Cancer OMOP, the global community and vision

13-Nov-2023

Asieh Golozar, MD PhD MHS MPH

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Northeastern University

ODYSSEUS DATA SERVICES INC

OHDSI is a Global Success

OMOP Data by the Numbers

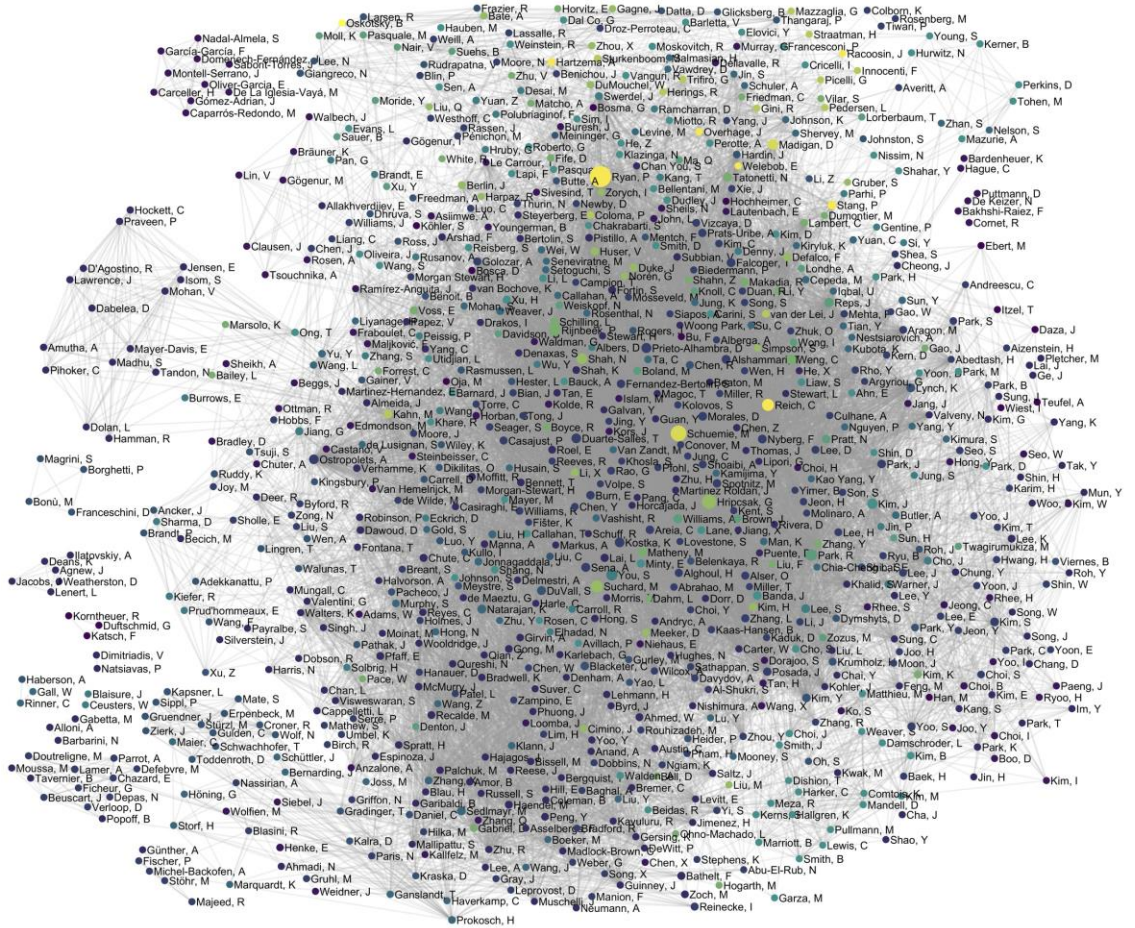
- 534 data sources
- 49 countries
- 956 million unique patient records
- approximately 12% of the world's population

OHDSI By The Numbers

- 3,758 collaborators
- 83 countries
- 21 time zones
- 6 continents
- 1 community

Collaborations in Scholarship

- More than 600 peer-reviewed studies relating to OMOP or OHDSI tools/methods/best practices have been published over the last decade
- Most papers have multiple authors, and OHDSI continues to see a rise in new authors collaborating on studies



The Secret Sources

Open-Source

Community

Standardization

Reproducibility

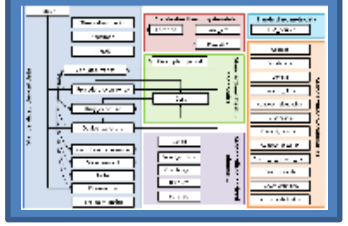
Open-Science

Scale

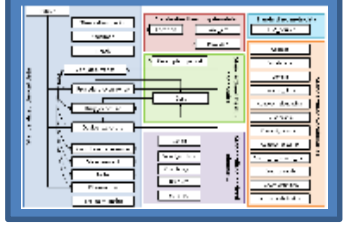
Journey to Evidence Streamlined with Standardization

Standardized data

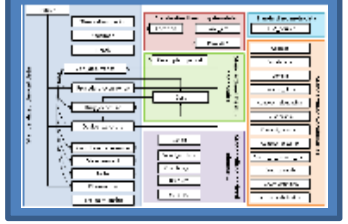
Source 1 CDM



Source 2 CDM



Source 3 CDM



Standardized inputs

Cohort definition: a specification to identify the set of persons satisfying one or more criteria for a duration of time

Standardized analytics

T, {E} → Treatment pathways

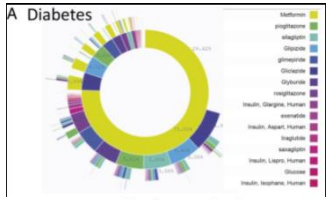
T, O → Incidence rate analysis

T, C, O → Comparative cohort design

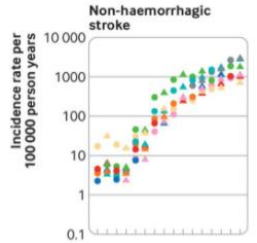
T, O → Self-controlled case series

T, O → Patient-level prediction

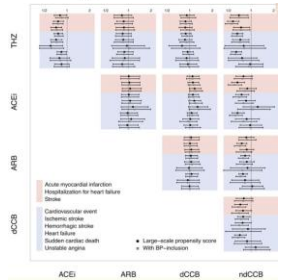
Impactful results



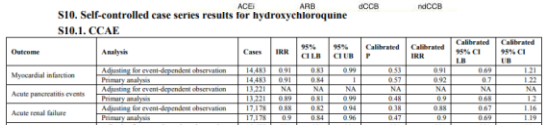
Hripcsak et al
PNAS 2016



Li et al
BMJ 2021



Suchard et al
Lancet 2019

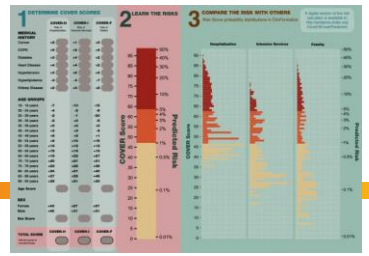


S10. Self-controlled case series results for hydroxychloroquine

S10.1. CCAE

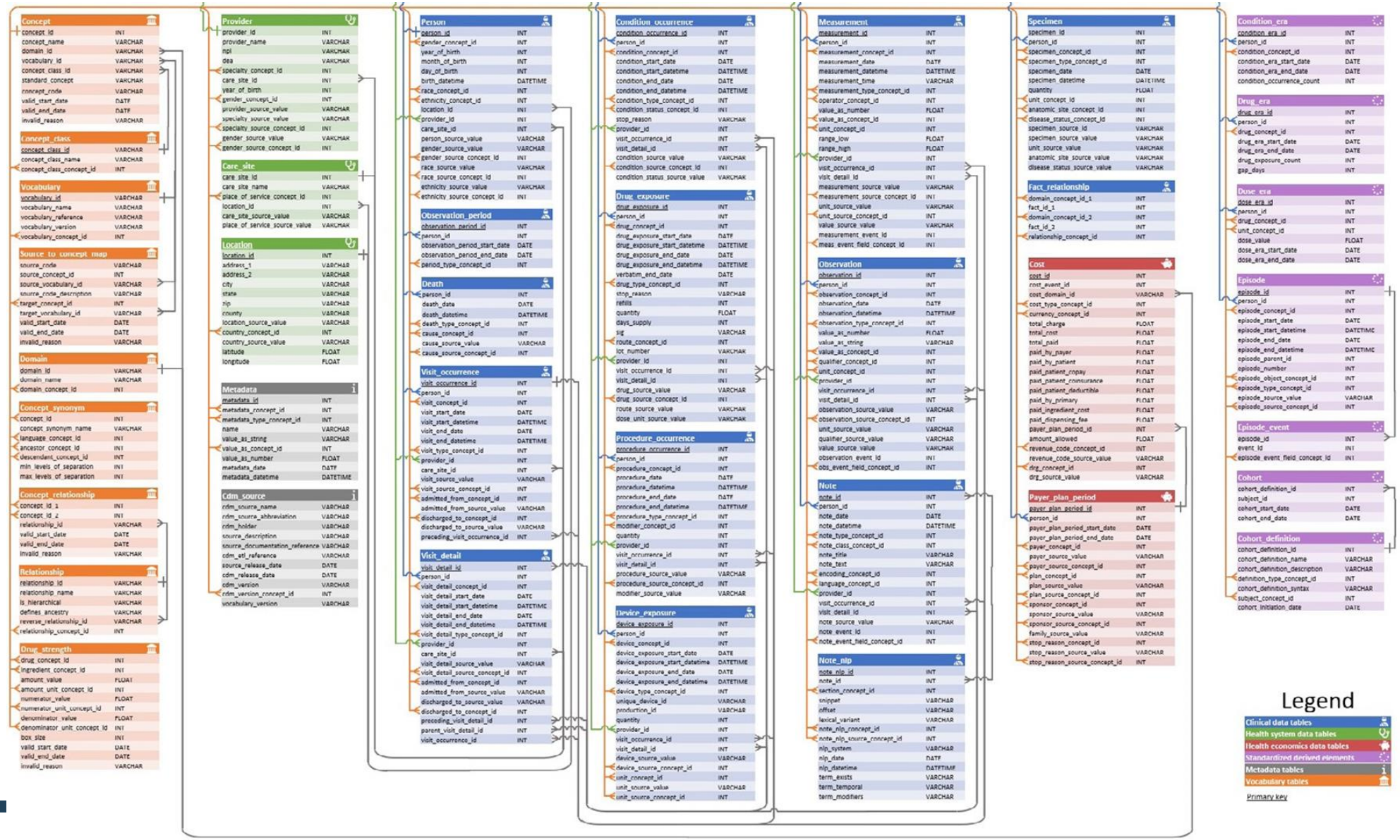
Outcome	Analysis	Cases	IRR	95% CI LB	95% CI UB	Calibrated P	Calibrated IRR	Calibrated 95% CI LB	Calibrated 95% CI UB
Myocardial infarction	Adjusting for event-dependent observation	14,483	0.91	0.83	0.99	0.53	0.91	0.69	1.21
	Primary analysis	14,483	0.91	0.84	1	0.57	0.92	0.7	1.22
Acute pancreatitis events	Adjusting for event-dependent observation	13,221	NA	NA	NA	NA	NA	NA	NA
	Primary analysis	13,221	0.89	0.81	0.99	0.48	0.9	0.68	1.2
Acute renal failure	Adjusting for event-dependent observation	17,178	0.88	0.82	0.94	0.38	0.88	0.67	1.16
	Primary analysis	17,178	0.9	0.84	0.96	0.47	0.9	0.69	1.19

Lane et al Lancet Rheumatology 2020



Williams et al
BMC MRM 2022

Data Standardization: OMOP CDM



Open-Source Community

- 262 Repositories
- 30 M+ lines of code
- 681 Developers
- 31 organizations
- 47,672 commits
- 2,838 GitHub Forks
- 4,168 GitHub Stars
- 5,547 GitHub Subscribers

Package	Version	Maintainer(s)	Availability	Open issues	pull-requests	Build status	Coverage
Achilles	v1.7.2	Frank DeFalco	CRAN	29	3	R check passing	codecov 2%
Andromeda	v0.6.3	Adam Black	CRAN	13	2	R check passing	codecov 89%
BigKnn	v1.0.2	Martijn Schuemie	GitHub	0	0	R check passing	codecov 96%
BrokenAdaptiveRidge	v1.0.0	Marc Suchard	CRAN	2	0	R check passing	codecov 96%
Capr	v2.0.7	Martin Lavallee	GitHub	2	0	R check passing	codecov 80%
Characterization	v0.1.2	Jenna Reps	GitHub	11	1	R check failing	codecov unknown
CirceR	v1.3.1	Chris Knoll	GitHub	3	1	R check passing	codecov 89%
CohortDiagnostics	v3.2.4	Jamie Gilbert	GitHub	56	2	R check passing	codecov 90%
CohortExplorer	v0.1.0	Gowtham Rao	CRAN	0	0	R check passing	codecov 100%
CohortGenerator	v0.8.1	Anthony Sena	GitHub	19	2	R check passing	codecov 98%
CohortMethod	v5.1.0	Martijn Schuemie	GitHub	14	1	R check failing	codecov 88%
Cyclops	v3.3.1	Marc Suchard	CRAN	18	0	R check failing	codecov 85%
DatabaseConnector	v6.2.4	Martijn Schuemie	CRAN	12	0	R check passing	codecov 59%
DataQualityDashboard	v2.4.1	Katy Sadowksi	GitHub	43	7	R check passing	codecov 86%
DeepPatientLevelPrediction	v2.0.0	Egill Fridgeirsson	GitHub	18	1	R check failing	codecov 100%
EmpiricalCalibration	v3.1.1	Martijn Schuemie	CRAN	1	0	R check passing	codecov 84%
EnsemblePatientLevelPrediction	v1.0.2	Jenna Reps	GitHub	5	0	R check passing	codecov unknown
Eunomia	v1.0.2	Frank DeFalco	GitHub	10	1	R check passing	codecov 72%
EvidenceSynthesis	v0.5.0	Martijn Schuemie	CRAN	3	0	R check passing	codecov 78%
FeatureExtraction	v3.3.1	Anthony Sena	GitHub	44	6	R check failing	codecov 93%
Hydra	v0.4.0	Anthony Sena	GitHub	6	7	R check failing	codecov 87%
IterativeHardThresholding	v1.0.2	Marc Suchard	CRAN	1	0	R check passing	codecov 92%
MethodEvaluation	v2.3.0	Martijn Schuemie	GitHub	1	0	R check passing	codecov 59%
OhdsiSharing	v0.2.2	Lee Evans	GitHub	0	1	R check passing	codecov 0%
OhdsiShinyModules	v2.0.0	Jenna Reps	GitHub	108	2	R check failing	codecov 79%
ParallelLogger	v3.3.0	Martijn Schuemie	CRAN	4	0	R check passing	codecov 81%
PatientLevelPrediction	v6.3.6	Jenna Reps & Peter Rijnbeek	GitHub	47	0	R check failing	codecov 89%
PhenotypeLibrary	v3.30.1	Gowtham Rao	GitHub	1	0	R check passing	codecov 100%

OHDSI Methods and Research

Three ideas

1. LEGEND Principles

METHODS RESEARCH

LEGEND in Principle

LEGEND (Large-scale Evidence Generation and Evaluation across a Network of Databases) applies high-level analytics to perform observational research on hundreds of millions of patient records within OHDSI's international database network.

LEGEND is based on 10 guiding principles that were published in JAMIA (August, 2020) and are listed below.

Journal of the American Medical Informatics Association, 27(8), 2020, 1331–1337
doi: 10.1093/jamia/ocaa103
Perspective

Perspective

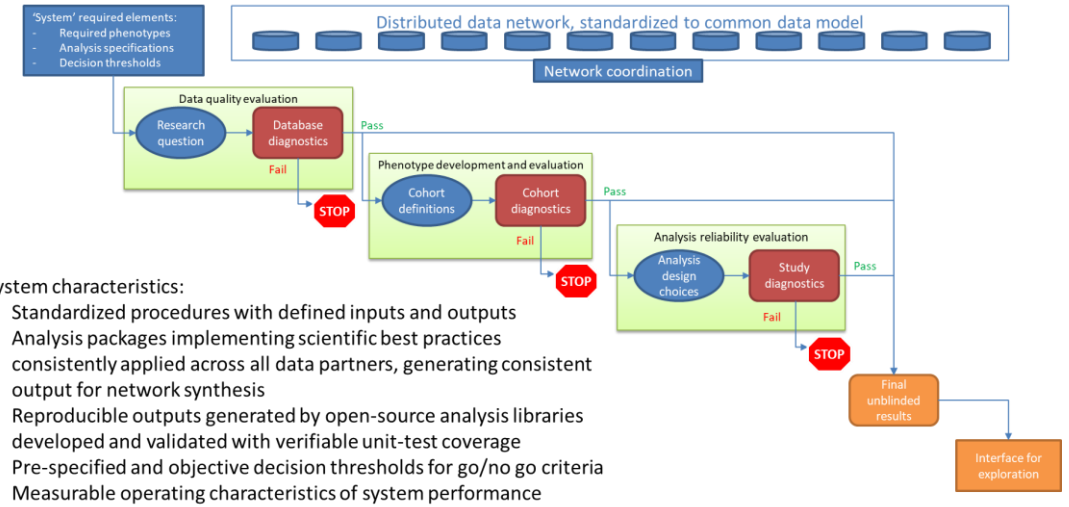
Principles of Large-scale Evidence Generation and Evaluation across a Network of Databases (LEGEND)

Martijn J. Schuemie[1,2], Patrick B. Ryan[1,3], Nicole Pratt[4], RuiJun Chen[3,5], Seng Chan You[6], Harlan M. Krumholz[7], David Madigan[8], George Hripcsak[3,9], and Marc A. Suchard[2,10]

1. **LEGEND will generate evidence at a large scale.** Instead of answering a single question at a time (eg, the effect of 1 treatment on 1 outcome), LEGEND answers large sets of related questions at once (eg, the effects of many treatments for a disease on many outcomes). Aim: Avoids publication bias, achieves comprehensiveness of results, and allows for an evaluation of the overall coherence and consistency of the generated evidence.
2. **Dissemination of the evidence will not depend on the estimated effects.** All generated evidence is disseminated at once. Aim: Avoids publication bias and enhances transparency.
3. **LEGEND will generate evidence using a prespecified analysis design.** All analyses, including the research questions that will be answered, will be decided prior to analysis execution. Aim: Avoids P hacking.
4. **LEGEND will generate evidence by consistently applying a systematic process across all research questions.** This principle precludes modification of analyses to obtain a desired answer to any specific question. This does not imply a simple one-size-fits-all process, rather that the logic for modifying an analysis for specific research questions should be explicated and applied systematically. Aim: Avoids P hacking and allows for the evaluation of the operating characteristics of this process (Principle 6).
5. **LEGEND will generate evidence using best practices.** LEGEND answers each question using current best practices, including advanced methods to address confounding, such as propensity scores. Specifically, we will not employ suboptimal methods (in terms of bias) to achieve better computational efficiency. Aim: Minimizes bias.
6. **LEGEND will include empirical evaluation through the use of control questions.** Every LEGEND study includes control questions. Control questions are questions where the answer is known. These allow for measuring the operating characteristics of our systematic process, including residual bias. We subsequently account for this observed residual bias in our P values, effect estimates, and confidence intervals using empirical calibration. [7,8] Aim: Enhances transparency on the uncertainty due to residual bias.
7. **LEGEND will generate evidence using open-source software that is freely available to all.** The analysis software is open to review and evaluation, and is available for replicating analyses down to the smallest detail. Aim: Enhances transparency and allows replication.
8. **LEGEND will not be used to evaluate new methods.** Even though the same infrastructure used in LEGEND may also be used to evaluate new causal inference methods, generating clinical evidence should not be performed at the same time as method evaluation. This is a corollary of Principle 5, since a new method that still requires evaluation cannot already be best practice. Also, generating evidence with unproven methods can hamper the interpretability of the clinical results. Note that LEGEND does evaluate how well the methods it uses perform in the specific context of the questions and data used in a LEGEND study (Principle 6). Aim: Avoids bias and improves interpretability.
9. **LEGEND will generate evidence across a network of multiple databases.** Multiple heterogeneous databases (different data capture processes, health-care systems, and populations) will be used to generate the evidence to allow an assessment of the replicability of findings across sites. Aim: Enhances generalizability and uncovers potential between-site heterogeneity.
10. **LEGEND will maintain data confidentiality; patient-level data will not be shared between sites in the network.** Not sharing data will ensure patient privacy, and comply with local data governance rules. Aim: Privacy.

2. Objective Diagnostics

Engineering open science systems that build trust into the real-world evidence generation and dissemination process



System characteristics:

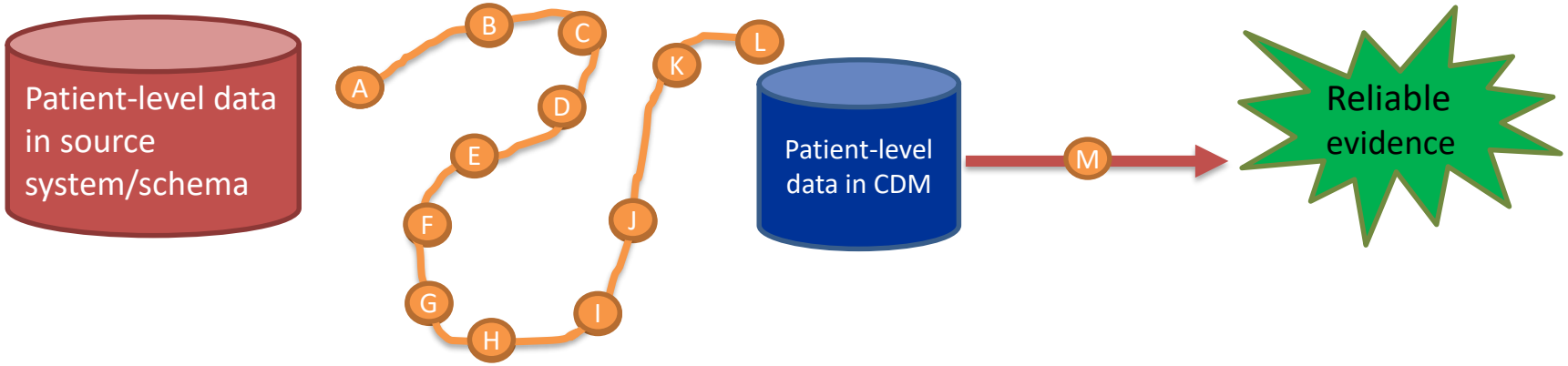
- Standardized procedures with defined inputs and outputs
- Analysis packages implementing scientific best practices consistently applied across all data partners, generating consistent output for network synthesis
- Reproducible outputs generated by open-source analysis libraries developed and validated with verifiable unit-test coverage
- Pre-specified and objective decision thresholds for go/no go criteria
- Measurable operating characteristics of system performance

3. Standardized software

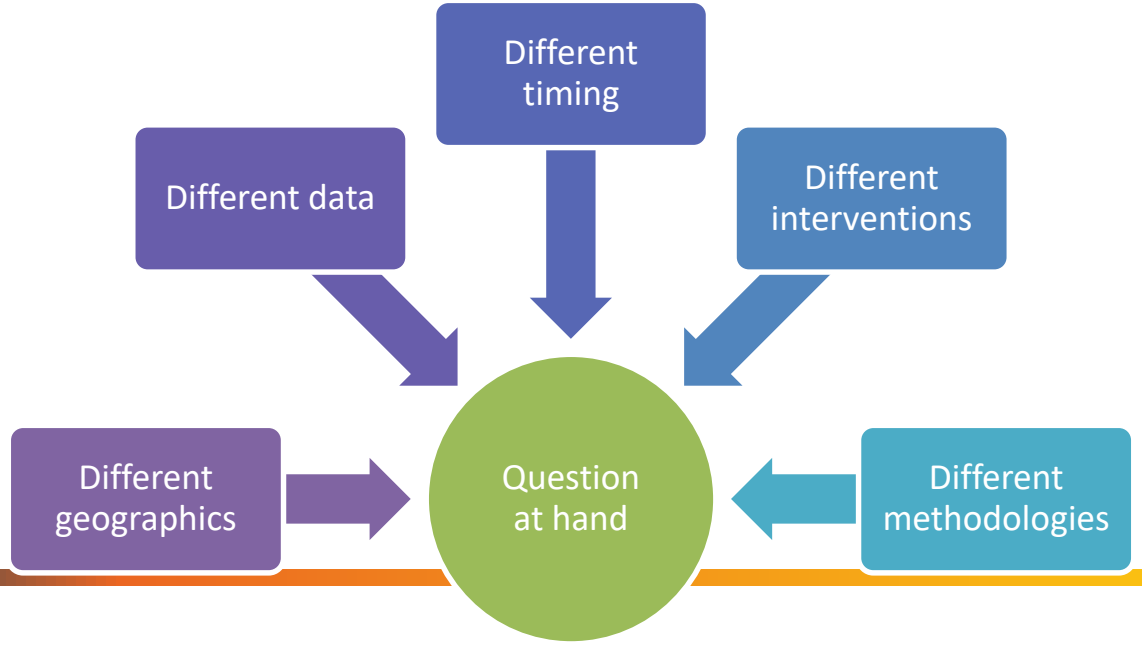
HADES
HEALTH ANALYTICS DATA-TO-EVIDENCE SUITE

Open-source software

Reproducibility and Trust



A Common Data Model enables standardized analytics to generate reliable evidence.



Towards Reliable Evidence …

Some current practices across the broader research community

Examining one target-comparator pair at a time

Not using appropriate methods to control for bias

Modify the design until significant results are found

Large-scale evidence generation across a network of databases (LEGEND)

* Pre-specified fixed design and dissemination of the results regardless of the estimates (avoid publication bias)
* Systematic process across all research questions
* Large-scale: looking at thousands of target-comparator pairs at a time
* Use of best practices: LSPS, extensive diagnostics, negative and positive controls

Some of the LEGEND Principles Step-by-Step

- 39 mono-drugs, 13 mono-classes
- 58 duo-drugs, 32 duo-classes
- 10,278 comparisons

	Theoretical	Observed (n > 2,500)
Single ingredients	58	39
Single ingredient comparisons	58 * 57 = 3,306	1,296
Single drug classes	15	13
Single class comparisons	15 * 14 = 210	156
Dual ingredients	58 * 57 / 2 = 1,653	58
Single vs duo drug comparisons	58 * 1,653 = 95,874	3,810
Dual classes	15 * 14 / 2 = 105	32
Single vs duo class comparisons	15 * 105 = 1,575	832
Duo vs duo drug comparisons	1,653 * 1,652 = 2,730,756	2,784
Duo vs duo class comparisons	105 * 104 = 10,920	992
...	...	...
Total comparisons	2,843,250	10,278

Select multiple target and comparator cohorts, for example, all drug in a drug class

↓

Carefully design the study, including sensitivity analyses

↓

Run on multiple databases

↓

Spend a lot of time on diagnostics: propensity score balance, covariate balance, calibration plots etc.

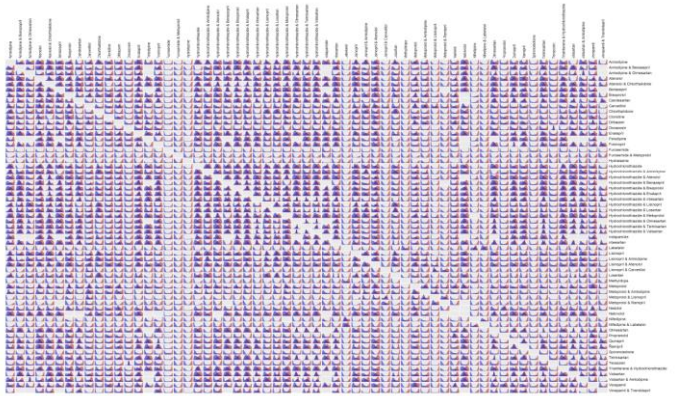
↓

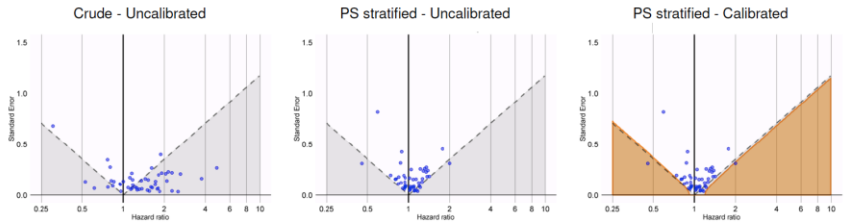
Publish all results

OHDSI

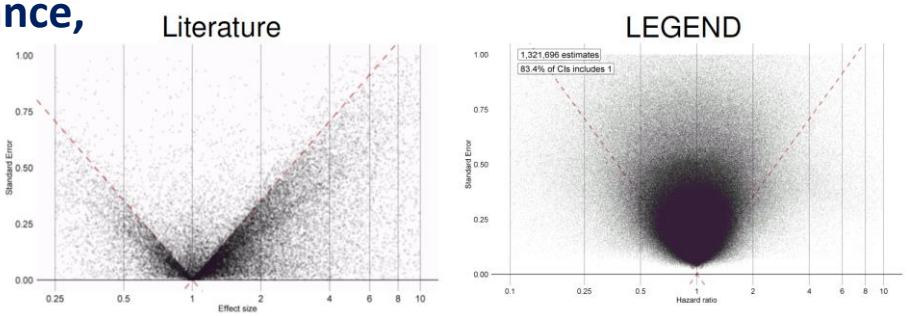
OHDSI Large-Scale Evidence Generation and Evaluation in a Network of Databases (LEGEND): Study of the Effects of Treatments for Hypertension

Version: 0.1





68% have $p < 0.05$ | 16% have $p < 0.05$ | 4% have $p < 0.05$



LEGEND in practice

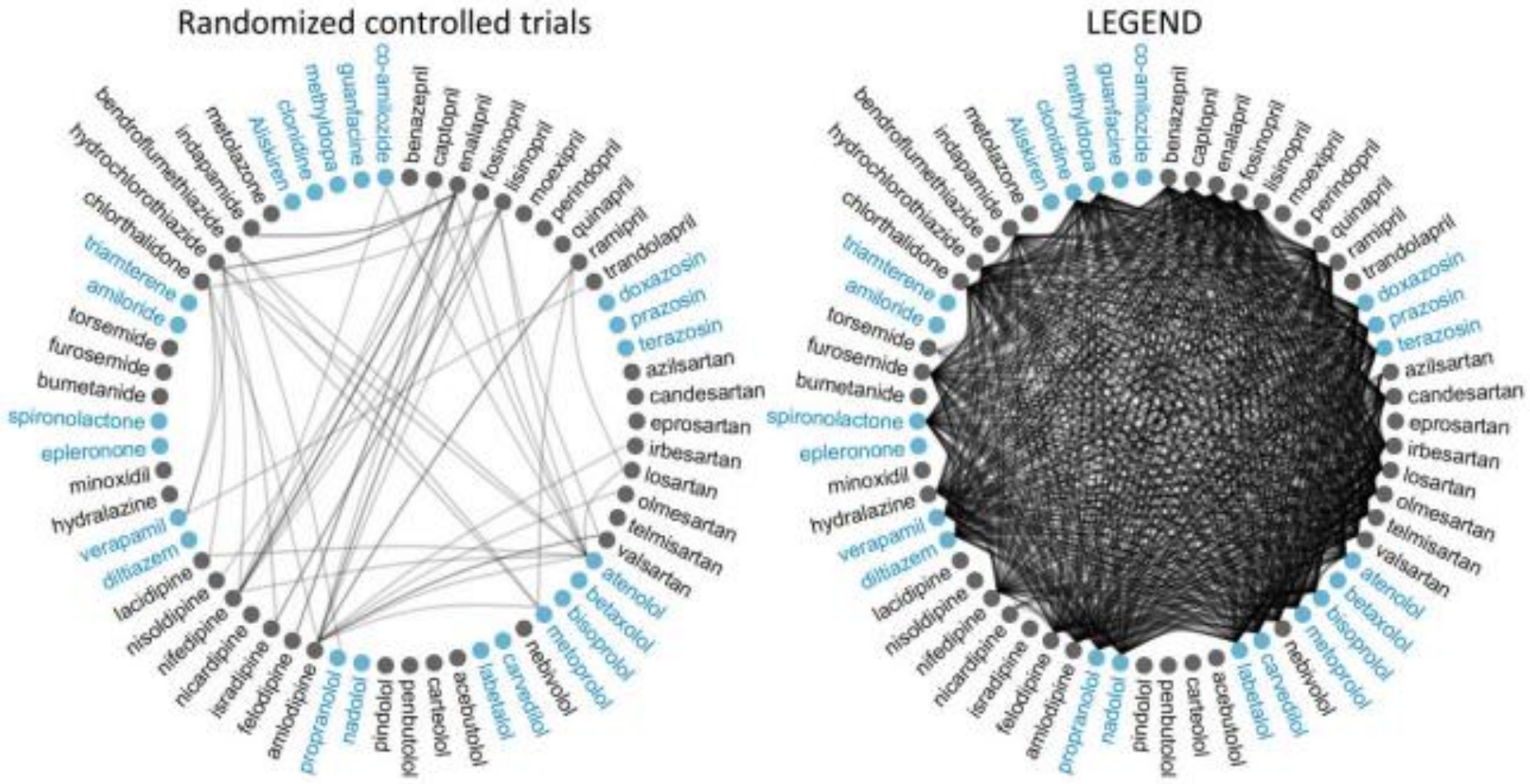


Figure 3. Comparisons of single-drug hypertension treatments in randomized controlled trials (left) and in LEGEND (right). Each circle represents an ingredient. Color groupings indicate drug classes. A line between circles indicates the 2 drugs are compared in at least 1 study.

OHDSI Scales

	Theoretical	Observed (n>2,500)
Single ingredients	58	39
Single ingredient comparisons	58 * 57 = 3,306	1,296
Single drug classes	15	13
Single class comparisons	15 * 14 = 210	156
Dual ingredients	58 * 57 / 2 = 1,653	58
Single vs duo drug comparisons	58 * 1,653 = 95,874	3,810
Dual classes	15 * 14 / 2 = 105	32
Single vs duo class comparisons	15 * 105 = 1,575	832
Duo vs duo drug comparisons	1,653 * 1,652 = 2,730,756	2,784
Duo vs duo class comparisons	105 * 104 = 10,920	992
…	…	…
Total comparisons	2,843,250	10,278
Outcomes of interest	58	58
Target-comparator-outcomes	2,843,250 * 58 = 164,908,500	587,020
Negative control outcomes	76	76
Target-comparator-neg controls	2,843,250 * 76=216,087,000	769,476
Positive control outcomes	76 * 3=228	228
Target-comparator-pos controls	2,843,250 * 228=648,261,00	662,484
Total comparisons	864,348,000	1,431,960
Total	**864,348,000 * 9= 7,779,132,000**	**1,431,960 * 9=12,887,640**

- **US Insurance databases**
- IBM® MarketScan® CCAE
- IBM® MarketScan® MDCD
- IBM® MarketScan® MDCR
- Optum© Clinformatics®
- **Japanese insurance databases**
- Japan Medical Data Center
- **Korean national insurance databases**
- NHIS-NSC
- **US EHR databases**
- Columbia University Medical Center
- Optum© PANTHER®
- **German EHR databases**
- QuintilesIMS Disease Analyzer (DA) Germany

Why is oncology any different than the rest of medicine?

Problem 1: Cancer is a rare disease

HIGHLIGHTS OF PRESCRIBING INFORMATION
These highlights do not include all the information needed to use TABRECTA safely and effectively. See full prescribing information for TABRECTA.

TABRECTA™ (capmatinib) tablets, for oral use
Initial U.S. Approval: 2020

----------INDICATIONS AND USAGE----------
TABRECTA is a kinase inhibitor indicated for the treatment of adult patients with metastatic non-small cell lung cancer (NSCLC) whose tumors have a mutation that leads to mesenchymal-epithelial transition (MET) exon 14 skipping as detected by an FDA-approved test.

This indication is approved under accelerated approval based on overall response rate and duration of response. Continued approval for this indication may be contingent upon verification and description of clinical benefit in confirmatory trial(s). (1)

----------DOSAGE AND ADMINISTRATION----------
- Select patients for treatment with TABRECTA based on presence of a mutation that leads to MET exon 14 skipping. (2.1)
- Recommended dosage: 400 mg orally twice daily with or without food. (2.2)

----------DOSAGE FORMS AND STRENGTHS----------
Tablets: 150 mg and 200 mg (3)

----------CONTRAINDICATIONS----------
None. (4)

----------WARNINGS AND PRECAUTIONS----------
- Interstitial Lung Disease (ILD)/Pneumonitis: Monitor for new or worsening pulmonary symptoms indicative of ILD/pneumonitis. Permanently discontinue TABRECTA in patients with ILD/pneumonitis. (2.3, 5.1)
- Hepatotoxicity: Monitor liver function tests. Withhold, dose reduce, or permanently discontinue TABRECTA based on severity. (2.3, 5.2)
- Risk of Photosensitivity: May cause photosensitivity reactions. Advise patients to limit direct ultraviolet exposure. (5.3)
- Embryo-Fetal Toxicity: Can cause fetal harm. Advise patients of the potential risk to a fetus and to use effective contraception. (5.4, 8.1, 8.3)

----------ADVERSE REACTIONS----------
The most common adverse reactions (≥ 20%) are peripheral edema, nausea, fatigue, vomiting, dyspnea, and decreased appetite. (6)

To report SUSPECTED ADVERSE REACTIONS, contact Novartis Pharmaceuticals Corporation at 1-888-669-6682 or FDA at 1-800-FDA-1088 or www.fda.gov/medwatch.

----------DRUG INTERACTIONS----------
Strong and Moderate CYP3A Inducers: Avoi

----------USE IN SPECIFIC POP
Lactation: Advise not to breastfeed. (8.2)

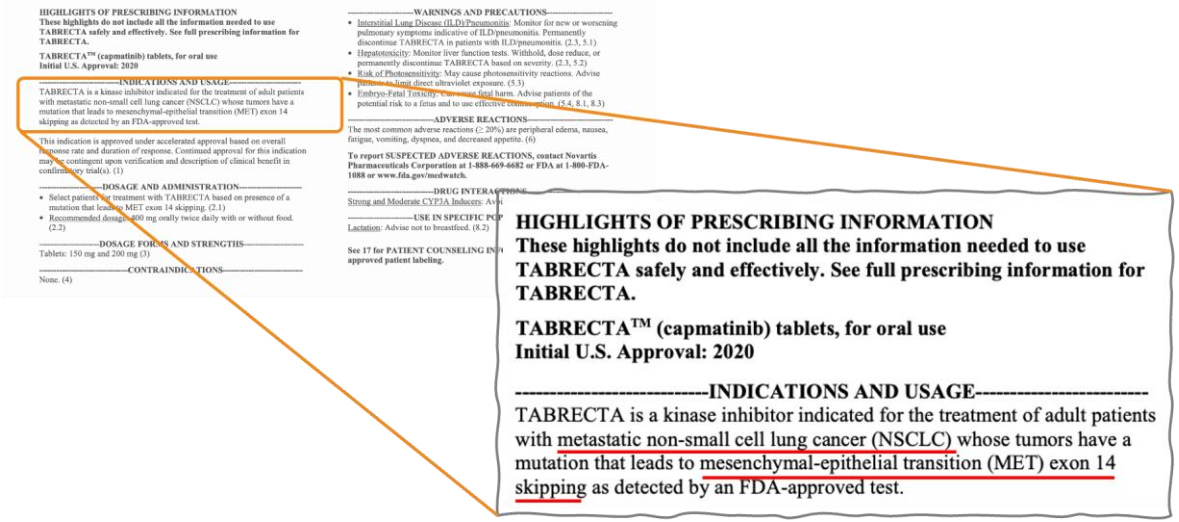
See 17 for PATIENT COUNSELING INF
approved patient labeling.

HIGHLIGHTS OF PRESCRIBING INFORMATION
These highlights do not include all the information needed to use TABRECTA safely and effectively. See full prescribing information for TABRECTA.

TABRECTA™ (capmatinib) tablets, for oral use
Initial U.S. Approval: 2020

----------INDICATIONS AND USAGE----------
TABRECTA is a kinase inhibitor indicated for the treatment of adult patients with metastatic non-small cell lung cancer (NSCLC) whose tumors have a mutation that leads to mesenchymal-epithelial transition (MET) exon 14 skipping as detected by an FDA-approved test.

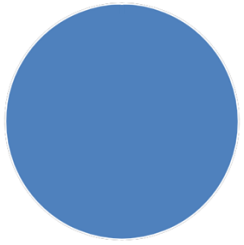
Problem 1: Cancer is a rare disease



HIGHLIGHTS OF PRESCRIBING INFORMATION
These highlights do not include all the information needed to use TABRECTA safely and effectively. See full prescribing information for TABRECTA.

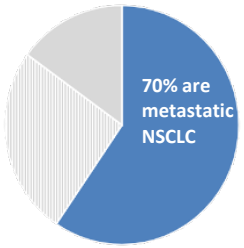
TABRECTA™ (capmatinib) tablets, for oral use
Initial U.S. Approval: 2020

----------INDICATIONS AND USAGE----------
TABRECTA is a kinase inhibitor indicated for the treatment of adult patients with metastatic non-small cell lung cancer (NSCLC) whose tumors have a mutation that leads to mesenchymal-epithelial transition (MET) exon 14 skipping as detected by an FDA-approved test.

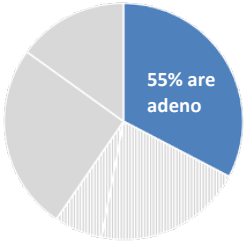


2,000,000 cases

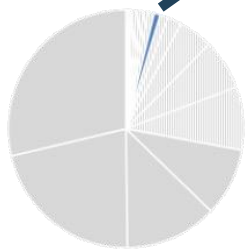
1,870,000 cases



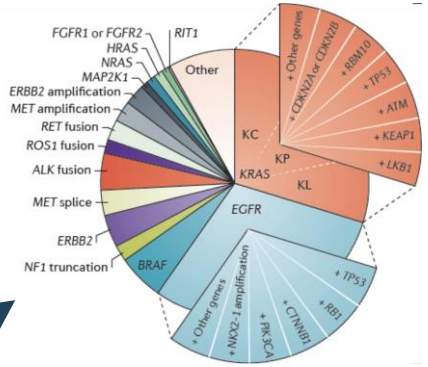
1,309,000 cases



719,950 cases



1% have MET ex14 mutation



7,200 cases

Problem 2: Cancer needs detail

“What is the Overall Survival or Progression-free Survival of patients with metastatic Non-small Cell Lung Cancer with confirmed MET exon 14 skipping who received oral capmatinib as first line?“

Problem 2: Cancer needs detail

“What is the Overall Survival or Progression-free Survival of patients with metastatic Non-small Cell Lung Cancer with confirmed MET exon 14 skipping who received oral capmatinib as first line?“

Concept	Category
Non-small Cell	Histology

Problem 2: Cancer needs detail

“What is the Overall Survival or Progression-free Survival of patients with metastatic Non-small Cell Lung Cancer with confirmed MET exon 14 skipping who received oral capmatinib as first line?“

Concept	Category
Non-small Cell	Histology
Lung	Anatomical site

Problem 2: Cancer needs detail

“What is the Overall Survival or Progression-free Survival of patients with metastatic Non-small Cell Lung Cancer with confirmed MET exon 14 skipping who received oral capmatinib as first line?“

Concept	Category
Non-small Cell	Histology
Lung	Anatomical site
Metastatic disease	Tumor attribute

Problem 2: Cancer needs detail

“What is the Overall Survival or Progression-free Survival of patients with metastatic Non-small Cell Lung Cancer with confirmed MET exon 14 skipping who received oral capmatinib as first line?“

Concept	Category
Non-small Cell	Histology
Lung	Anatomical site
Metastatic disease	Tumor attribute
MET exon 14 skipping	Genomic Variant

Problem 2: Cancer needs detail

"What is the Overall Survival or Progression-free Survival of patients with metastatic Non-small Cell Lung Cancer with confirmed MET exon 14 skipping who received oral capmatinib as first line?"

Concept	Category
Non-small Cell	Histology
Lung	Anatomical site
Metastatic disease	Tumor attribute
MET exon 14 skipping	Genomic Variant
First line treatment	Treatment Episode

Problem 2: Cancer needs detail

"What is the Overall Survival or Progression-free Survival of patients with metastatic Non-small Cell Lung Cancer with confirmed MET exon 14 skipping who received oral capmatinib as first line?"

Concept	Category
Non-small Cell	Histology
Lung	Anatomical site
Metastatic disease	Tumor attribute
MET exon 14 skipping	Genomic Variant
First line treatment	Treatment Episode
Capmatinib	Regimen

Problem 3: No standards

There are no common or even good terminologies

Concept	Category	
Non-small Cell	Histology	ICDO, SNOMED
Lung	Anatomical site	ICDO, SNOMED
Metastatic disease	Tumor attribute	
MET exon 14 skipping	Genomic Variant	CiVIC, OncoKB, ClinVar, NCIt, CAP, LOINC, SNOMED
First line treatment	Treatment Episode	
Capmatinib	Regimen	RxNorm, HemOnc

Problem 3: No standards

Vocabularies are badly curated

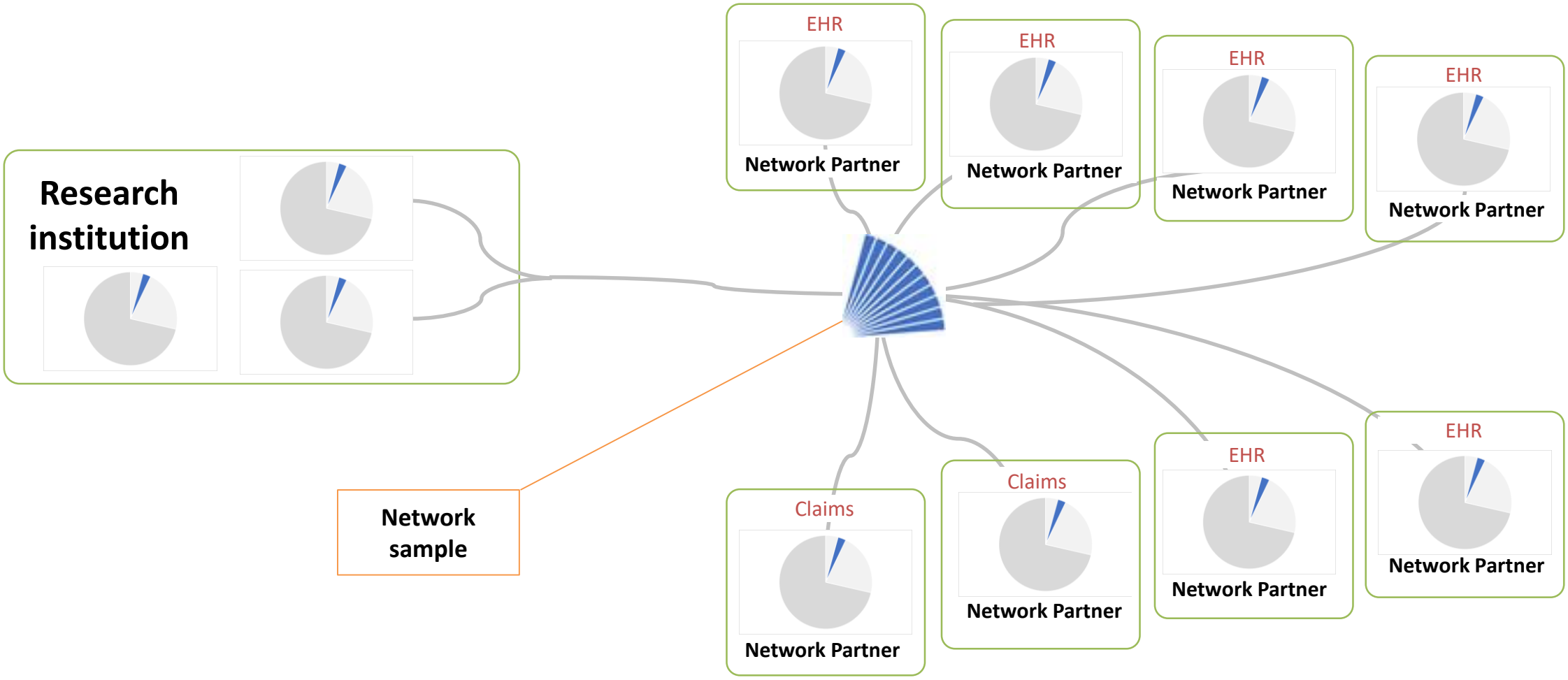
- Lymph Node Status:
  - Lymph Node Status
  - Nodal Status: Para-Aortic, Mediastinal, Pelvic, Femoral Inguinal and Distant (Mediastinal, Scalene)
  - LN Status: Femoral-Inguinal, Para-Aortic, Pelvic
  - Clinical Status of Lymph Node Mets
  - Clinical Status of Lymph Nodes
- Lymph node size:
  - LN Size
  - Size of Lymph Nodes
- Mets at DX-Distant LN
- LN Distant: Mediastinal, Scalene
- Adenopathy
- Nodal Stations Involved
- Laterality
  - Laterality
  - LN Laterality
  - Regional Lymph Node – Laterality

The OHDSI Oncology Working Group Has Worked on the Solution

- Oncology Network
- Oncology Module for the OMOP CDM
- Example studies

OHDSI Oncology Network

Data from many institutions can be analyzed together



OMOP CDM: Oncology Module

Solves all problems of oncology research

1. **Cancer Disease Model**
   - Cancer Diagnosis: Base Diagnosis + Diagnostic Modifiers
   - (One-to-many connection between them)

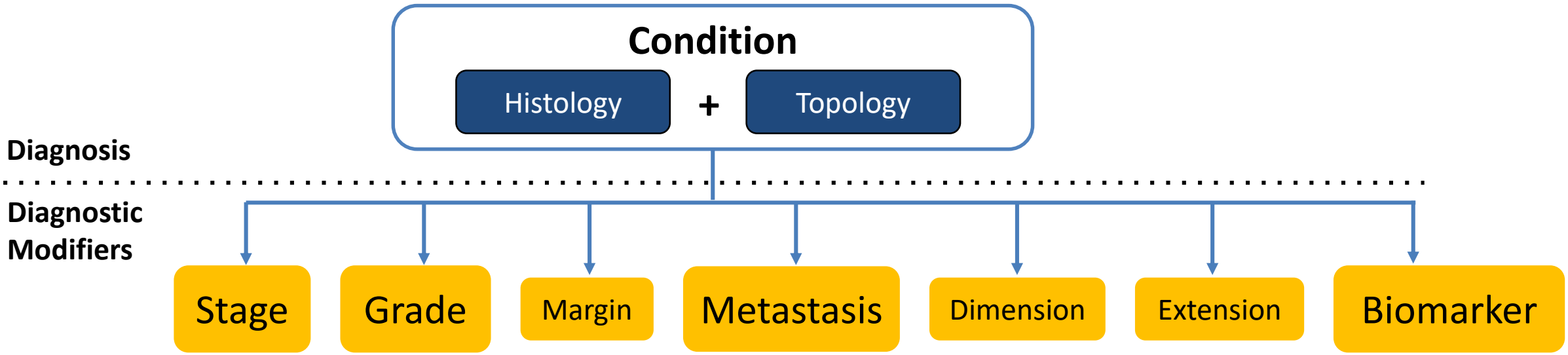
2. **Cancer Treatment Model**
   - Composite Level (Treatment Episodes) or Individual Level (standard OMOP)

3. **Cancer Episode Model**
   - Continuous periods of disease or treatment with distinct clinical meaning
   - Composed of multiple events
   - Essential for conducting cancer research

1 OMOP Oncology: Cancer Disease Model

Cancer Disease Model

Cancer Diagnosis: Base Diagnosis + Diagnostic Modifiers

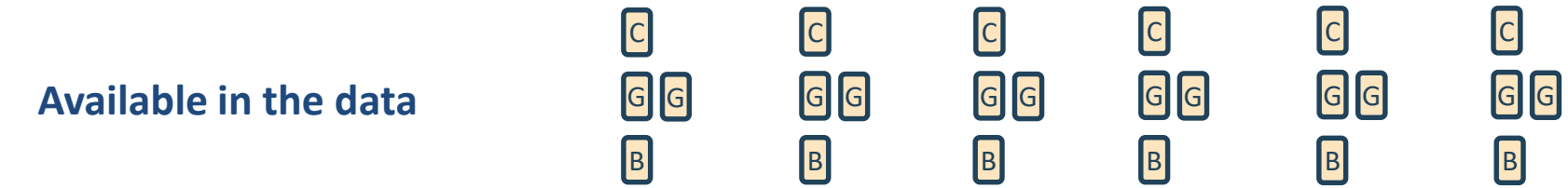


2 OMOP Oncology: Cancer Treatment Model

Abstracted **chemotherapy** regimens rarely available

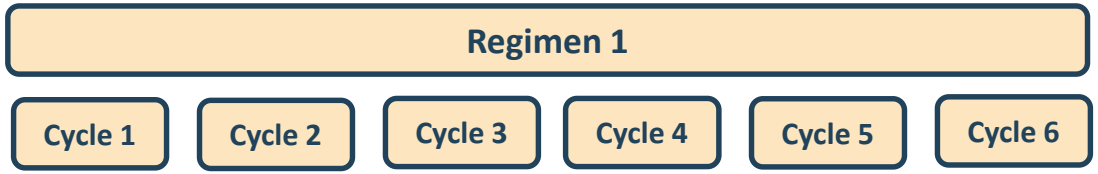
Metastatic non-squamous NSCLC

Cisplatin+Gemcitabine (GC)+Bevacizumab
21-day cycle for up to 6 cycles



Available in the data

Needed for research but mostly not available



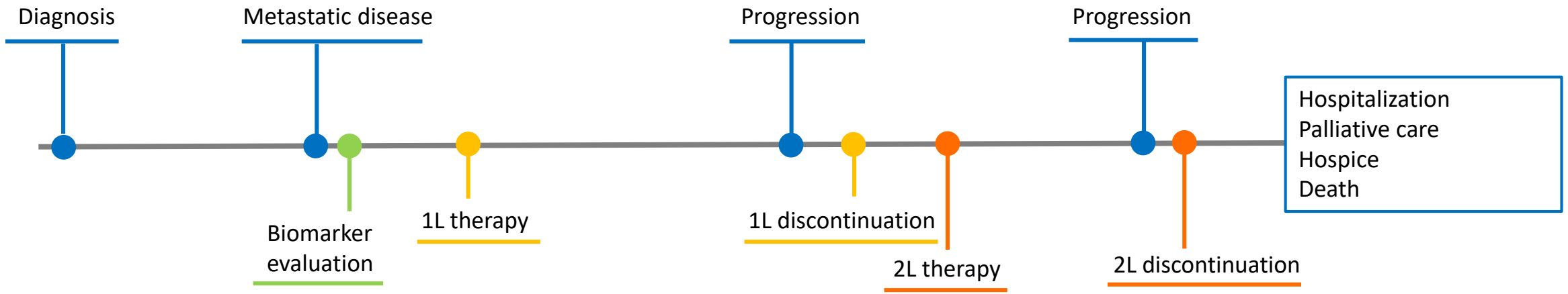
3 OMOP Oncology: Cancer Episode Model

Episodes

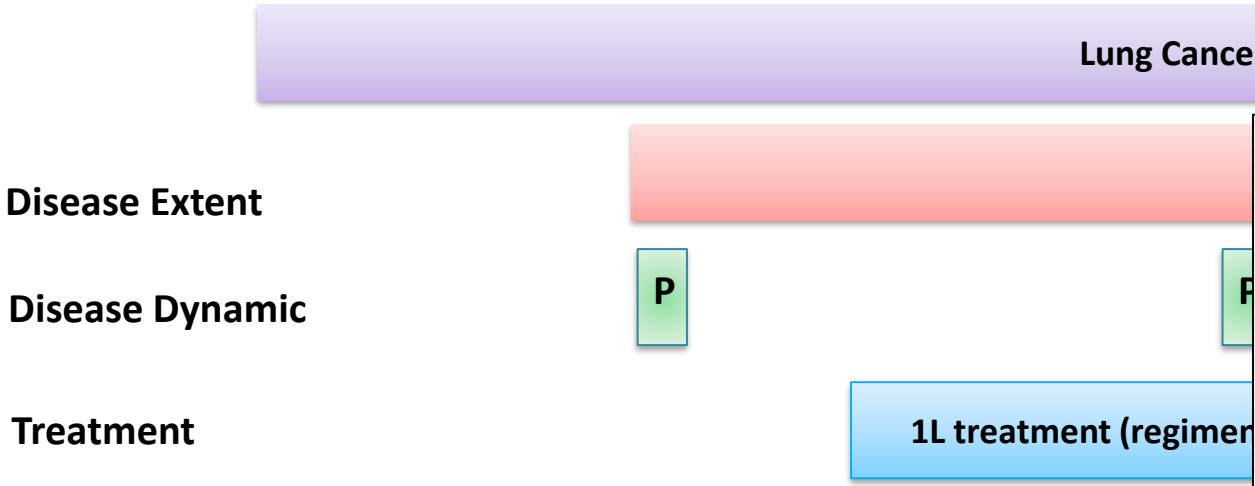
- ✓ Continuous periods of disease or treatment with distinct clinical meaning
- ✓ Composed of multiple events
- ✓ Essential for conducting cancer research
- ✓ Obtained directly from source data (e.g., registries) or algorithmically derived

- **Parent Episode:**
  - **Overarching disease episode:** Covers the entire cancer duration
- **Children Episodes:**
  - **Disease dynamic** (remission, stable, progression)
  - **Disease extent** (confined, invasive, metastatic)

Cancer Episode Model: Schematic Patient Journey



Episodes



Model allows addressing a variety of questions, including but not limited to:

- Overall survival
- Progression free survival
- Time to treatment discontinuation
- Time to next line of treatment
- Predictive and prognostic value of specific biomarkers

Cancer Disease Model: Terminologies

Solves all problems of oncology research

1 **Cancer Disease Model**

Cancer Diagnosis: Base Diagnosis + Diagnostic Modifiers

ICD-O Cancer Modifiers + OMOP Genomics

2 **Cancer Treatment Model**

Composite Level (Treatment Episodes) or Individual Level (standard OMOP)

HemOnc

3 **Cancer Episode Model**

Overarching disease episode

Disease dynamic (remission, stable, progression)

Disease extent (confined, invasive, metastatic)

De novo vocabularies

Genomic Variants are not Features

Without features, epidemiological methods cannot work

- Genomic markers need to be turned into **features** of an **analytical dataset**:

Met exon 14 skipping mutation

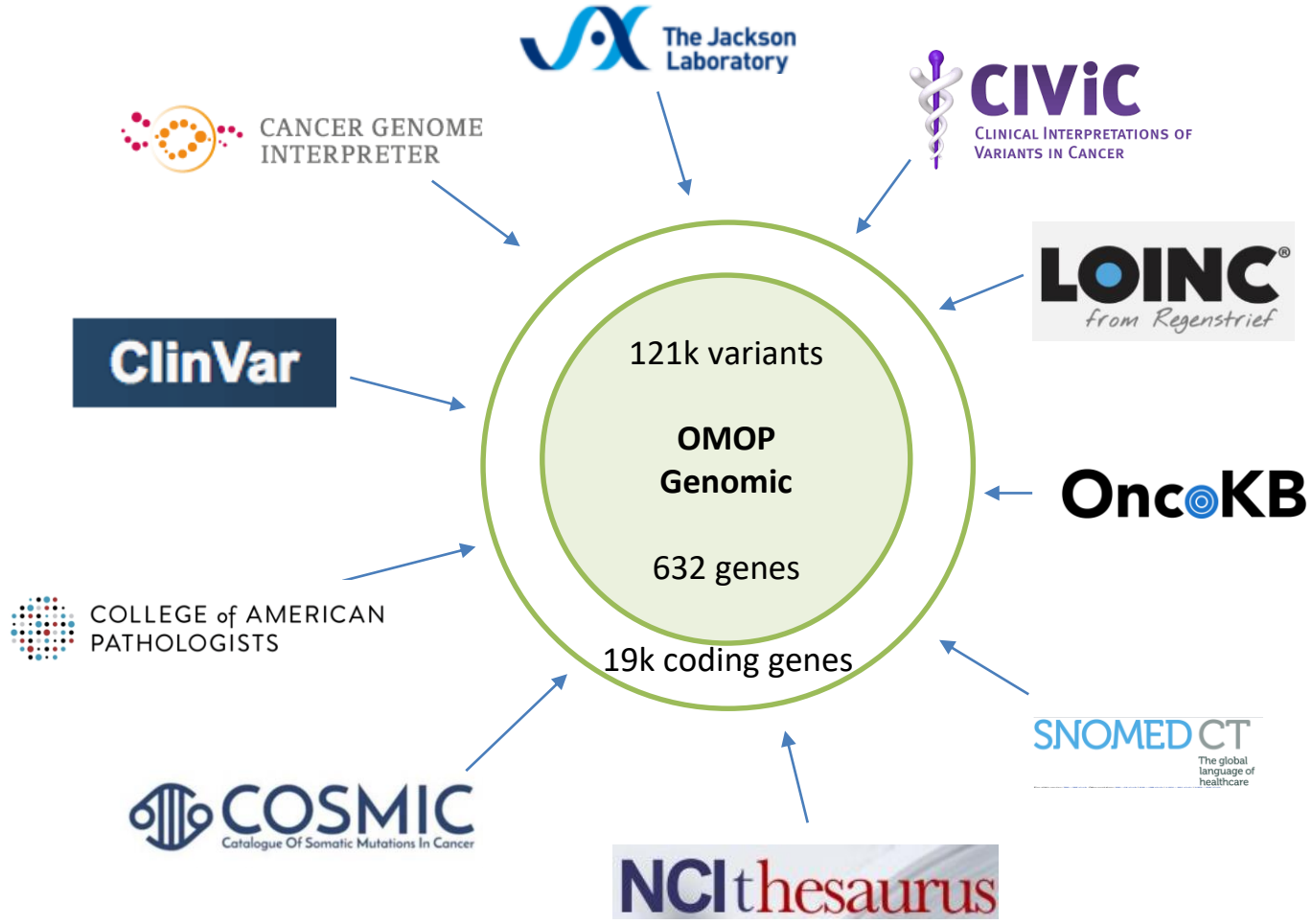
- Splice region variant
- Chromosome 7q31.2
- Exon 14 missing
- Location 116411884-116411895

	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6	Feature 7	Feature 8	Feature 9	Feature 10	Feature 11	Feature 12
Person 1												
Person 2												
Person 3												
Person 4												
Person 5												
Person 6												
Person 7												
Person 8					1							
Person 9												
Person 10												
Person 11												
Person 12												
Person 13											1	
Person 14												
Person 15												
Person 16												
Person 17		1										
Person 18												

OMOP Genomic is built from relevant sources

… by

- Combining public repositories
- Deduping them



Oncology module enables observational cancer study in a network setting

We can study

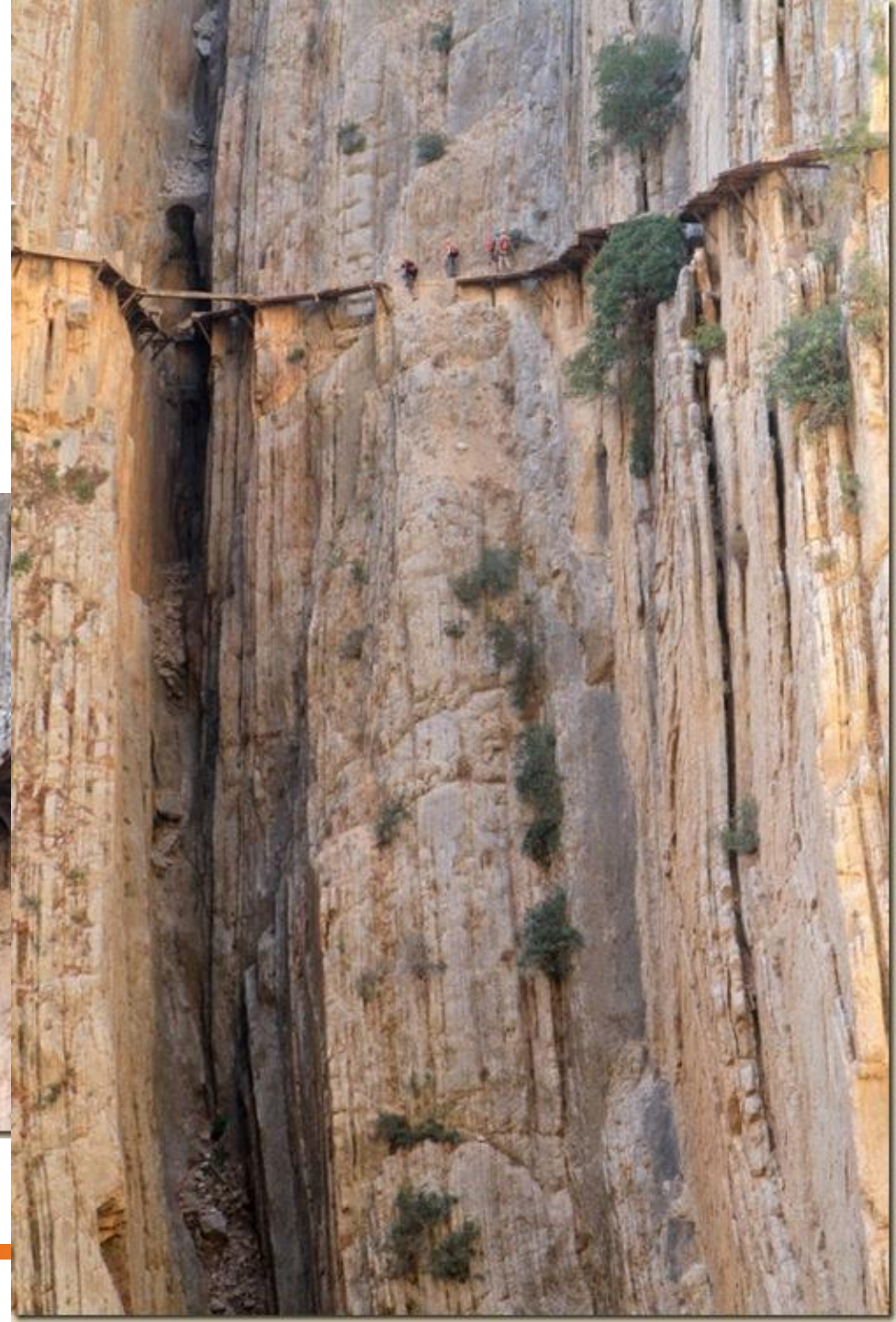
Disease biology	• Incidence • Prevalence • Tumor burden • Tumor evolution
Diagnosis	• Screening utility
Prognosis	• Biomarker significance • Mortality
Tx outcome	• Response rate • Overall survival • Progression-free survival
Utilization	• Treatment utilization • Adherence to guidelines • Uptake of new treatments • Utilization of new tests

In populations with..

- Stage
- Grade
- Dimension of tumor
- Extension of tumor
- Tumor margin
- Remission, stable or progressive disease
- Regimen
- Lines of therapy
- Diagnostic biomarker
- Prognostic biomarker
- Predictive biomarker

.. with speed, at scale

Don’t create your own data model



Building the Future of Observational Cancer Research Together

Open Research Network at inception

Open Research Network at scale