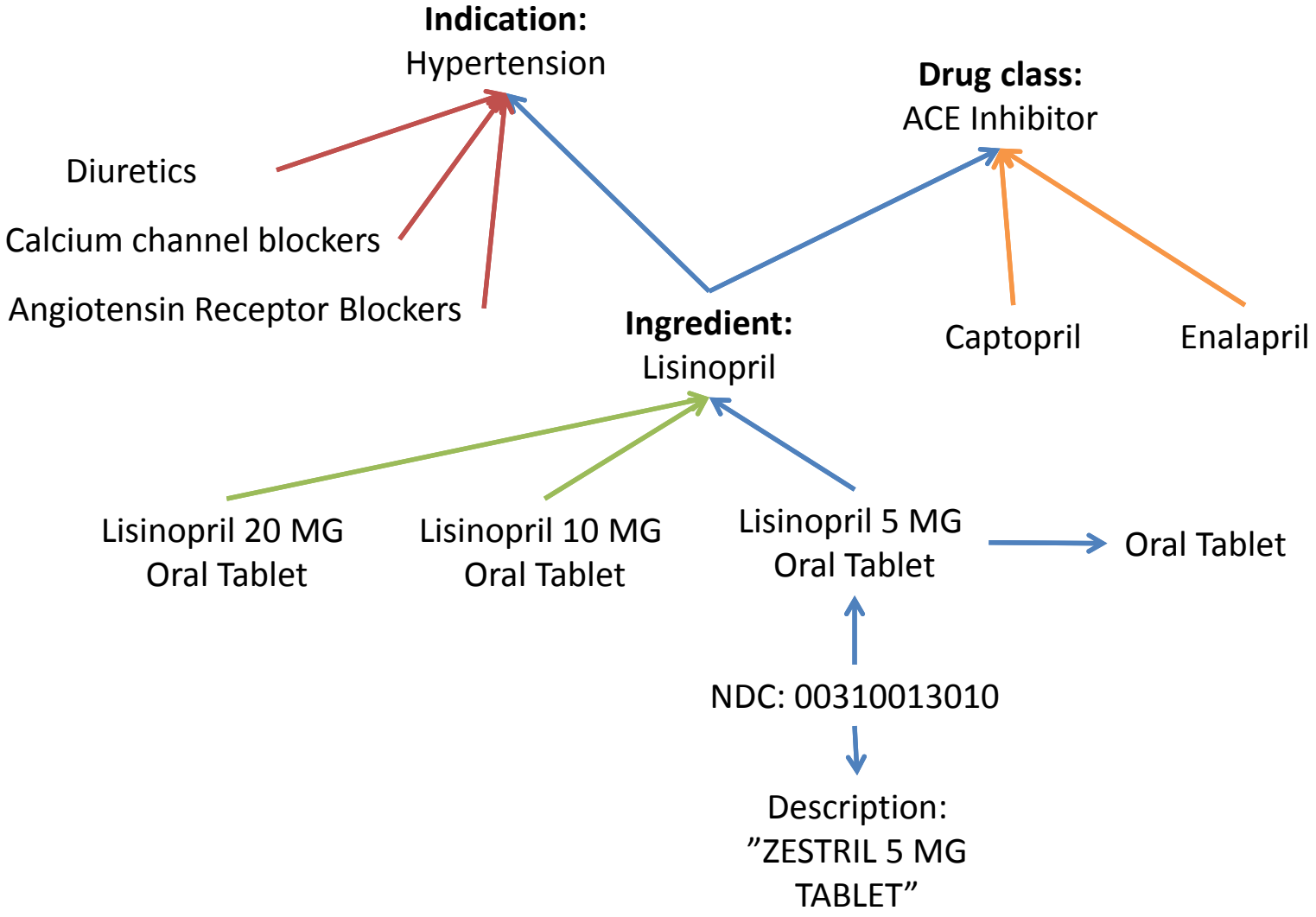
Does it Work that Way with Drugs?

- Codes
  - NDC, GPI, Multilex, HCPCS, etc.
- Concepts
  - Drug products (Generic and Brand)
  - Drug ingredients
  - Drug Classes
- Relationships
- Ancestry

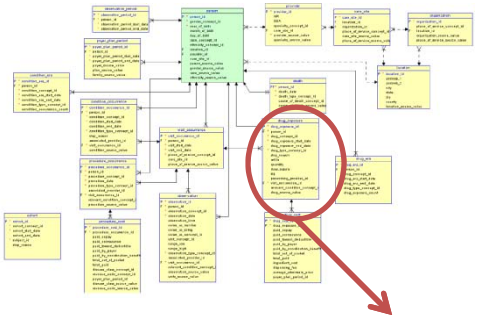
Drug Hierarchy

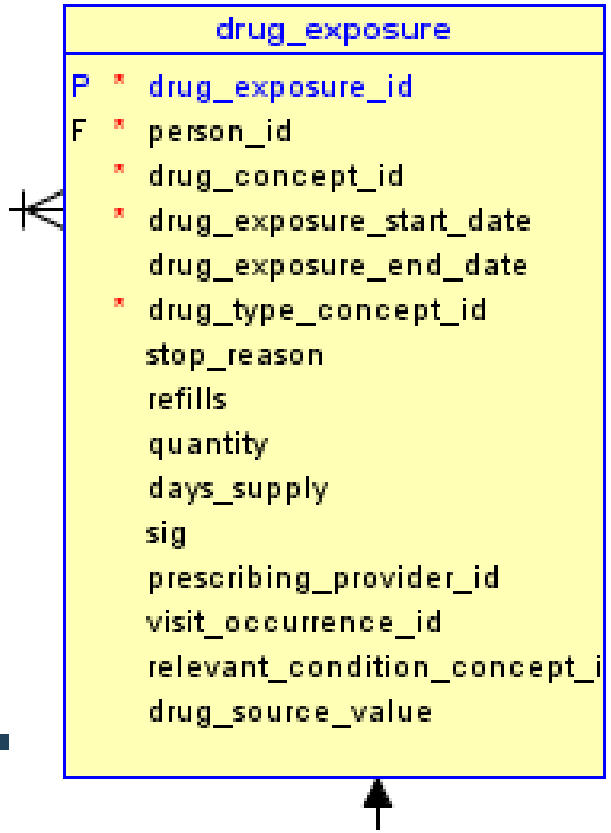


Turning NDCs into something useful



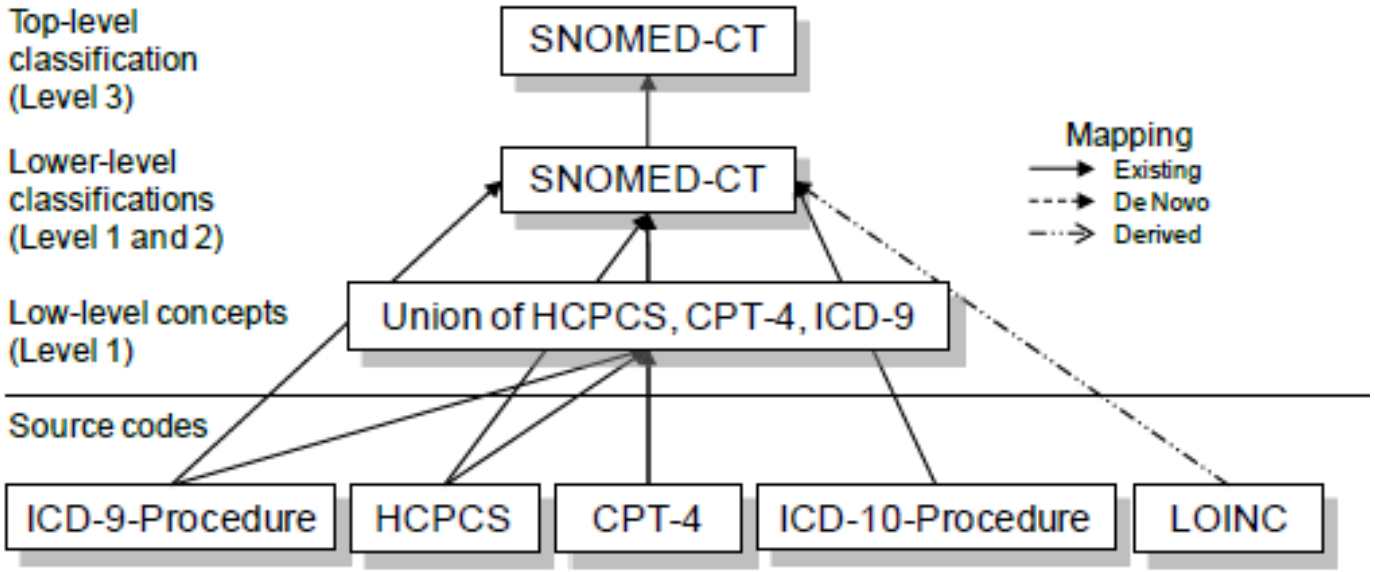
How vocabulary integrates with CDM: Drug exposure





- DRUG_SOURCE_VALUE would contain the code originally found in raw database
  - ex. NDC for US administrative claims
- DRUG_CONCEPT_ID would contain concept from standard vocabulary
  - ex. corresponding RxNorm clinical drug concept that NDC mapped to in SOURCE_TO_CONCEPT_MAP
- DRUG_TYPE_CONCEPT_ID would contain concept for provenance of the data
  - ex. pharmacy dispensing

Procedure vocabularies

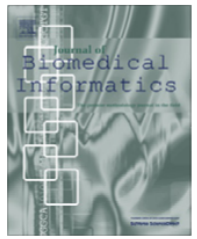


Lessons learned on standardized vocabulary

Contents lists available at SciVerse ScienceDirect

Journal of Biomedical Informatics

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



Evaluation of alternative standardized terminologies for medical conditions within a network of observational healthcare databases ☆

Christian Reich [a,*], Patrick B. Ryan [a,b,1], Paul E. Stang [a,b,1], Mitra Rocca [c,2]

- SNOMED and MedDRA provide useful standards to harmonize data, with minimal information loss in translation from ICD-9-CM

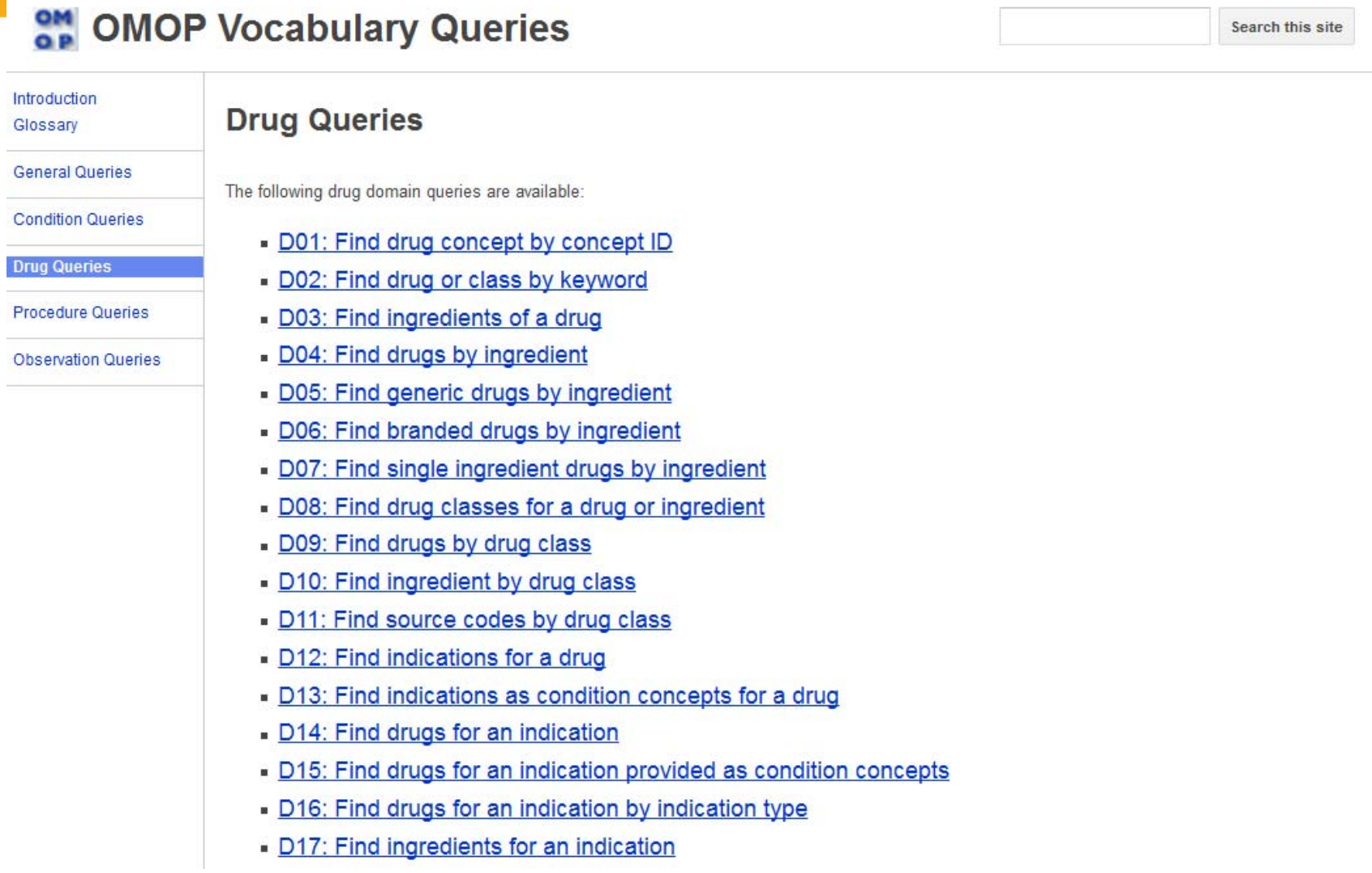
Health Serv Outcomes Res Method (2013) 13:58–67
DOI 10.1007/s10742-012-0102-1

Applying standardized drug terminologies to observational healthcare databases: a case study on opioid exposure

Frank J. DeFalco · Patrick B. Ryan · M. Soledad Cepeda

- Use of multiple standardized drug vocabularies can improve the quality of observational analysis

Example queries for many use cases



http://vocabqueries.omop.org/

Classification systems to support analyses

- UMLS:
  - http://www.nlm.nih.gov/research/umls/
- RxNorm
  - https://www.nlm.nih.gov/research/umls/rxnorm/
- WHO Anatomical Therapeutic Chemical (ATC)
  - http://www.whocc.no/atc/structure_and_principles/
- First DataBank Enhanced Therapeutic Classification (ETC)
  - http://www.fdbhealth.com/fdb-medknowledge-foundations/
- VA National Drug File Reference Terminology (NDF-RT)
  - http://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/NDFRT/
- Generic Product Identifier (GPI)
  - http://www.medi-span.com/medi-span-electronic-drug-file.aspx
- VA Classification system
  - http://www.pbm.va.gov/NationalFormulary.asp

Other resources

- Video presentations of tutorials on how to query OMOP CDM and vocabulary: http://omop.org/2013Symposium
- CDM specifications and example ETLs from organizations converting their data to the OMOP CDM: http://omop.org/CDM
- Vocabulary specifications and release notes: http://omop.org/Vocabularies
- Publications using the OMOP CDM: http://omop.org/PublicationList

Join the journey

Interested in OHDSI?

Questions or comments?

Contact:

ryan@ohdsi.org

OMOP Analysis Use Cases

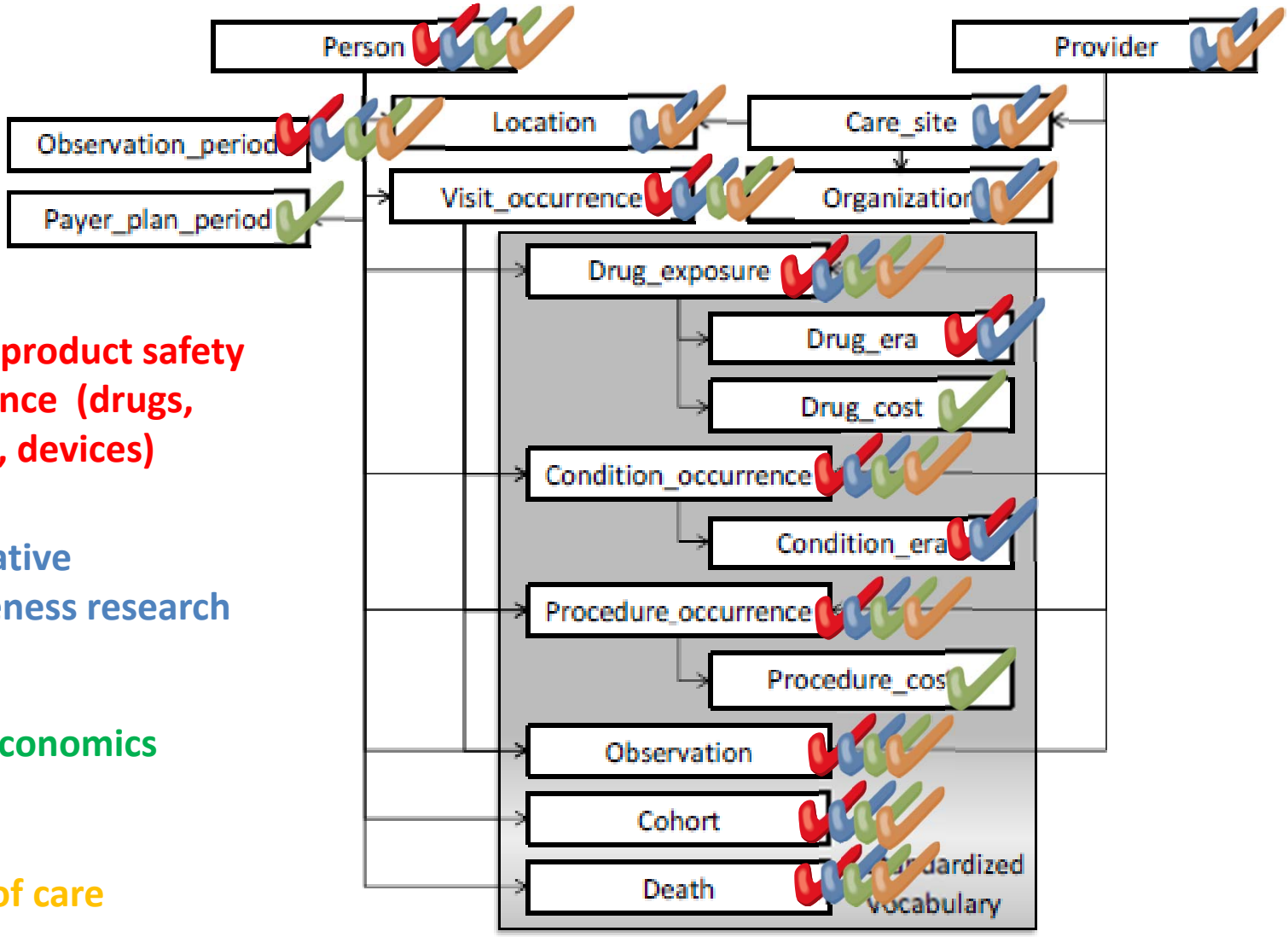
Patrick Ryan, PhD
Janssen Research and Development
on behalf of OHDSI team
25 March 2014

OMOP Common Data Model is only means to an end

CDM design is driven by analysis use cases:

- Descriptive statistics about disease natural history and health service utilization
- Epidemiology study designs to study the effects of medical products
- Informatics solutions to enable real-time exploration of patient-level data, ex:
  - Standardized reporting
  - Risk identification
  - Clinical trial feasibility assessment

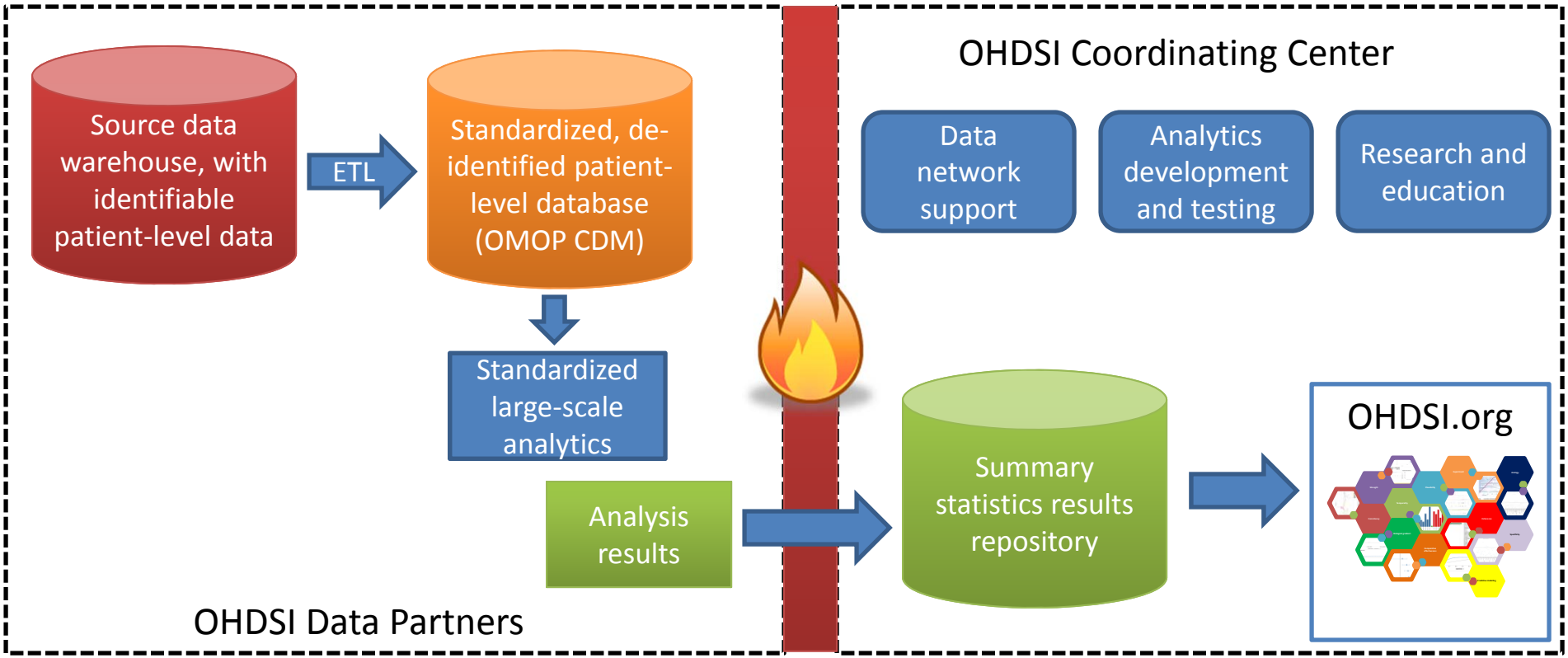
One model, multiple use cases



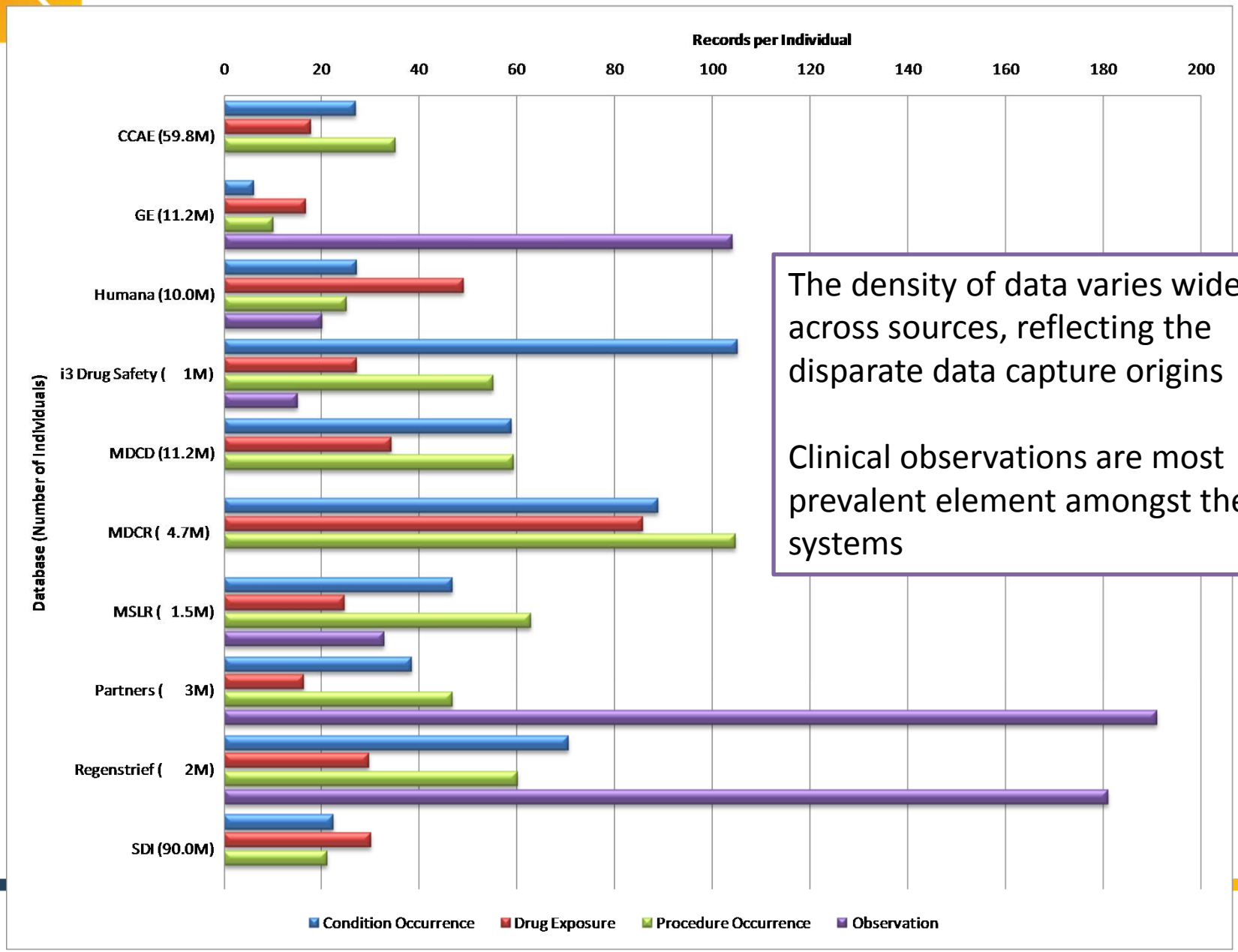
- **Medical product safety surveillance (drugs, vaccines, devices)**
- **Comparative effectiveness research**
- **Health economics**
- **Quality of care**

http://omop.org/CDM

How standardized analytics work across OHDSI network



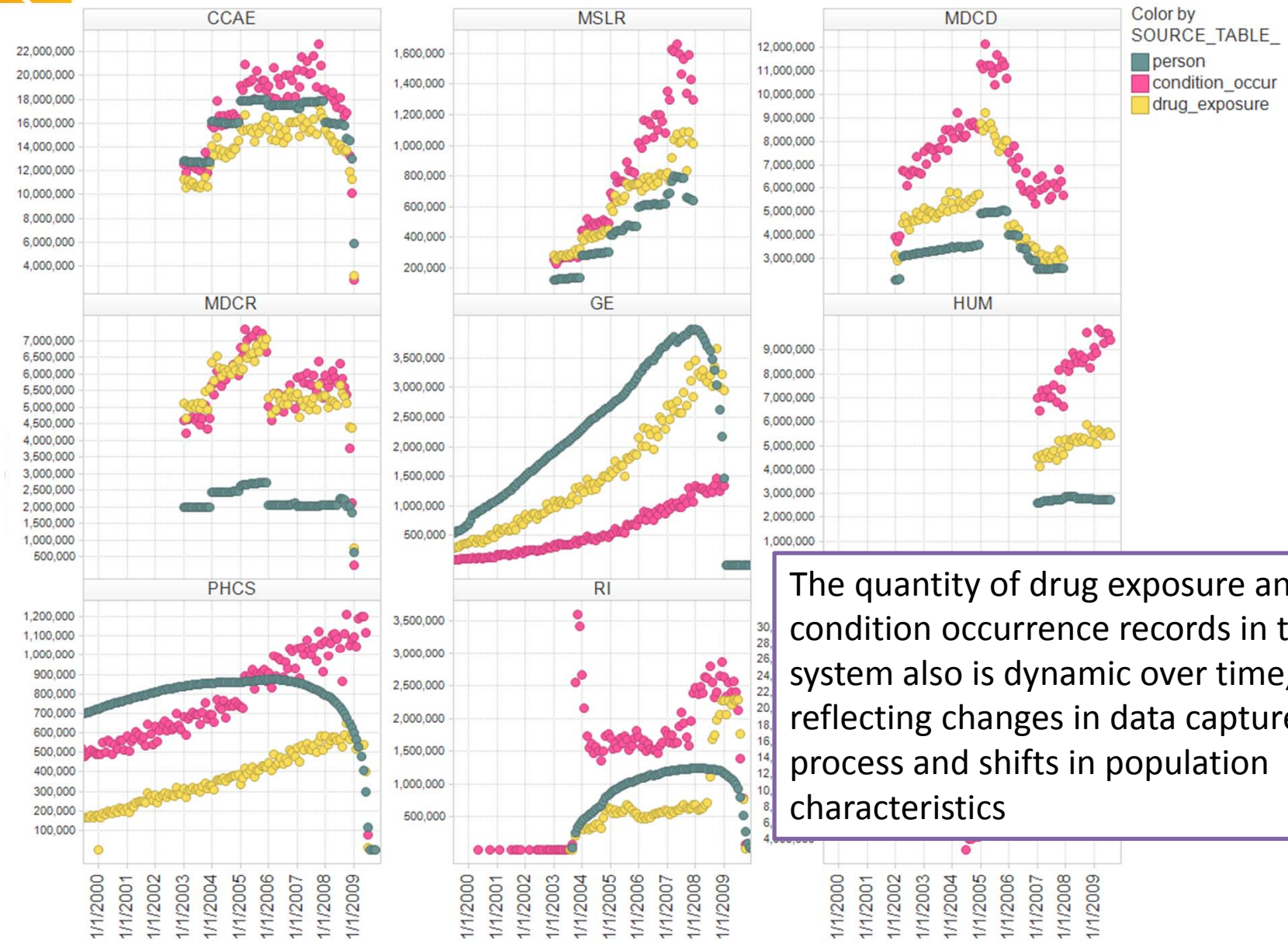
Data density across the OMOP community



The density of data varies widely across sources, reflecting the disparate data capture origins

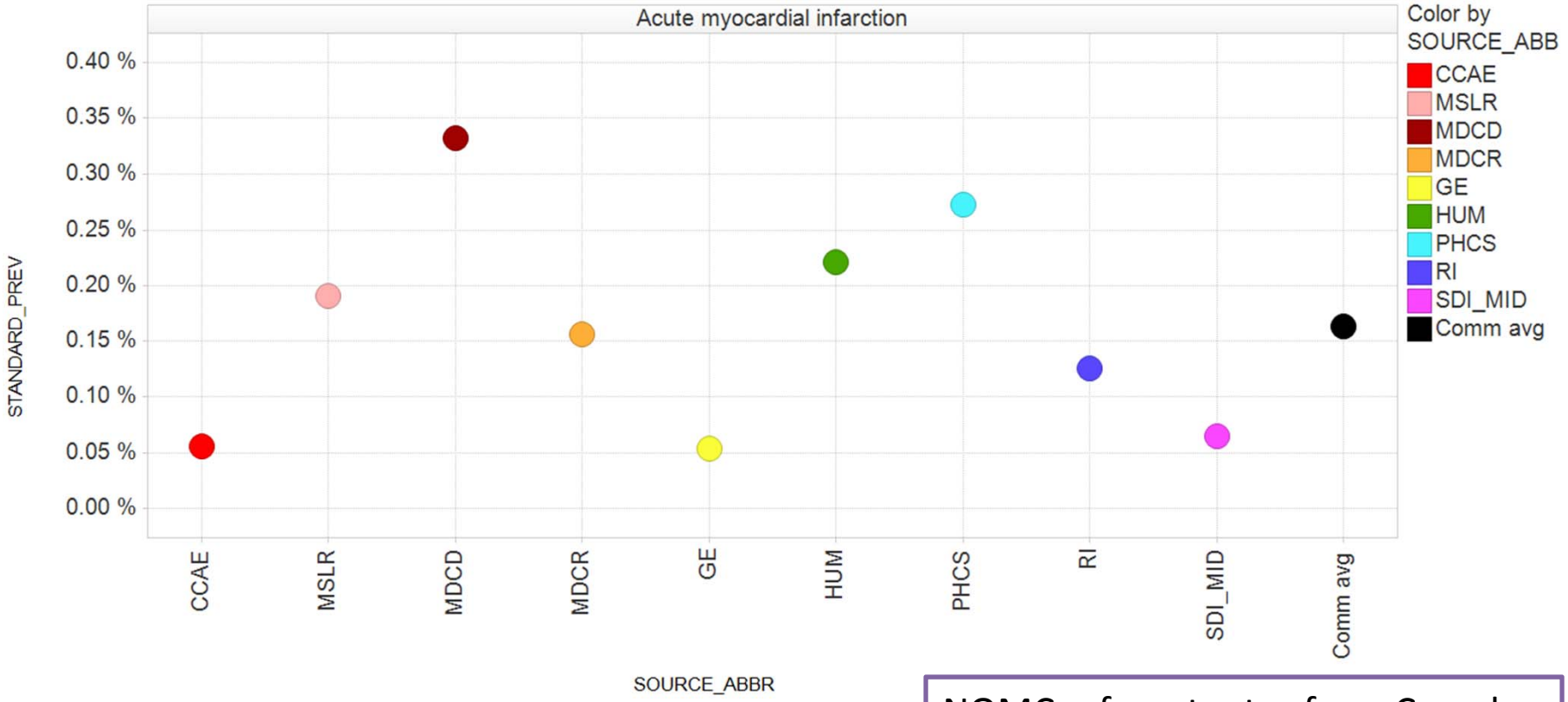
Clinical observations are most prevalent element amongst the EHR systems

Standardized Records Over Time



The quantity of drug exposure and condition occurrence records in the system also is dynamic over time, reflecting changes in data capture process and shifts in population characteristics

Exploring prevalence of disease: ex: Acute Myocardial Infarction



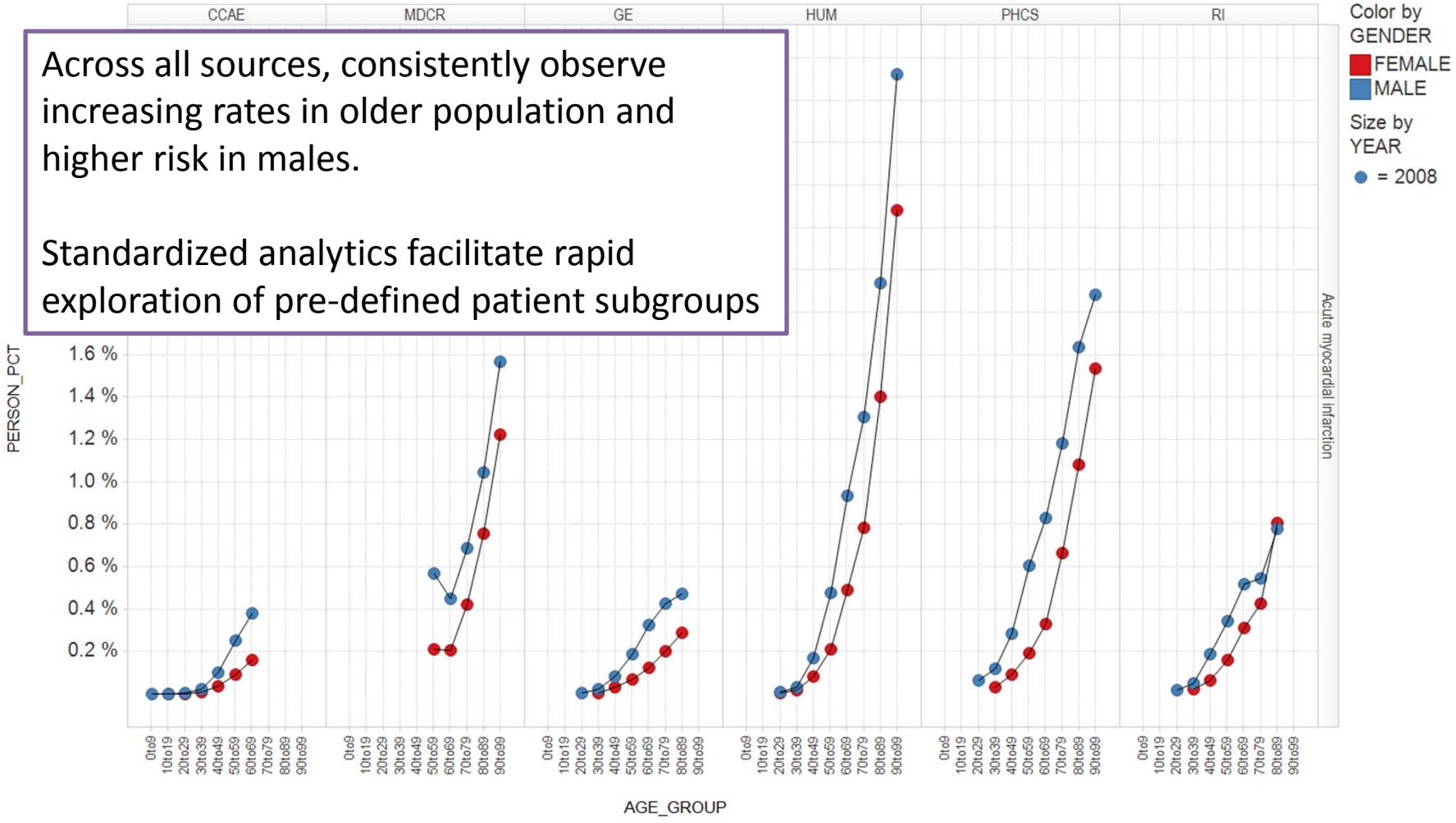
NQMC referent rates from Canada:
0.25% in 2003/4 decreased to
0.22% in 2008/9

Standardized condition prevalence = 5-yr annualized pre
stratified by age and gender, standardized to US Census

Exploring prevalence of disease: ex: Acute Myocardial Infarction

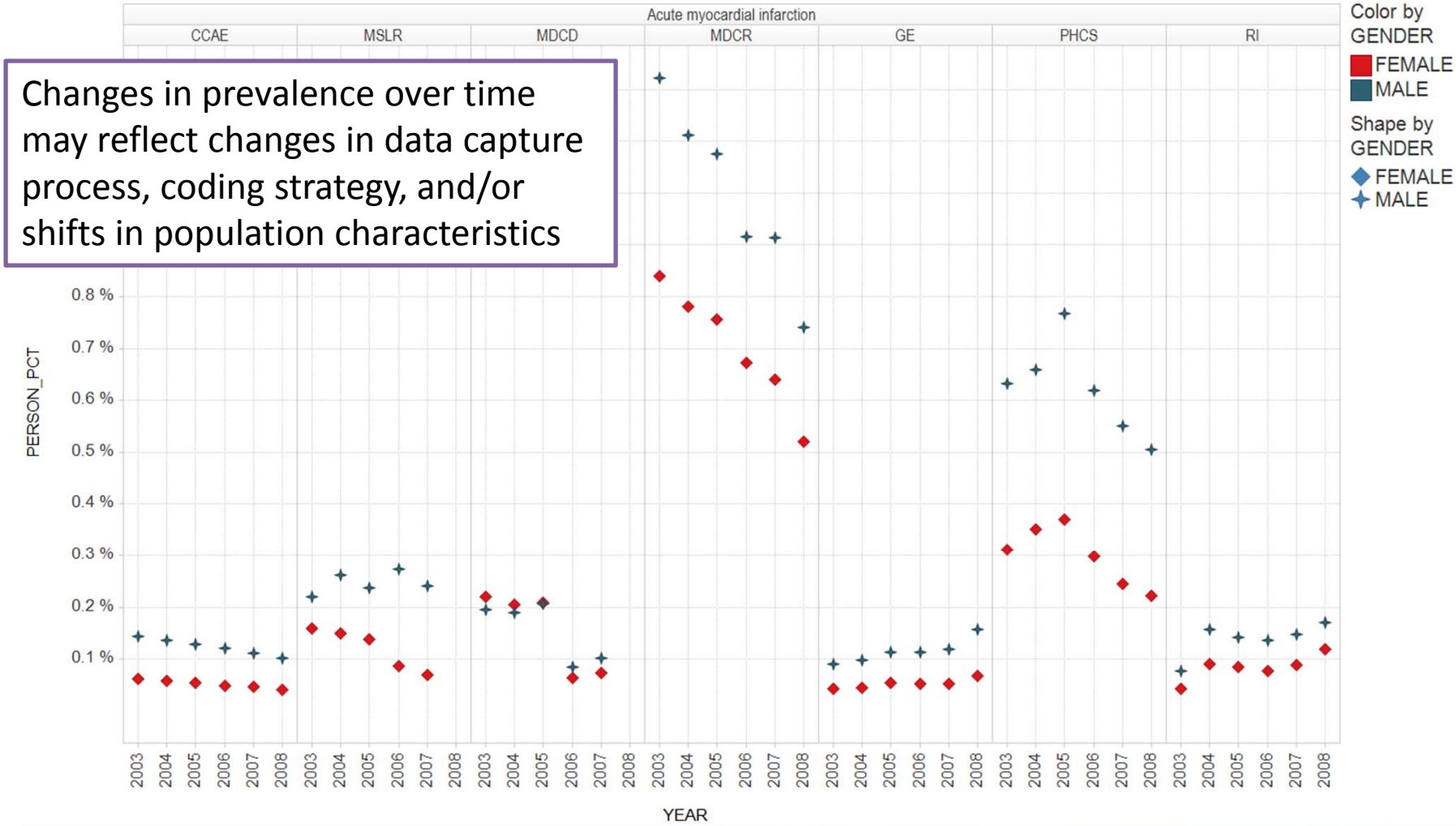
Across all sources, consistently observe increasing rates in older population and higher risk in males.

Standardized analytics facilitate rapid exploration of pre-defined patient subgroups



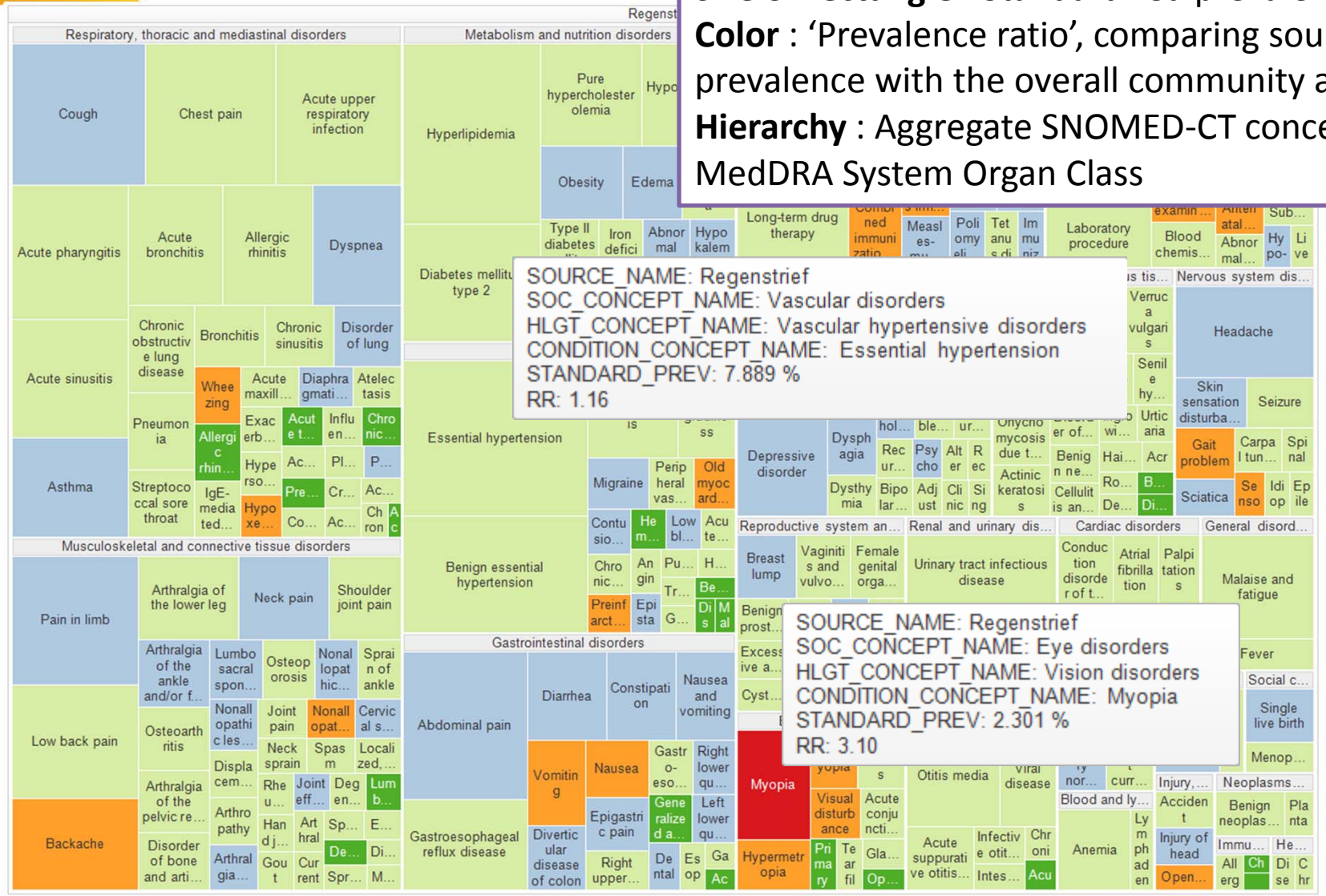
Exploring prevalence of disease: ex: Acute Myocardial Infarction

Changes in prevalence over time may reflect changes in data capture process, coding strategy, and/or shifts in population characteristics

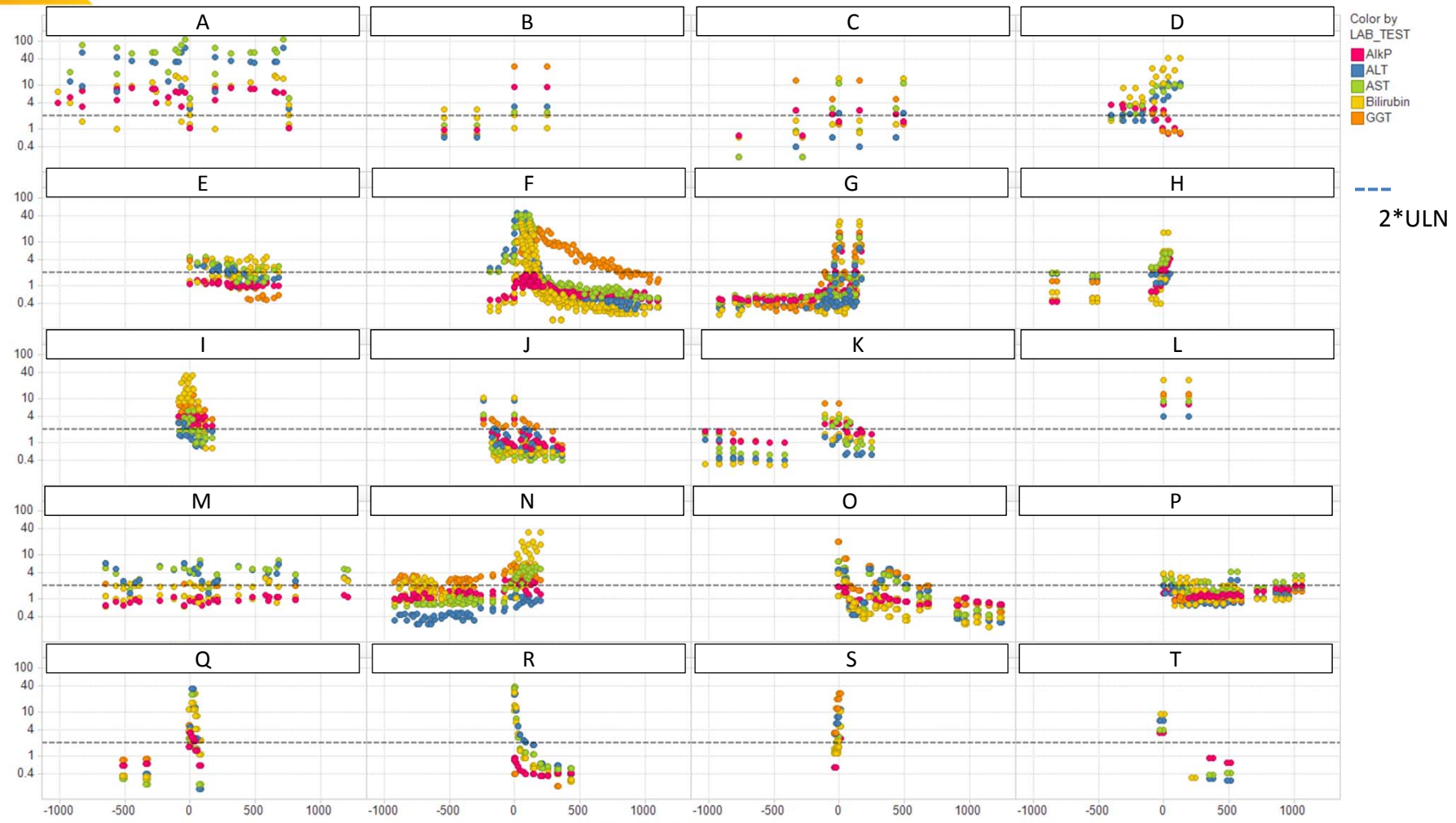


Exploring prevalence of all diseases

Treemap displays 3 dimensions:
Size of rectangle : Standardized prevalence
Color : 'Prevalence ratio', comparing source prevalence with the overall community average
Hierarchy : Aggregate SNOMED-CT concepts by MedDRA System Organ Class

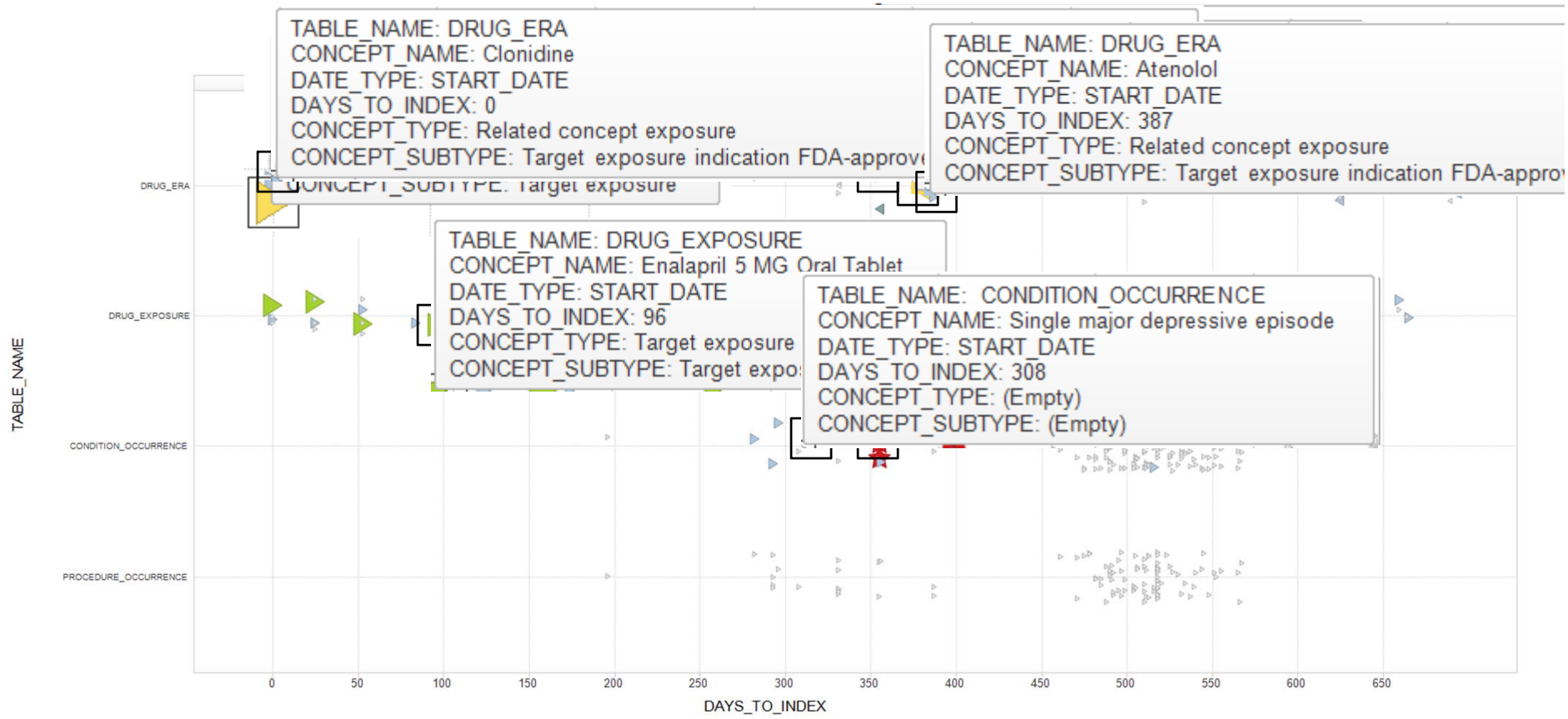


Exploring clinical patterns in laboratory results
ex: Acute Liver Injury



Katz AJ1, Ryan PB, Racoosin JA, Stang PE. Assessment of case definitions for identifying acute liver injury in large observational databases. Drug Saf. 2013 Aug;36(8):651-61. doi: 10.1007/s40264-013-0060-8.

Patient-level exploration:
ex: ACE inhibitor and Angioedema



Target exposure
Target era
Target outcome
Related to exposure
Related to outcome

Start Date
End Date
Outcome

The NEW ENGLAND JOURNAL of MEDICINE
N Engl J Med 2013; 368:1272-1274 | April 4, 2013 | DOI: 10.1056/NEJMp1302834

Dabigatran and Postmarketing Reports of Bleeding

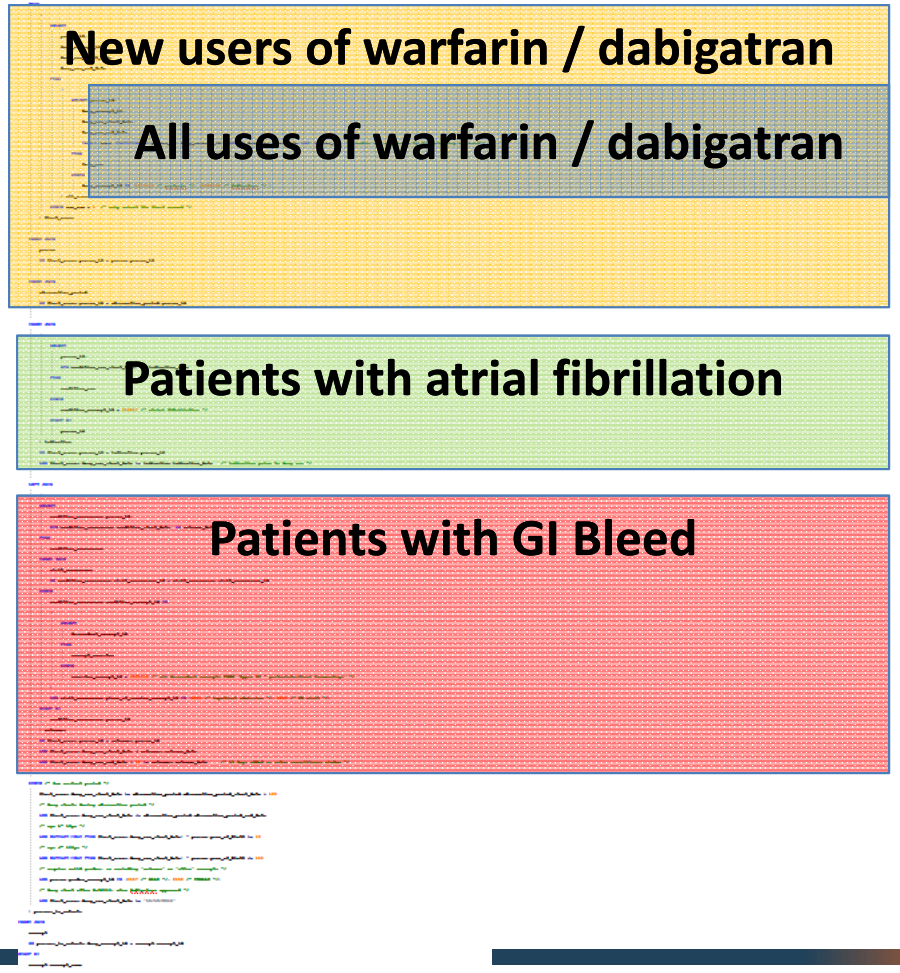
Mary Ross Southworth, Pharm.D., Marsha E. Reichman, Ph.D., and Ellis F. Unger, M.D.

Questions we are regularly asked:
1) What does it take to do an analysis like this?
2) How can this be done against the OMOP CDM?

	Patients	Events	100,000 days at risk)	Patients	Events	100,000 days at risk)
Gastrointestinal hemorrhage						
Analysis with required diagnosis of atrial fibrillation	10,599	16	1.6	43,541	160	3.5
Sensitivity analysis without required diagnosis of atrial fibrillation	12,195	19	1.6	119,940	338	3.1
Intracranial hemorrhage						
Analysis with required diagnosis of atrial fibrillation	10,587	8	0.8	43,594	109	2.4
Sensitivity analysis without required diagnosis of atrial fibrillation	12,182	10	0.9	120,020	204	1.9

* Patients were included in the cohorts if, in the 183 days before the index dispensing of dabigatran or warfarin, they were enrolled in plans for drug and medical coverage and had been given a diagnosis of atrial fibrillation in any care setting. Patients were excluded from the cohorts if, in the 183 days before the index dispensing, they had a claim for an event of interest in an inpatient or emergency department setting or a claim for dispensing of dabigatran or warfarin. Events were assessed during drug exposure, from inpatient or emergency department settings only.

Comparing warfarin vs. dabigatran for risk of GI bleed

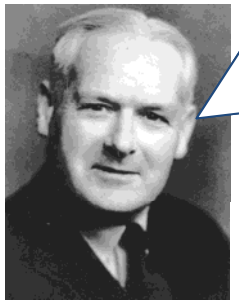


Full MiniSentinel analysis implemented in OMOP Lab:

- 1 SQL Statement
- 117 lines
- <3 minutes execution

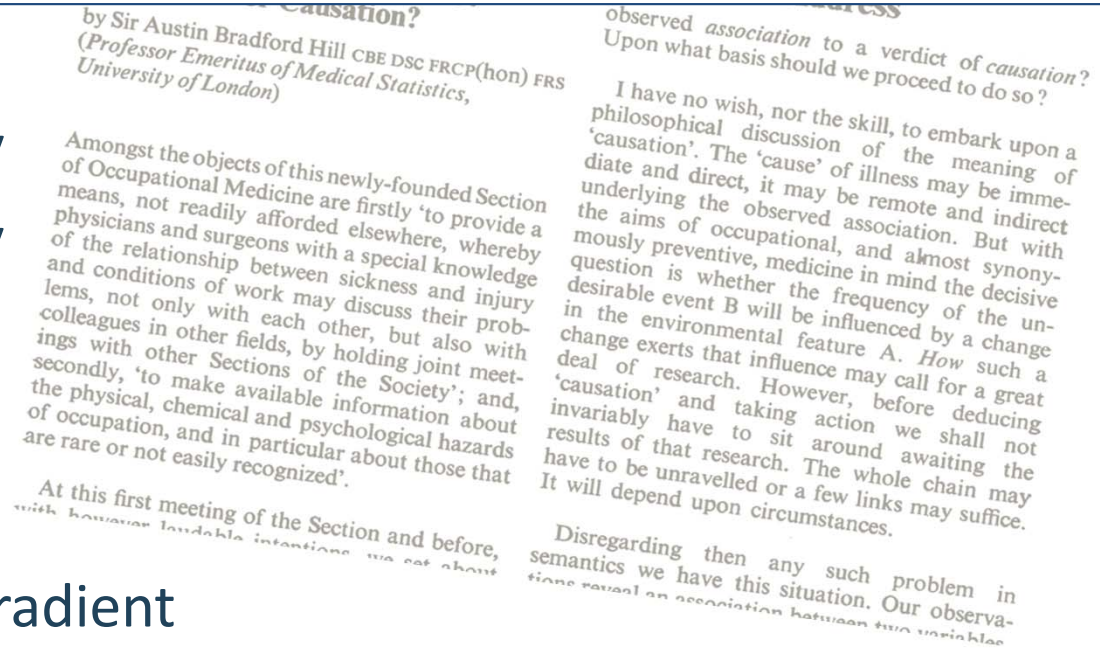
http://omop.org/2013Symposium

Studying causal effects of medical products



"What aspects of that association should we especially consider before deciding that the most likely interpretation of it is causation?"

- Strength
- Consistency
- Temporality
- Plausibility
- Experiment
- Coherence
- Biological gradient
- Specificity
- Analogy



...Causation?

by Sir Austin Bradford Hill CBE DSC FRCP(hon) FRS
(*Professor Emeritus of Medical Statistics, University of London*)

Amongst the objects of this newly-founded Section of Occupational Medicine are firstly 'to provide a means, not readily afforded elsewhere, whereby physicians and surgeons with a special knowledge of the relationship between sickness and injury and conditions of work may discuss their problems, not only with each other, but also with colleagues in other fields, by holding joint meetings with other Sections of the Society'; and, secondly, 'to make available information about the physical, chemical and psychological hazards of occupation, and in particular about those that are rare or not easily recognized'.

At this first meeting of the Section and before,

...observed *association* to a verdict of *causation*? Upon what basis should we proceed to do so?

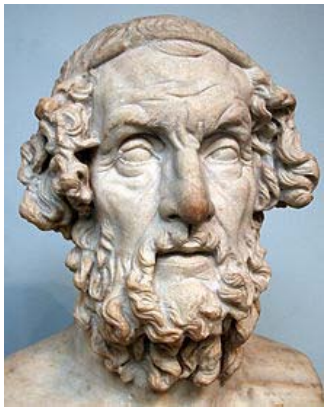
I have no wish, nor the skill, to embark upon a philosophical discussion of the meaning of 'causation'. The 'cause' of illness may be immediate and direct, it may be remote and indirect underlying the observed association. But with the aims of occupational, and almost synonymously preventive, medicine in mind the decisive question is whether the frequency of the undesirable event B will be influenced by a change in the environmental feature A. *How* such a change exerts that influence may call for a great deal of research. However, before deducing 'causation' and taking action we shall not invariably have to sit around awaiting the results of that research. The whole chain may have to be unravelled or a few links may suffice. It will depend upon circumstances.

Disregarding then any such problem in semantics we have this situation. Our observa-

Austin Bradford Hill, "The Environment and Disease: Association or Causation?," *Proceedings of the Royal Society of Medicine*, 58 (1965), 295-300.

Introducing HOMER

- Health Outcomes and Medical Effectiveness Research (HOMER) system



- Live, interactive evidence exploration system with fully functional implementations of all of the components of Sir Bradford Hill's viewpoints for risk identification and assessment, plus some additional components designed by the OMOP team

HOMER implementation of Hill's viewpoints

Demos of applications built within OHDSI community

- Concept Explorer
- EARS
- PSYCHIC

Join the journey

Interested in OHDSI?

Questions or comments?

Contact:

ryan@ohdsi.org

OMOP CDM Tips/Tricks for ETL design and development

Patrick Ryan, PhD
Janssen Research and Development
on behalf of OHDSI team
25 March 2014

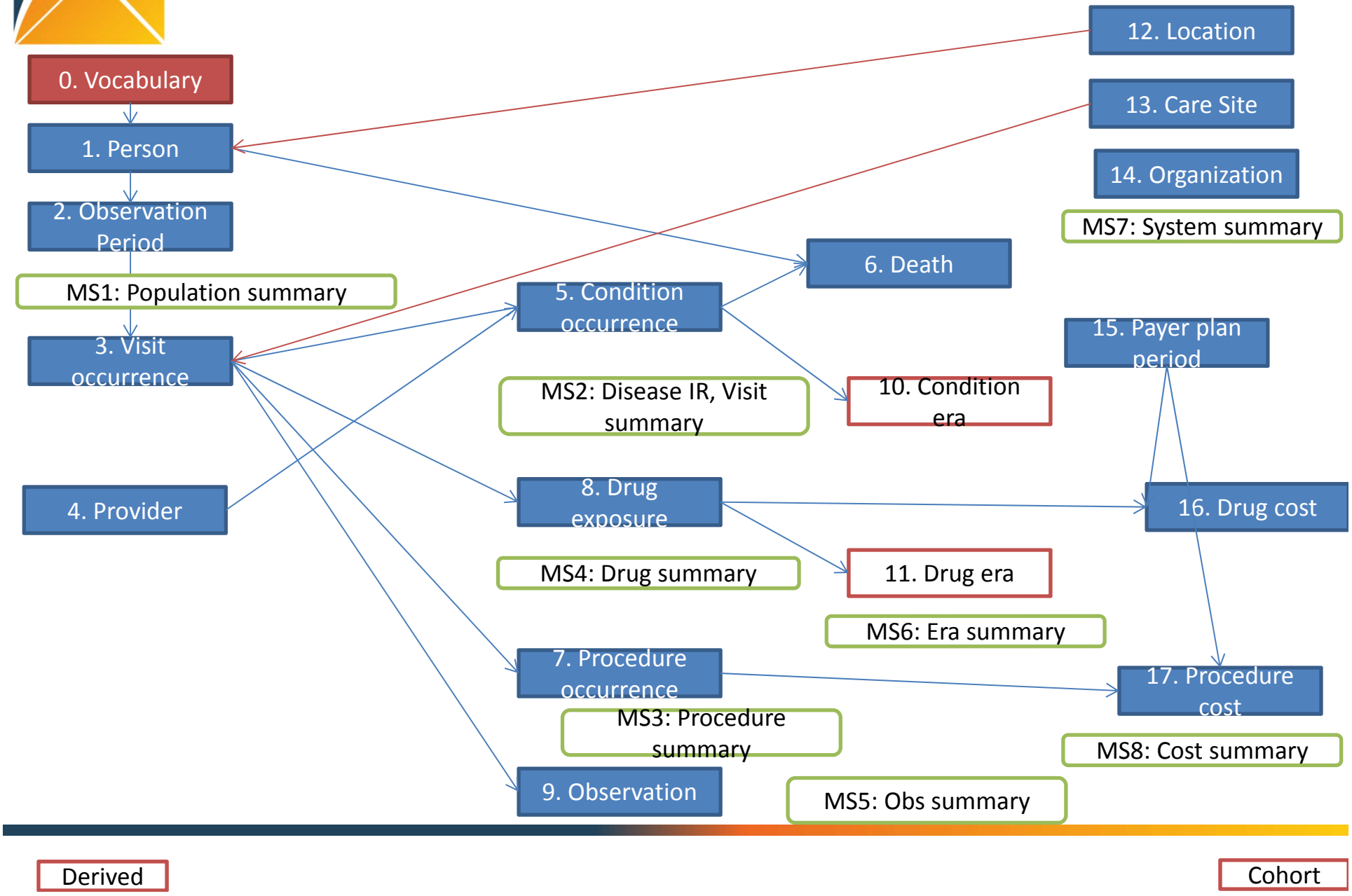
What is the biggest challenge in the area of working across various data sources or with the common data model (CDM)?

- Successful use of an observational data network requires a collaborative, interdisciplinary approach
  - **Local knowledge** of the source data: underlying data capture process and its role in the healthcare system
  - **Clinical understanding** of medical products and disease
  - **Domain expertise** in the analytical use cases: epidemiology, pharmacovigilance, health economics and outcomes research
  - Command of **advanced statistical techniques** for large-scale modeling and exploratory analysis
  - **Informatics experience** with ontology management and leveraging standard terminologies for analysis
  - **Technical/programming skills** to implement design and develop a scalable solution

Recommended steps for effective OMOP CDM ETL design

✔1. Training on OMOP CDM and Vocabulary
2. Discuss analysis opportunities (why are we doing this? What do you want to be able to do once CDM is done…)
3. Evaluate technology requirements/environment
4. Discuss data dictionary/documentation on raw database
5. Systematic scan of raw database
6. Drafting Business Logic
   - table level
   - variable level
   - value level (mapping)
   - capture what's 'lost'
7. Create data sample to allow initial development
8. DON'T START IMPLEMENTING UNTIL DESIGN IS COMPLETE

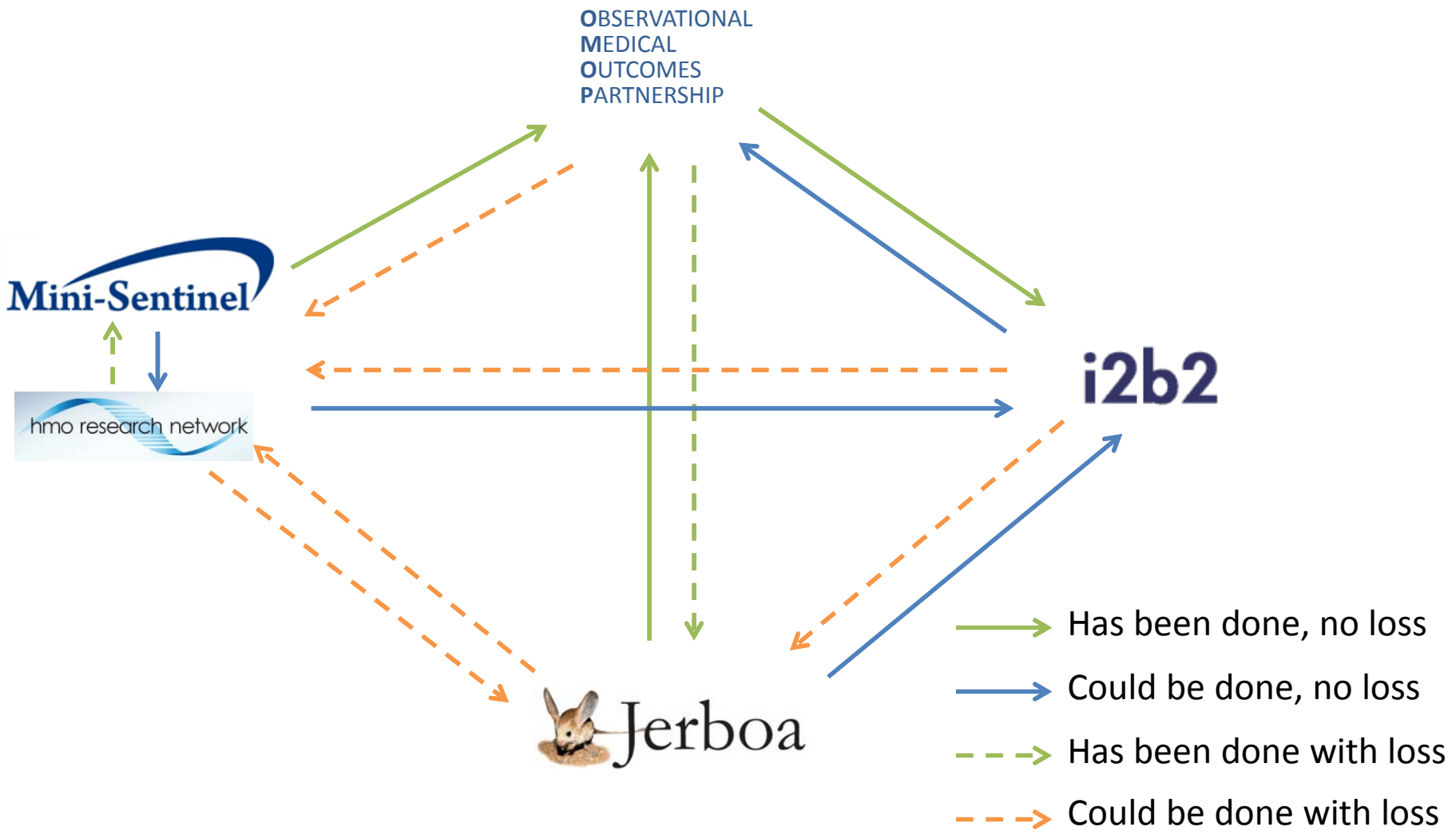
Design progression of OMOP CDM tables



Lessons learned (the hard way)

- A successful ETL requires a village, don't make one person try to be the hero and do it all themselves
  - Team design
  - Team implementation
  - Team testing
- Document early and often…the more details the better
- Data quality checking is required at every step of the process
- Don't make assumptions about data based on documentation, verify by looking at the data
- Good design and comprehensive specifications should save unnecessary iterations and thrash during implementation
- ETL design/documentation/implementation is a living process, it'll never be done and it can always be better…but don't let the perfect be the enemy of the good

Compatibility across data models through common ETLs



Join the journey

Interested in OHDSI?

Questions or comments?

Contact:

ryan@ohdsi.org