

# **Research infrastructure of the Observational Health Data Sciences and Informatics (OHDSI) consortium: Institutional and researcher's perspectives**

MEDINFO 2017

Vojtech Huser  
Christian Reich  
Rae Woong Park



# Speakers overview



- **Vojtech Huser, MD, PhD**
  - Staff scientist
  - National Institute of Health, U.S. National Library of Medicine
  
- **Rae Wong Park, MD, PhD**
  - Director of Korean Society of Medical Informatics
  - Director of Department of Biomedical Informatics, Ajou University School of Medicine
  
- **Christian Reich, MD, PhD**
  - VP Real World Insights, QuintilesIMS, USA
  - Principle Investigator OHDSI

# Agenda

- Research network description
- Case studies
  - European implementation case study
  - Asian implementation case study
  - Data quality research study
- Questions

# Agenda

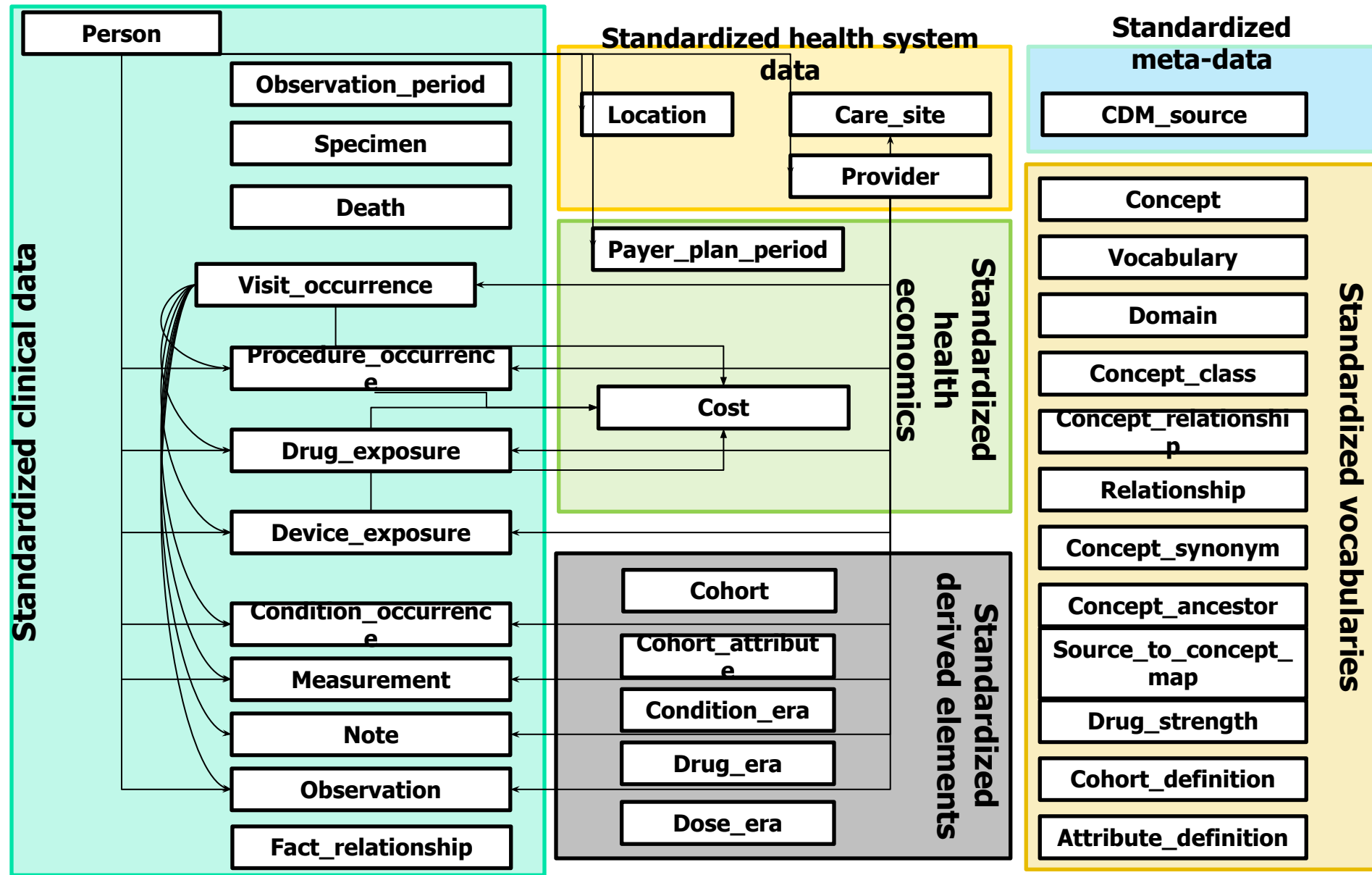
- Research network description
- Case studies
  - European implementation case study
  - Asian implementation case study
  - Data quality research study
- Questions

# History of OHDSI

- OMOP (2008-2013) [www.omop.org](http://www.omop.org)
  - OMOP = Observational Medical Outcomes Partnership
  - Research on methods for drug safety evaluation
    - Methods library developed; positive/negative drug outcome pairs
  - Common Data Model (then, was a byproduct)
  - Foundation for the NIH
    - Transition to Reagan Udall Foundation for the Food and Drug Administration
- OHDSI (after 2013) [www.ohdsi.org](http://www.ohdsi.org)
  - OHDSI = Observational Health Data Science and Informatics
  - Continues to use the name 'OMOP CDM'
  - Community of researchers; public; non-pharma funded



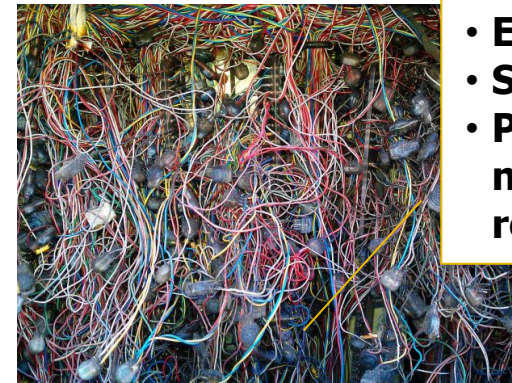
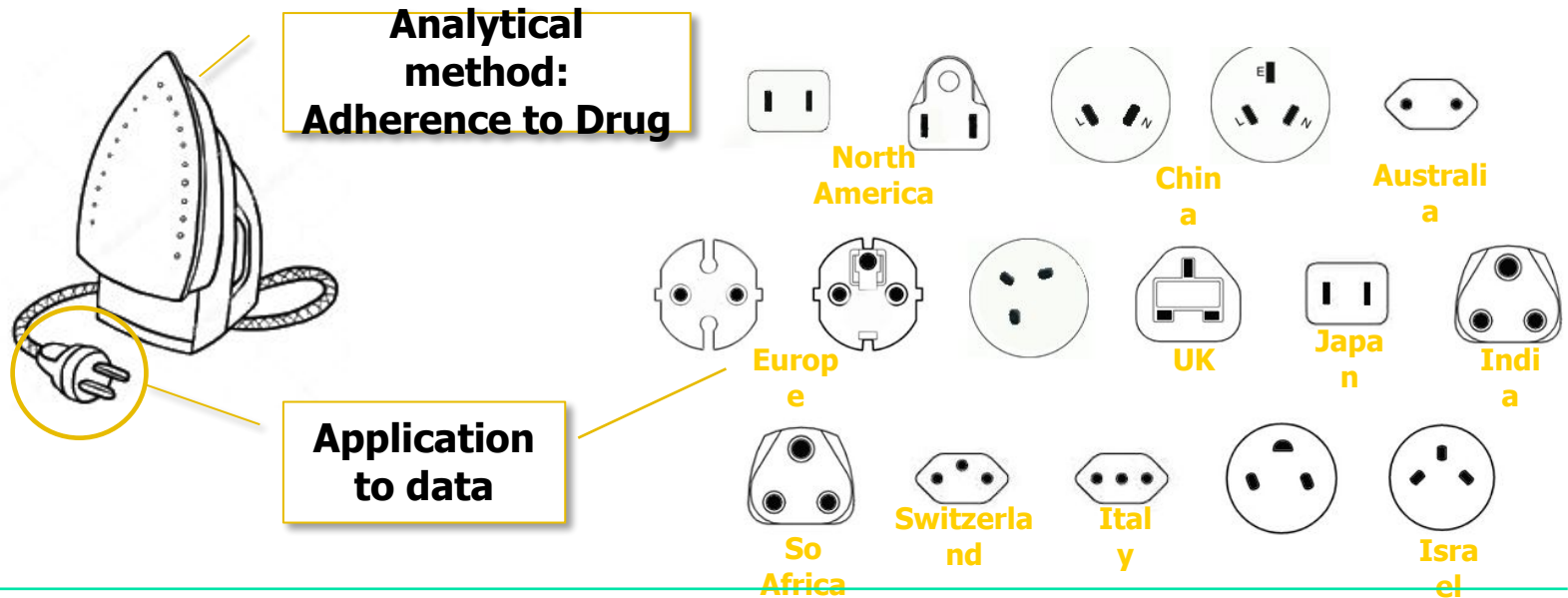
# Common Data Model



# Current Approach:

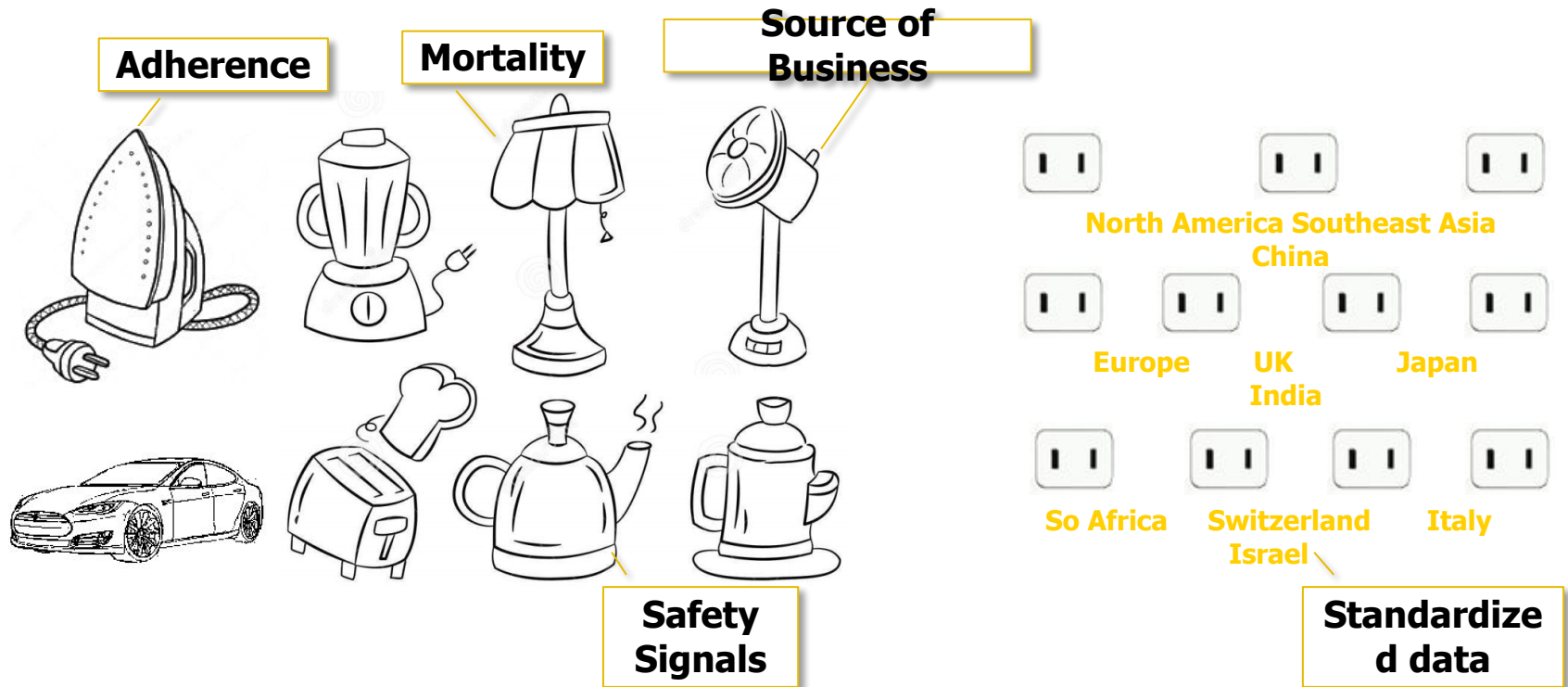
## "New, script based input data mapping for every study"

"What's the adherence to my drug in the data assets I can analyze?"



- Not scalable
- Expensive
- Slow
- Prohibitive to non-expert routine use

# Data Standardization Enables Systematic Research



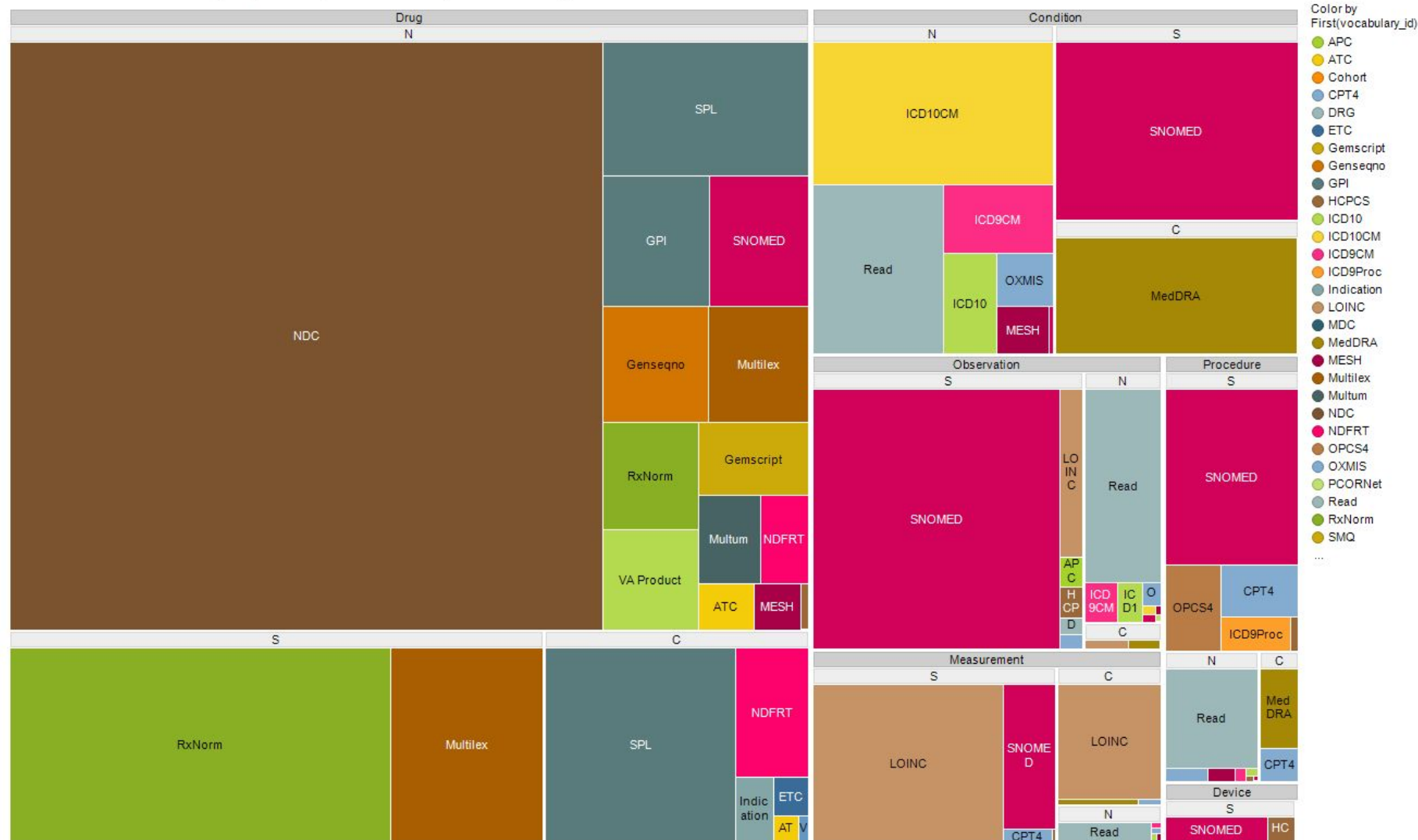
OHDSI Tools

OMOP CDM



# Standard content: OMOP Vocabularies

Breakdown of OHDSI concepts by domain, standard class, and vocabulary



# Standardized methods: ATLAS

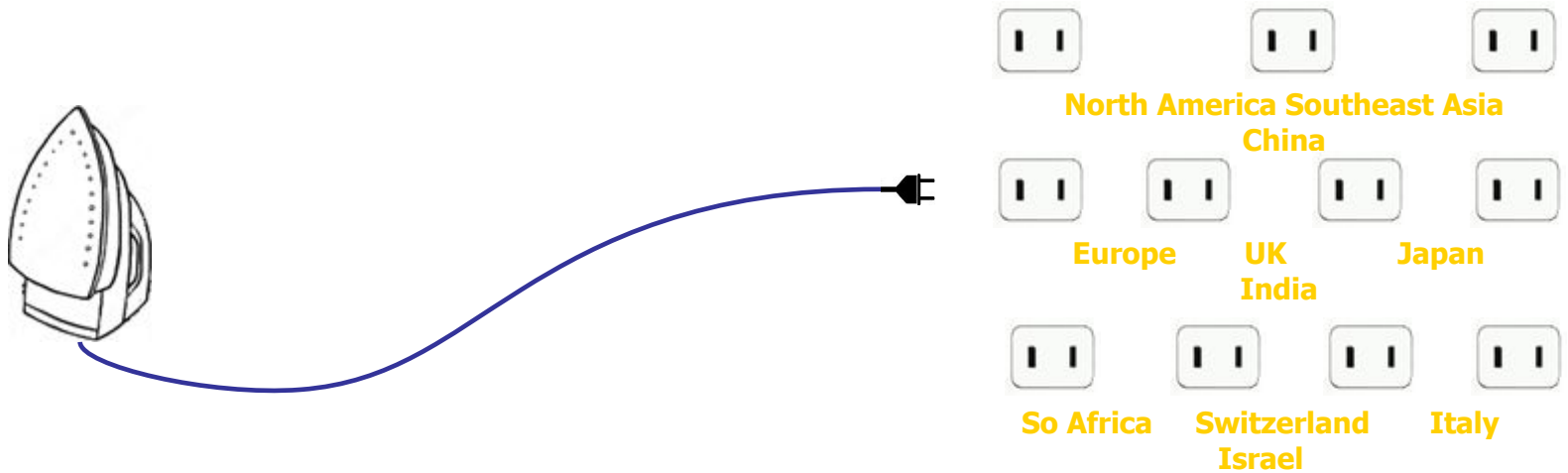
ATLAS	Home
Home	Welcome to ATLAS.
Data Sources	ATLAS is an open source application developed
Vocabulary	Documentation
Concept Sets	The ATLAS user guide can be found <a href="#">here</a> .
Cohorts	Getting Started
Incidence Rates	Define a New Cohort
Profiles	Search the Vocabulary
Estimation	Release Notes
Jobs	ATLAS Version 1.2.0 Current Release Notes <ul style="list-style-type: none"><li>Cohort Definition End Dates</li><li>New Feature: Estimation</li><li>Concept Set Copy Enabled</li></ul>
Configuration	ATLAS Version 1.1.0 Current Release Notes <ul style="list-style-type: none"><li>Fix: Cohort Definition UI bugfixes</li><li>Fix: Data Sources report fixes (Measurement reports and trellis graph rendering NaNs)</li><li>Feature: Cohort Definitions (CIRCE) UI and Backend Changes: Inclusion rules and Inclusion Rule Impact Reports</li><li>Feature: Patient Profiles</li></ul>
R Services	WebAPI Version 1.2.0 Current Release Notes <ul style="list-style-type: none"><li>Feature: Cohort generation update to support inclusion rules</li><li>Feature: Inclusion rule impact analysis</li></ul>
Feedback	

## Design your study

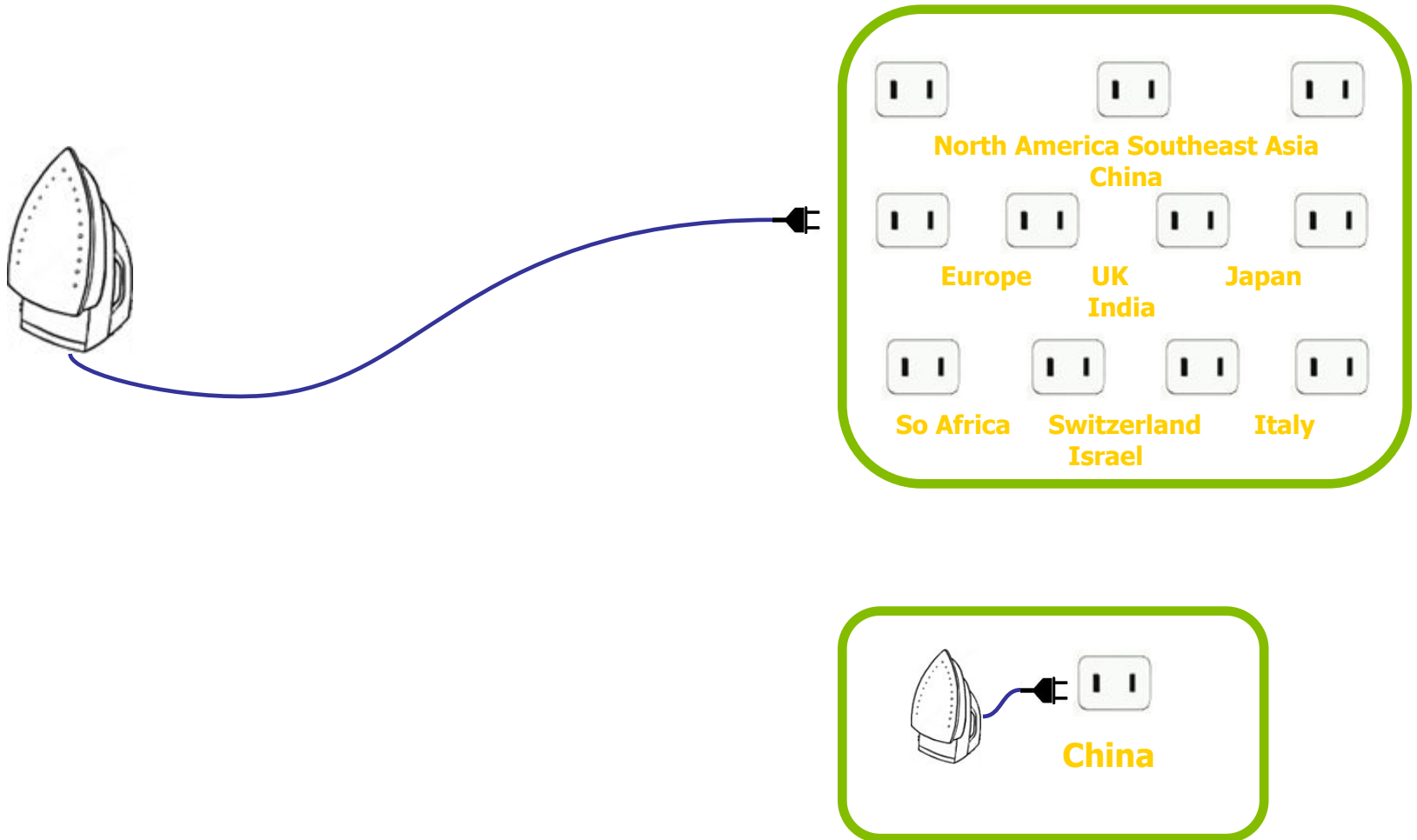
- What's your target cohort?
- What's your compactor cohort?
- What's your outcome cohort?
- What's your time-at-risk?
- What's your model specification?
- What's your covariate adjustment strategy?

## Run

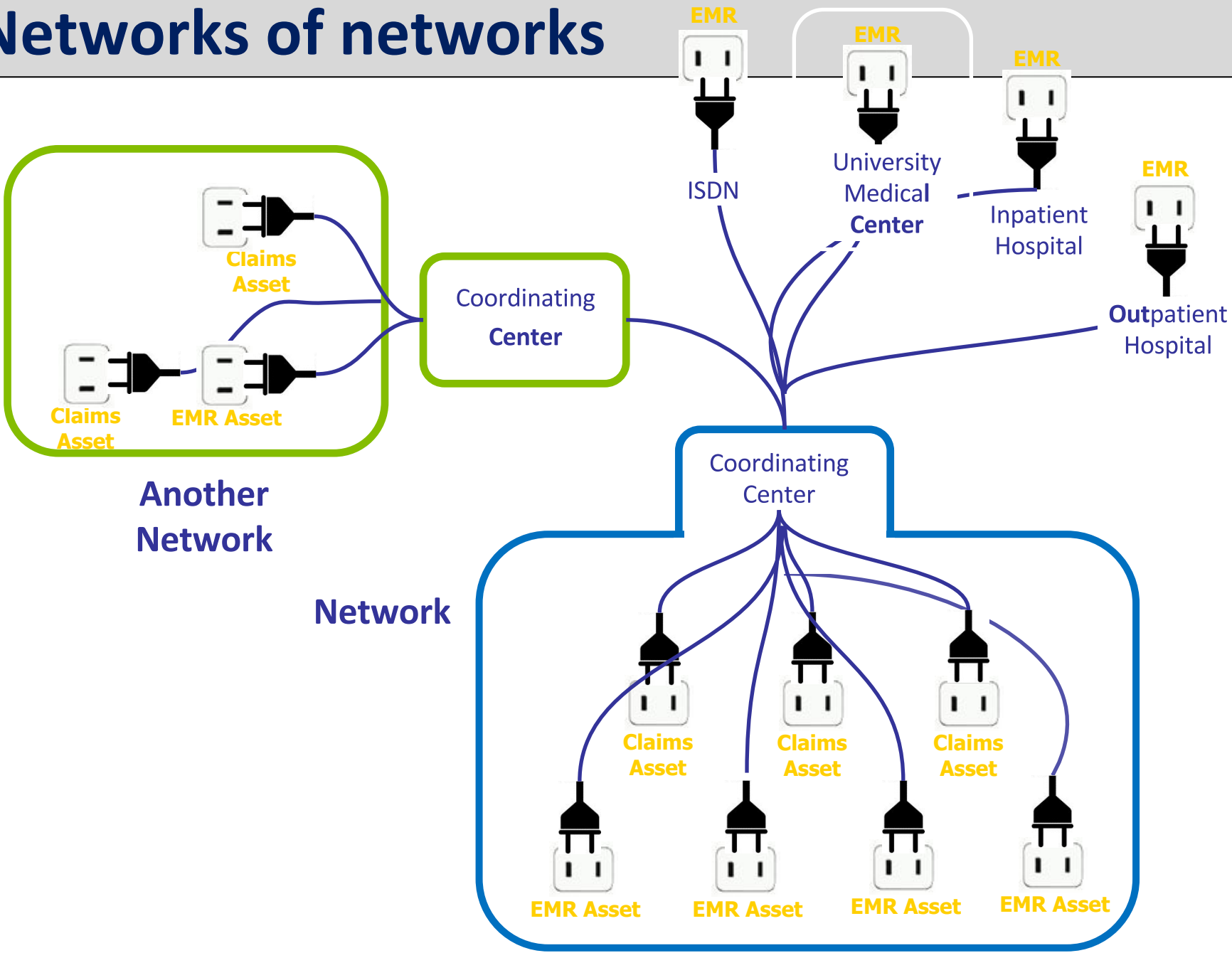
# Analytics can be remote



# Analytics can be behind firewall



# Networks of networks



# Past network studies



## ■ Clinical studies

- 2015
  - Treatment Pathway Study (diabetes, hypertension, depression)
- 2016
  - Levetiracetam vs. phenytoin in epilepsy
  - Comparison of combination treatment in hypertension\*
- 2017
  - Sisyphus challenge (Alendronate vs. Raloxifene for osteoporosis)\*
- Other
  - Anticoagulants, Prediction, Celecoxib vs. nsNSAIDs

## ■ Informatics studies

- 2015: Pediatric drug use epidemiology study
- 2016: Achilles Heel Evaluation study
- 2017: Data Quality

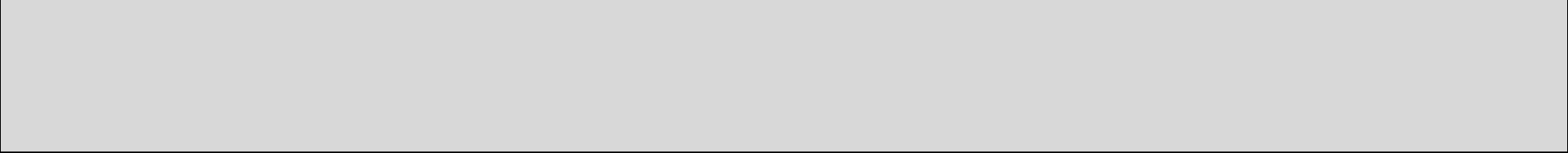
# Focus of the panel

- Institutional perspective on OHDSI
  - Case studies 1 and 2
- Researcher's perspective on OHDSI
  - Case study 3

# Agenda

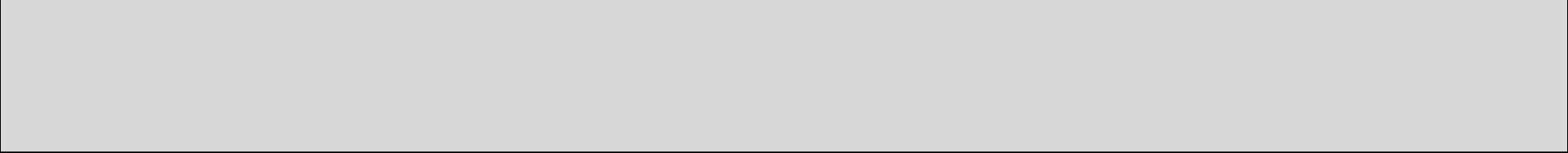
- Research network description
- Case studies
  - European implementation case study
  - Asian implementation case study
  - Researcher's perspective: Data quality study
- Questions





# Agenda

- Research network description
- Case studies
  - European implementation case study
  - Asian implementation case study
  - Researcher's perspective: Data quality study
- Questions



**Vojtech Huser**

# Agenda

- Research network description
- Case studies
  - European implementation case study
  - Asian implementation case study
  - **Researcher's perspective: Data quality study**
- Questions

# OHDSI: Researcher's perspective

## ■ Strengths

- Analysis portability
  - Analysis written at one site can possibly be executed by other partners within the consortium
- Common Data Model, OMOP Vocabularies
- Tools + R packages
- Community of researchers, past studies are open source

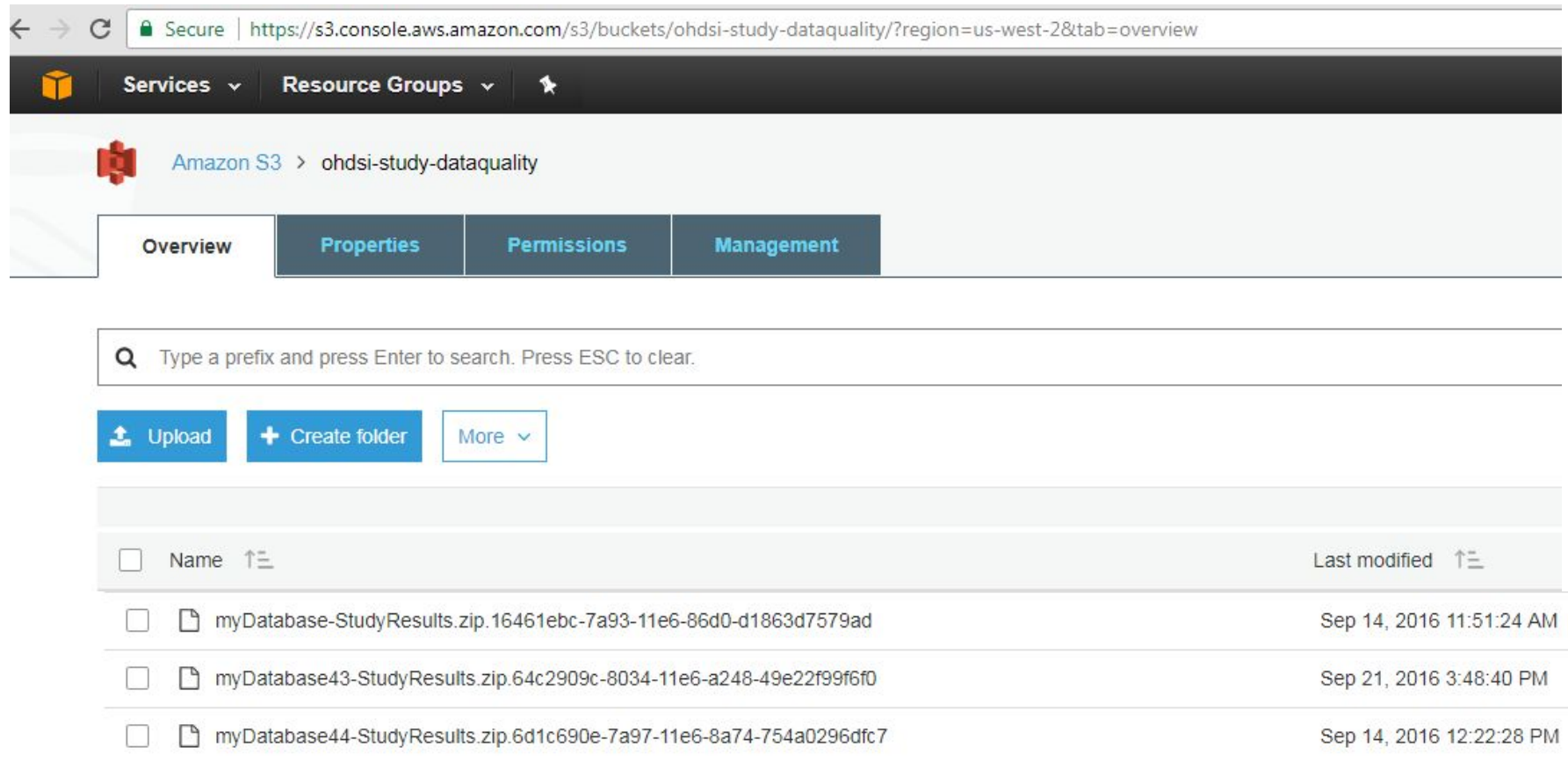
## ■ Weaknesses

- Must have resources for data transformation to CDM
- Expertise to install/use OHDSI tools and packages






# Network results aggregation

- Study conventions (R package) (STEP 1: Local Execution)
  - `install_github("ohdsi/StudyProtocolSandbox/DataQuality")`
  - `executeDQ(connectionDetails,cdmDatabaseSchema,resultsDatabaseSchema,workFolder='c:/mystudy')`
- Package results (STEP 2)
  - .zip file which a site researcher inspects closely
  - `install_github("ohdsi/OhdsiSharing")`
  - `packageResults(...,workFolder,dbName)`
  - `submitResults(...,studyBucketName,studyKey,studySecret)`
- STEP 3: Aggregated data analysis
- Full example
  - <https://github.com/OHDSI/StudyProtocolSandbox/tree/master/DataQuality#2participate-on-dataquality-study>

# Data logistics (example)



The screenshot shows the Amazon S3 console interface. The browser address bar displays the URL: <https://s3.console.aws.amazon.com/s3/buckets/ohdsi-study-dataquality/?region=us-west-2&tab=overview>. The console header includes the Amazon logo, 'Services', and 'Resource Groups'. The breadcrumb navigation shows 'Amazon S3 > ohdsi-study-dataquality'. Below this, there are four tabs: 'Overview' (selected), 'Properties', 'Permissions', and 'Management'. A search bar is present with the placeholder text 'Type a prefix and press Enter to search. Press ESC to clear.' Below the search bar are three buttons: 'Upload', 'Create folder', and 'More'. The main content area displays a table of objects in the bucket.

<input type="checkbox"/>	Name 	Last modified 
<input type="checkbox"/>	 myDatabase-StudyResults.zip.16461ebc-7a93-11e6-86d0-d1863d7579ad	Sep 14, 2016 11:51:24 AM
<input type="checkbox"/>	 myDatabase43-StudyResults.zip.64c2909c-8034-11e6-a248-49e22f99f6f0	Sep 21, 2016 3:48:40 PM
<input type="checkbox"/>	 myDatabase44-StudyResults.zip.6d1c690e-7a97-11e6-8a74-754a0296dfc7	Sep 14, 2016 12:22:28 PM



# OMOP Vocabularies

- Common framework
  - CONCEPT, CONCEPT\_RELATIONSHIP, CONCEPT\_ANCESTOR
- Benefit: pre-build infrastructure [+ mapping]
- Browser
  - <http://www.ohdsi.org/web/atlas/#/concept/21600381>
- Example of a researcher benefit
  - ICD10CM -> SNOMED CT (after Oct 1<sup>st</sup>, 2015)
  - ICD9CM -> SNOMED CT

# Acute renal failure

← → ↻ [www.ohdsi.org/web/atlas/#/concept/197320](http://www.ohdsi.org/web/atlas/#/concept/197320)

## ATLAS

🔍 Acute renal failure syndrome

Details Related Concepts Hierarchy Record Counts

Column visibility Copy CSV Show All entries Filter:

Showing 1 to 18 of 18 entries

Previous 1 Next

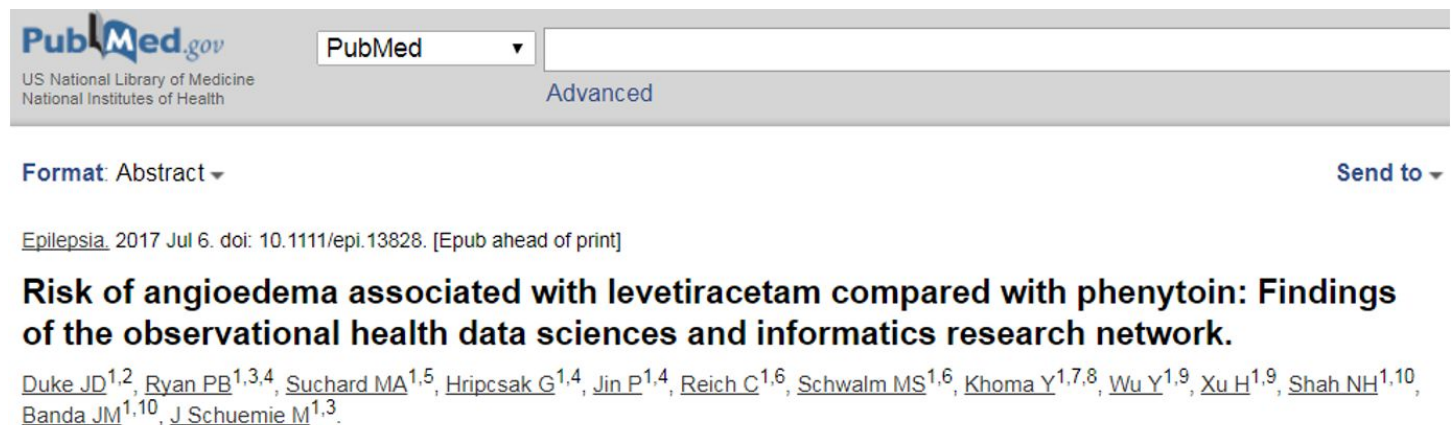
	Id	Code	Name	Class	RC	DRC	Distance	Domain	Vocabulary
▼ Vocabulary									
ICD9CM (10)	44833392	283.11	Hemolytic-uremic syndrome	5-dig billing code	0	0	1	Condition	ICD9CM
ICD10 (9)	44825529	572.4	Hepatorenal syndrome	4-dig billing code	0	0	1	Condition	ICD9CM
ICD10CM (8)	44833658	584	Acute kidney failure	3-dig nonbill code	0	0	1	Condition	ICD9CM
OXMIS (4)	44837189	584.6	Acute kidney failure with lesion of renal cortical necrosis	4-dig billing code	0	0	1	Condition	ICD9CM
Cohort (4)	44820969	584.8	Acute kidney failure with other specified pathological lesion in kidney	4-dig billing code	0	0	1	Condition	ICD9CM
▼ Standard Concept	44826731	584.9	Acute kidney failure, unspecified	4-dig billing code	0	0	1	Condition	ICD9CM
Standard (95)	44823297	669.3	Acute kidney failure following labor and delivery	4-dig nonbill code	0	0	1	Condition	ICD9CM
Non-Standard (89)	44827985	669.30	Acute kidney failure following labor and delivery, unspecified as to episode of care or not applicable	5-dig billing code	0	0	1	Condition	ICD9CM
Classification (5)	44837293	669.32	Acute kidney failure following labor and delivery, delivered, with mention of postpartum complication	5-dig billing code	0	0	2	Condition	ICD9CM
▼ Invalid Reason	44824443	669.34	Acute kidney failure following labor and delivery, postpartum condition or complication	5-dig billing code	0	0	2	Condition	ICD9CM
Valid (173)	35206734	D59.3	Hemolytic-uremic syndrome	4-char billing code	0	0	1	Condition	ICD10CM
Invalid (16)	35208366	K76.7	Hepatorenal syndrome	4-char billing code	0	0	1	Condition	ICD10CM
▼ Class	45552960	K91.83	Postprocedural hepatorenal syndrome	5-char billing code	0	0	2	Condition	ICD10CM
Clinical Finding (108)	1571485	N17	Acute kidney failure	3-char nonbill code	0	0	1	Condition	ICD10CM
Read (42)	35209270	N17.1	Acute kidney failure with acute cortical necrosis	4-char billing code	0	0	1	Condition	ICD10CM
ICD10 code (8)	35209272	N17.8	Other acute kidney failure	4-char billing code	0	0	1	Condition	ICD10CM
4-char billing code (6)	35209273	N17.9	Acute kidney failure, unspecified	4-char billing code	0	0	1	Condition	ICD10CM
Cohort (4)	35210387	O90.4	Postpartum acute kidney failure	4-char billing code	0	0	1	Condition	ICD10CM
▼ Domain									
Condition (171)									
Measurement (9)									
Procedure (5)									
Observation (2)									
Spec Anatomic Site (1)									
▼ Relationship									
Has relation to descendant of: Non-standard to Standard map (OMOP) (71)									
Has ancestor of (25)									
Has relation to descendant of: Non-standard to Standard map (OMOP) Has									

Showing 1 to 18 of 18 entries

Previous 1 Next

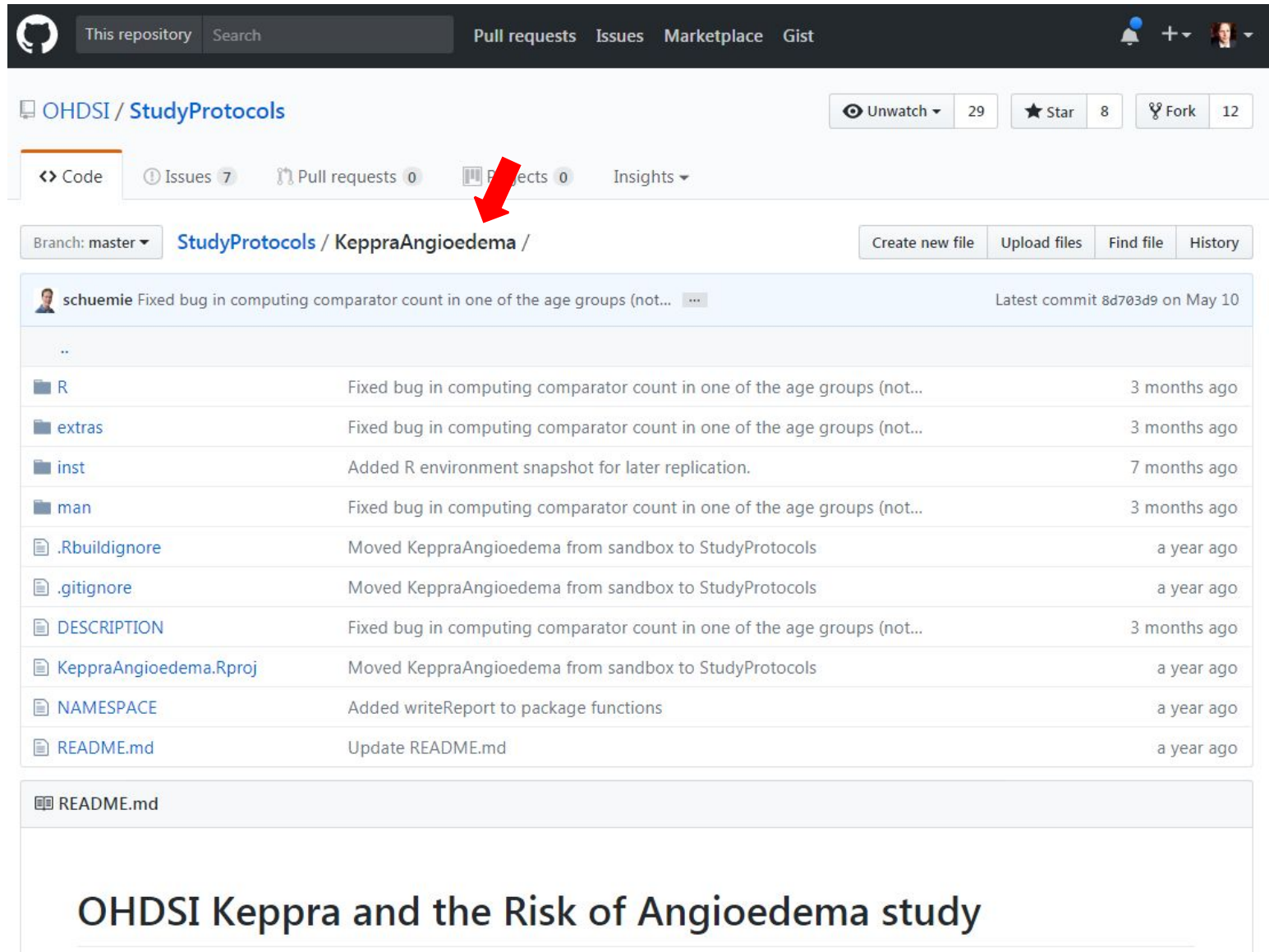
# Learning from past studies

- Evolution over time
  - increasing sophistication (SQL, SQL+R packages, portable phenotypes)
- Anti-epilepsy drug analysis (levetiracetam) (“second generation”)



- <https://github.com/OHDSI/StudyProtocols>
- <https://github.com/OHDSI/StudyProtocolSandbox/>

# Study package (in R language)



This repository Search Pull requests Issues Marketplace Gist

OHDSI / StudyProtocols Unwatch 29 Star 8 Fork 12

<> Code Issues 7 Pull requests 0 Projects 0 Insights

Branch: master StudyProtocols / KeppraAngioedema / Create new file Upload files Find file History

**schuemie** Fixed bug in computing comparator count in one of the age groups (not... Latest commit 8d703d9 on May 10


..		
R	Fixed bug in computing comparator count in one of the age groups (not...	3 months ago
extras	Fixed bug in computing comparator count in one of the age groups (not...	3 months ago
inst	Added R environment snapshot for later replication.	7 months ago
man	Fixed bug in computing comparator count in one of the age groups (not...	3 months ago
.Rbuildignore	Moved KeppraAngioedema from sandbox to StudyProtocols	a year ago
.gitignore	Moved KeppraAngioedema from sandbox to StudyProtocols	a year ago
DESCRIPTION	Fixed bug in computing comparator count in one of the age groups (not...	3 months ago
KeppraAngioedema.Rproj	Moved KeppraAngioedema from sandbox to StudyProtocols	a year ago
NAMESPACE	Added writeReport to package functions	a year ago
README.md	Update README.md	a year ago

README.md

## OHDSI Keppra and the Risk of Angioedema study

# Value Set definition (“second generation” study example)

Branch: master ▾ StudyProtocols / KeppraAngioedema / inst / sql / sql\_server / Angioedema.sql Find file Copy path

 schuemie Moved KeppraAngioedema from sandbox to StudyProtocols 0177874 on May 9, 2016

1 contributor

109 lines (91 sloc) 3.75 KB Raw Blame History

```
1 select codeset_id, concept_id
2 INTO #Codesets
3 FROM
4 (
5   SELECT 0 as codeset_id, c.concept_id FROM (select distinct I.concept_id FROM
6   (
7     select DISTINCT concept_id from @cdm_database_schema.CONCEPT where concept_id in (711584) and invalid_reason is null
8     UNION
9
10    select c.concept_id
11    from @cdm_database_schema.CONCEPT c
12    join @cdm_database_schema.CONCEPT_ANCESTOR ca on c.concept_id = ca.descendant_concept_id
13    and ca.ancestor_concept_id in (711584)
14    and c.invalid_reason is null
15  ) I
16 ) C
17 ) C
18 UNION
19 SELECT 1 as codeset_id, c.concept_id FROM (select distinct I.concept_id FROM
20 (
21   select DISTINCT concept_id from @cdm_database_schema.CONCEPT where concept_id in (4029498) and invalid_reason is null
22   UNION
23
24   select c.concept_id
25   from @cdm_database_schema.CONCEPT c
26   join @cdm_database_schema.CONCEPT_ANCESTOR ca on c.concept_id = ca.descendant_concept_id
```



# “Third generation” study

The screenshot shows a web browser displaying the OHDSI website. The address bar shows the URL: [www.ohdsi.org/web/wiki/doku.php?id=research:bisphosp...](http://www.ohdsi.org/web/wiki/doku.php?id=research:bisphosp...). The OHDSI logo is on the left, and the text "Observational Health Data Sciences and Informatics" is next to it. A search bar is on the right with the word "Search" inside. Below the search bar are links for "Recent Changes", "Media Manager", and "Sitemap". A "Log In" link is also visible. A breadcrumb trail shows "Trace: • bisphosphonates\_and\_hip\_fracture". A sidebar on the left is labeled "Sidebar". The main content area shows the title "Comparative effectiveness of alendronate and raloxifene in reducing the risk of hip fracture" in large, bold letters. Below the title, there is a box containing the "Objective" and "Rationale" sections.

Trace: • [bisphosphonates\\_and\\_hip\\_fracture](#)

Search

[Recent Changes](#) [Media Manager](#) [Sitemap](#) [Log In](#)

research:bisphosphonates\_and\_hip\_fracture

## Comparative effectiveness of alendronate and raloxifene in reducing the risk of hip fracture

**Objective:** To compare the effectiveness in reducing the risk of hip fracture between alendronate and raloxifene.

**Rationale:** Osteoporosis is characterized by decreased bone mass and deterioration of bone tissue, resulting in reduced bone strength and increased fracture risk. Approved therapies for osteoporosis include bisphosphonates, calcitonin, raloxifene and teriparatide. Among these drugs, alendronate and raloxifene are the most popular osteoporosis medication and a burden of prescription are performed annually.

# “Third generation” study

Branch: master ▾

[StudyProtocols](#) / [AlendronateVsRaloxifene](#) / [inst](#) / [settings](#) / [CohortsToCreate.csv](#)

 msuchard move into sub

1 contributor

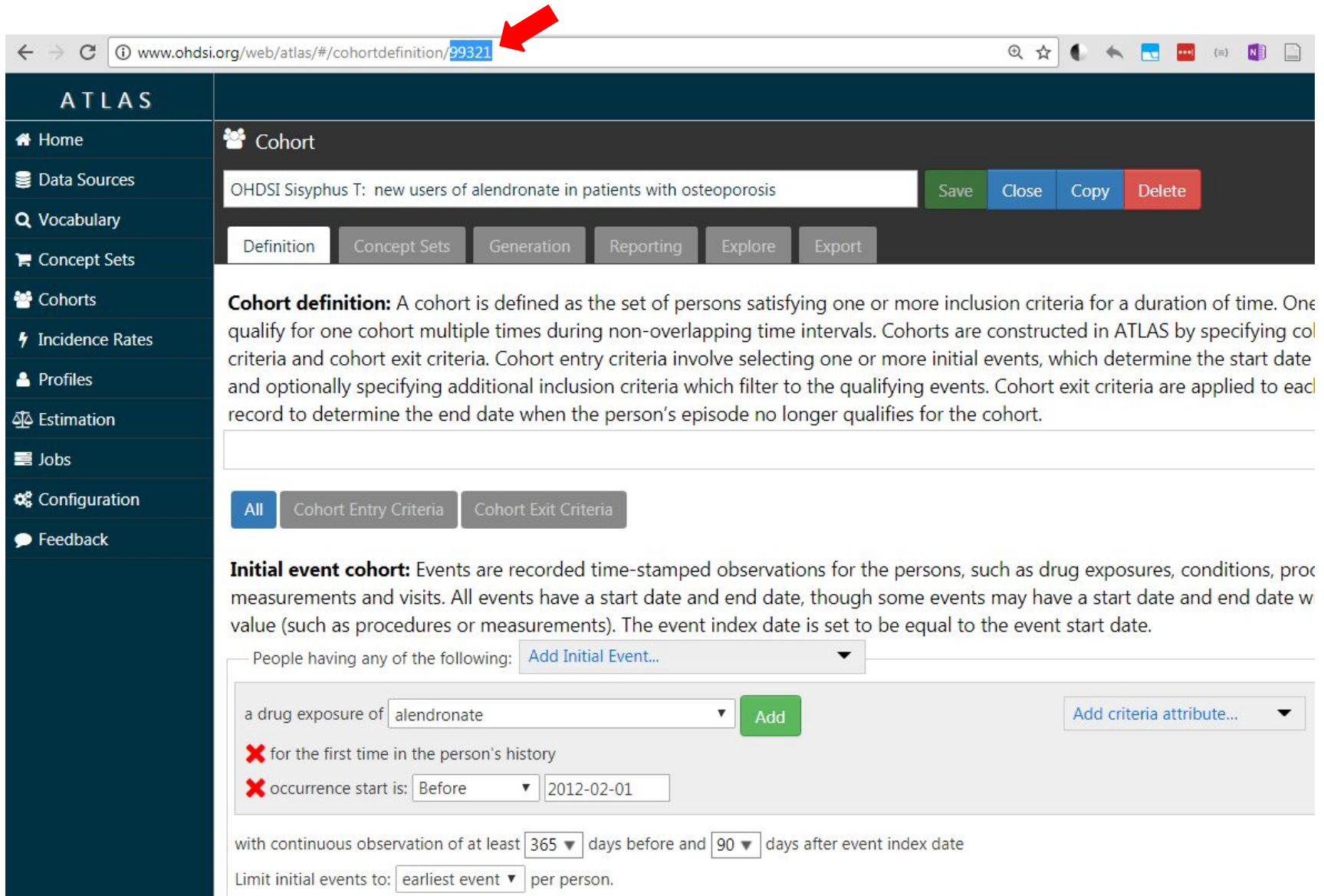
10 lines (9 sloc) | 268 Bytes

Raw

🔍 Search this file...

	cohortId	atlasId	name
1			
2	99321	99321	Alendronate
3	99322	99322	Raloxifene
4	99323	99323	HipFracture
5	100791	100791	VertebralFracture
6	100792	100792	NonHipNonVertebralFracture
7	100793	100793	OsteonecrosisOfJaw
8	100794	100794	EsophagealCancer
9	100795	100795	AtypicalFemuralFracture

# Web-based phenotype definition



ATLAS

Home

Data Sources

Vocabulary

Concept Sets

Cohorts

Incidence Rates

Profiles

Estimation

Jobs

Configuration

Feedback

Cohort

OHDSI Sisyphus T: new users of alendronate in patients with osteoporosis

Save Close Copy Delete

Definition Concept Sets Generation Reporting Explore Export

**Cohort definition:** A cohort is defined as the set of persons satisfying one or more inclusion criteria for a duration of time. One qualify for one cohort multiple times during non-overlapping time intervals. Cohorts are constructed in ATLAS by specifying col criteria and cohort exit criteria. Cohort entry criteria involve selecting one or more initial events, which determine the start date and optionally specifying additional inclusion criteria which filter to the qualifying events. Cohort exit criteria are applied to each record to determine the end date when the person's episode no longer qualifies for the cohort.

All Cohort Entry Criteria Cohort Exit Criteria

**Initial event cohort:** Events are recorded time-stamped observations for the persons, such as drug exposures, conditions, procedures, measurements and visits. All events have a start date and end date, though some events may have a start date and end date with no value (such as procedures or measurements). The event index date is set to be equal to the event start date.

People having any of the following: Add Initial Event...

a drug exposure of alendronate Add

Add criteria attribute...

✗ for the first time in the person's history

✗ occurrence start is: Before 2012-02-01

with continuous observation of at least 365 days before and 90 days after event index date

Limit initial events to: earliest event per person.



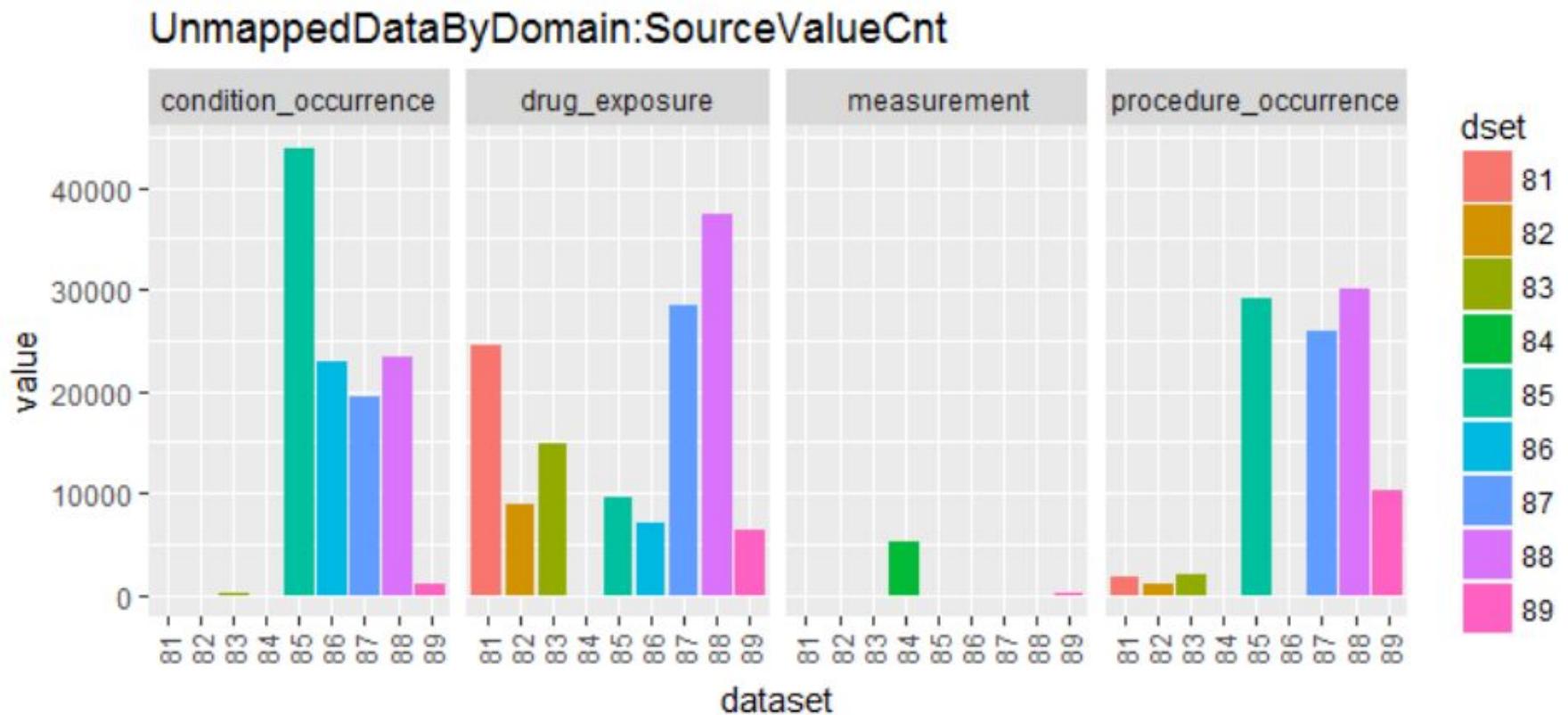
# Data Quality Study

- 12+ datasets (from 7 sites)
  - dataset metadata (least aggressive data view)
    - re-using dataset characterization pre-computations (from Achilles OHDSI tool;
      - Number of distinct procedure concepts per person
  - Ethical review; US: IRB (=institutional review board)
    - OHDSI central IRB
- Empirical comparison
  - DQ rules vs. empirical thresholds
    - % of patients with at least one visit

	median	percentile10	min	max
ach_2003:Percentage (1+ visit)	89.96	62.74	37.82	100

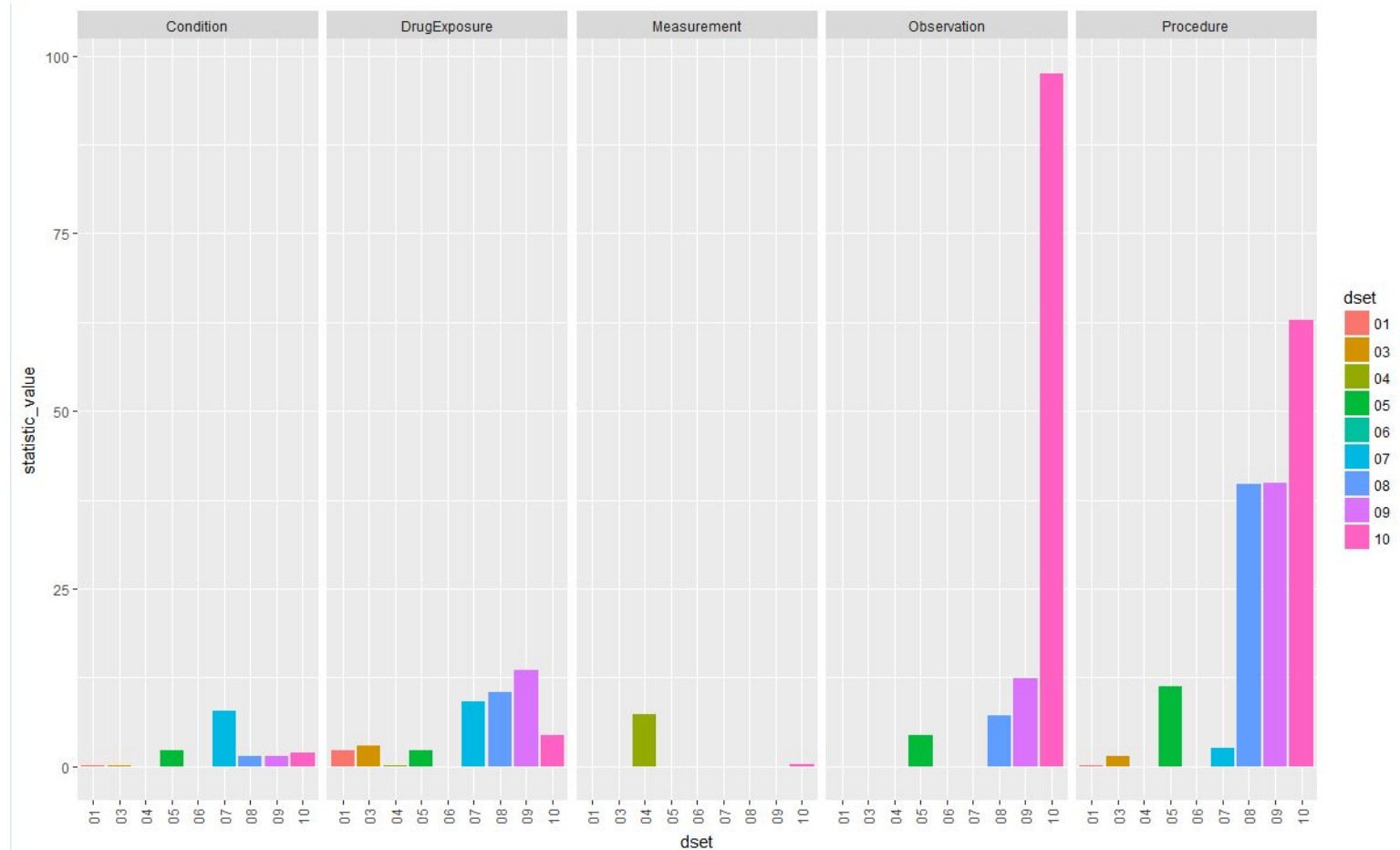
# Unmapped data (example #1)

- Count of local codes not mapped to a standard



# Unmapped data (example #2)

- Percentage of unmapped data by domain



## Questions

# Agenda

- Research network description
- Case studies
  - European implementation case study
  - Asian implementation case study
  - Researcher's perspective: Data quality study
- Questions