Oncology Research Spotlight: Advancing Cancer Care in OHDSI

OHDSI Community Meeting

11 August 2026

Disclosure

Asieh Golozar

OHDSI Oncology WG

FALCON

Oncology Research Network

nemesis

Let’s take a real-life example

Breast Cancer in Males

Men also get breast cancer

It is rare

- **0.5-1%** of all breast cancer
- **2670** new cases vs compared to **321,910** cases in women

It is deadlier

- 510 deaths in men compared to >42,000 deaths in women
- 5-year overall survival rate for men is ~**77.6%**, compared to **86.4%** for women.

Trials left men out, and guidelines extrapolate

- 79.4% of pivotal phase III breast cancer trials enrolled women only and 0.087% of all participants were men
- Male inclusion was 2.5% in endocrine therapy trials vs 28.6% in non-endocrine trials
- Treatment guidelines rest on extrapolation from evidence in women

ASCO®

"There are substantial knowledge gaps concerning the optimal management of breast cancer in men. To date, approaches to treating men with breast cancer have been extrapolated largely from research conducted in women with breast cancer."

NCCN

"Few males have been included in breast cancer trials. Therefore, recommendations regarding management of breast cancer in males are generally extrapolated from findings of clinical trials focusing on breast cancer in females."

ESMO

"Knowledge of this disease is limited and mainly derives from small single-institution retrospective studies with contradictory results. The management of male patients with BC is generalised from the management of BC in women."

The default treatment comes at a cost

- Tamoxifen is the default endocrine therapy for men.
- ~50% of men discontinue tamoxifen within 5 years.
- ~1 in 5 discontinue because of toxicity.
- Thromboembolism is more frequent in men receiving tamoxifen.
- Alternatives exist, but evidence to guide treatment selection in men is limited.

Are we giving men the right treatment, or simply the treatment best studied in women?

SOURCES Corrigan et al., Oncologist 2020. American Cancer Society, Cancer Facts & Figures 2025. Hassett et al., J Clin Oncol 2020. Giordano, N Engl J Med 2018. Oke et al., Cancer 2022. Pemmaraju et al., Ann Oncol 2012. Eggemann et al., Br J Cancer 2019. US FDA, August 2019.

How can we help in this case?

- Guidelines
- Treatment selection
- Effectiveness and safety
- Access and reimbursement
- Regulatory decisions
- Clinical trial design
- Disease characterization and unmet need

What‘s going on in Cancer RWE?

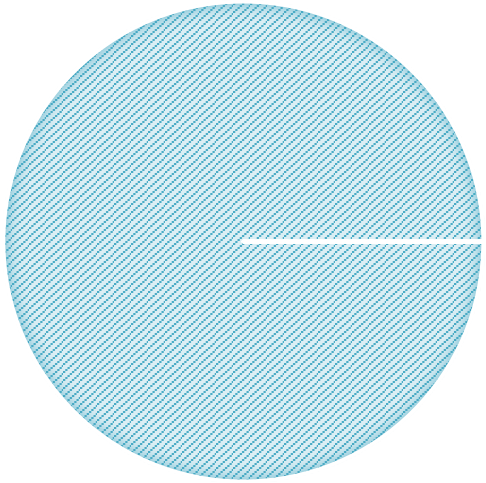
Conclusions: The number of RWE publications on TT for solid tumours is increasing, but studies are heterogeneous, mostly retrospective, and published in low IF journals. International collaboration and promotion of standardised data sources is imperative to enhance the relevance of RWE research to complement clinical guidelines and impact clinical practice.

What is to blame for this?

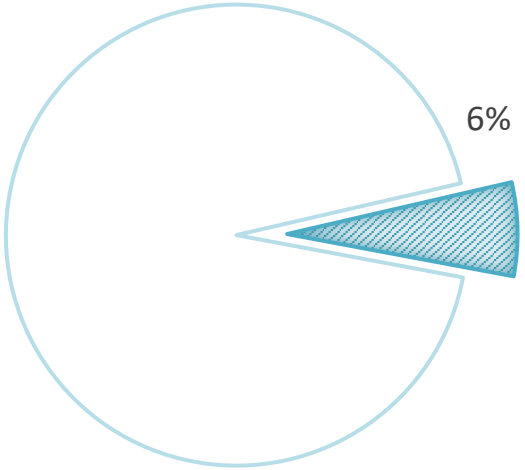
What is to blame?

The Cancer

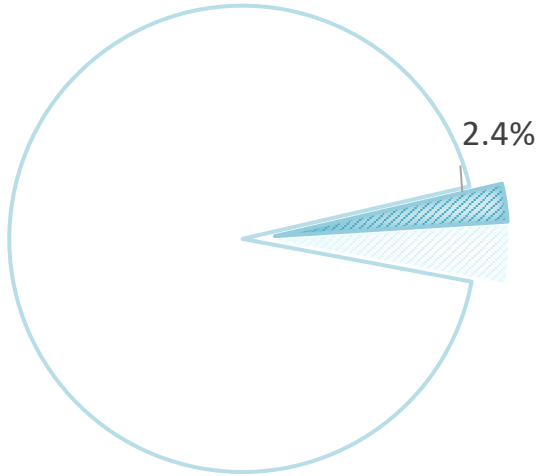
A) it‘s a rare disease



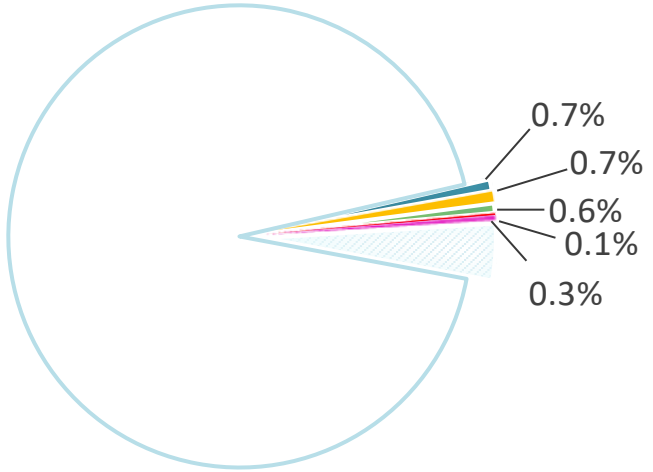
~600,000 bladder cancers



Metastatic



Treated



Regimens

What is to blame?

The Cancer

B) it needs a lot of detail

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.

What constitutes a cancer diagnosis?

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.

So: Cancers Diagnoses have **Detail**

with metastatic non-small cell lung cancer (NSCLC) whose tumors have a mutation that leads to mesenchymal-epithelial transition (MET) exon 14

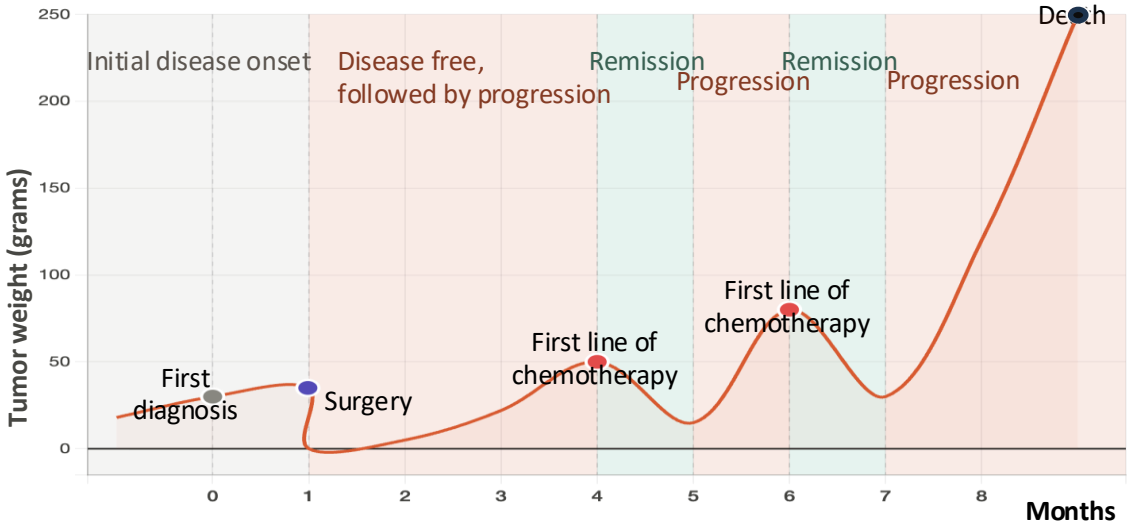
- Metastases
- Morphologies
- Topographies
- Biomarkers
- Stages
- Grades
- Invasion/Extensions
- **Disease dynamics**

Metastatic

non-small cell

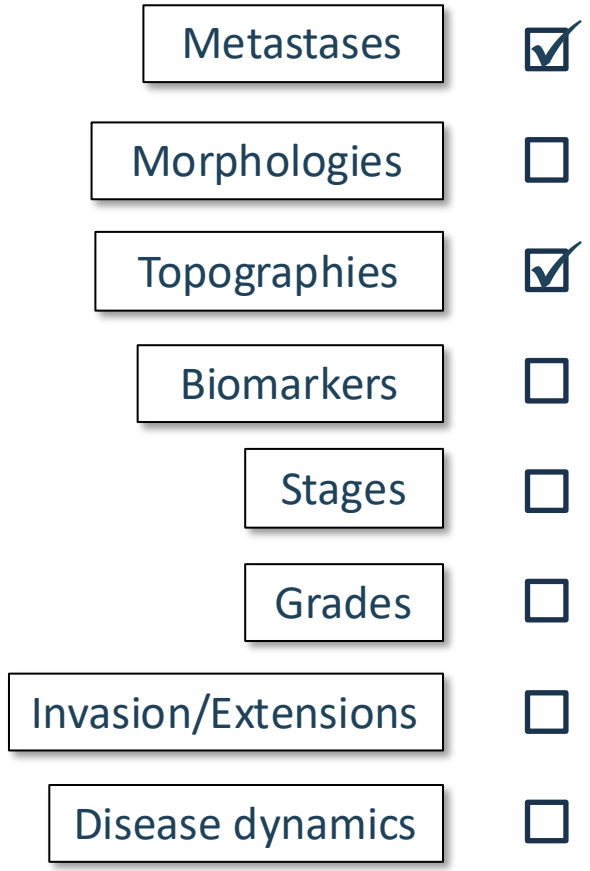
lung cancer

MET ex14 skipping



Can we Represent all the Detail?

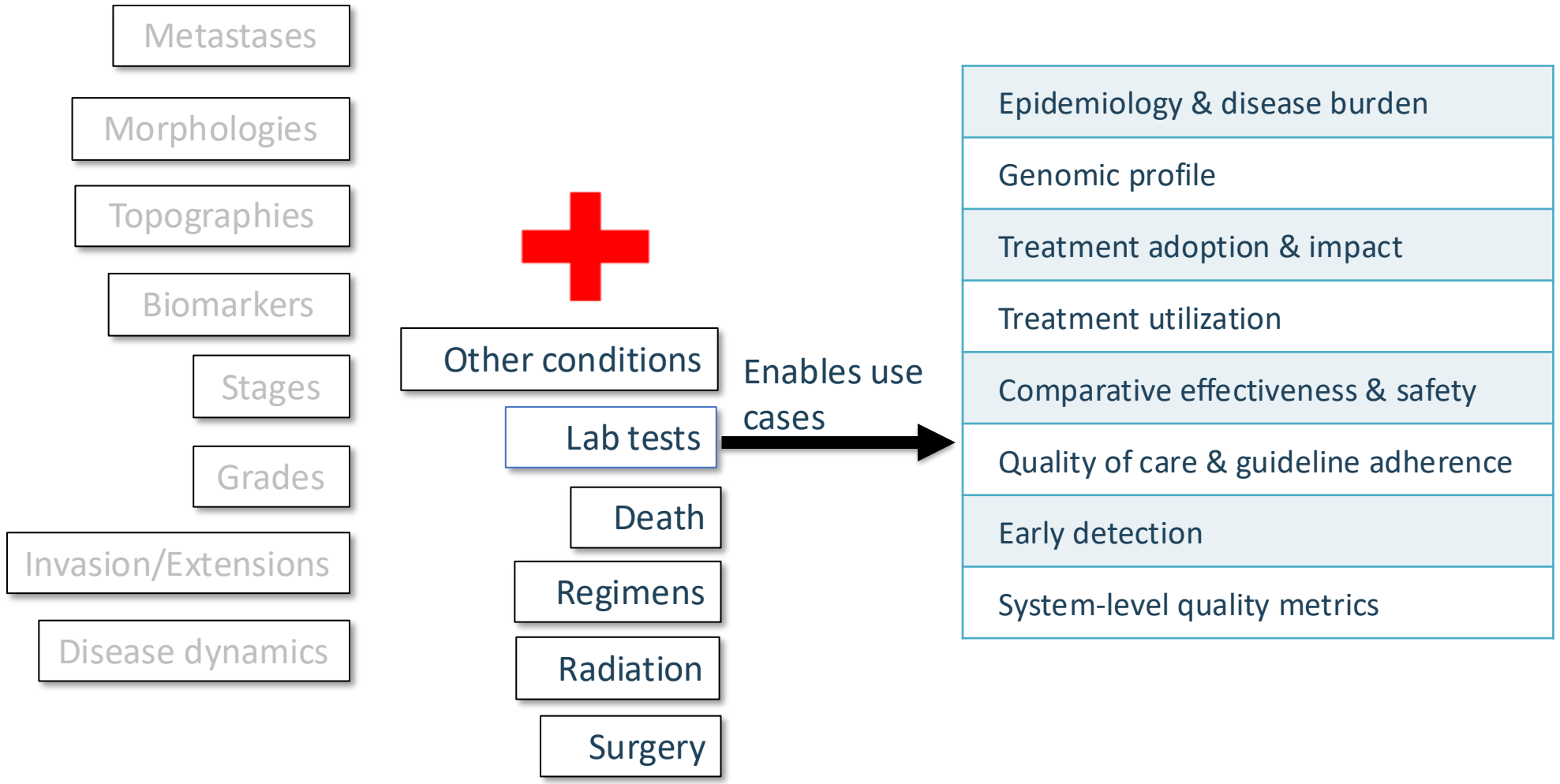
Usually, diseases get coded in ICD10.



ICD10 only has 55 mets and a decent set of cancer topographies

ICD10CM (not ICD10) has 206 cancers "in remission" or "not in remission" (no progression)

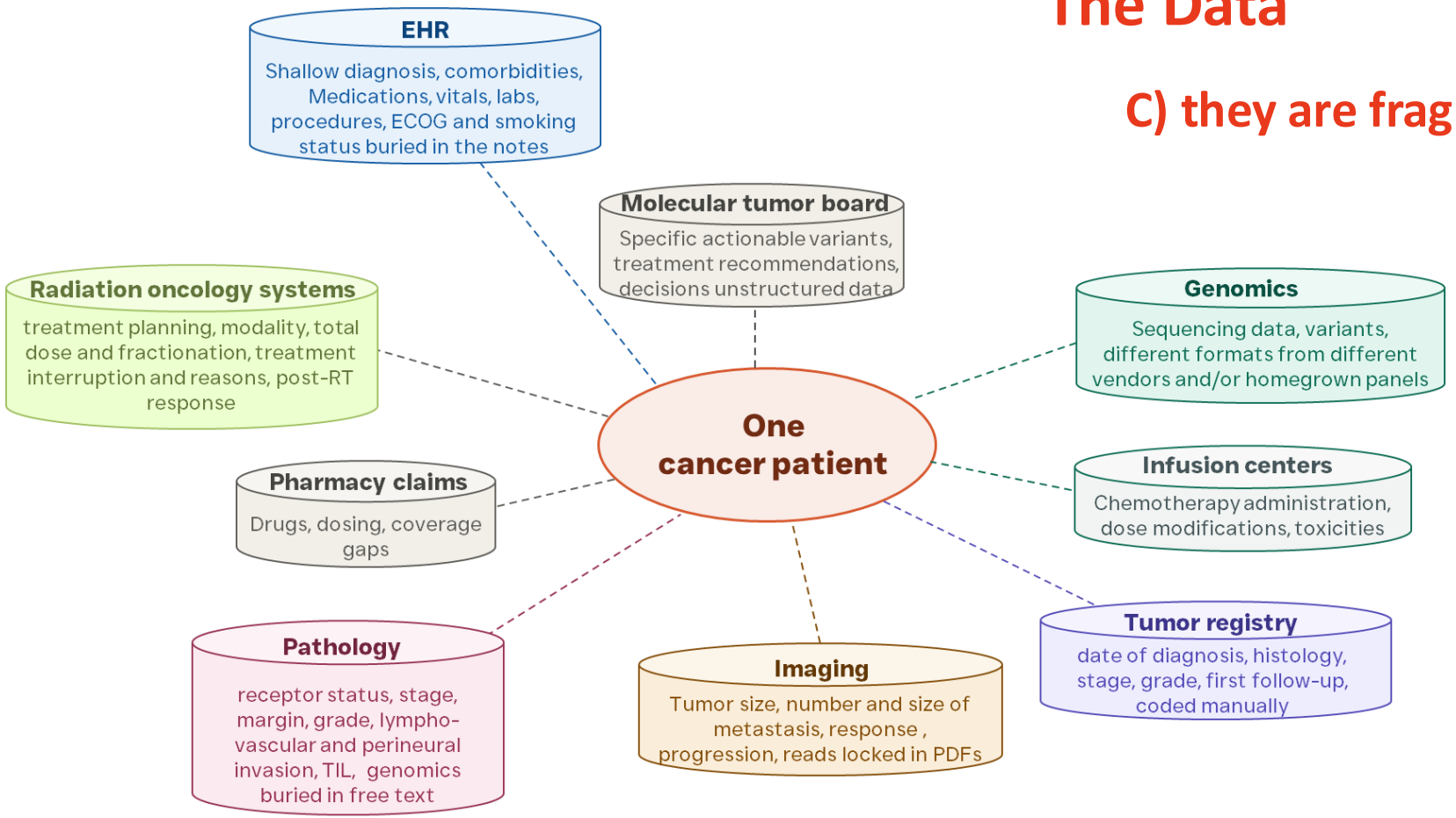
There is more detail beyond Condition



What is to blame?

The Data

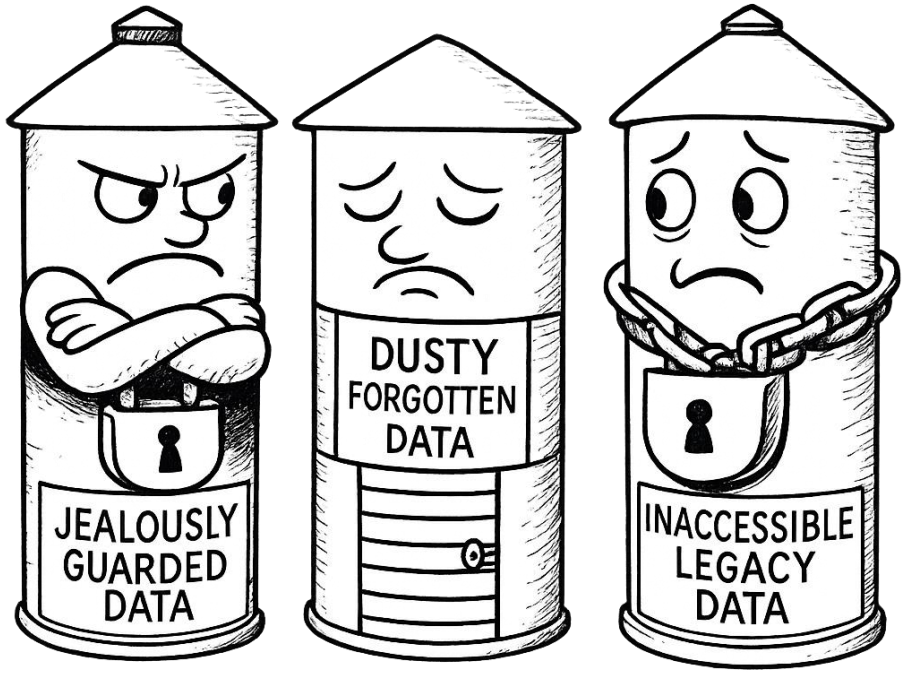
C) they are fragmented – over different systems



What is to blame?

The Data

C) they are fragmented – and not easy to bring together



What is to blame?

The Data

D) their quality is bad

ESMO daily REPORTER

An ESMO study reveals that poor quality of real-world data from oncology studies is still an issue

Journal of Clinical Oncology
An American Society of Clinical Oncology Journal

"Poor-quality of real-world data in oncology limits the reliability and impact of oncology RWE."

We need to solve these problems

The Cancer

A) it‘s a rare disease
B) it needs a lot of detail

The Data

C) they are fragmented
D) their quality is bad

The Solution Space

The Cancer

A Build a network of cancer centers

Move from fragmented single-center studies to federated, multi-site evidence generation at scale

B Define what research needs

Identify and standardize oncology related details required for reliable evidence

The Data

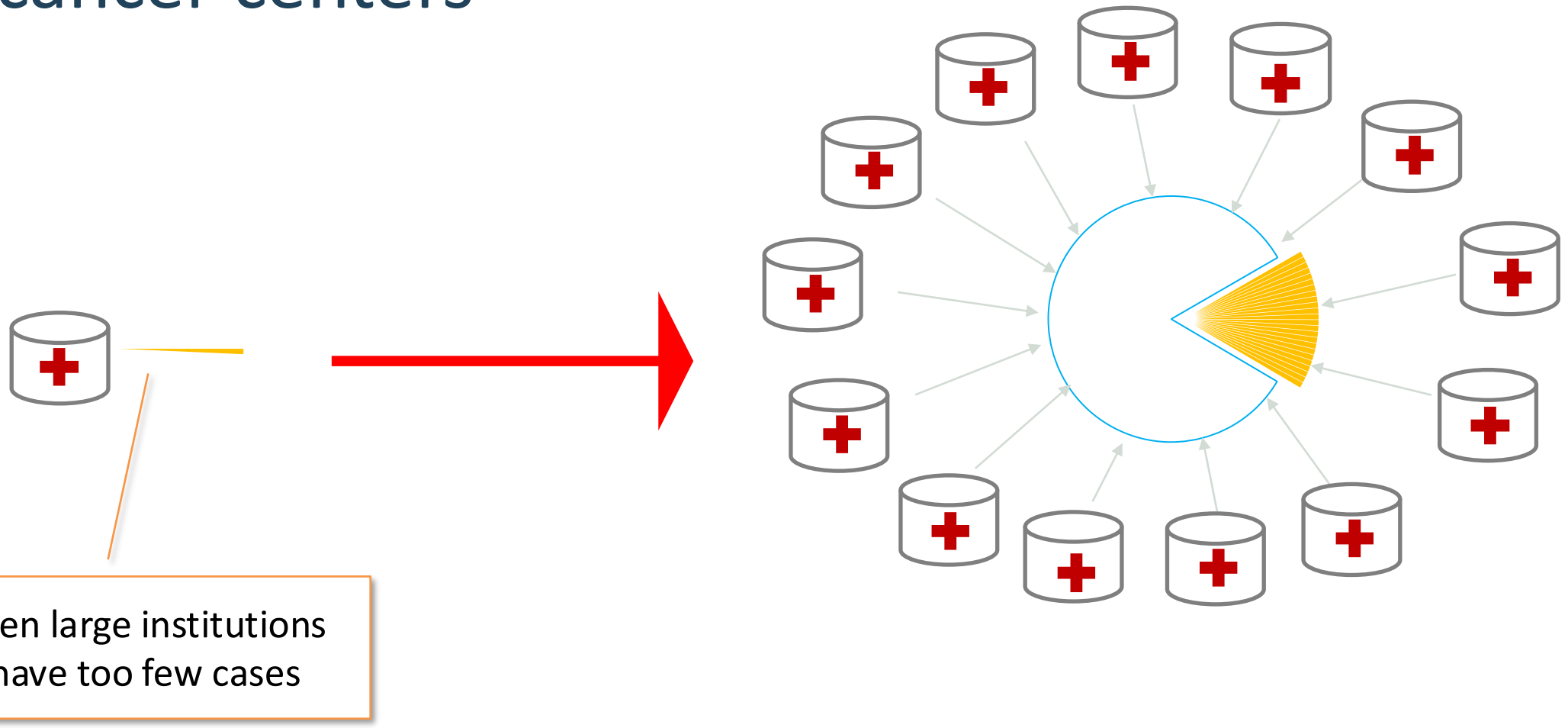
C Connect the silos

Aggregate patient data from the silos inside an organization and take care of most of the fragmentation

D Fix the quality

Systematically assess, clean, and validate data through structured diagnostics

A network of cancer centers

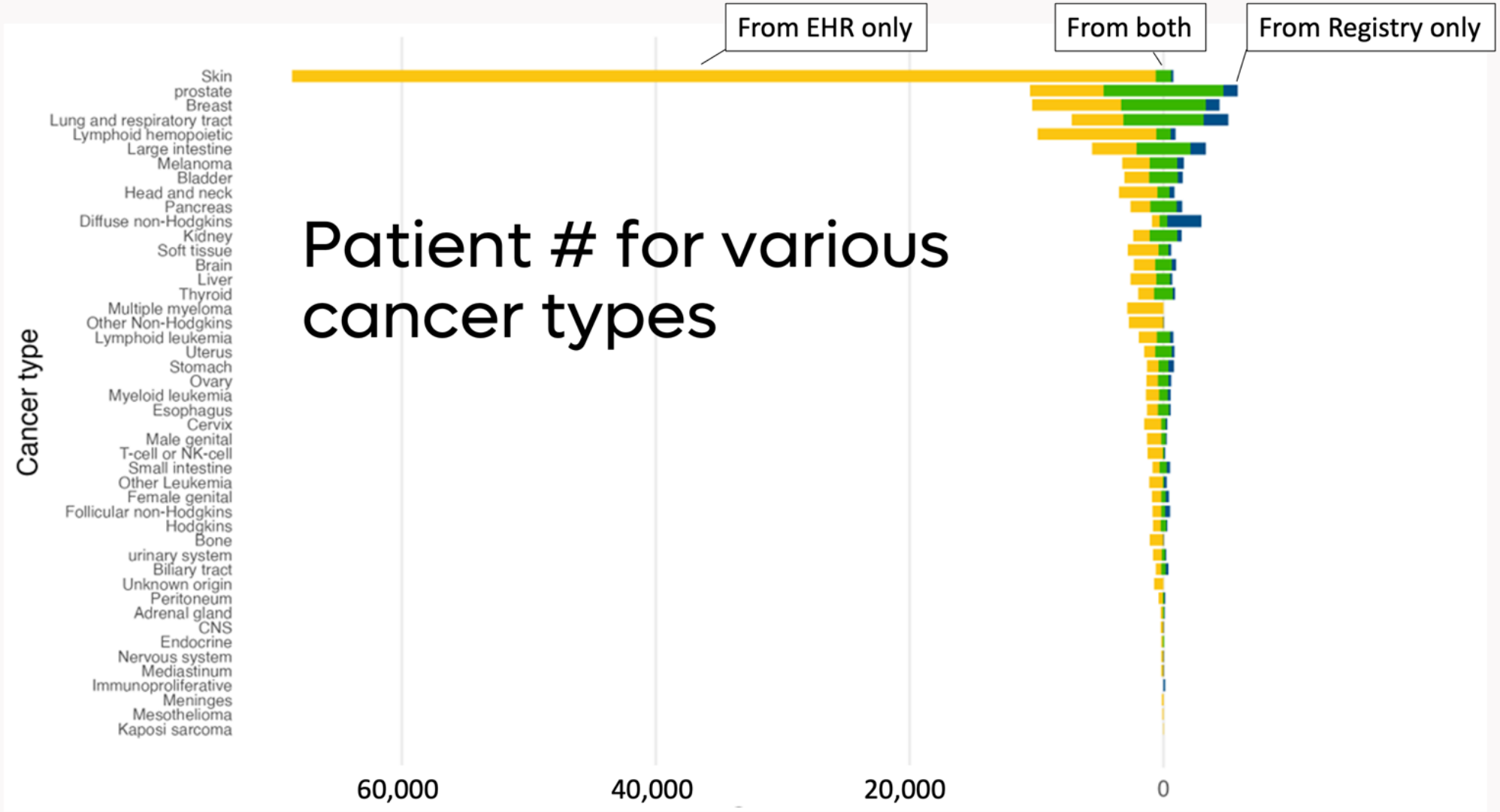


The
FALCON Research Network

- The first large-scale federated oncology RWE network

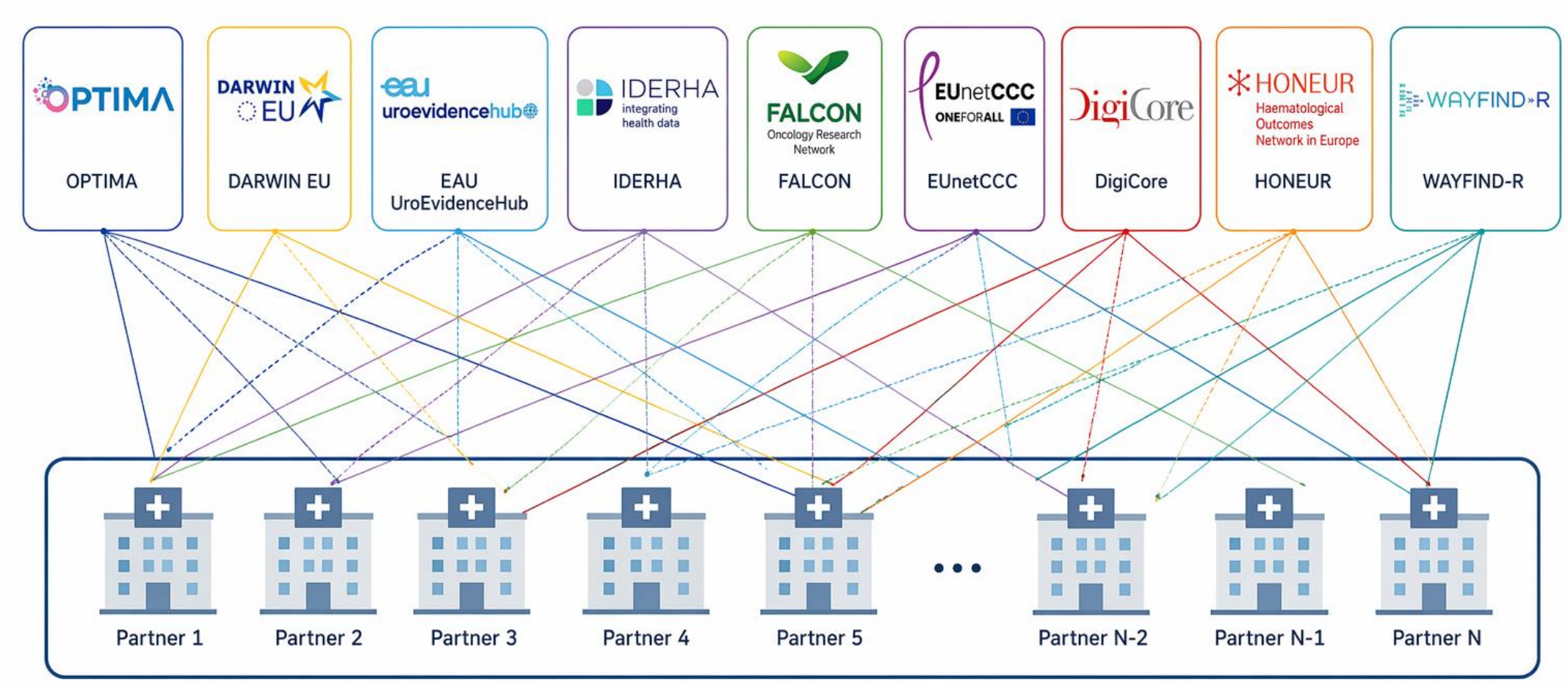
- +50 data partners
- 19 countries
- 194,904,822 lives

FALCON aggregates Fragmented Data

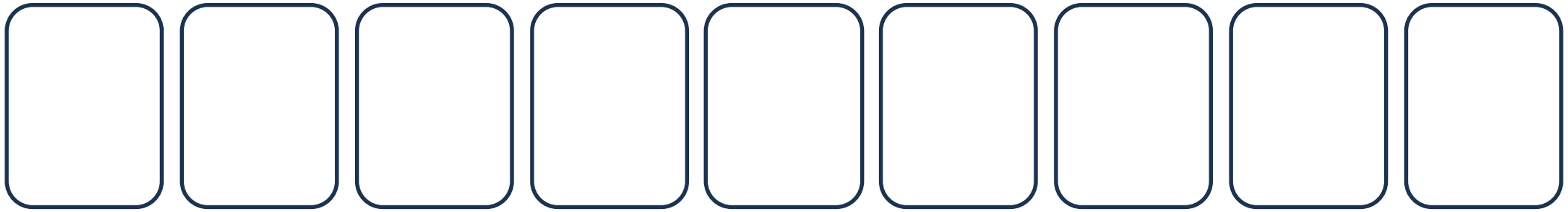


Oncology Networks in OHDSI

9 oncology networks, similar data partners



Different Coordinating Centers



Different

- Studies
- Sponsorship
- Governance
- Logos

In common

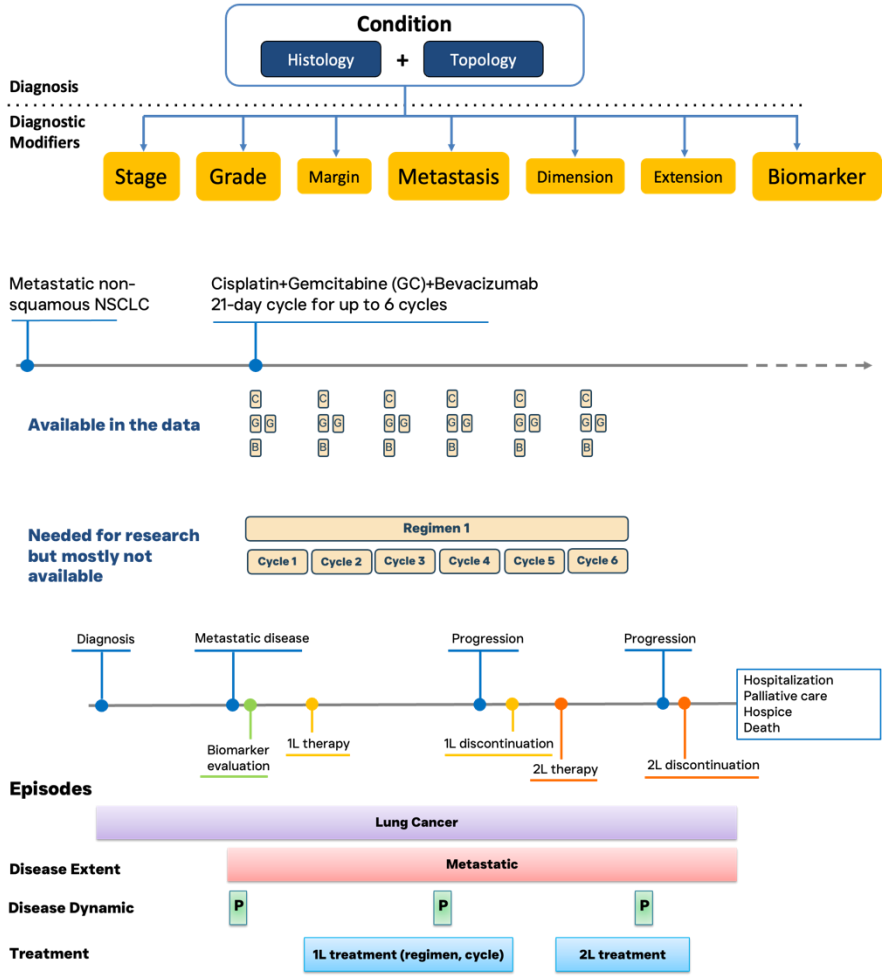
- OMOP CDM
- Vocabularies
- Conventions
- Quality

Data harmonization for cancer

1 **Cancer Disease Model**

2 **Cancer Treatment Model**

3 **Cancer Episode Model**



Cancer terminologies

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

Upcoming vocabulary release

- Briefly:
  - NAACCR mapping
  - SNOMED, LOINC de-standardization and mapping
  - Small genomic update
- Now in good shape:
  - Mets, lymph nodes
  - Stages, grades
- More details during the Sep 1 Community call- Vocab refresh

Documentation

- ETL OnRamp: https://github.com/OHDSI/OncologyWG/wiki/OnRamp
- Book of OHDSI
  - Oncology
  - Genomic

What about the other problems facing cancer RWE?

Start with the Use Cases

Use case categories
Epidemiology & disease burden
Genomic profile
Treatment adoption & impact
Treatment utilization
Comparative effectiveness & safety
Quality of care & guideline adherence
Early detection
System-level quality metrics

Full list of oncology use cases:
https://github.com/OHDSI/OncologyWG/wiki/Use-Case-Repository

.. And drive the Variables

Use case categories
Epidemiology & disease burden
Genomic profile
Treatment adoption & impact
Treatment utilization
Comparative effectiveness & safety
Quality of care & guideline adherence
Early detection
System-level quality metrics

Clinical data	Frequency of being required
Base diagnosis	0.93
Other conditions	0.80
Stage	0.66
Metastases	0.57
Lab tests	0.56
Death	0.56
Regimens	0.46
Disease dynamic	0.39
-Omics	0.38
Radiation Therapy	0.16
Grade	0.13
Extent	0.11
Disease Episode	0.1
Surgery	0.08
Lymph nodes	0.02

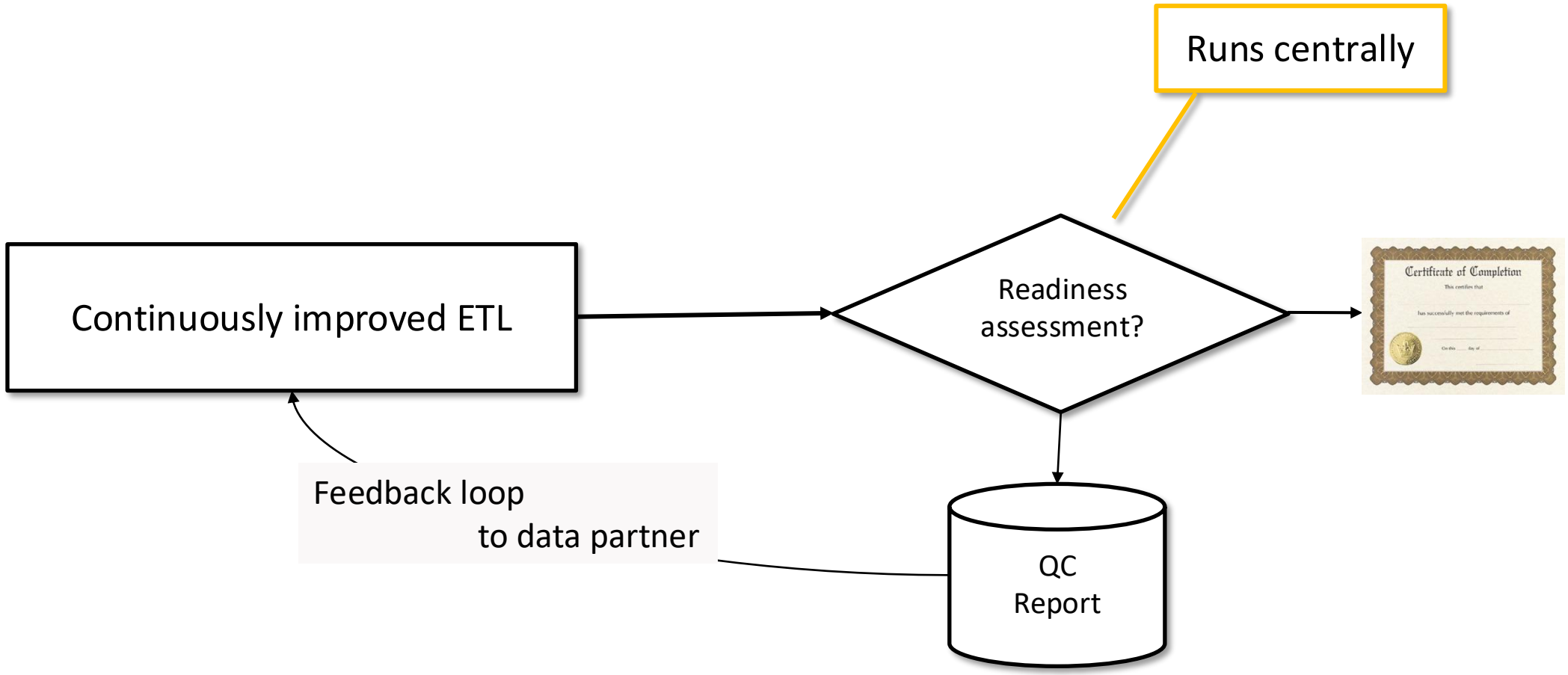
.. And drive the Variables

Use case categories
Epidemiology & disease burden
Genomic profile
Treatment adoption & impact
Treatment utilization
Comparative effectiveness & safety
Quality of care & guideline adherence
Early detection
System-level quality metrics

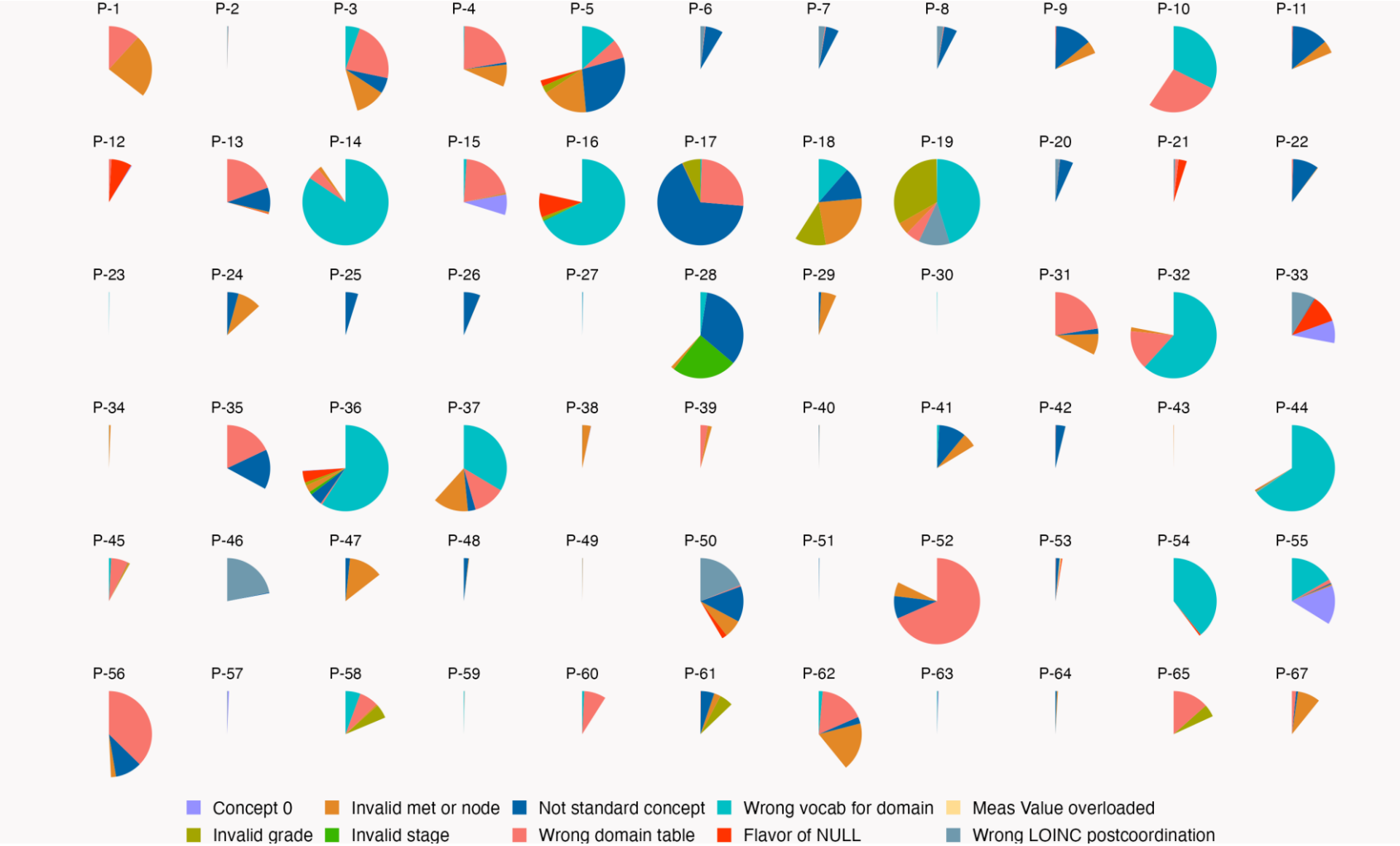
Variables other than cancer diagnoses

Clinical data	Frequency of being required
Base diagnosis	0.93
Other conditions	0.80
Stage	0.66
Metastases	0.57
Lab tests	0.56
Death	0.56
Regimens	0.46
Disease dynamic	0.39
-Omics	0.38
Radiation Therapy	0.16
Grade	0.13
Extent	0.11
Disease Episode	0.1
Surgery	0.08
Lymph nodes	0.02

Quality Management System



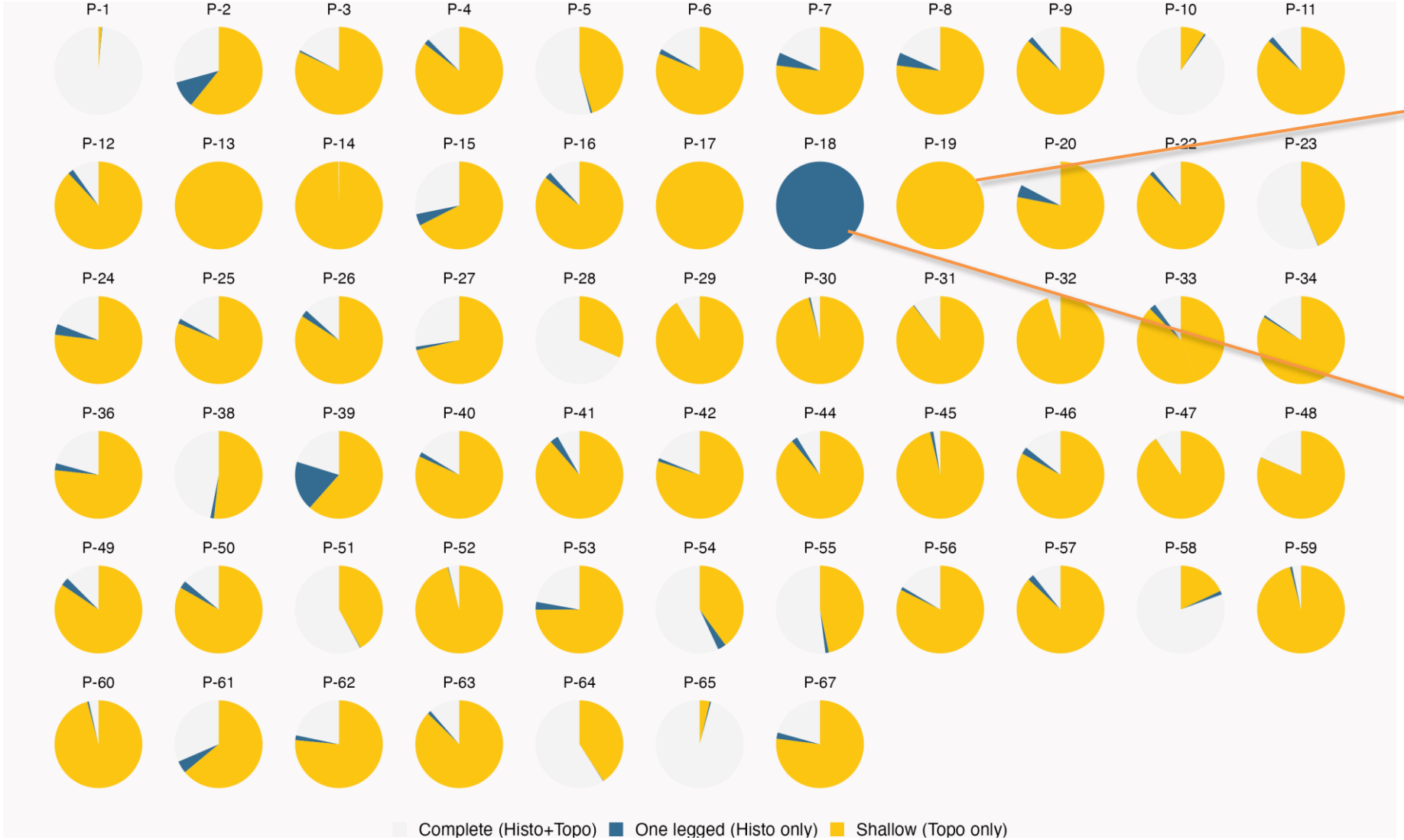
Problems with Standard Concepts



Data partners picking their own coding **is not ok!**

We cannot fix that through custom cohort definitions.

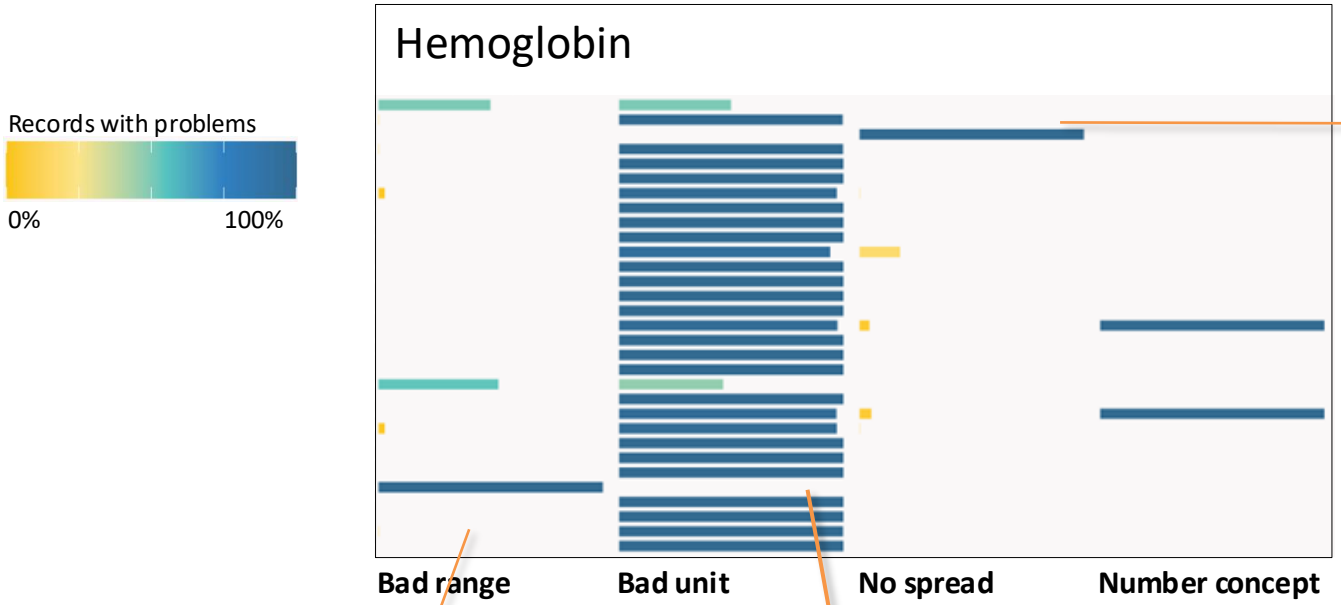
Problems with Cancer Conditions



ICD-10 derived cancer concepts are insufficient

If the path lab doesn't combine histologies with topographies, we must

Problems with Labs



Very few data or test not working

Non-numerical values, usually qualitative

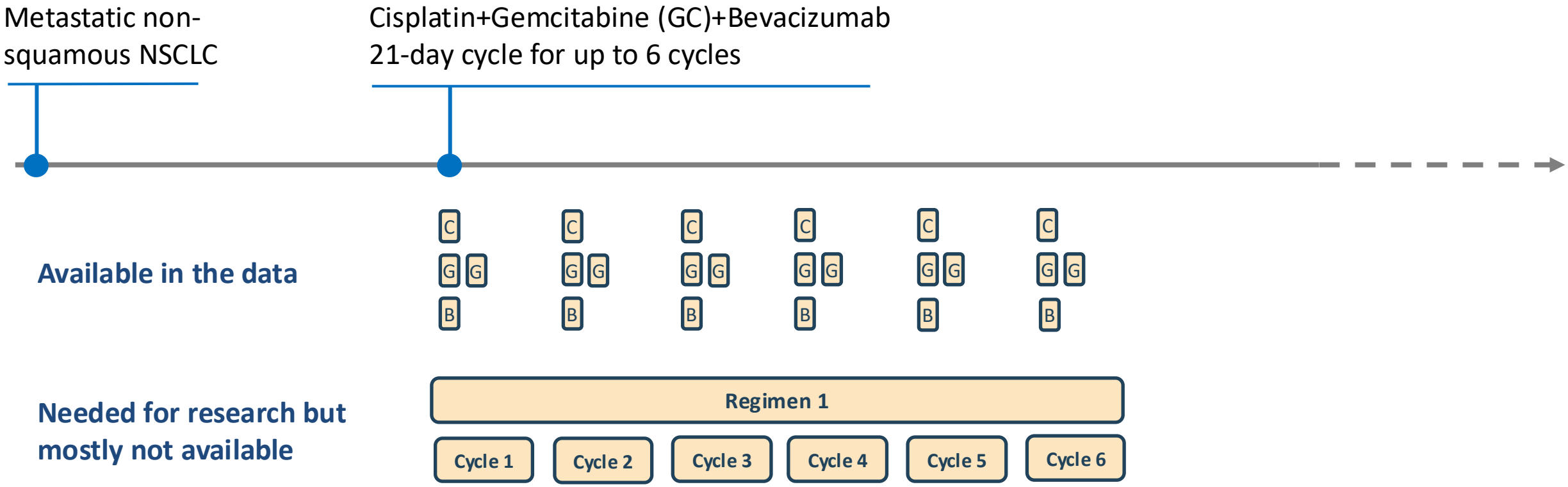
Ranges not expected for the test

Unit are notoriously unreliable, leaving us to rely on ranges

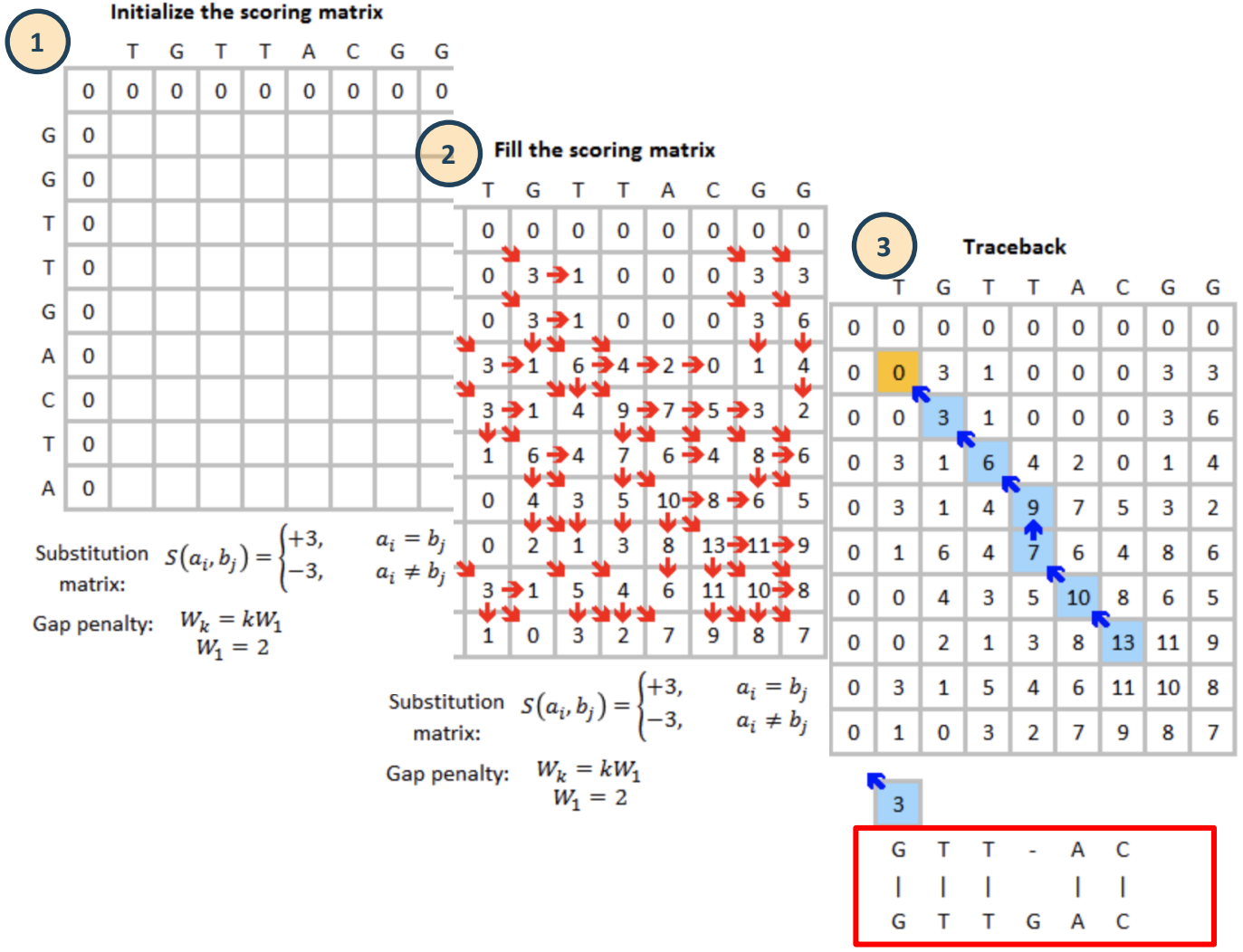
Covered lab tests

ALT
ANC
aPTT
AST
CrCl
Creatinine
Direct bilirubin
GFR
Hemoglobin
HbA1c
INR
Platelets
PTT
Total bilirubin

OMOP Oncology: Cancer Treatment Model



How does ARTEMIS do it?



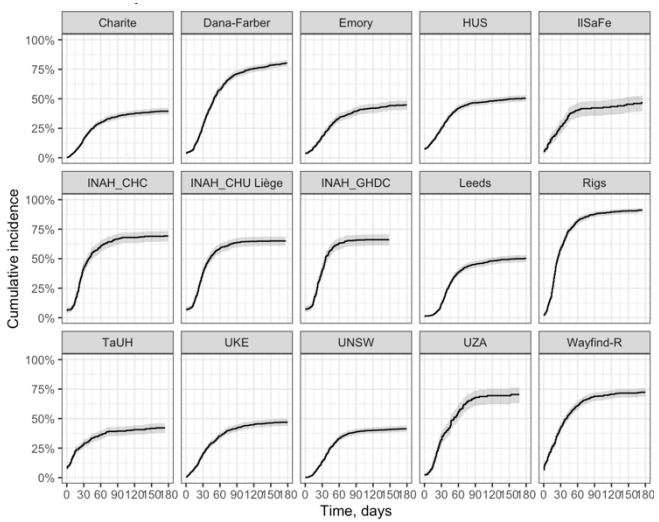
- ARTEMIS uses an adapted Smith–Waterman sequence-alignment approach to derive chemotherapy regimens.
- Standard sequence alignment accounts for matches, mismatches, and gaps, but not the relative timing of treatment events.
- ARTEMIS extends the alignment scoring to incorporate temporal differences, allowing treatment delays and deviations from expected schedules to be considered.

Ongoing research

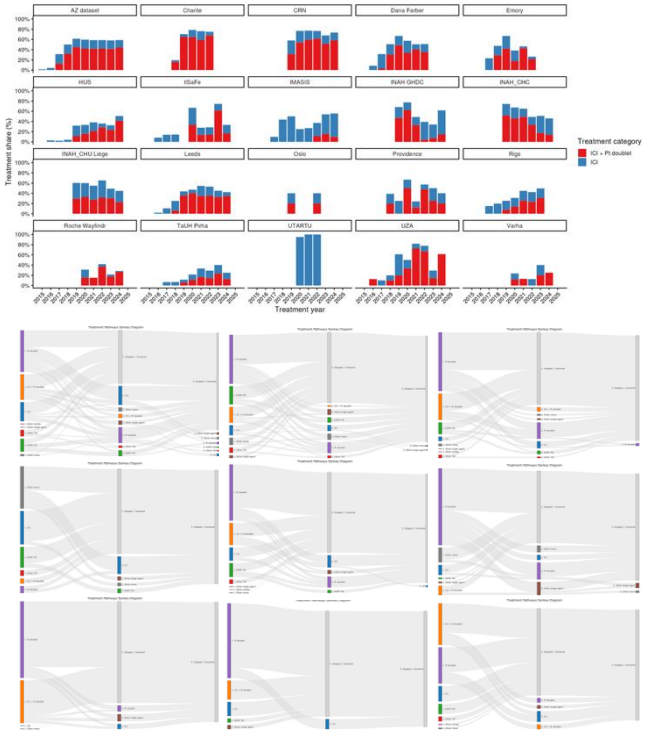
FALCON-Lung

23 data partners from 9 countries with >100,000 NSCLC

Time to systemic treatment initiation

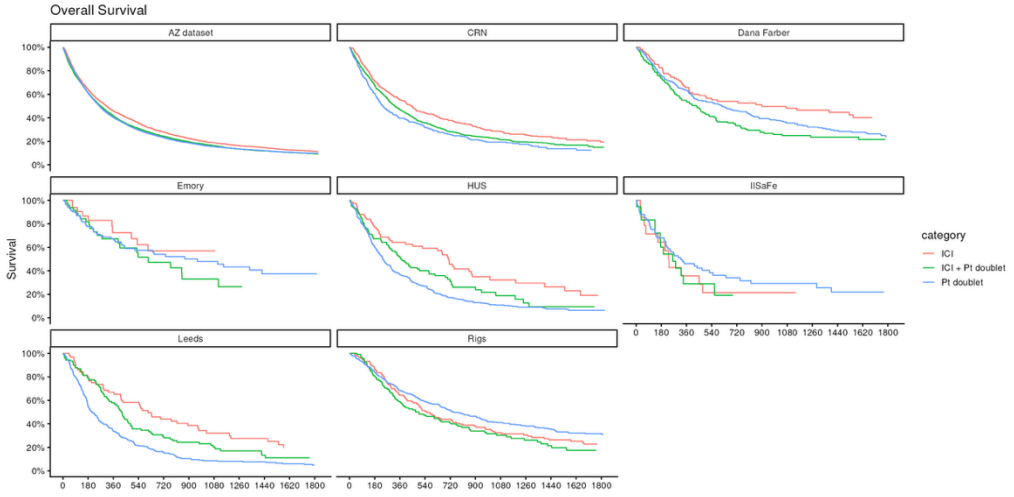


Treatment share and pattern since 2015



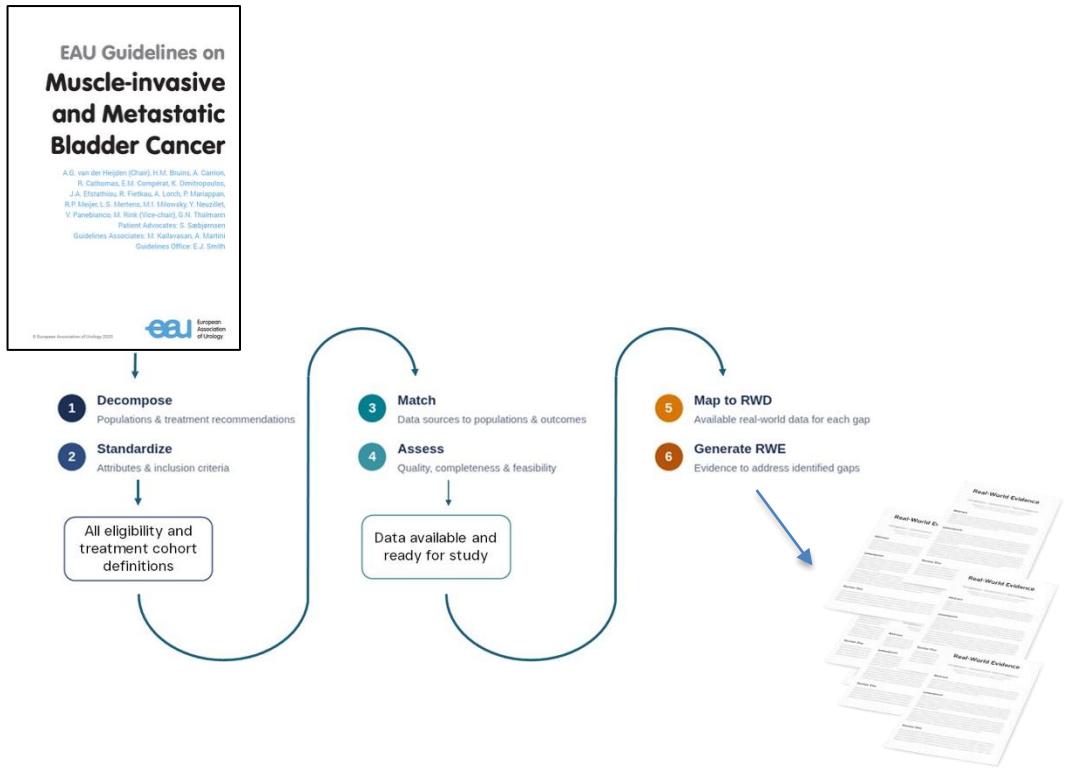
Explore results through an interactive dashboard. Stratify by regimen, treatment line, site, age and sex.
https://oncology.ohdsi.org/hus-nsclc/

Overall survival by 1st line treatment categories

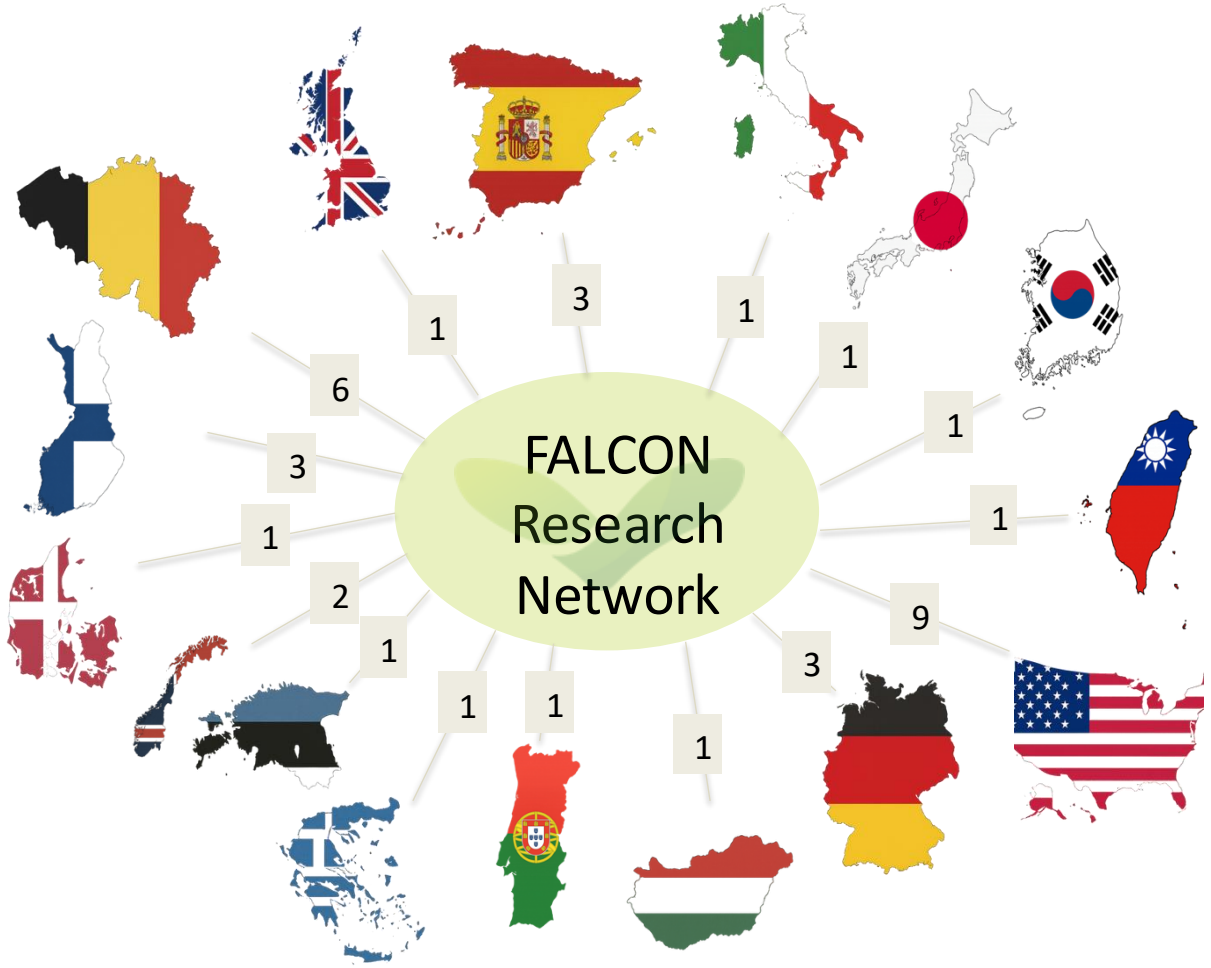


1867P FALCON-Lung: Evolving global ICI treatment patterns and outcomes in the era of personalized mNSCLC treatment, Verbiest, A.T.L. et al. Annals of Oncology, Volume 36, S1075

FALCON Bladder:
RWE to influence Bladder Cancer Guideline



Guideline-relevant RWE



16 countries, 40 sites

Studyathon: Sep 29 – Oct 1, 2026
Antwerp, Belgium
www.falcon-network.org

FALCON-Male Breast

- Landscape study: Characterize patients, treatments, and available data across FALCON.
- Evidence opportunities: Prioritize the most important and answerable clinical questions.
- Study requirements: Define required data and identify ready sites.
- Build what is missing: Address key data and quality gaps with partners.
- Evidence at scale: Launch prioritized studies
- Build the program: Create a collaborative, long-term male breast cancer research program with interested organizations and partners.

- **Kick-off September 2026**

Summary

- Cancer RWE is an imperative
- The Oncology WG is laying the foundation
- Research started
- More research is to come:
  - Prostate cancer
  - Multiple myeloma
- Bring your needs and research to the WG

golozar@ohdsi.org,
oncology@ohdsi.org