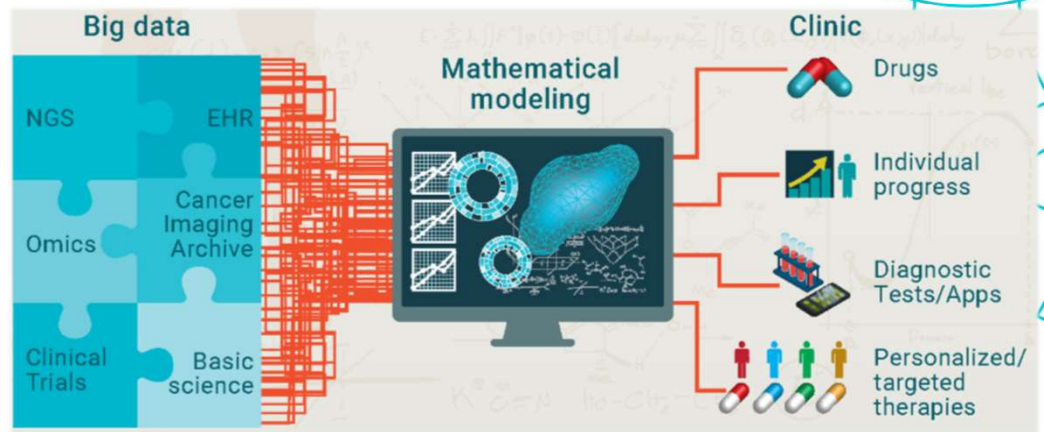


OMOP-tietomallin käyttöönotto HUS- tietoaltaalla



eCare for Me
CleverHealth Network

tieto *EVRY*

aiforia®

BC Platforms

BUSINESS
FINLAND

productivity leap

HYKS-INSTITUUTTI

Kimmo Porkka

Yliääkäri, kehityspäällikkö
HUS Syöpäkeskus, hematologian klinikka ja
HUS-tietohallinto

Professori
Hematology Research Unit Helsinki
Helsingin yliopisto

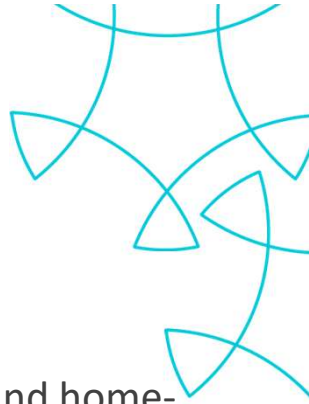


eCare for Me
CleverHealth Network

HUS
Helsinki
University
Hospital

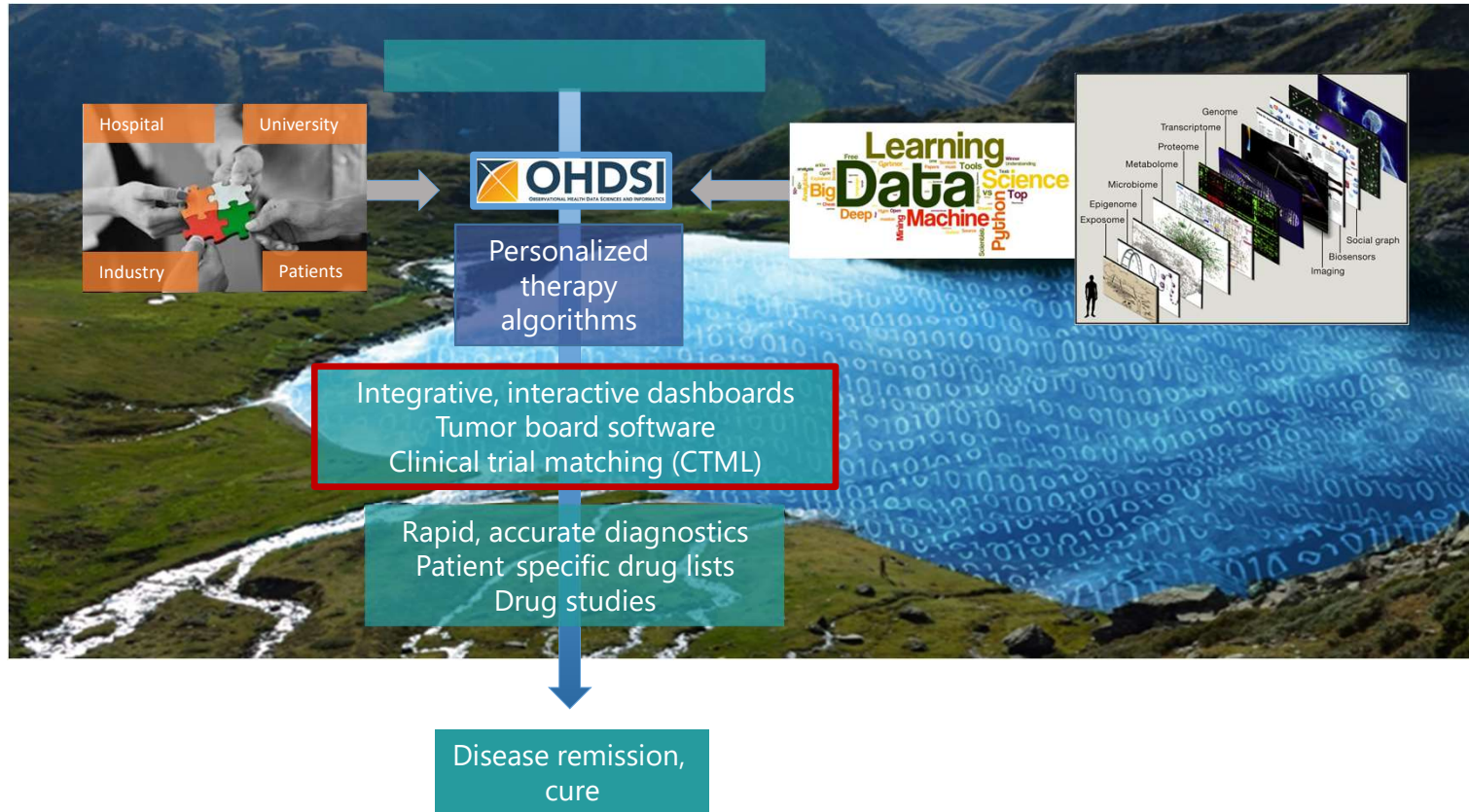
Use case: eCare for Me project

- Part of the CleverHealth ecosystem (Helsinki University Hospital - HUS)
 - Data-driven solutions to improve early diagnostics, automated treatment selection and home-based care
- A public-private partnership model
- Main funding from Business Finland
 - Finnish government organization for innovation funding and trade, travel and investment promotion
- Several SMEs, medium- and large-size companies as partners



Block 2: automated diagnostics, treatment selection (POC: acute leukemia)

PI: prof. Kimmo Porkka



General aims

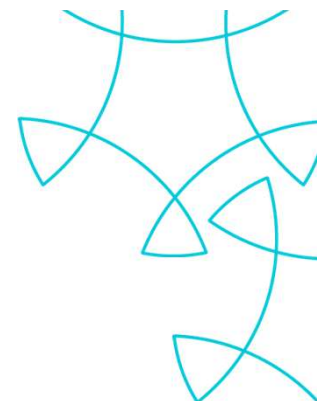
- Provide robust, reliable, reproducible evidence for research and decision-making using extended, harmonized RWD
- Implement precision medicine for cancer
 - Deep tumor profiling: genomics, transcriptomics
 - Novel data sources: image analysis, epigenetics (early diagnosis), functional assays (e.g. ex vivo drug testing), immunoprofiling, pharmacogenomics
 - Utilize public datasets (genomics, transcriptomics) for *in silico* discovery and validation
 - Collaborate/integrate with standardized EHR datasets (OHDSI OMOP, HL7 FHIR, Flatiron)
- Data integration and modeling for scalable clinical applications



HUS datalake 2018: no common data model

- Data structuring ad hoc, project-based; not into a common standard
- Overlap of data parsing efforts – not an optimal use of scarce resources
- Commercial registries (e.g. BCB) suboptimally implemented, require manual data entry
- Merging Uranus and Apotti datasets a big unknown

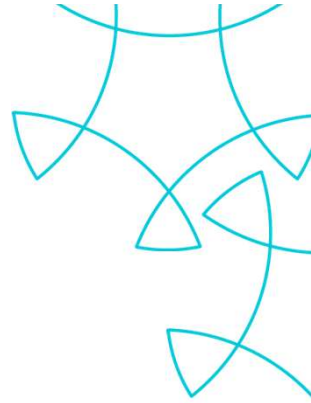
- National, international projects: most of the time used for data structuring, conversion, not analytical work (high costs, long duration)
- Applications/algorithms built on HUS datalake poorly scalable
- No standard analytical tools



HUS datalake 2019=>

- Query of CDM standards for observational research on EHR data
 - HL7 FHIR
 - OHDSI OMOP
 - Interoperable
- OMOP selected as a pilot (BCPlatforms BCRQuest; eCare4Me B2)
- eCare4Me/B2
 - Comprehensive mapping and ETL work at the main datalake
 - Close collaboration between clinical domain experts, HUS IT and ETL experts
 - Uranus, Apotti

Observational Health Data Sciences and Informatics (OHDSI, as “Odyssey”)

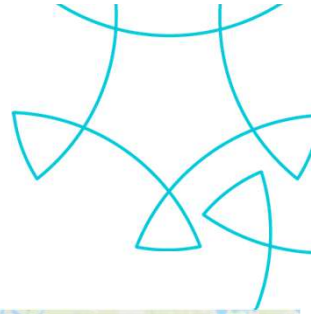


- Mission: To improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care
- An academic, multi-stakeholder, interdisciplinary, international collaborative with a coordinating center at Columbia University, NY

<http://ohdsi.org>

OMOP: **O**bservational **M**edical **O**utcomes **P**artnership
common data model implemented by the OHDSI
community

OHDSI's global research community

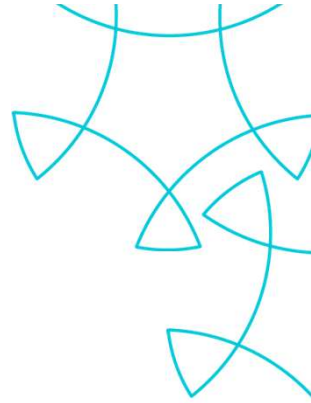


- >300 collaborators from 30 different countries
- Experts in informatics, statistics, epidemiology, clinical sciences
- Active participation from academia, government, industry, providers
- Records on about 600 million unique patients in >100 databases



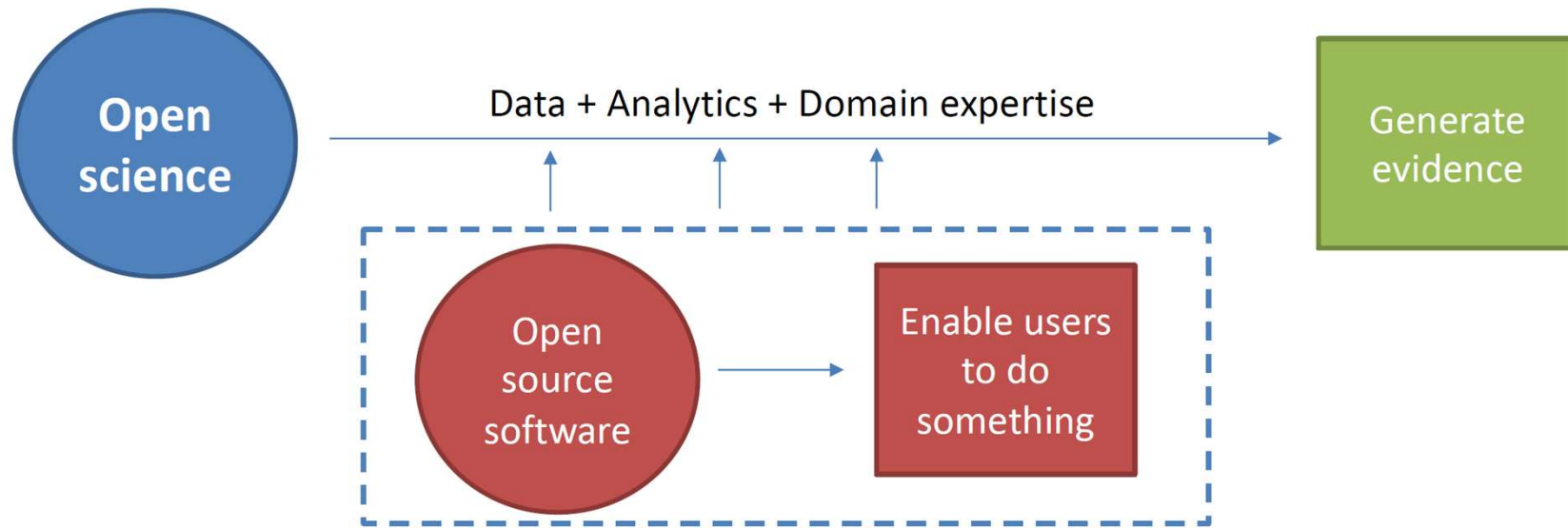
<http://ohdsi.org/who-we-are/collaborators>

Evidence OHDSI seeks to generate from observational data



- **Clinical characterization – tally**
 - Natural history: Who has diabetes, and who takes metformin?
 - Quality improvement: What proportion of patients with diabetes experience complications?
- **Population-level estimation – cause**
 - Safety surveillance: Does metformin cause lactic acidosis?
 - Comparative effectiveness: Does metformin cause lactic acidosis more than glyburide?
- **Patient-level prediction – predict**
 - Precision medicine: Given everything you know about me, if I take metformin, what is the chance I will get lactic acidosis?
 - Disease interception: Given everything you know about me, what is the chance I will develop diabetes?

Open Science



Standardized, transparent, reproducible workflows

Database
summary

Cohort
definition

Cohort
summary

Compare
cohorts

Exposure-
outcome
summary

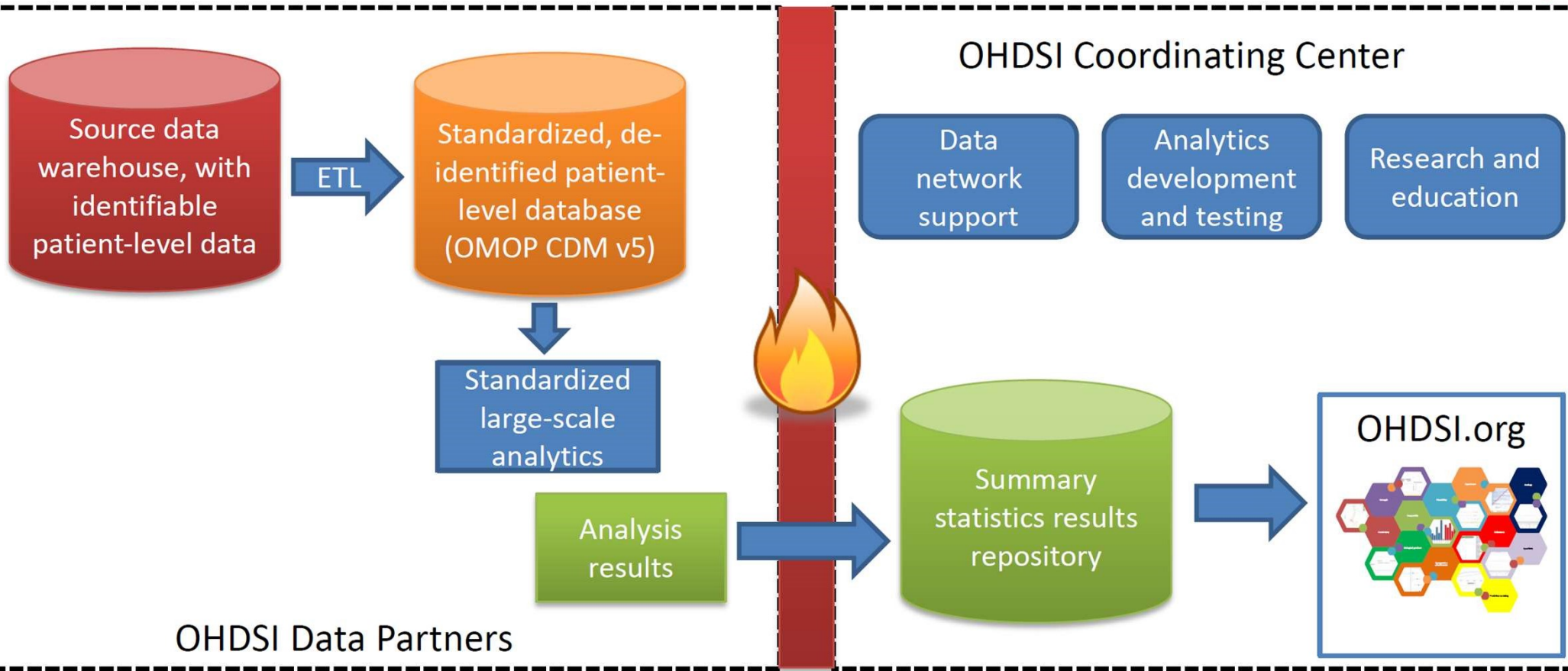
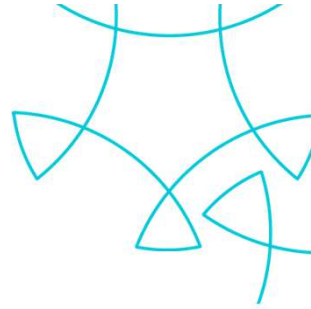
Effect
estimation
&
calibration

Compare
databases



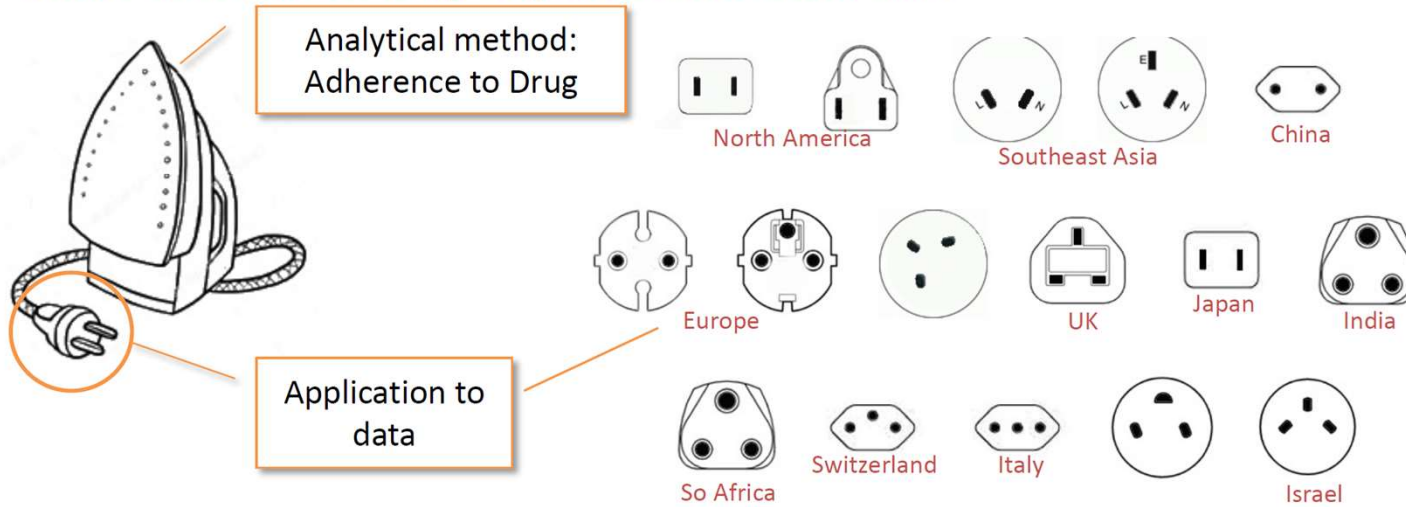
eCare for Me
CleverHealth Network

How OHDSI Works

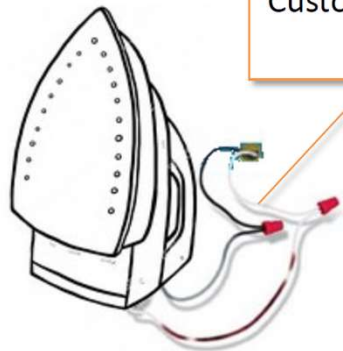


Current Approach: "One Study – One Script"

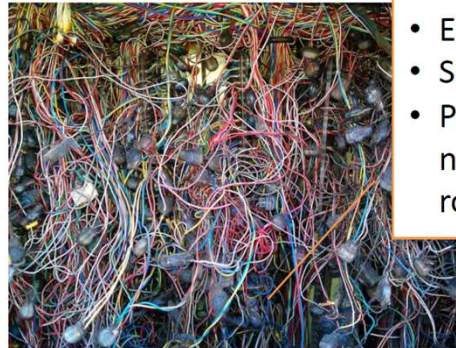
"What's the adherence to my drug in the data assets I own?"



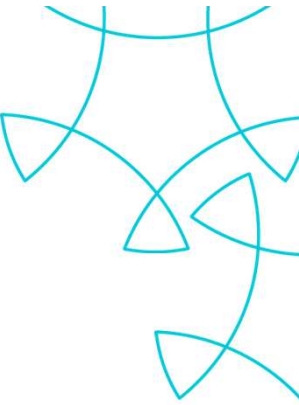
Current solution:



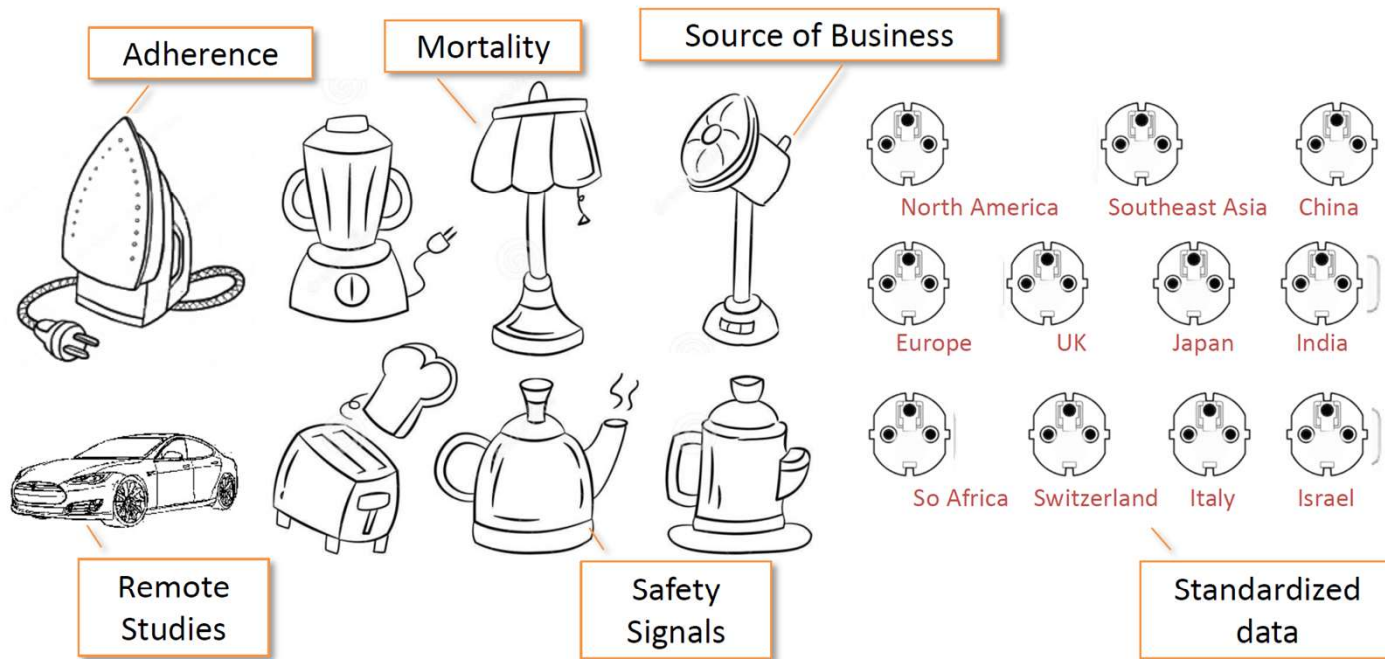
Custom script for each study



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



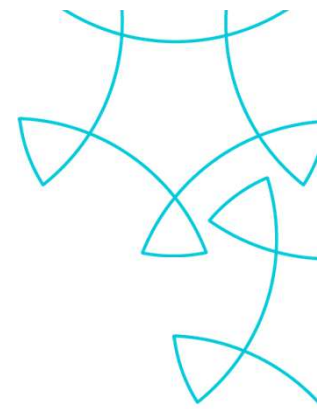
Solution: Standardized Data and Analytics



1. ATLAS, Remote Studies
 - Standard Cohorts
 - Standardized Analytics

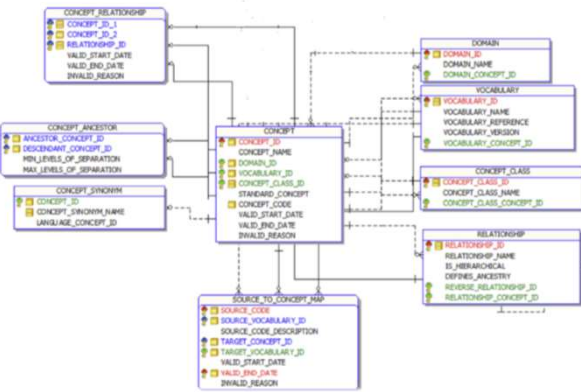
2. OMOP CDM
 - Standardized Format
 - Standardized Coding

OMOP-CDM: core structure



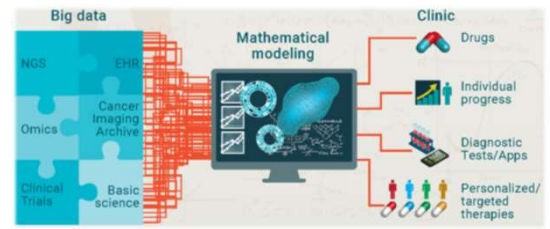
syntactic operability:
common underlying data structure
(standard grammar)

semantic operability:
common understanding required to
exchange interchange information
(standard vocabulary)

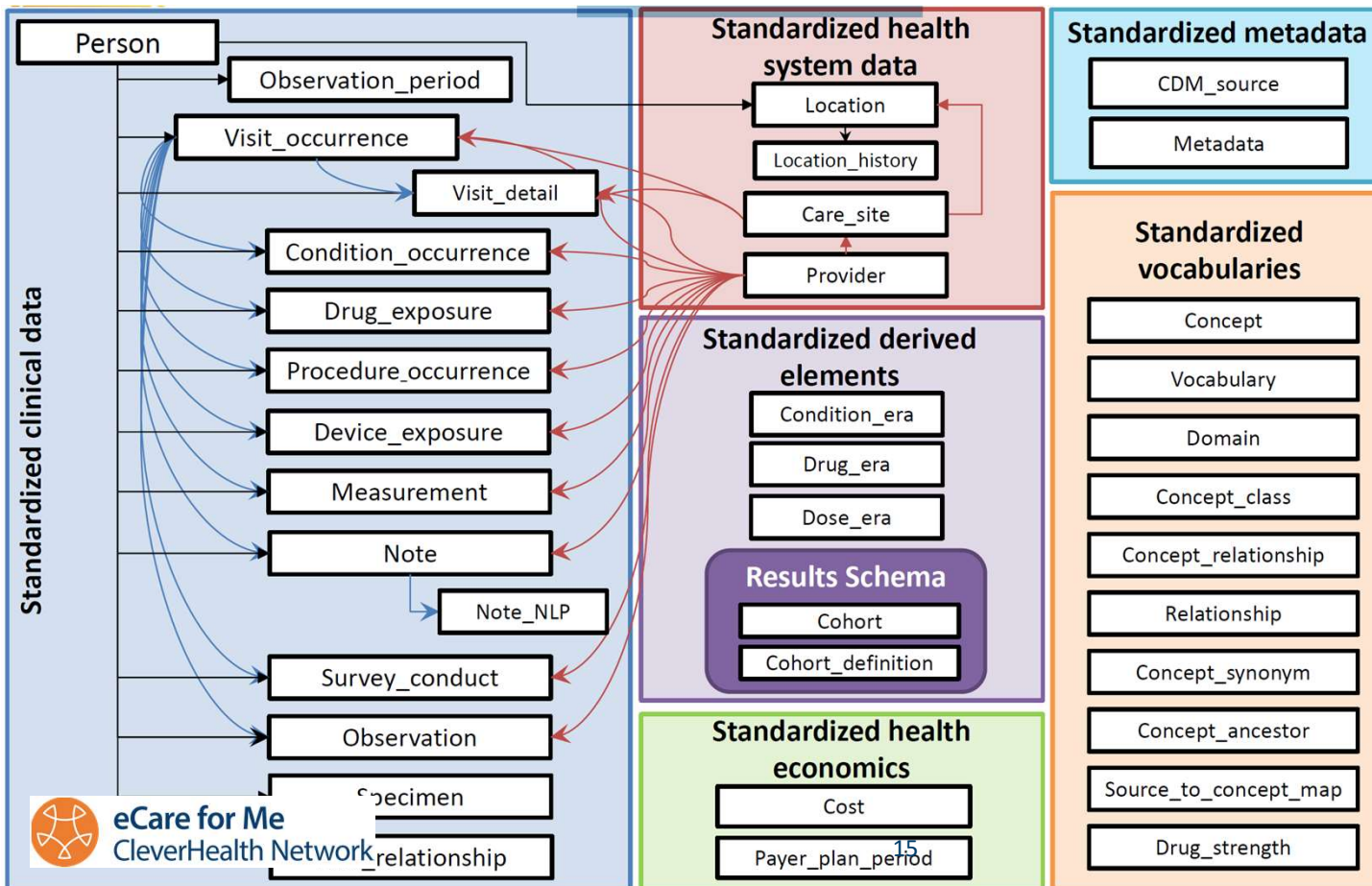
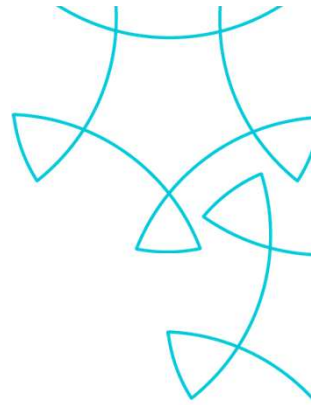


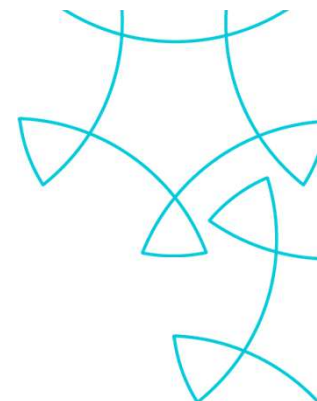
analytical operability:
robust evidence generation
(standard analytics)

- 1 SNOMED
 - 2 ICD9CM
 - 3 ICD9Proc
 - 4 CPT4
 - 5 HCPCS
 - 6 LOINC
 - 7 NDCPRT
 - 8 RxNorm
 - 9 NDC
 - 10 GPI
 - 11 UCLUM
 - 12 Gender
 - 13 Race
 - 14 Place of Service
 - 15 MedDRA
 - 16 Multum
 - 17 Read
 - 18 ODMIS
 - 19 Indication
 - 20 ETC
 - 21 ATC
 - 22 Multilex
 - 28 VA Product
 - 31 SMQ
 - 32 VA Class
 - 33 Cohort
 - 34 ICD10
- Systematic Nomenclature of Medicine - Clinical Terms (SNOMED)
International Classification of Diseases, Ninth Revision, Clinical Modification, Volume 1 and 2 (ICD9CM)
International Classification of Diseases, Ninth Revision, Clinical Modification, Volume 3 (ICD9Proc)
Current Procedural Terminology version 4 (AMA)
Healthcare Common Procedure Coding System (HCPCS)
Logical Observation Identifiers Names and Codes (Regenstrief Institute)
National Drug File - Reference Terminology (VA)
RxNorm (NLM)
National Drug Code (FDA and manufacturers)
Medi-Span Generic Product Identifier (Wolters Kluwer Health)
Unified Code for Units of Measure (Regenstrief Institute)
OHOP Gender
Race and Ethnicity Code Set (USSC)
Place of Service Codes for Professional Claims (CMS)
Medical Dictionary for Regulatory Activities (MSDR)
Center Multum (Center)
NHS UK Read Codes Version 2 (HSCIC)
Oxford Medical Information System (OXMIS)
Indications and Contraindications (FDB)
Enhanced Therapeutic Classification (FDB)
WHO Anatomical Therapeutic Chemical Classification Multilex (FDB)
VA National Drug File Product (VA)
Standardised MedDRA Queries (MSSO)
VA National Drug File Class (VA)
Legacy OHOP HDI or DOI cohort
International Classification of Diseases, 10th Revision, (WHO)



Deep information model: OMOP CDM



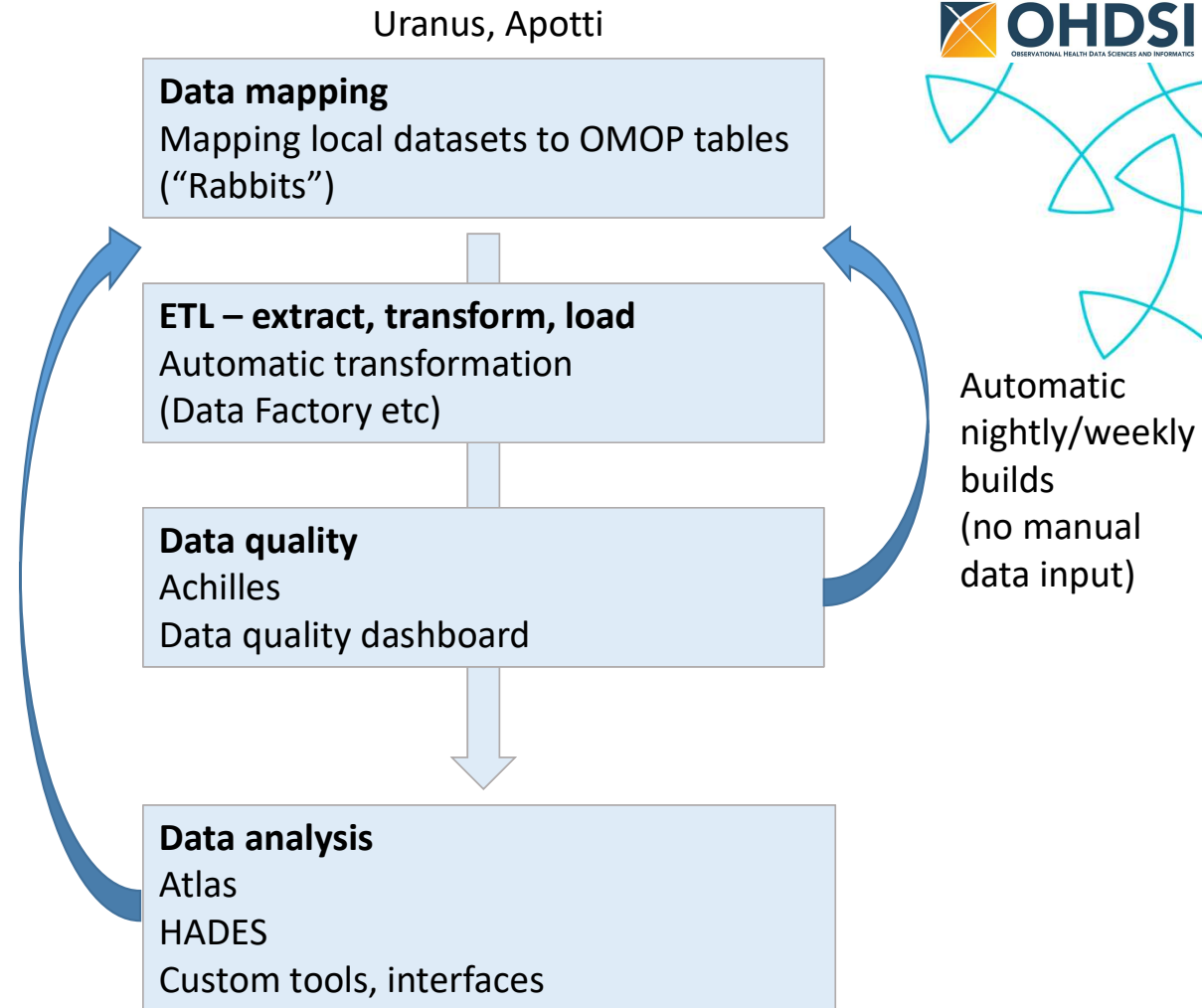


OMOP's standardized vocabularies

- 153 Vocabularies across 41 domains
 - MU3 standards: SNOMED, RxNorm, LOINC
 - Disparate sources: ICD9CM, ICD10(CM), Read, NDC, Gemscript, CPT4, HCPCS...
- >9 million concepts
 - >3.3 million standard concepts
 - >5.1 million source codes
 - >629,000 classification concepts
- >55 million concept relationships
- >84 million ancestral relationships

Publicly available for download at: <http://athena.ohdsi.org/>

OMOP-CDM: implementation



OMOP-teams/HUS



Oscar Brück
Perttu Koskenvesa
Ilona Siljander
Otso Brummer
Kimmo Porkka

Annu Kaila
Niilo Aho
Jani Salmi
Juha Muinonen

Anja Kajanne
Mirka Tammi

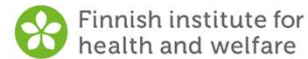


Juha Koski
Pasi Rikala
Pekka Stigell
Tero-Matti Kinanen



Antti Larjo
Sergei Häyrynen
Stanislav Muhametsin

Miikka Ermes
Matti Ristimäki



Esa-Matti Tolppanen
Jonna Salonen



FINNGEN

Javier Gracia
Tabuenca



Caroline Heckman
Esa Pitkänen
Andrea Ganna

Imre Västrik

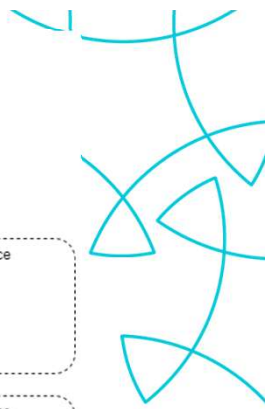
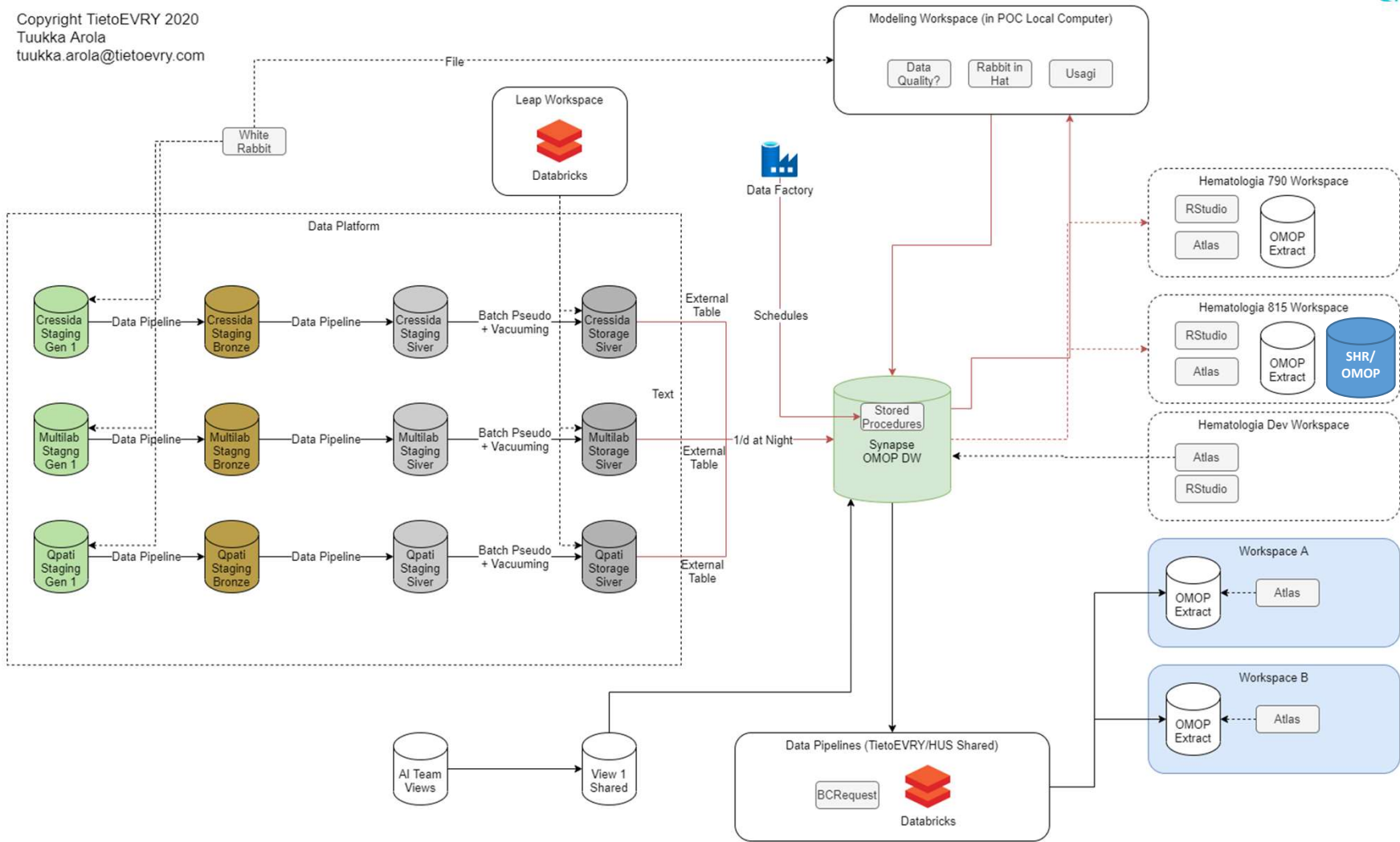


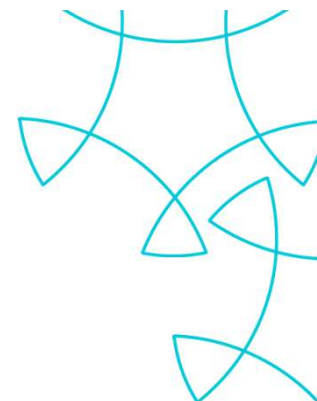
Kari Pitkänen
Sami Blom



BUSINESS FINLAND

Copyright TietoEVRY 2020
 Tuukka Arola
 tuukka.arola@tietoevry.com

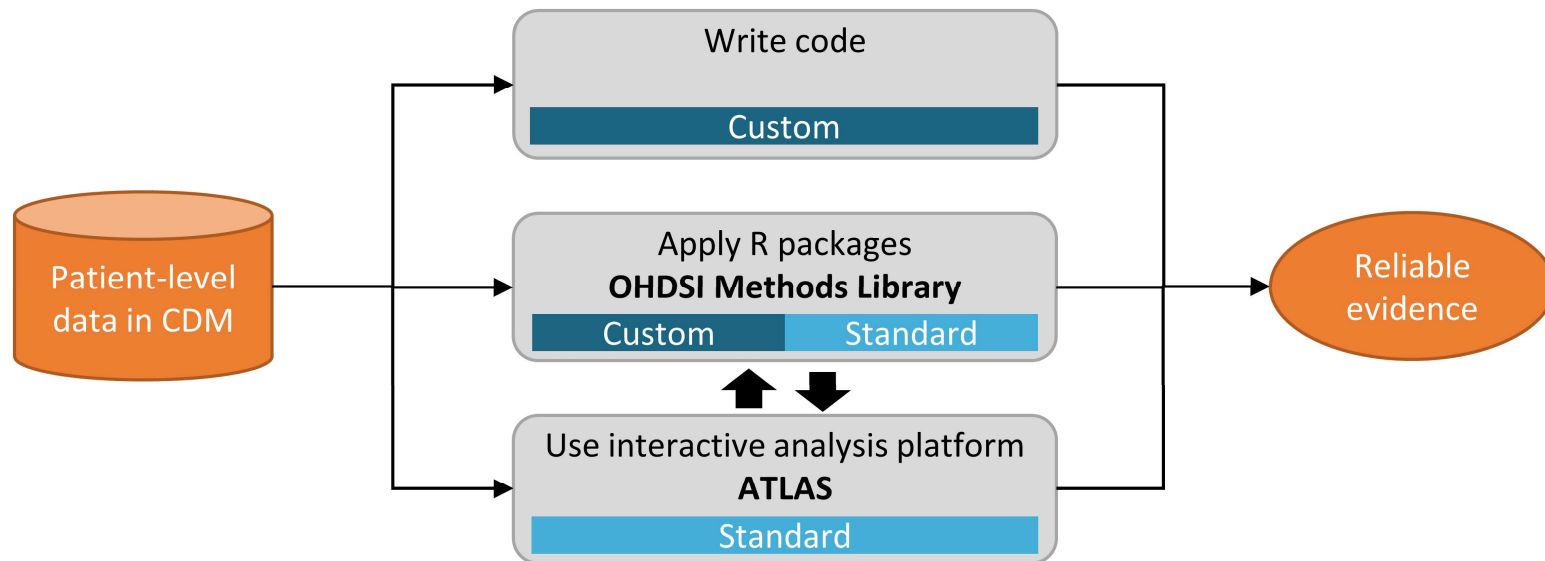




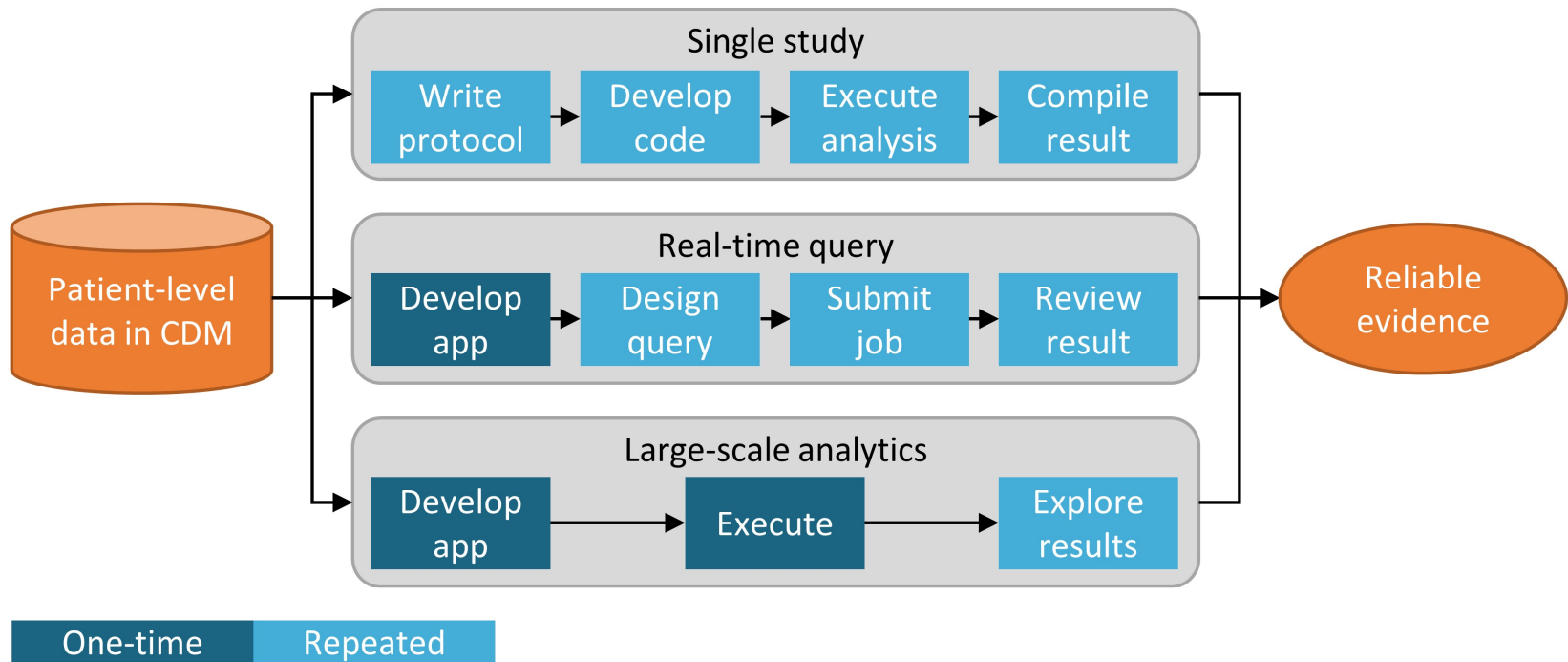
Data Analytics Use Cases

- Characterization
 - “What happened to them?”
 - Count or percentage, Averages, Descriptive statistics, Incidence, Prevalence, Phenotype
- Population-level estimation
 - “What are the causal effects?”
 - Relative risk, Hazards/Odds ratio, Average treatment effect, Causal effect, Correlation
- Patient-level prediction
 - “What will happen to me?”
 - Probability for an individual, Prediction model, High/low risk groups, Probabilistic phenotype

OMOP Analysis Implementation: reproducible research



Some Analysis Strategies




ARTICLE

<https://doi.org/10.1038/s41467-020-18849-z>

OPEN



Deep phenotyping of 34,128 adult patients hospitalised with COVID-19 in an international network study

Edward Burn  et al.[#]

Comorbid conditions appear to be common among individuals hospitalised with coronavirus disease 2019 (COVID-19) but estimates of prevalence vary and little is known about the prior medication use of patients. Here, we describe the characteristics of adults hospitalised with COVID-19 and compare them with influenza patients. We include 34,128 (US: 8362, South Korea: 7341, Spain: 18,425) COVID-19 patients, summarising between 4811 and 11,643 unique aggregate characteristics. COVID-19 patients have been majority male in the US and Spain, but predominantly female in South Korea. Age profiles vary across data sources. Compared to 84,585 individuals hospitalised with influenza in 2014-19, COVID-19 patients have more typically been male, younger, and with fewer comorbidities and lower medication use. While protecting groups vulnerable to influenza is likely a useful starting point in the response to COVID-19, strategies will likely need to be broadened to reflect the particular characteristics of individuals being hospitalised with COVID-19.

Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis

Marc A Suchard, Martijn J Schuemie, Harlan M Krumholz, Seng Chan You, Ruijun Chen, Nicole Pratt, Christian G Reich, Jon Duke, David Madigan, George Hripcsak, Patrick B Ryan

Summary

Background Uncertainty remains about the optimal monotherapy for hypertension, with current guidelines recommending any primary agent among the first-line drug classes thiazide or thiazide-like diuretics, angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, dihydropyridine calcium channel blockers, and non-dihydropyridine calcium channel blockers, in the absence of comorbid indications. Randomised trials have not further refined this choice.

Methods We developed a comprehensive framework for real-world evidence that enables comparative effectiveness and safety evaluation across many drugs and outcomes from observational data encompassing millions of patients, while minimising inherent bias. Using this framework, we did a systematic, large-scale study under a new-user cohort design to estimate the relative risks of three primary (acute myocardial infarction, hospitalisation for heart failure, and stroke) and six secondary effectiveness and 46 safety outcomes comparing all first-line classes across a global network of six administrative claims and three electronic health record databases. The framework addressed residual confounding, publication bias, and p-hacking using large-scale propensity adjustment, a large set of control outcomes, and full disclosure of hypotheses tested.

Findings Using 4.9 million patients, we generated 22,000 calibrated, propensity-score-adjusted hazard ratios (HRs) comparing all classes and outcomes across databases. Most estimates revealed no effectiveness differences between classes; however, thiazide or thiazide-like diuretics showed better primary effectiveness than angiotensin-converting enzyme inhibitors: acute myocardial infarction (HR 0.84, 95% CI 0.75–0.95), hospitalisation for heart failure (0.83, 0.74–0.95), and stroke (0.83, 0.74–0.95) risk while on initial treatment. Safety profiles also favoured thiazide or thiazide-like diuretics over angiotensin-converting enzyme inhibitors. The non-dihydropyridine calcium channel blockers were significantly inferior to the other four classes.

Interpretation This comprehensive framework introduces a new way of doing observational health-care science at scale. The approach supports equivalence between drug classes for initiating monotherapy for hypertension—in keeping with current guidelines, with the exception of thiazide or thiazide-like diuretics superiority to angiotensin-converting enzyme inhibitors and the inferiority of non-dihydropyridine calcium channel blockers.

Funding US National Science Foundation, US National Institutes of Health, Janssen Research & Development, IQVIA, South Korean Ministry of Health & Welfare, Australian National Health and Medical Research Council.

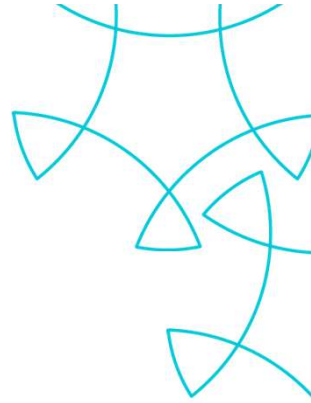


Published Online
October 24, 2019
[https://doi.org/10.1016/S0140-6736\(19\)32317-7](https://doi.org/10.1016/S0140-6736(19)32317-7)
See Online/Comment
[https://doi.org/10.1016/S0140-6736\(19\)32461-4](https://doi.org/10.1016/S0140-6736(19)32461-4)

Department of Biostatistics, Fielding School of Public Health (Prof M A Suchard MD, M J Schuemie PhD), and Department of Biomathematics, David Geffen School of Medicine at UCLA (Prof M A Suchard), University of California, Los Angeles, CA, USA; Epidemiology Analytics, Janssen Research & Development, Titusville, NJ, USA (M J Schuemie, P B Ryan PhD); Department of Medicine, Yale University School of Medicine, New Haven, CA, USA (Prof H M Krumholz MD); Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, South Korea (S C You MD); Department of Medicine, Weill Cornell Medical College, New York, NY, USA (R Chen MD); Department of Biomedical Informatics, Columbia University Medical Center, New York, NY, USA (R Chen, Prof G Hripcsak MD, P B Ryan); Division of

Atlas

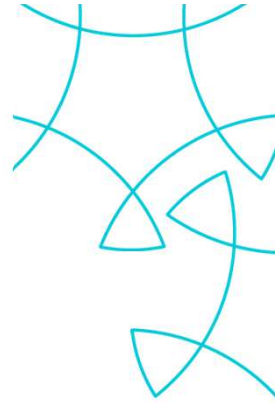
ATLAS is a free, publicly available, web-based tool developed by the OHDSI community that facilitates the design and execution of analyses on standardized, patient-level, observational data in the CDM format.



The screenshot displays the ATLAS web interface for defining a cohort. The left sidebar contains navigation options: Home, Data Sources, Search, Concept Sets, Cohort Definitions (selected), Characterizations, Cohort Pathways, Incidence Rates, Profiles, Estimation, Prediction, Jobs, Configuration, and Feedback. The main content area is titled 'Cohort #1770710' and shows the definition 'New users of ACE inhibitors as first-line monotherapy for hypertension'. Below this, there are tabs for Definition, Concept Sets, Generation, Reporting, Export, and Messages. A text input field prompts the user to 'enter a cohort definition description here'. The 'Cohort Entry Events' section is active, showing criteria for events: 'a drug exposure of ACE inhibitors for the first time in the person's history'. It includes a 'Restrict initial events' button and a 'Limit initial events to: earliest event per person' setting. The 'Inclusion Criteria' section lists two criteria: 'has hypertension diagnosis in 1 yr prior to treatment' and 'Has no prior antihypertensive drug exposures in medical history'. The footer includes the Apache 2.0 license, OHDSI logo, and eCare for CleverHealth logo.



ATLAS: Cohort building



- Optimized for observational research
 - Time series: *who* and *when* (vs classification)
 - Observation period, event timing
 - Assume a complex definition – Linearized AND-OR group

ATLAS

www.ohdsi.org/web/atlas/#/cohortdefinition/82352

applied to each cohort entry 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 the same 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 visit occurrence of Any Visit [Add](#) [Add criteria attribute...](#) [Delete Criteria](#)

with continuous observation of at least 0 days before and 5 days after event index date
Limit initial events to: all events per person.

Initial event inclusion criteria: From among the initial events, include:

People having all of the following criteria: [Add New Criteria...](#)

with exactly 0 using all occurrences of:
a condition occurrence of C Diff Diagnoses [Add](#) [Add criteria attribute...](#) [Delete Criteria](#)
starting between All days Before and 1 days After event index date and ending any time.

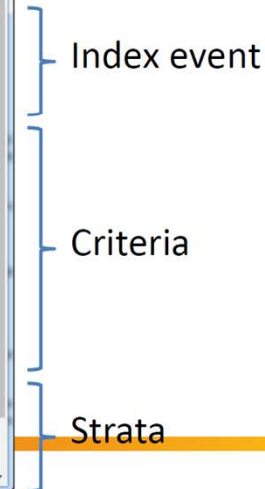
and with at least 1 using all occurrences of:
a condition occurrence of C Diff Diagnoses [Add](#) [Add criteria attribute...](#) [Delete Criteria](#)
starting between 2 days After and 30 days After event index date and ending any time.

Limit cohort of initial events to: earliest event per person.
[Remove initial event inclusion criteria](#)

Additional qualifying inclusion criteria: The qualifying cohort will be defined as all persons who have an initial event, satisfy the initial event inclusion criteria, and fulfill all additional qualifying inclusion criteria. Each qualifying inclusion criteria will be evaluated to determine the impact of the criteria on the attrition of persons from the initial cohort.
[New qualifying inclusion criteria](#) Please select a qualifying inclusion criteria to edit.

Limit qualifying cohort to: earliest event per person.

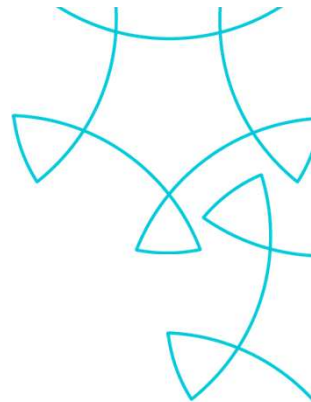
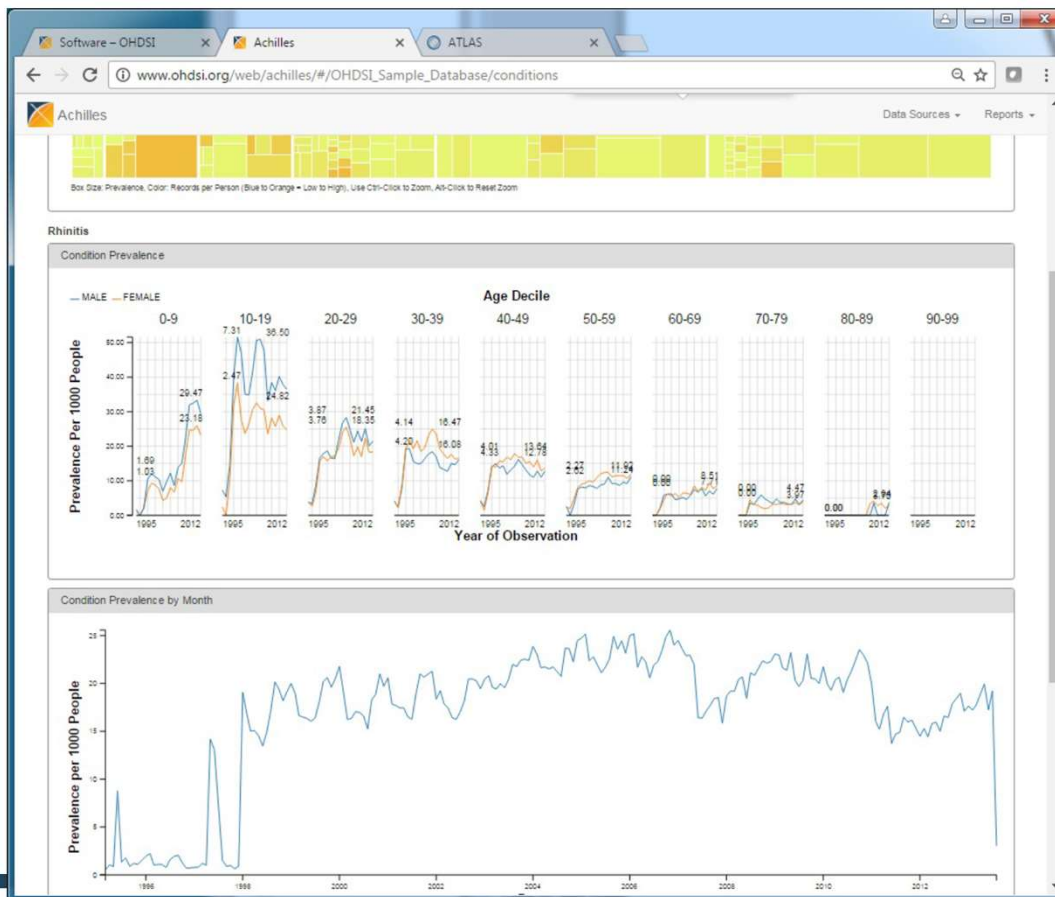
Cohort Exit Criteria
Cohort exit criteria: For all persons who entered the cohort, there must be a specification of when each person exits the cohort. A person must exit the cohort at the end of the observation period spanning the qualifying initial event start date, but additional cohort exit criteria may be also considered.
[Add a cohort exit criteria](#)





ATLAS: Visualization

- Tables
- Graphs





ATLAS: Analysis (observational)

- Approach: log regression, Poisson regression, survival
- Confounder: regularized-regression propensity score
- Residual confounding: calibration
- Diagnostics

The screenshot shows the ATLAS web interface for 'Population Level Effect Estimation'. The main configuration area includes the following fields and options:

- MEEST** (Search bar)
- Specify the statistical model used to estimate the risk of outcome between target and comparator cohorts:** Logistic regression
- Define the time-at-risk window start, relative to target/comparator cohort entry:** 0 days from cohort start date
- Define the time-at-risk window end:** 180 days from cohort end date
- Minimum washout period applied to target and comparator cohorts:** 0
- Minimum required days at risk, applied to target and comparator cohorts:** 0
- Remove patients who enter both cohorts?** Yes
- Remove patients who have observed the outcome prior to cohort entry?** Yes
- Use propensity score adjustment as a confounding adjustment strategy for baseline covariates?** Yes
- Which types of baseline covariates do you want to include in the propensity score model?**
 - Demographics
 - Gender
 - Age group (Cohort bands)

OHDSI standardized analytical tools: training

- Book of OHDSI: <https://ohdsi.github.io/TheBookOfOhdsi/>
- EH DEN academy: <https://academy.ehden.eu/>
- Training videos: <https://www.youtube.com/user/OHDSIJoinTheJourney>

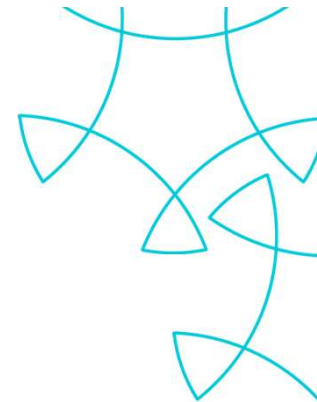
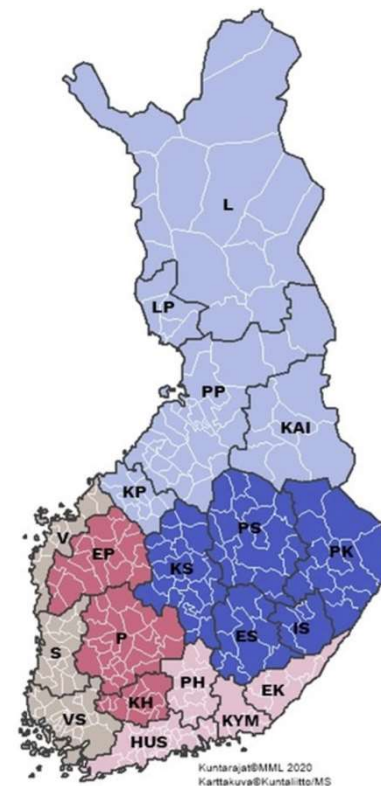


- Study-a-thons:
- Local seminars



FIN-OMOP: national, population-based OMOP CDM database of EHR and registry data

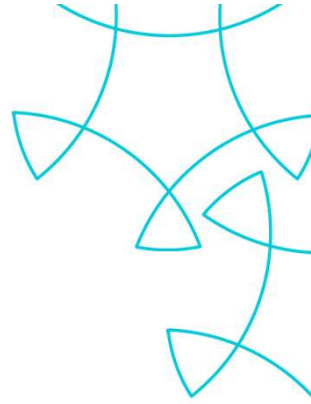
- Started 2019, project lead prof Tarja Laitinen, TAYS
- 5 university hospital districts, currently 3 actively mapping
- Inclusion of national governmental registries discussed (THL; e.g. drug prescriptions and reimbursement, cancer registry)
- Federated and centralized analysis
- First versions in production Q3 2021
- Funded by Business Finland, hospital districts, EU (EHDEN)



**BUSINESS
FINLAND**



Vision for Finland: all key observational datasets under a CDM



- All hospital datalakes
- THL registries
- Other high-quality, disease-specific registries

- A unique, population-based data resource for observational research and generating reliable evidence for decision making
- A highly desirable partner for international collaboration