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

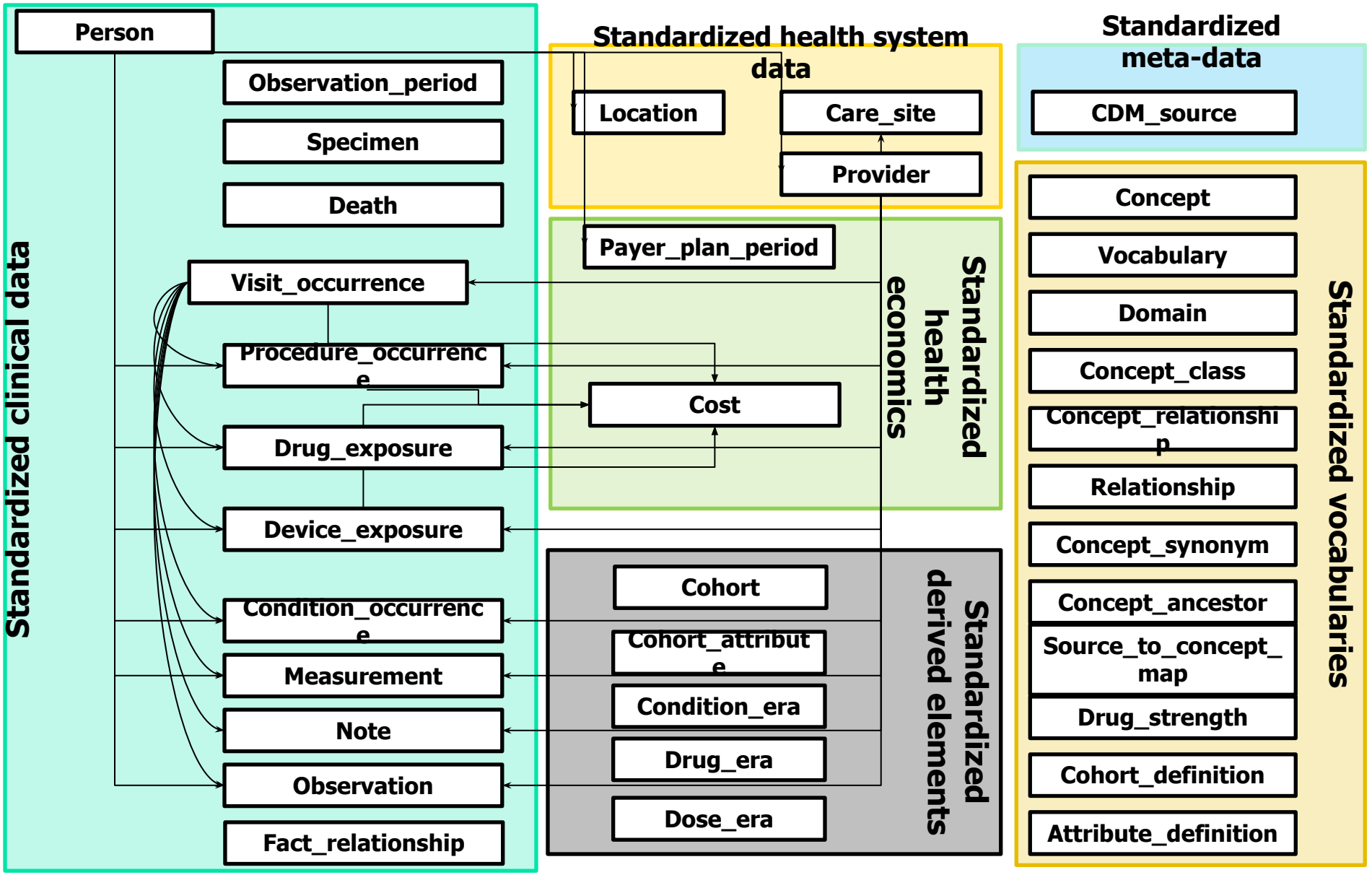
Agenda

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

History of OHDSI

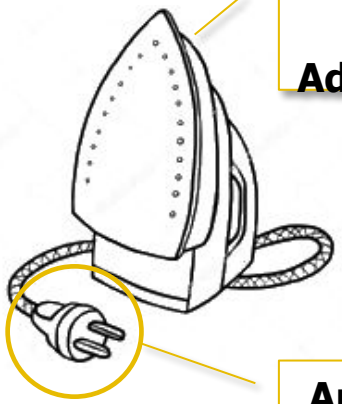
- OMOP (2008-2013) www.omop.org
  - OMOP = Observational Medical Outcomes Partnership
  - Research on methods for drug safety evaluation
    - Methods library developed; positive/negative drug outcome pairs
  - Common Data Model (then, was a byproduct)
  - Foundation for the NIH
    - Transition to Reagan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

Common Data Model



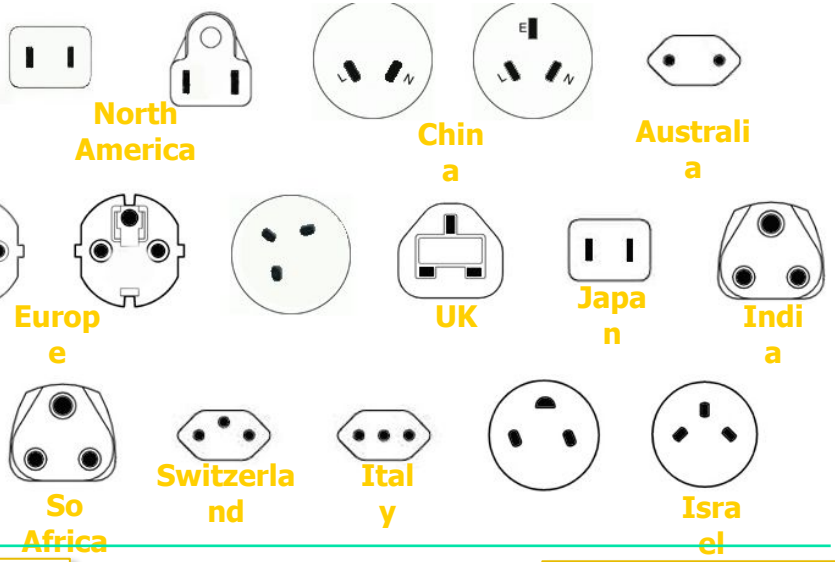
Current Approach:
“New, script based input data mapping for every study“

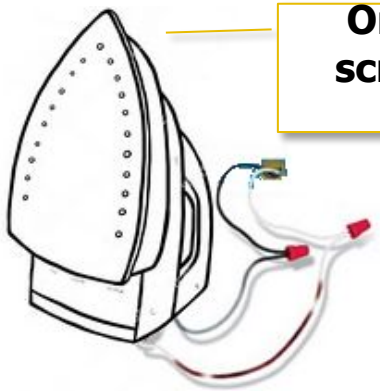
"What's the adherence to my drug in the data assets I can analyze?"



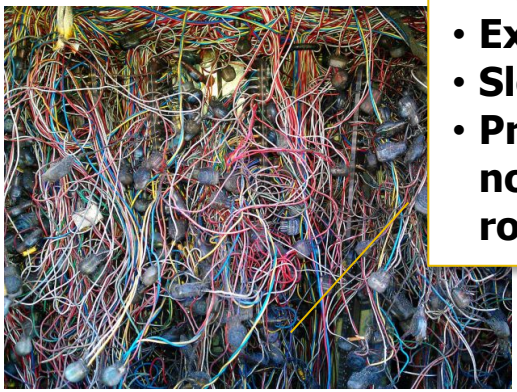
Analytical method: Adherence to Drug

Application to data



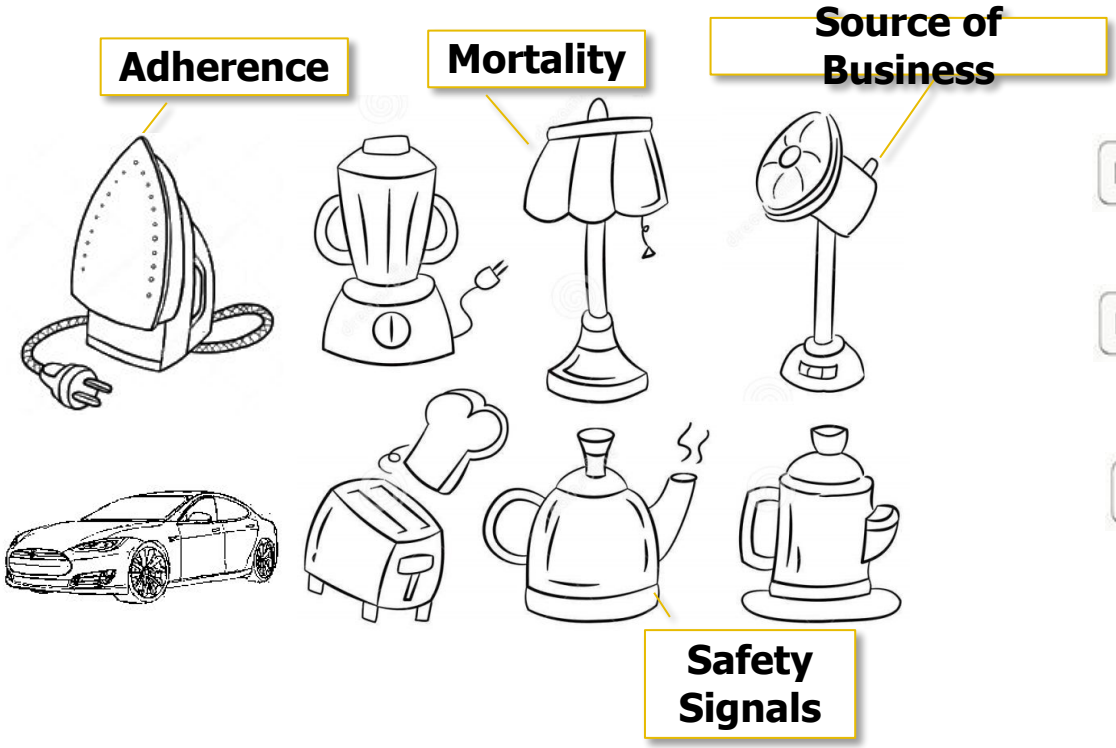


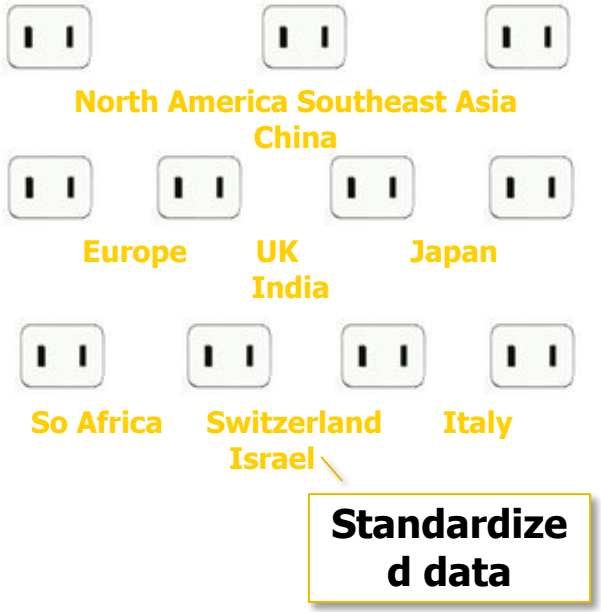
One SAS or R script for each study



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

Data Standardization Enables Systematic Research

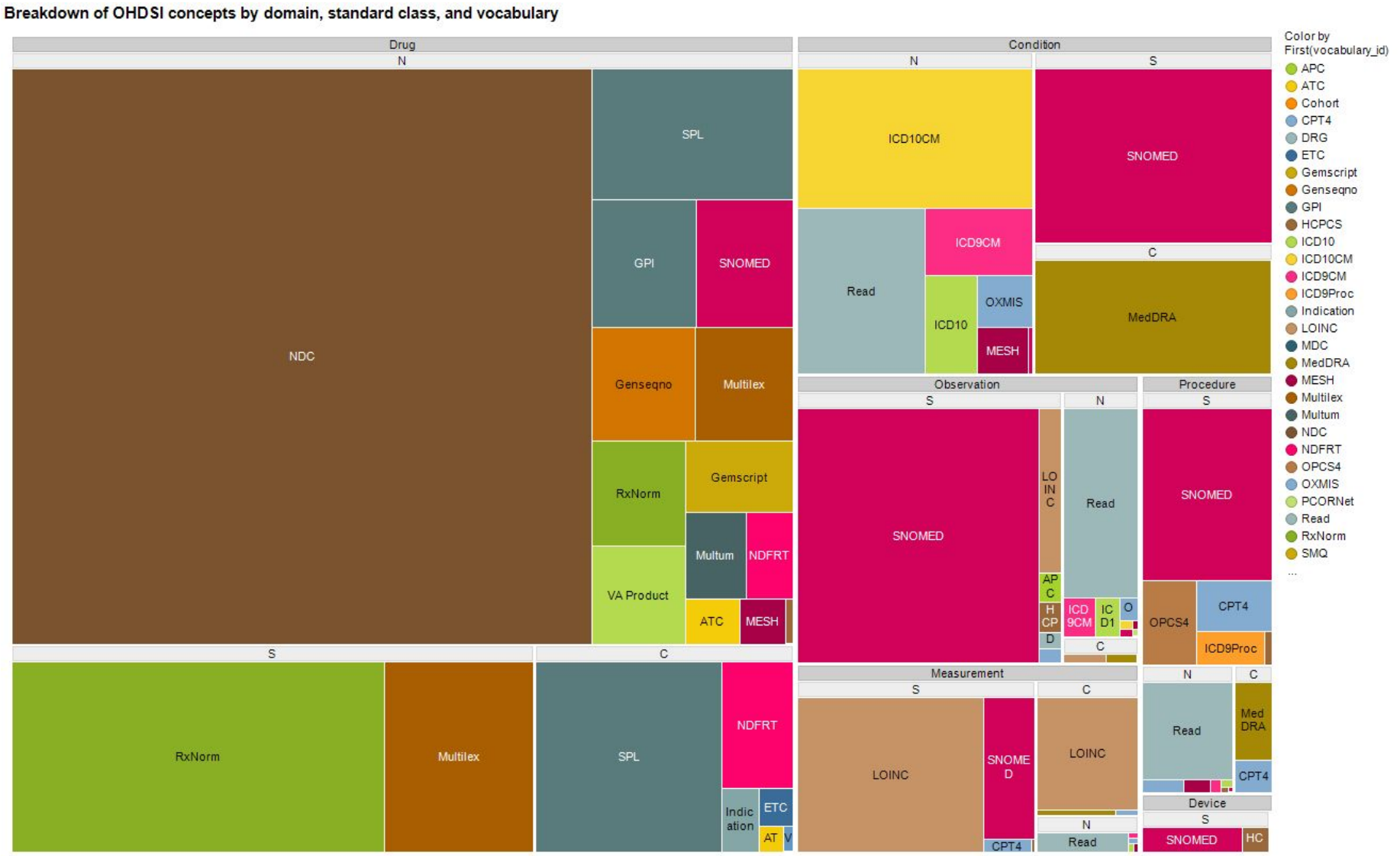




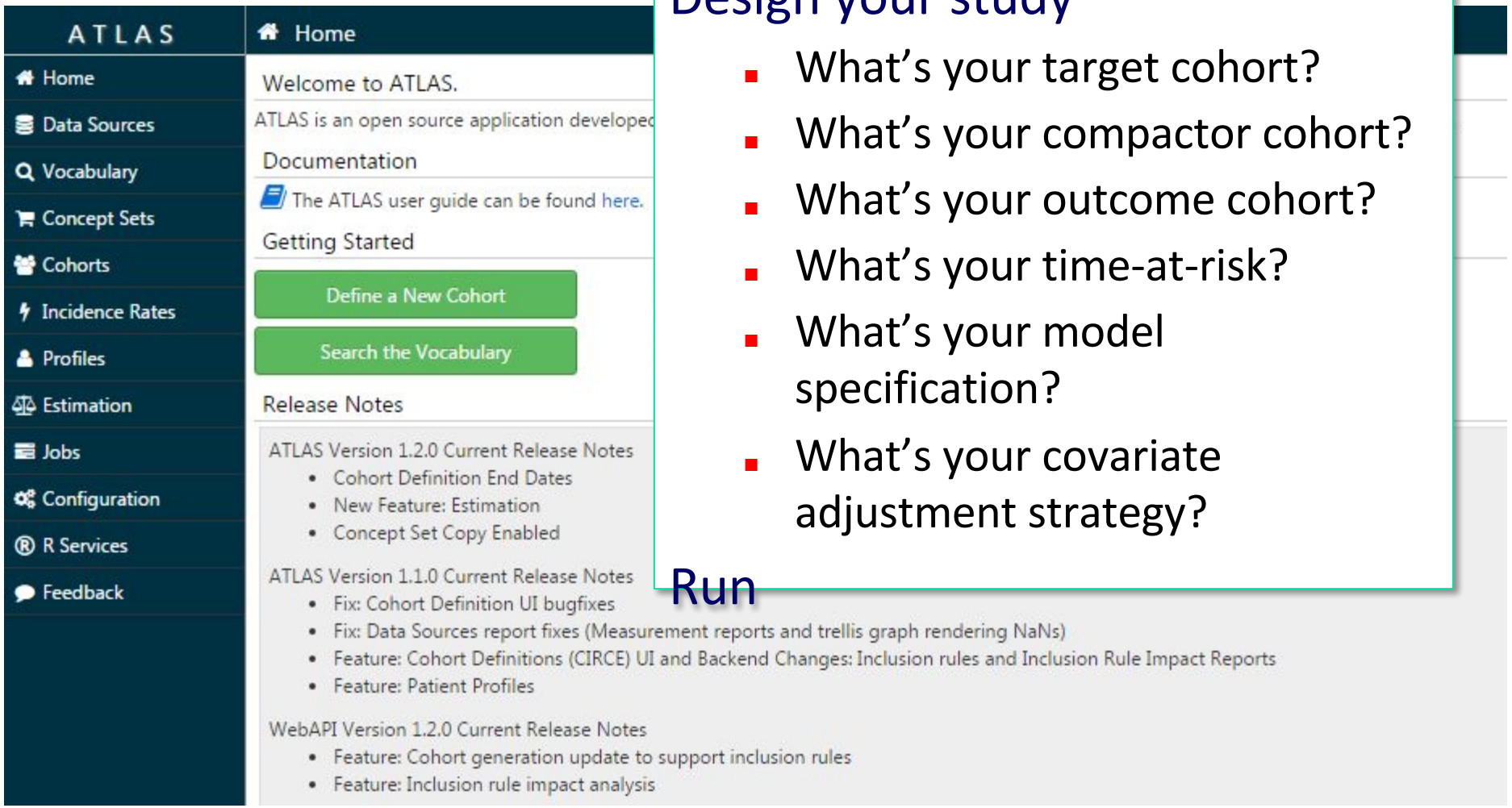
OHDSI Tools

OMOP CDM

Standard content: OMOP Vocabularies



Standardized methods: ATLAS

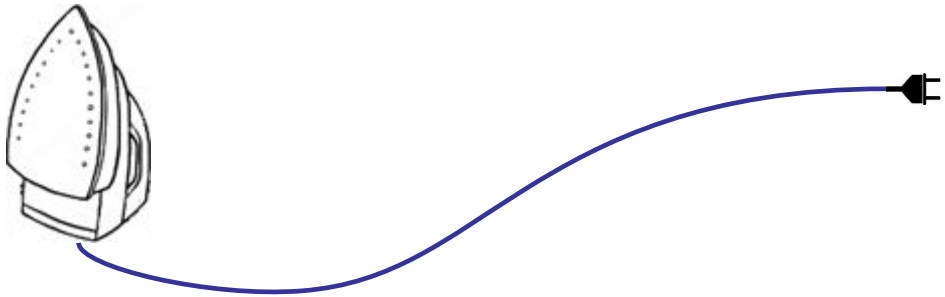


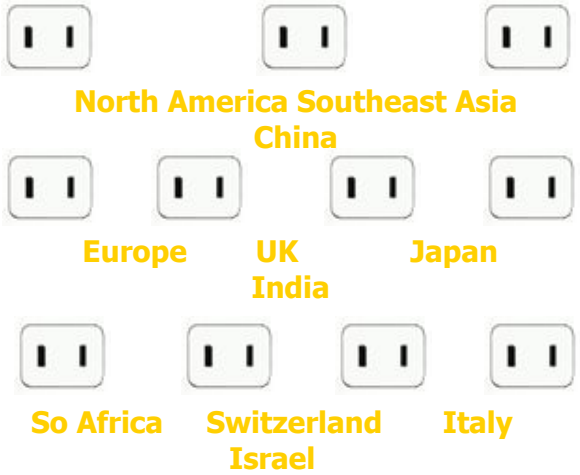
Design your study

- 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?

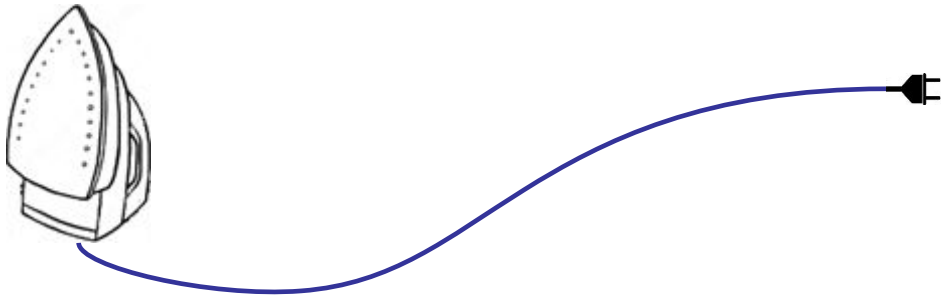
Run

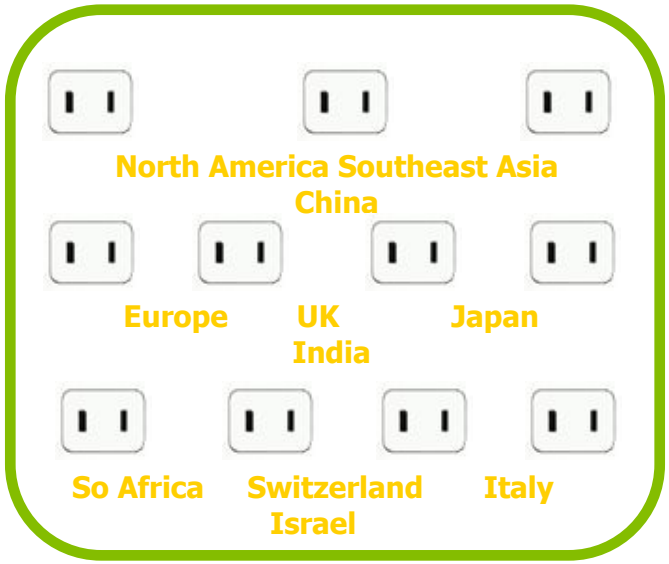
Analytics can be remote

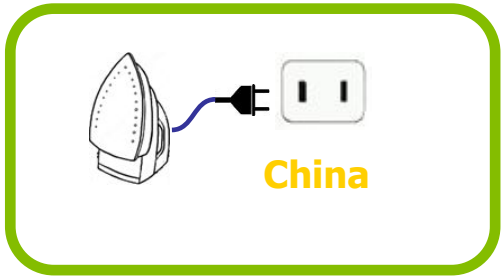




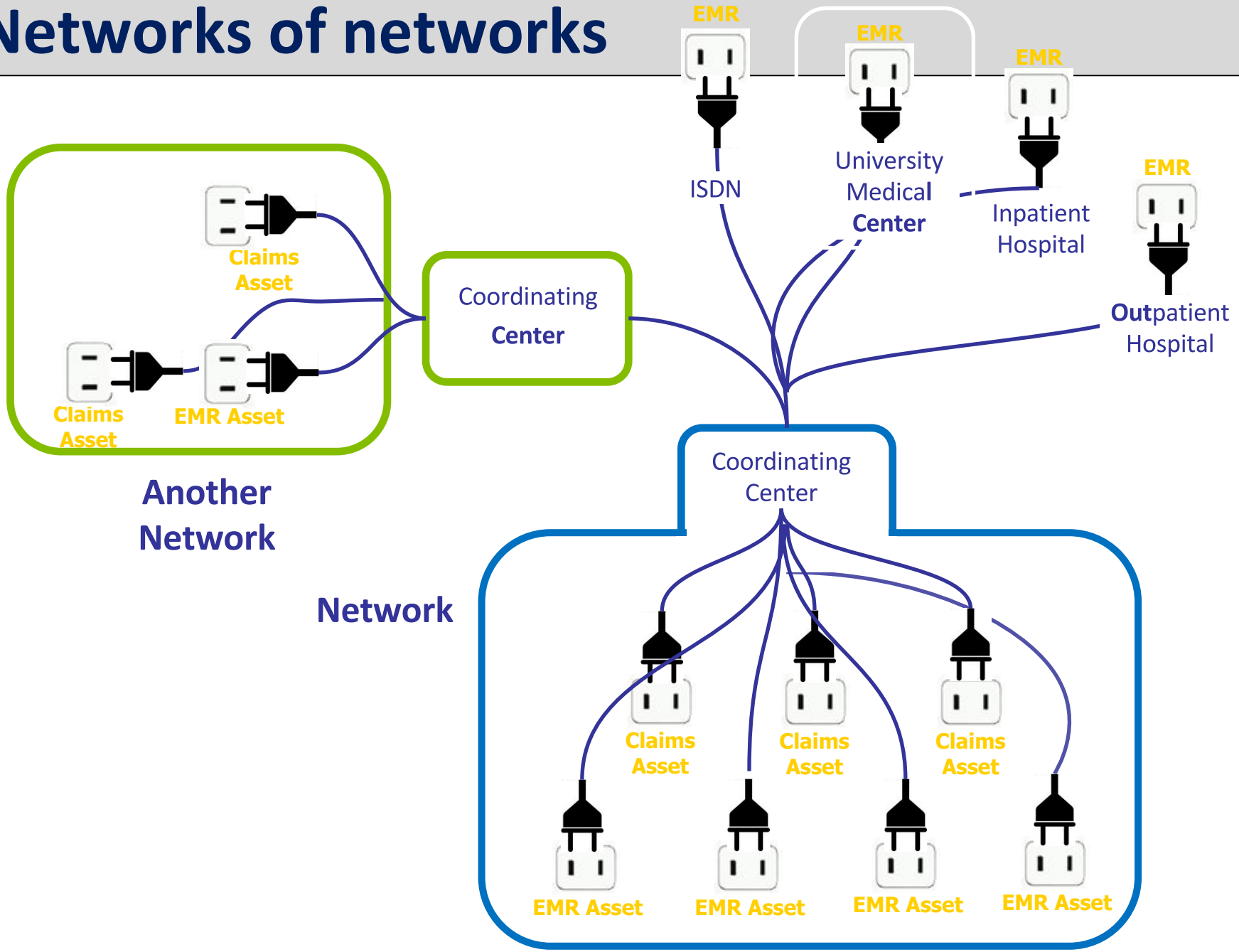
Analytics can be behind firewall







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

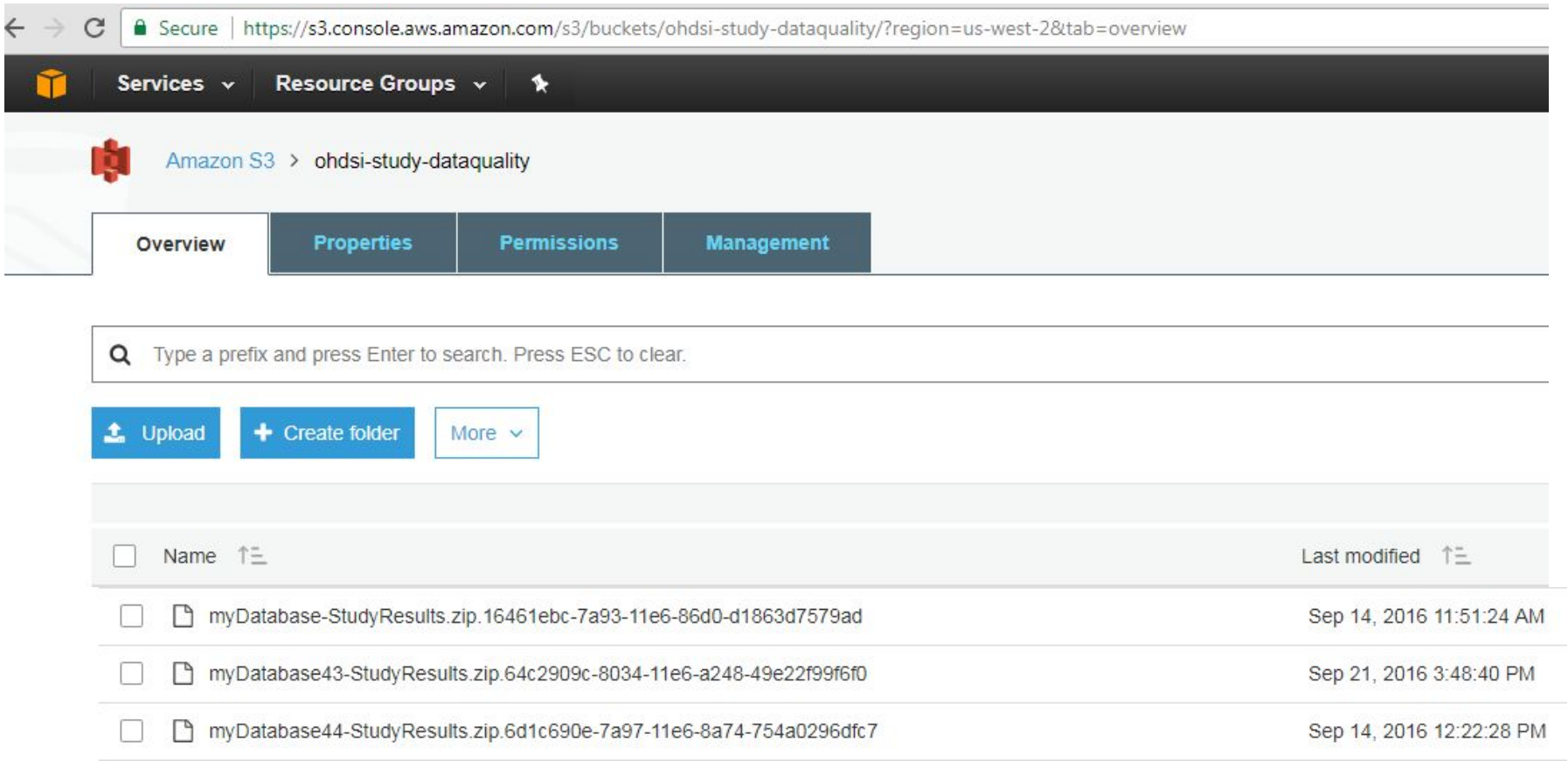
OHDSI: Researcher's perspective

- 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 recources 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

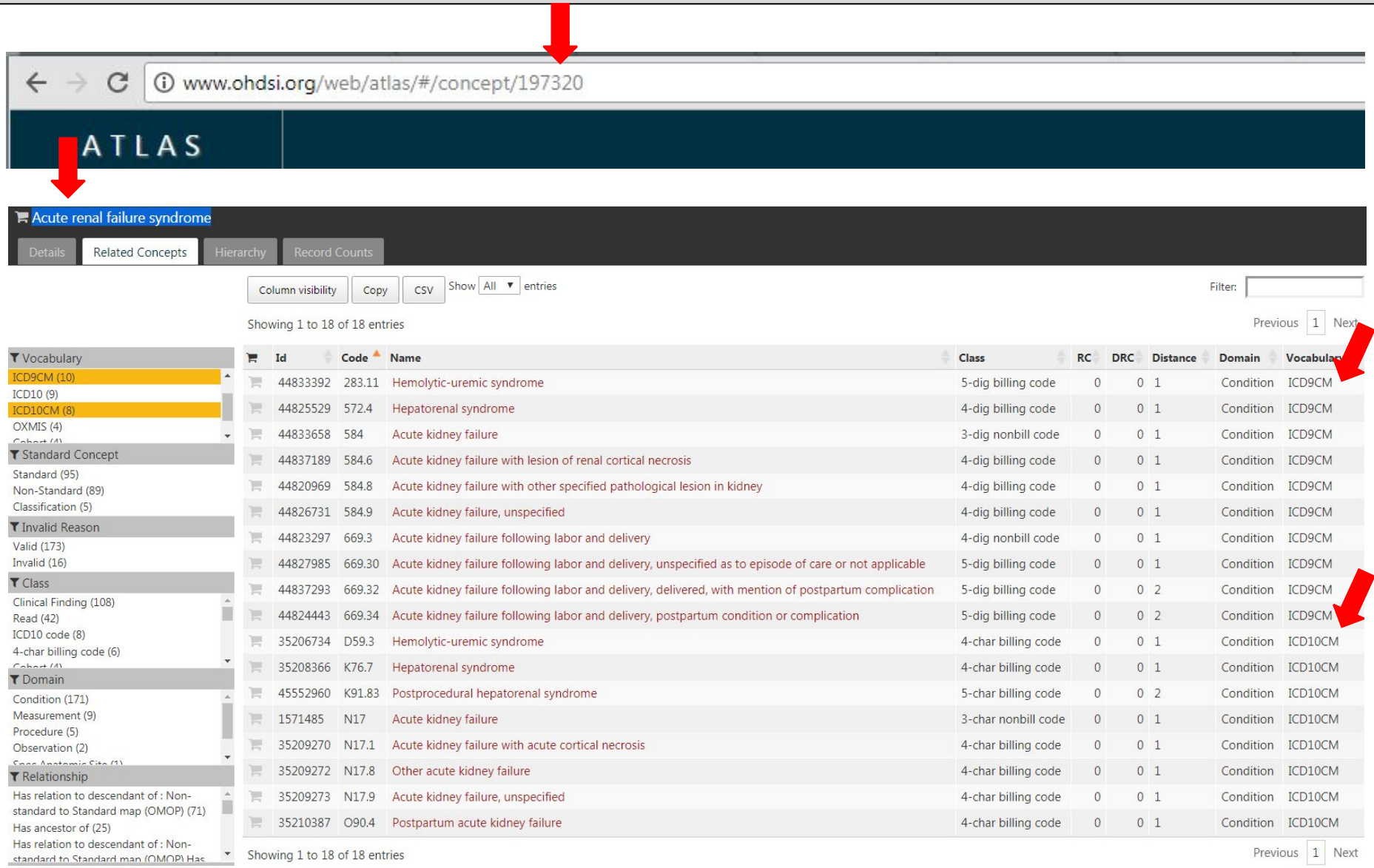
Data logistics (example)



OMOP Vocabularies

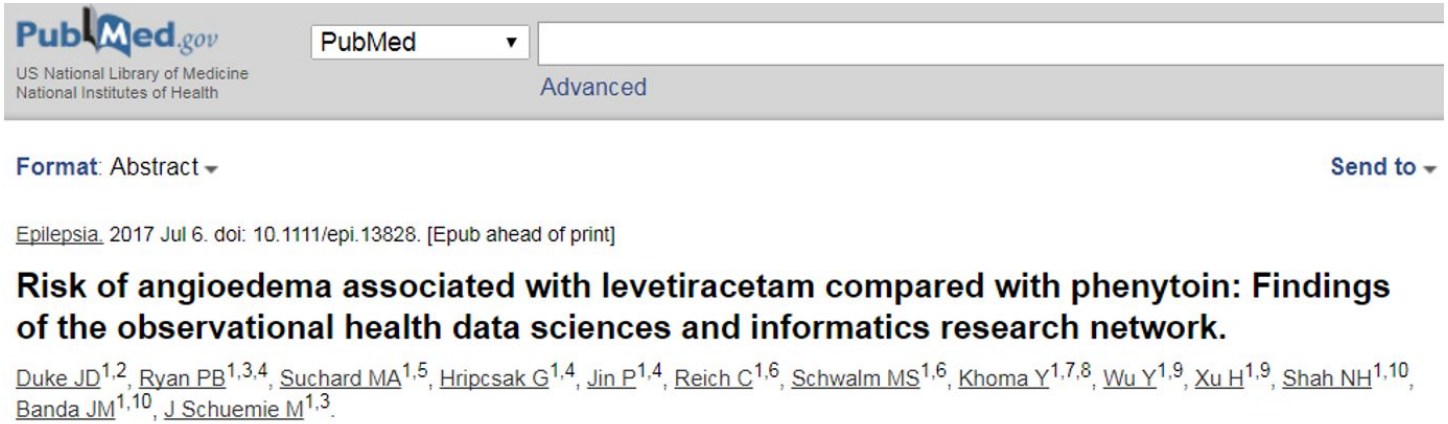
- Common framework
  - CONCEPT, CONCEPT_RELATIONSHIP, CONCEPT_ANCESTOR
- Benefit: pre-build infrastructure [+ mapping]
- Browser
  - http://www.ohdsi.org/web/atlas/#/concept/21600381
- Example of a researcher benefit
  - ICD10CM -> SNOMED CT (after Oct 1st, 2015)
  - ICD9CM -> SNOMED CT

Acute renal failure



Learning from past studies

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



PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed ▾ Advanced

Format: Abstract ▾ Send to ▾

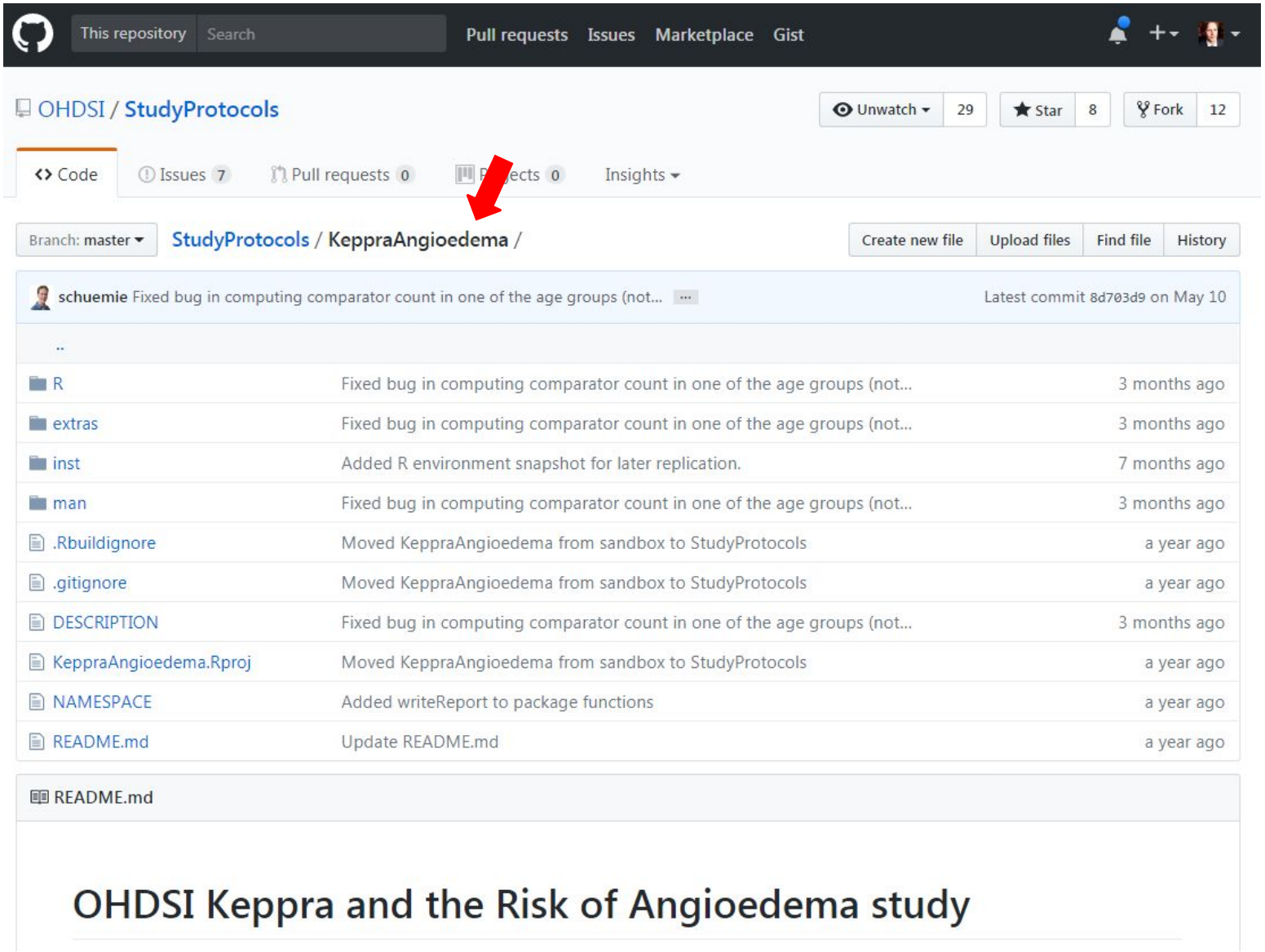
Epilepsia. 2017 Jul 6. doi: 10.1111/epi.13828. [Epub ahead of print]

Risk of angioedema associated with levetiracetam compared with phenytoin: Findings of the observational health data sciences and informatics research network.

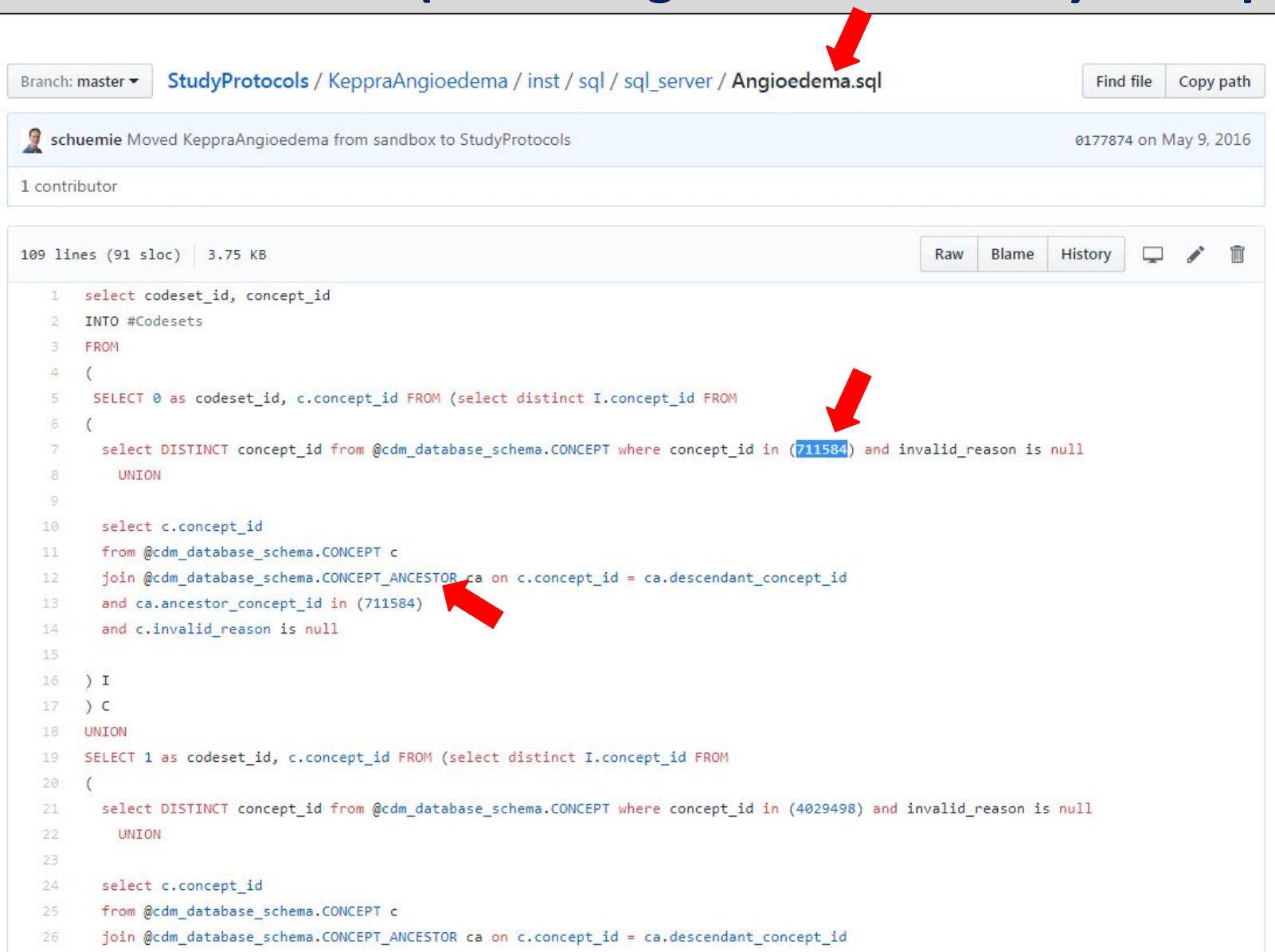
Duke JD[1,2], Ryan PB[1,3,4], Suchard MA[1,5], Hripcsak G[1,4], Jin P[1,4], Reich C[1,6], Schwalm MS[1,6], Khoma Y[1,7,8], Wu Y[1,9], Xu H[1,9], Shah NH[1,10], Banda JM[1,10], J Schuemie M[1,3].

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

Study package (in R language)



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

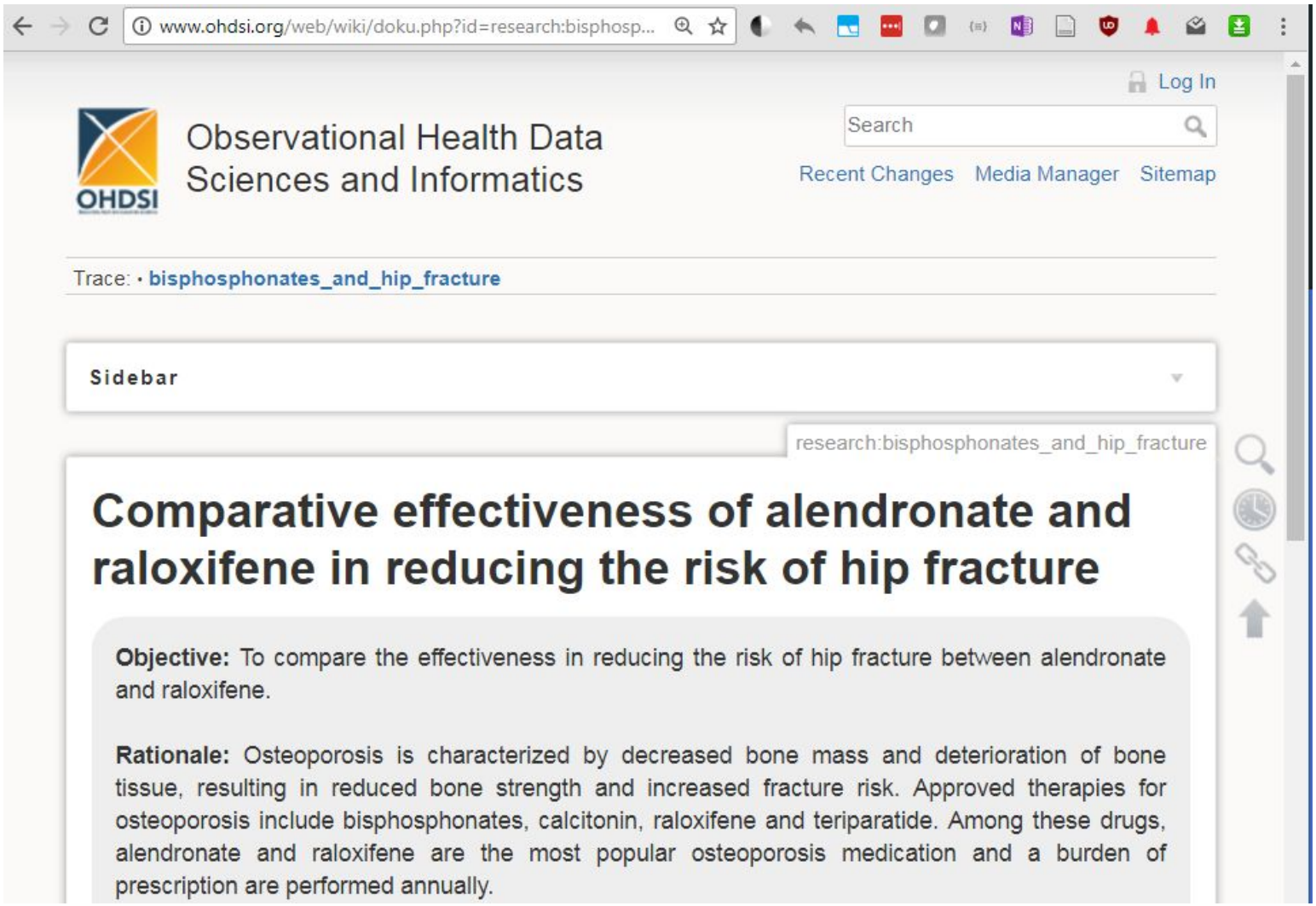
```
select codeset_id, concept_id
INTO #Codesets
FROM
(
 SELECT 0 as codeset_id, c.concept_id FROM (select distinct I.concept_id FROM
(
  select DISTINCT concept_id from @cdm_database_schema.CONCEPT where concept_id in (711584) and invalid_reason is null
    UNION

  select c.concept_id
  from @cdm_database_schema.CONCEPT c
  join @cdm_database_schema.CONCEPT_ANCESTOR ca on c.concept_id = ca.descendant_concept_id
  and ca.ancestor_concept_id in (711584)
  and c.invalid_reason is null

) I
) C
UNION
SELECT 1 as codeset_id, c.concept_id FROM (select distinct I.concept_id FROM
(
  select DISTINCT concept_id from @cdm_database_schema.CONCEPT where concept_id in (4029498) and invalid_reason is null
    UNION

  select c.concept_id
  from @cdm_database_schema.CONCEPT c
  join @cdm_database_schema.CONCEPT_ANCESTOR ca on c.concept_id = ca.descendant_concept_id
```

“Third generation” study



“Third generation” study

Branch: master ▾ StudyProtocols / AlendronateVsRaloxifene / inst / settings / CohortsToCreate.csv

msuchard move into sub

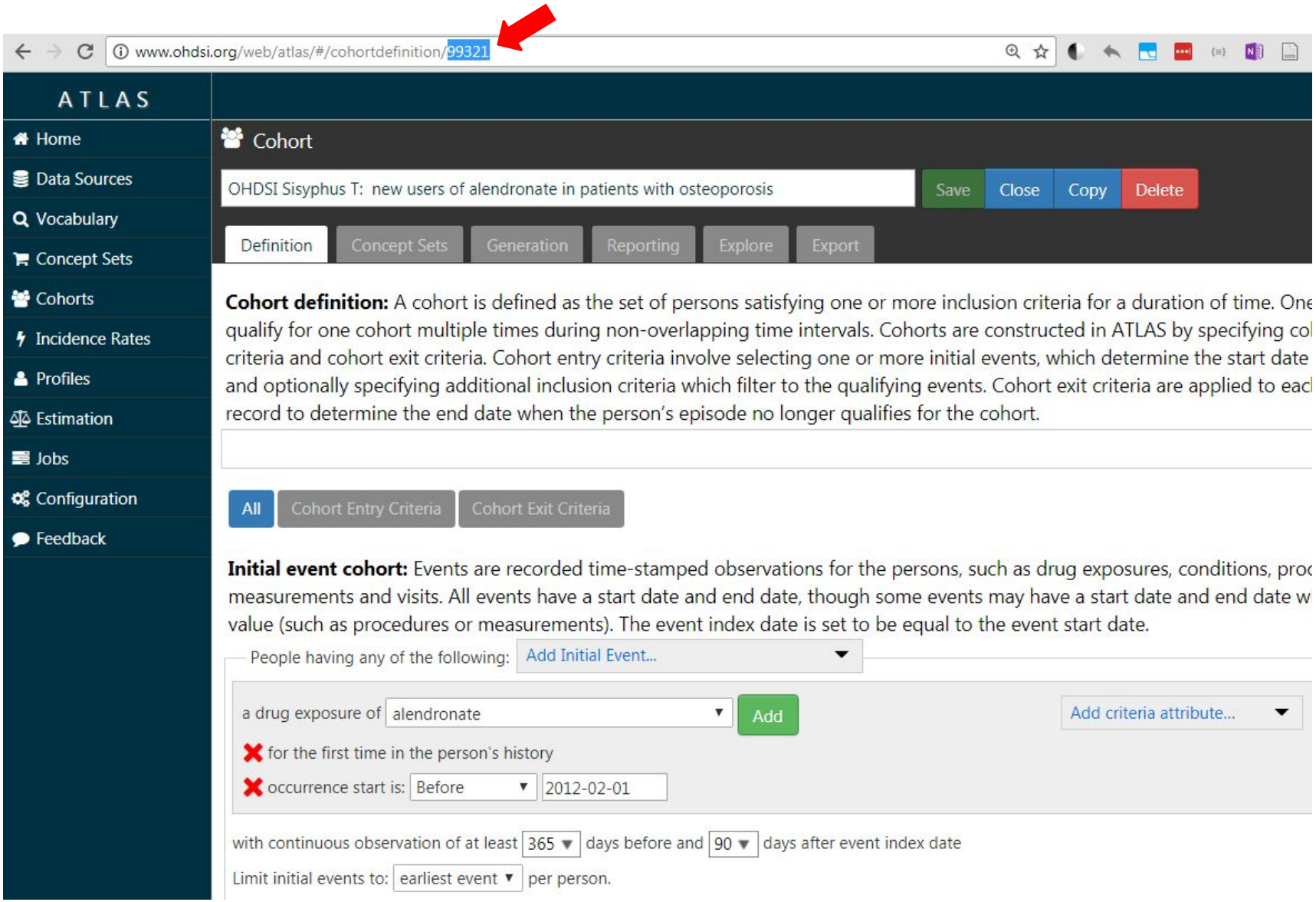
1 contributor

10 lines (9 sloc) | 268 Bytes Raw

Search this file...

	cohortId	atlasId	name
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



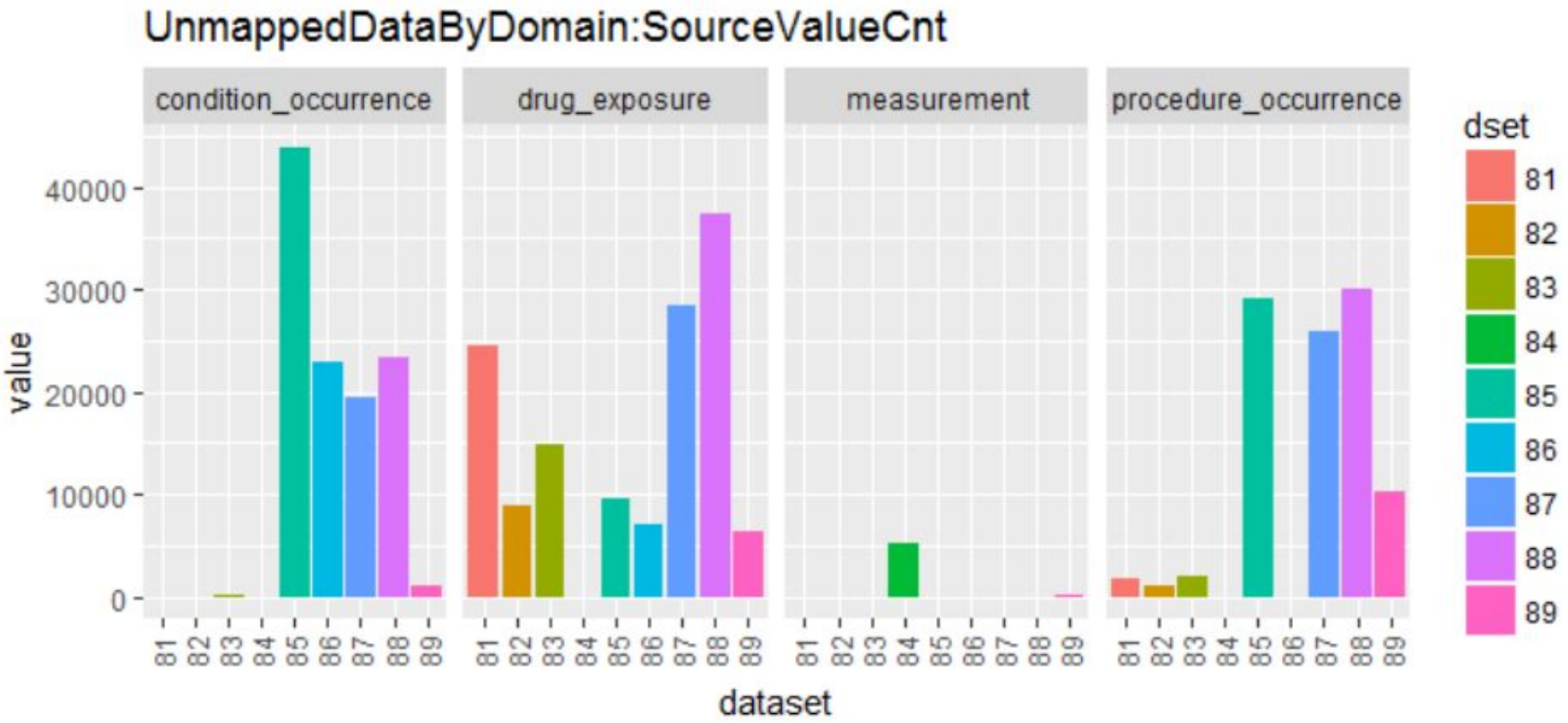
Data Quality Study

- 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

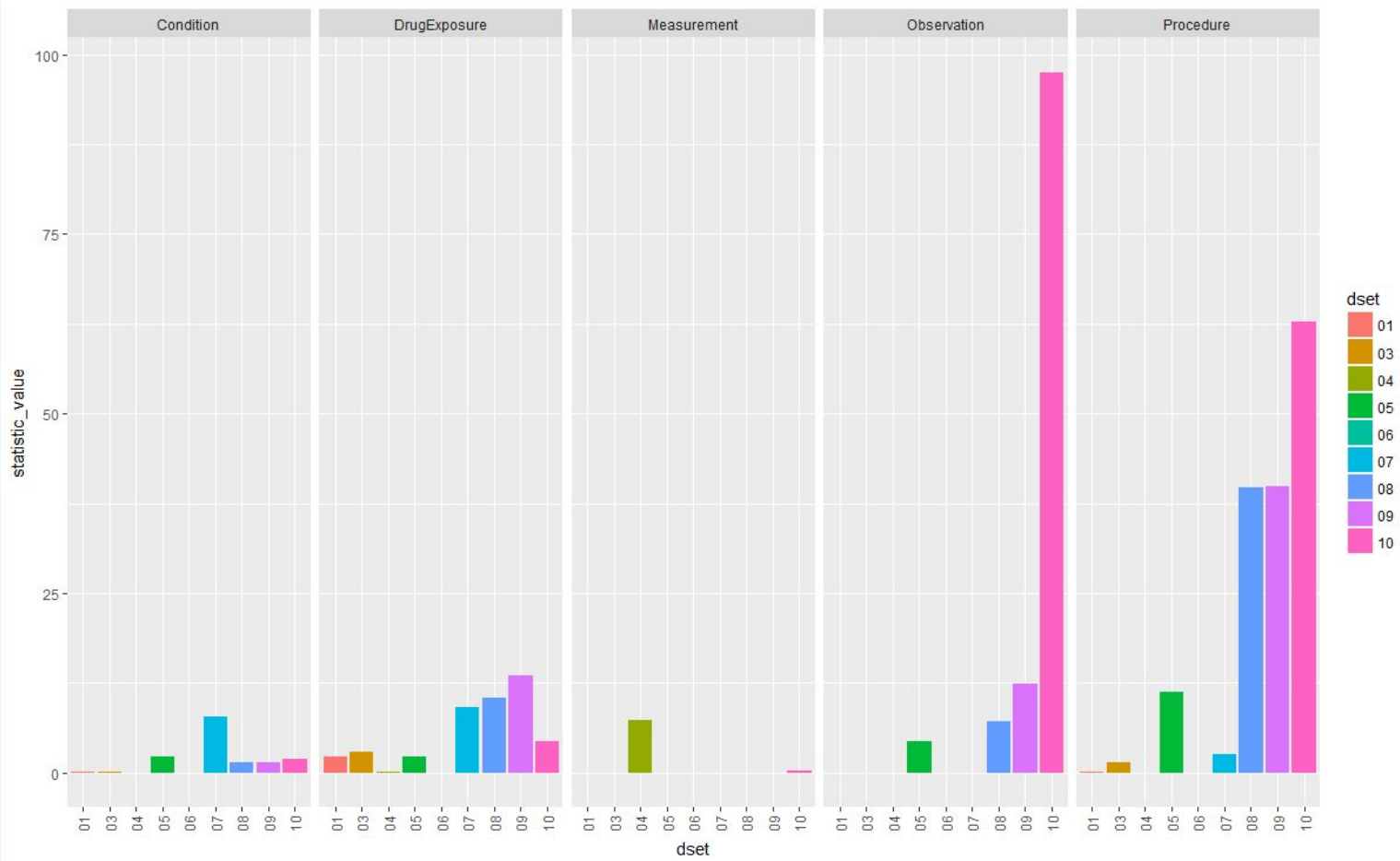
Unmapped data (example #1)

- 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