OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Ensuring Data Fitness for Oncology Research

Asieh Golozar

25-Mar-2025

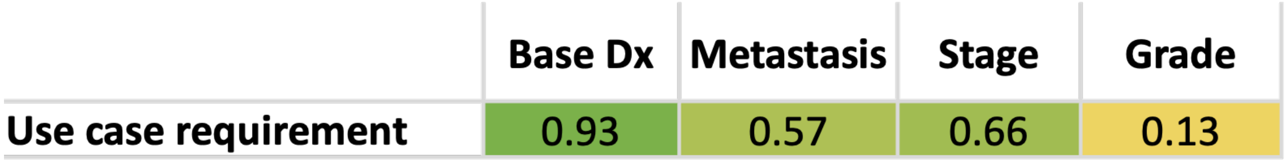
Overview

- Enabling data network with data fit for oncology studies
- HUS Studyathon
- Guidelinathon

What does it mean “ready for oncology studies”?

	Base Dx	Metastasis	Stage	Grade	Lymph nodes	Others (specify)	-Omics	Regimens	Radiation	Surgery	Extent	Dynamic	Episode of care	Death
Use case requirement	0.93	0.57	0.66	0.13	0	0	0.38	0.46	0.16	0.08	0.11	0.39	0.1	0.56
Vocab readiness	1	1	1	1	0.5	0.5	1	1	0.3	0.5	0.9	0.9	1	1
Model readiness	1	1	1	1	1	1	1	1	0.1	1	1	1	1	1
Available data/algorithm	0.77	0.65	0.79	0.69	0.48	0.58	0.40	0.69	0.50	0.62	0.46	0.35	0.31	0.69
Data Partners with data	20	17	20.5	18	12.5	15	10.5	18	13	16	12	9	8	18

Oncology Data Readiness- Approach



	Base Dx	Metastasis	Stage	Grade
Use case requirement	0.93	0.57	0.66	0.13

How do we get to that?

1. Query
2. Assess
3. Patch or fix
4. Iterate 1-3.

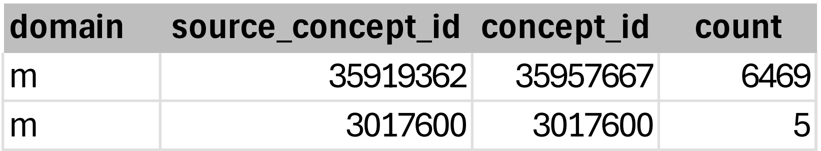
Self-service on https://oncology.ohdsi.org/

Query

Recap: Data Query

- Queries:
  - general.sql: for general cancer concepts: diagnoses, treatments, other mgt, 284,958 concepts
  - genomic.sql: for genomic concepts: small (usually SNPs), large (e.g. fusion proteins), DNA, RNA, protein level, 593,220 concepts
  - episode.sql: for disease (progression, remission) and treatment (regimen) episodes, 8,052 concepts

- Output:
  - All source-standard concept pairs, their domains, and their total counts
  - No patient related information



domain	source_concept_id	concept_id	count
m	35919362	35957667	6469
m	3017600	3017600	5

Content of the Data Query

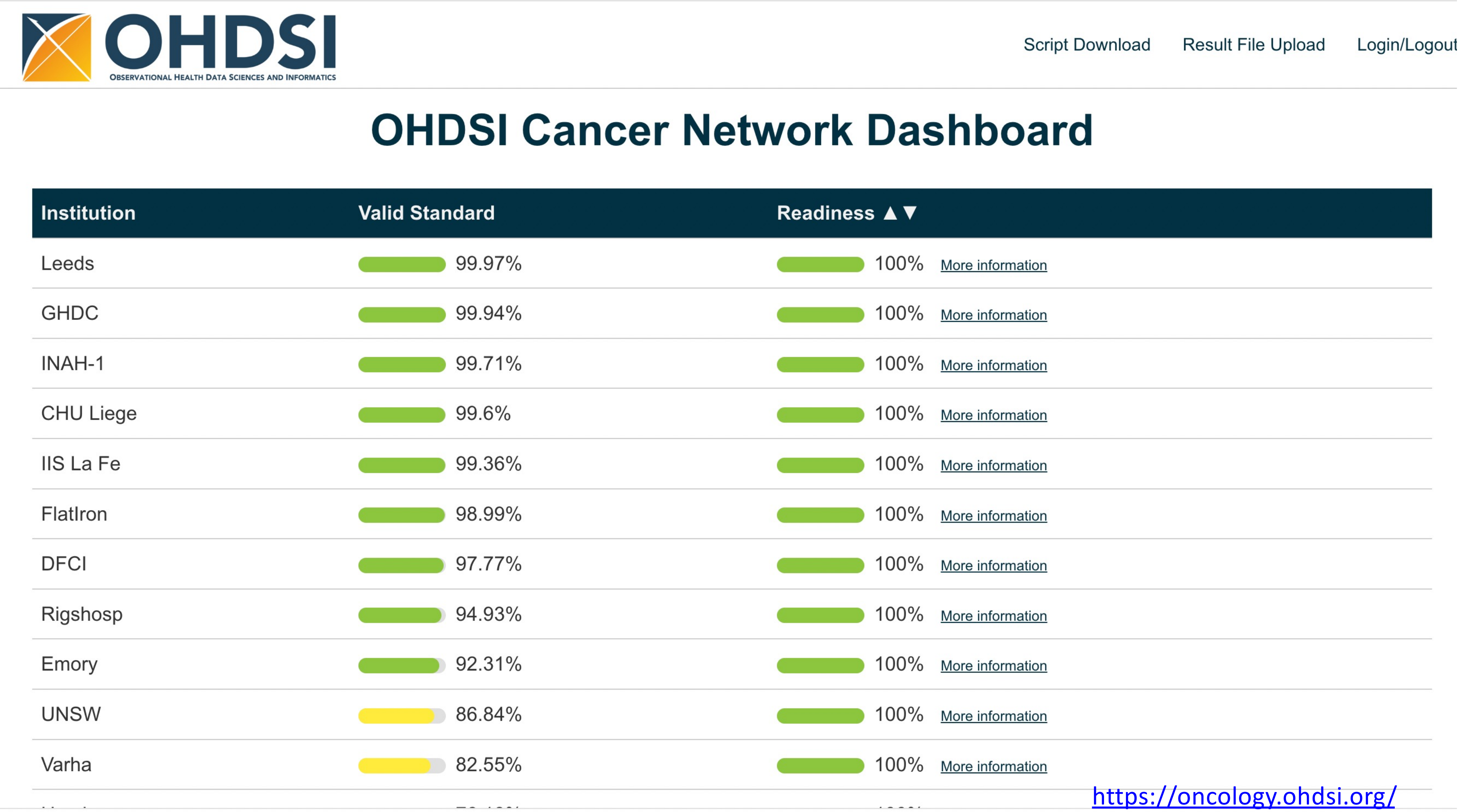
```
select 'd' as domain, drug_source_concept_id as source, drug_concept_id as standard, count(*) as cnt
from (
  select drug_exposure_id
  from drug_exposure
  join concepts on concept_id=drug_source_concept_id
union
  select drug_exposure_id
  from drug_exposure
  join concepts on concept_id=drug_concept_id
) a
join drug_exposure using(drug_exposure_id)
group by drug_source_concept_id, drug_concept_id

select 'e' as domain, device_source_concept_id, device_concept_id
select 'p' as domain, procedure_source_concept_id, procedure_concept_id
select 'c' as domain, condition_source_concept_id, condition_concept_id
select 'o' as domain, observation_source_concept_id, observation_concept_id
select 'm' as domain, measurement_source_concept_id, measurement_concept_id
select 'v' as domain, null, value_as_concept_id
select 'i' as domain, episode_source_concept_id, episode_concept_id
```

Records with hits in drug_source_concept_id

Records with hits in drug_concept_id

Long list of cancer/genomic/ episode concepts

Same query to the other tables



OHDSI Cancer Network Dashboard

Institution	Valid Standard	Readiness ▲▼	
Leeds	🟩 99.97%	🟩 100%	More information
GHDC	🟩 99.94%	🟩 100%	More information
INAH-1	🟩 99.71%	🟩 100%	More information
CHU Liege	🟩 99.6%	🟩 100%	More information
IIS La Fe	🟩 99.36%	🟩 100%	More information
FlatIron	🟩 98.99%	🟩 100%	More information
DFCI	🟩 97.77%	🟩 100%	More information
Rigshosp	🟩 94.93%	🟩 100%	More information
Emory	🟩 92.31%	🟩 100%	More information
UNSW	🟨 86.84%	🟩 100%	More information
Varha	🟨 82.55%	🟩 100%	More information

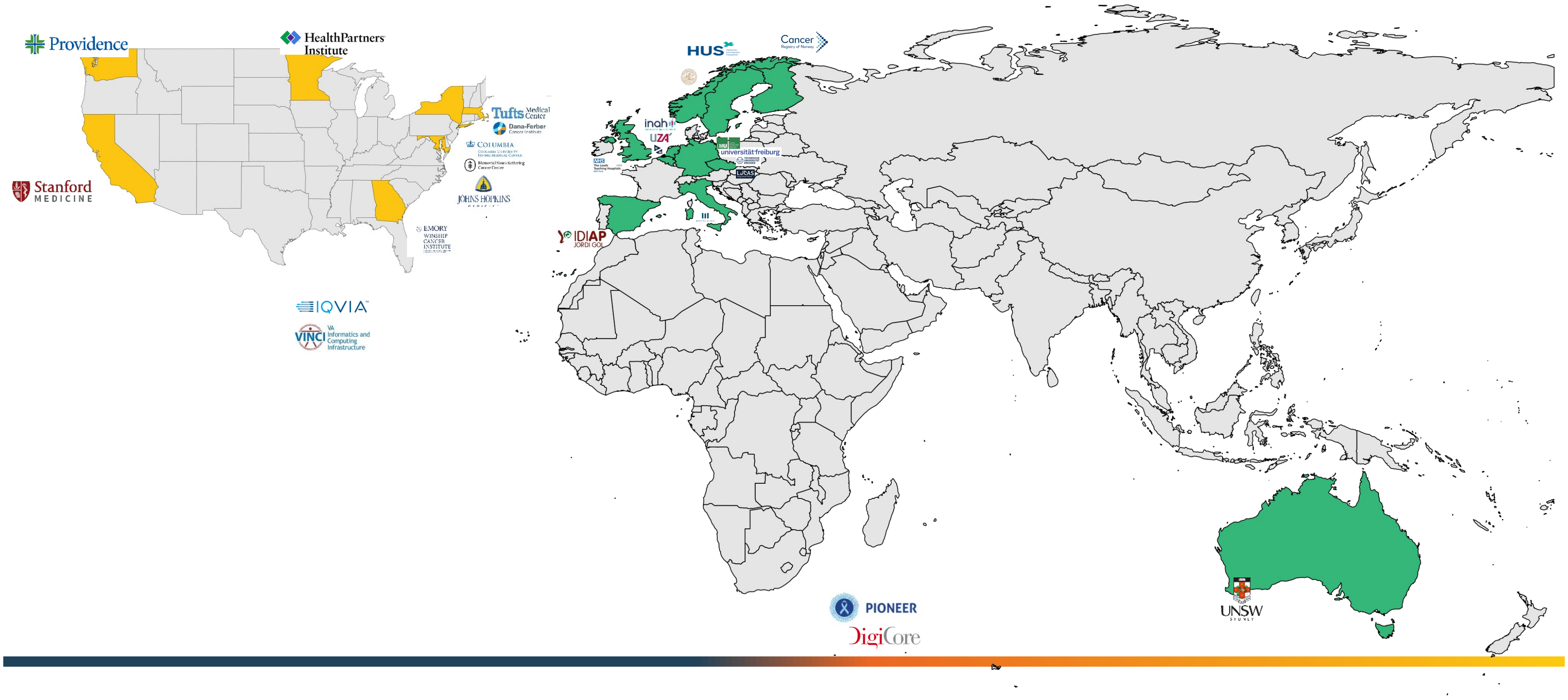
https://oncology.ohdsi.org/

Assess

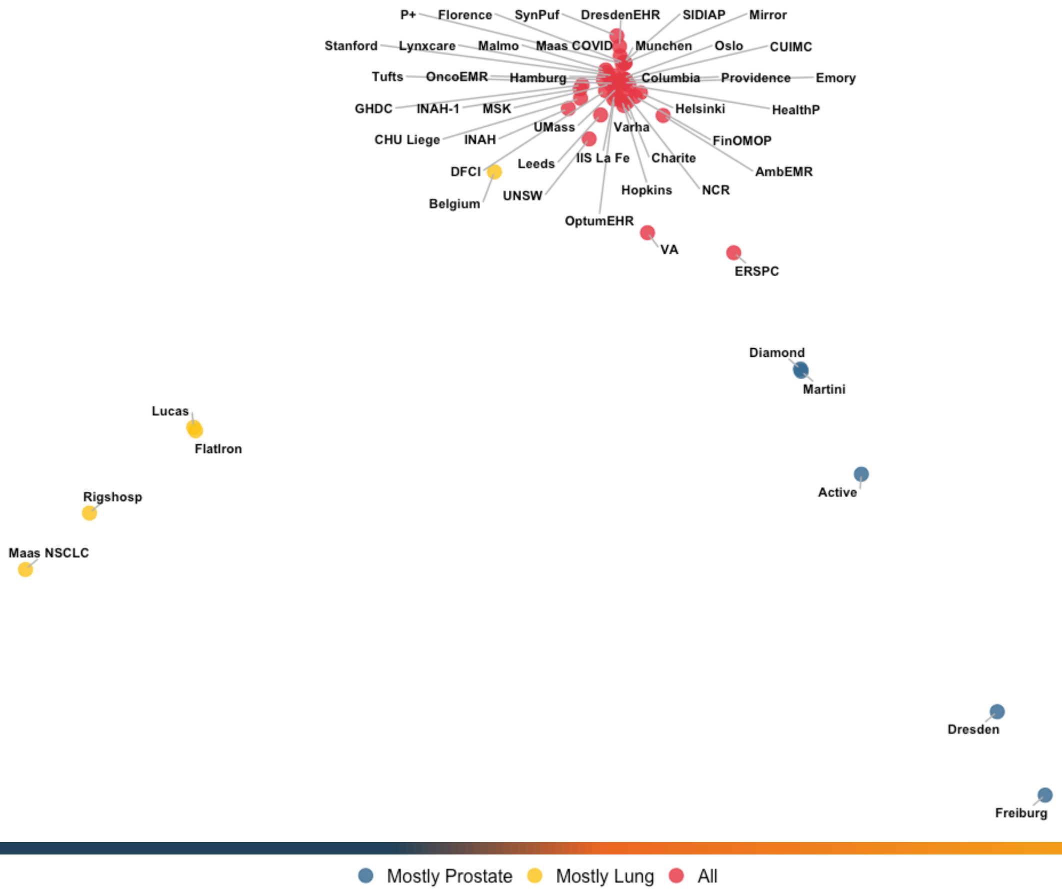
Returned Query Results

- 367,697 general records from 50 partners
- 3,872 genomic records from 26 partners
- 28,049 episodes records from 16 partners

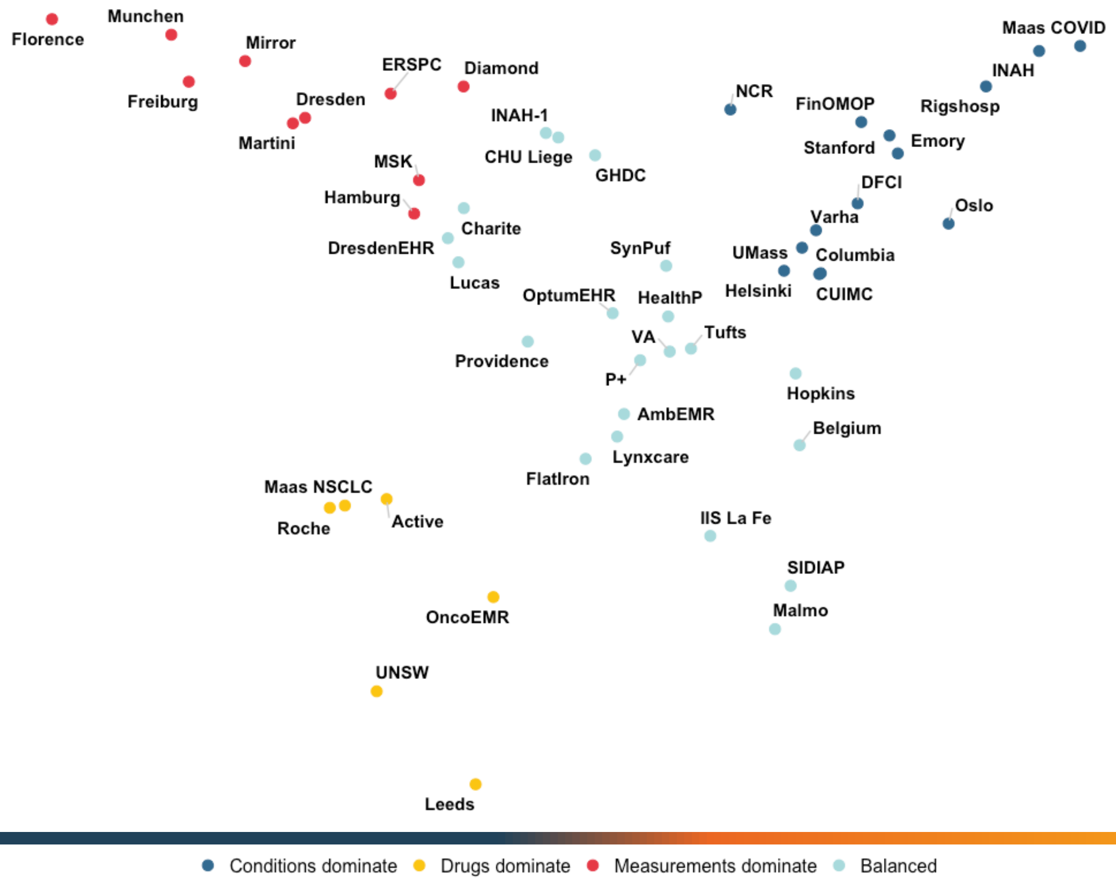
Origin of Sites


Providence
HealthPartners Institute
Stanford MEDICINE
Tufts Medical Center
Dana-Farber Cancer Institute
COLUMBIA
Columbia University Irving Medical Center
Memorial Sloan Kettering Cancer Center
JOHNS HOPKINS
EMORY WINSHIP CANCER INSTITUTE
IQVIA
VINCI
VA Informatics and Computing Infrastructure
HUS
Helsinki University Hospital
Cancer Registry of Norway
inah
UZA
NHS
The Leeds Teaching Hospitals NHS Trust
LMU
universität freiburg
TECHNISCHE UNIVERSITÄT DRESDEN
LUCAS
lung cancer focus
MARTINI-KLINIK
IDIAP JORDI GOL
PIONEER
DigiCore
UNSW SYDNEY

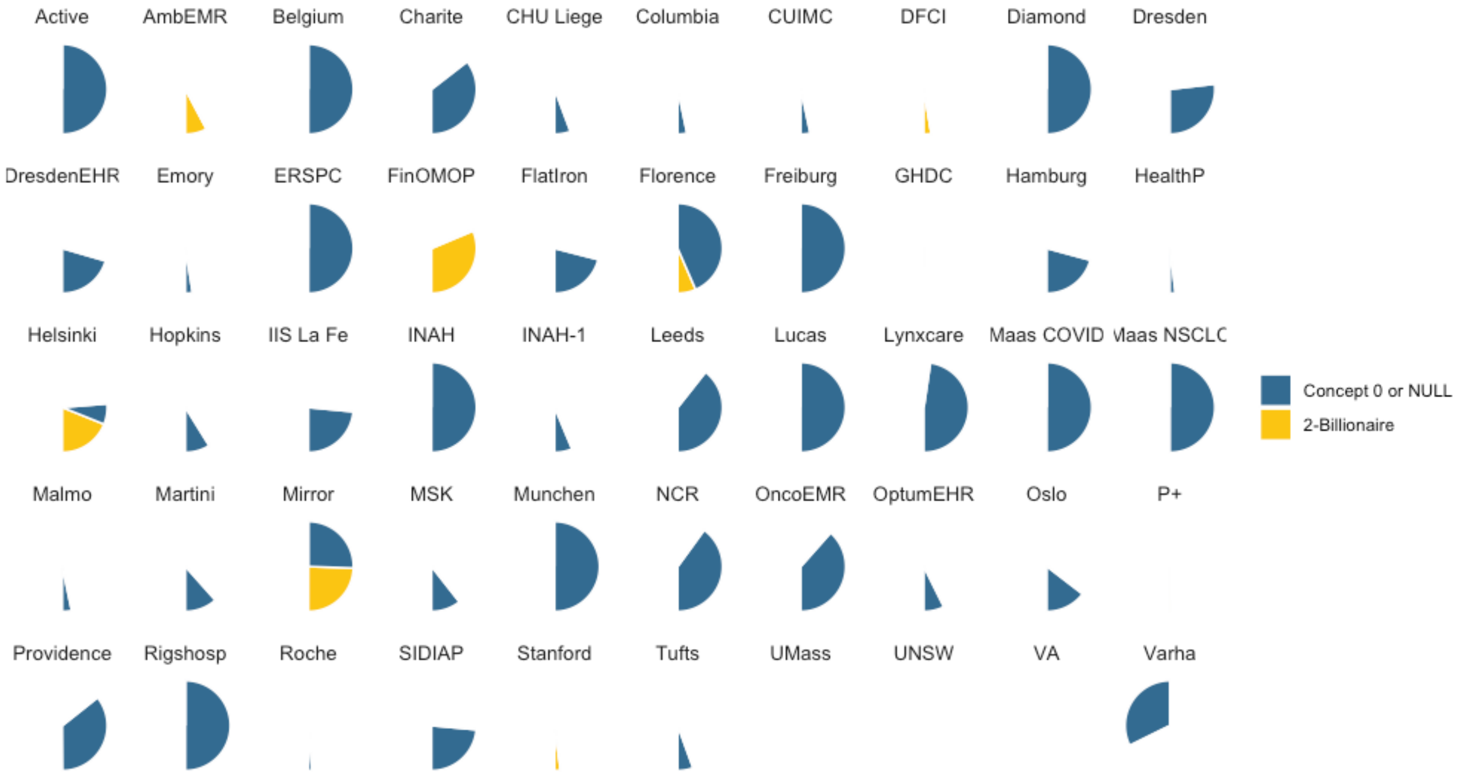
Distribution of Cancer Types



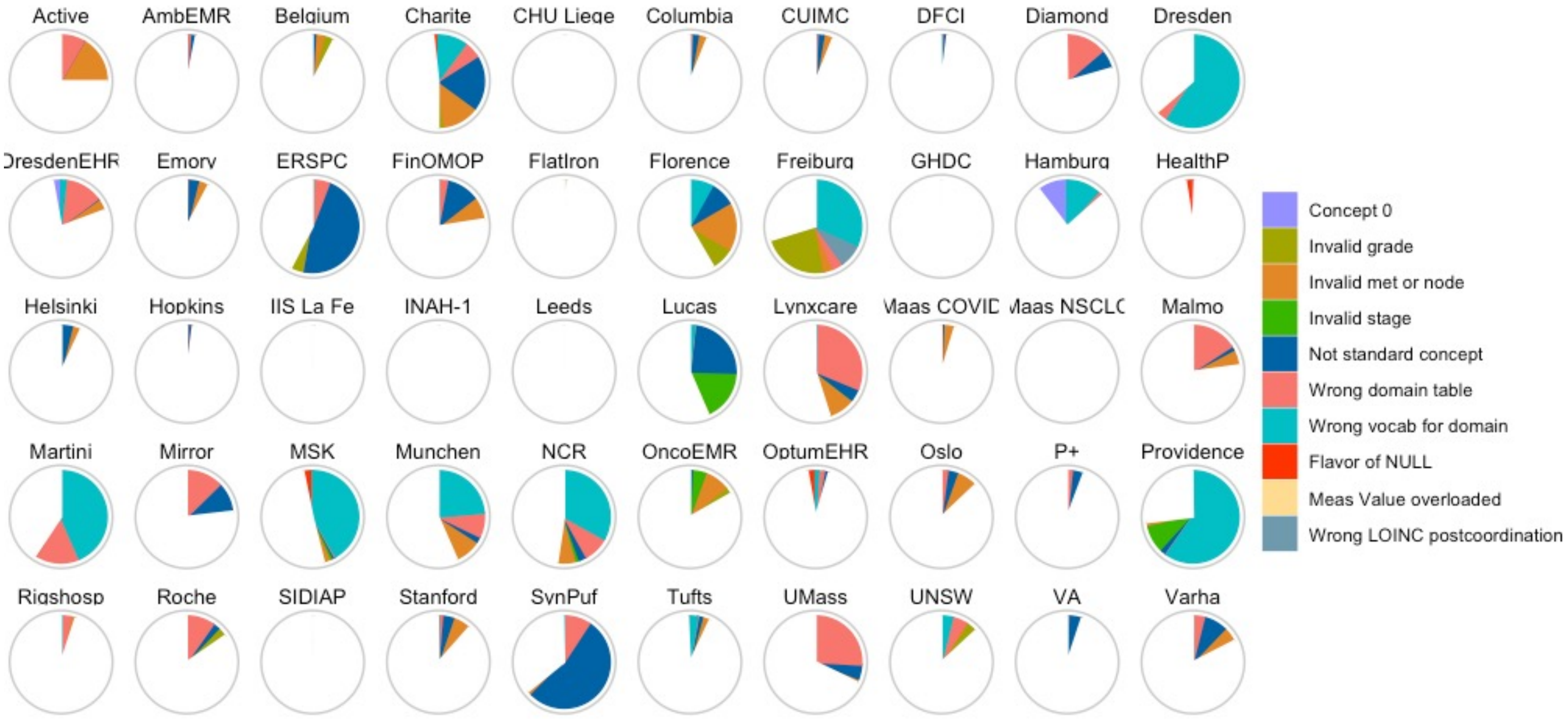
Information Distribution per Domain



Source Concepts – Misdemeanors



Standard Concepts – Felonies



Patch or Fix

Patches

Patches are a temporary short-term fix!!
Will be made available on Github for ETL purposes

Fix of concepts

- Mets, stages, grades
  - NAACCR -> Cancer Modifiers
  - LOINC -> Cancer Modifiers
- Conditions
  - SNOMED -> SNOMED

Combine histology+topography

- ICDO, SNOMED histology concepts
- SNOMED conditions concepts without topography

- ICDO, SNOMED topography concepts
- SNOMED conditions with generic histology (malignant neoplasm)

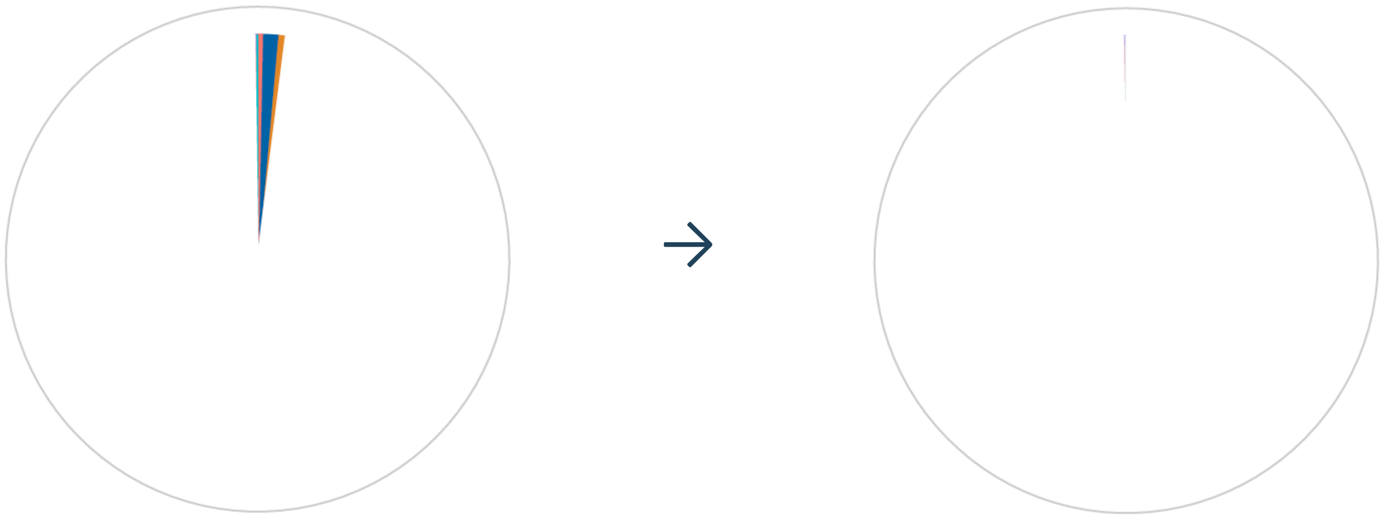
Fix

- New Vocabulary release for re-running the ETL
  - Only oncology fixes
  - This spring
  - Dissemination through Athena or https://oncology.ohdsi.org

→ This is an exception!! We will not establish a new process separate from OHDSI.

Iterate

Before and after patching



iCAN mNSCLC Studyathon

March 25-28, 2025
Helsinki, Finland

iCAN Digital Precision Cancer Medicine
HUS
HUS & HELSINGIN YLIOPISTO YHTEISTYÖSSÄ
HUS & HELSINGFORS UNIVERSITET I SAMARBETE
A COLLABORATION BETWEEN HUS & UNIVERSITY OF HELSINKI

Exploring the Real-World Treatment Landscape of mNSCLC

In this studyathon, we are **characterizing real-world treatment patterns of metastatic NSCLC**, with a focus on **the adoption and impact of immune checkpoint inhibitors (ICIs) across different regions**.

🔗 **Study GitHub Repository:** https://github.com/ohdsi-studies/MNSCLCStudyathon

Data Partner Status

Country	Institution	Data readiness	Patch	Rerun	Diagnostics	Diagnostics rerun	Analysis-1
Finland	HUS	☑	☑	☑	☑	☑	☑
	Varha	☑	☑	☑			
	Pirha	☑	☑	☑			
Norway	OUS	☑	☑	☑			
	CRN	☐	☐	☑			
Belgium	UZL	☑	☑	☑		☑	☑
	INAH-1	☑	☑				
	CHU Liege	☑	☑				
	GHDC	☑	☑				
Germany	Hamburg	☑	☑	☑			☑
	Dresden	☑	☑	☑			
	Charite	☑	☑	☑		☐	
UK	Leeds	☑	☑	☑		☑	☑
Spain	IIS La Fe	☑	☑	☑		☐	
Australia	UNSW	☑	☑	☐		☐	
Denmark	Risgshosp	☑	☑	☑		☑	
US	DFCI	☑	☑	☑	☑	☐	☑
	Providence	☑	☐	☐			
	Emory	☑	☑	☑	☑	☑	
	OptumEHR-Oncology	☑	☐	☐			
	FlatIron	☑	☑	☐		☐	
Fglobal	Wayfind-R	☑	☑	☑	☑	☑	☑

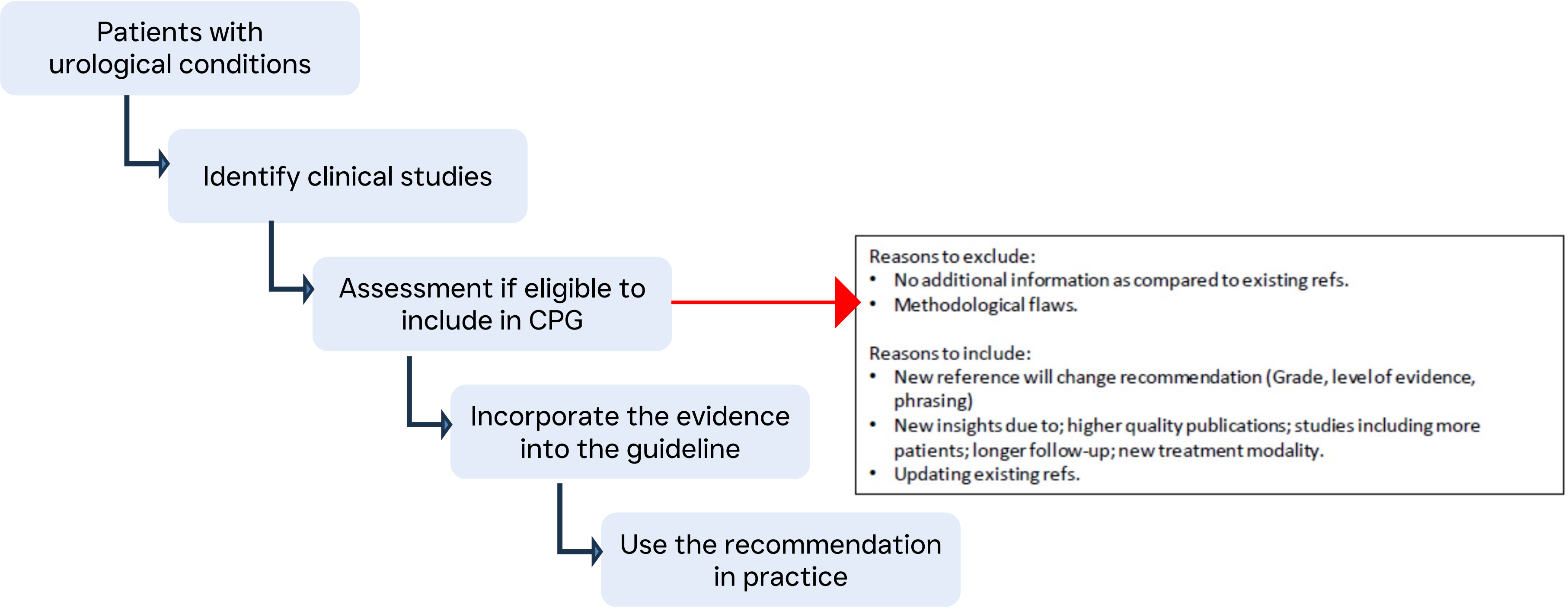
OHDSI

Observational Health Data Sciences and Informatics

Guidelinathon

How do we make RWE impactful?

How is guideline development done today?



Patients with urological conditions

Identify clinical studies

Assessment if eligible to include in CPG

Incorporate the evidence into the guideline

Use the recommendation in practice

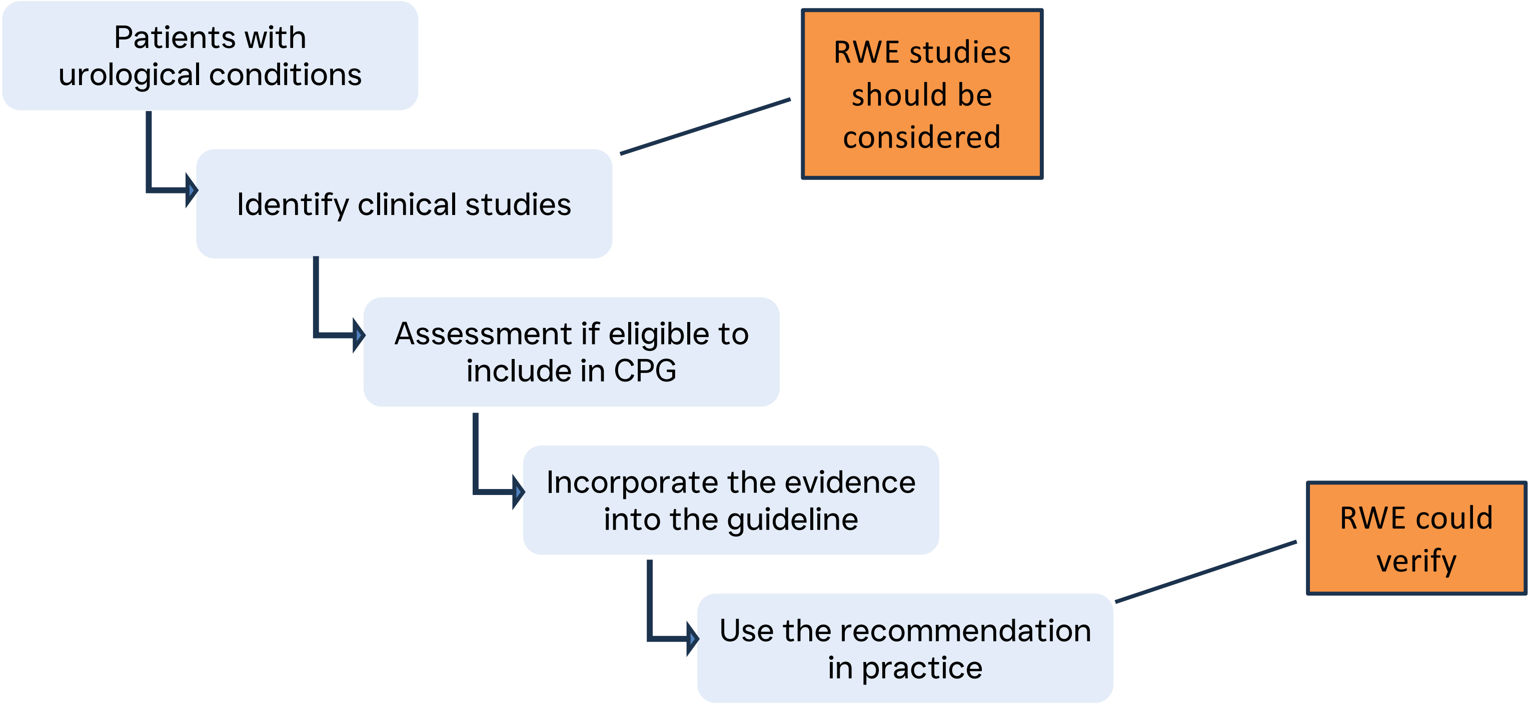
Reasons to exclude:

- No additional information as compared to existing refs.
- Methodological flaws.

Reasons to include:

- New reference will change recommendation (Grade, level of evidence, phrasing)
- New insights due to; higher quality publications; studies including more patients; longer follow-up; new treatment modality.
- Updating existing refs.

RWE is missing from this process.



Patients with urological conditions

Identify clinical studies

Assessment if eligible to include in CPG

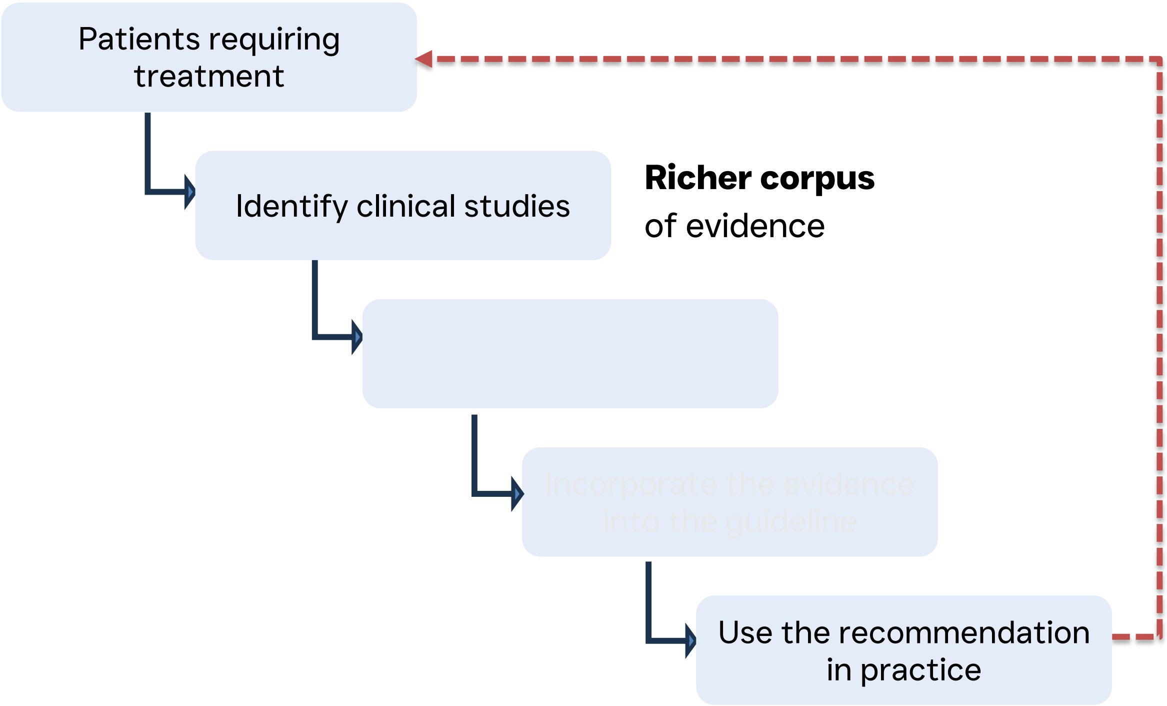
Incorporate the evidence into the guideline

Use the recommendation in practice

RWE studies should be considered

RWE could verify

What can RWE help with?



Patients requiring treatment

Identify clinical studies

Richer corpus of evidence

Incorporate the evidence into the guideline

Use the recommendation in practice

1. **Relevance:** Are there real-world patients who fit the criteria for each treatment recommendation?
2. **Adherence:** To which degree are clinical guideline recommendations applied in practice?
3. **Generalizability:** Do the recommended treatments achieve the desired outcomes in diverse patient populations?
4. **Unmet need:** Are there gaps in guidelines where RWE can improve recommendations?

Guideline Development Process Today

Currently

(non-RWE) **Study eligibility form high level evidence topics**				
Guideline Panel:		**Year of update:**		
Q1	Type of study - is the study design one of the following?	Yes ⇩	Unclear ⇩	No ⇩
Q2	Participants in the study	Yes ⇩	Unclear ⇩	No ⇩
Q3	Interventions and comparisons or tests in the study	Yes ⇩	Unclear ⇩	No ⇩
Q4	Outcomes in the study	Yes ⇩	Unclear ⇩	No ⇩
	Final decision (subject to clarification of 'unclear' points)	Include	Unclear	Exclude

Add RWE

RWE study eligibility	
Guideline Panel:	
Q1	Retrospective non-interventional study on data from point of care?
Q2	Selected cohorts in the study
Q3	Comparisons or tests in the study
Q4	Outcomes in the study
	Final decision

Problem: RWE studies are challenging

RCT	RWE studies
• Controlled	• Healthcare driven
• Randomized	• Prone to bias and confounding
• Designed for question	• Design often follows poor data
• Methodology well established for achieving study result	• Methodology for achieving study result and confounding control demanding

→ RWE studies need proper assessment

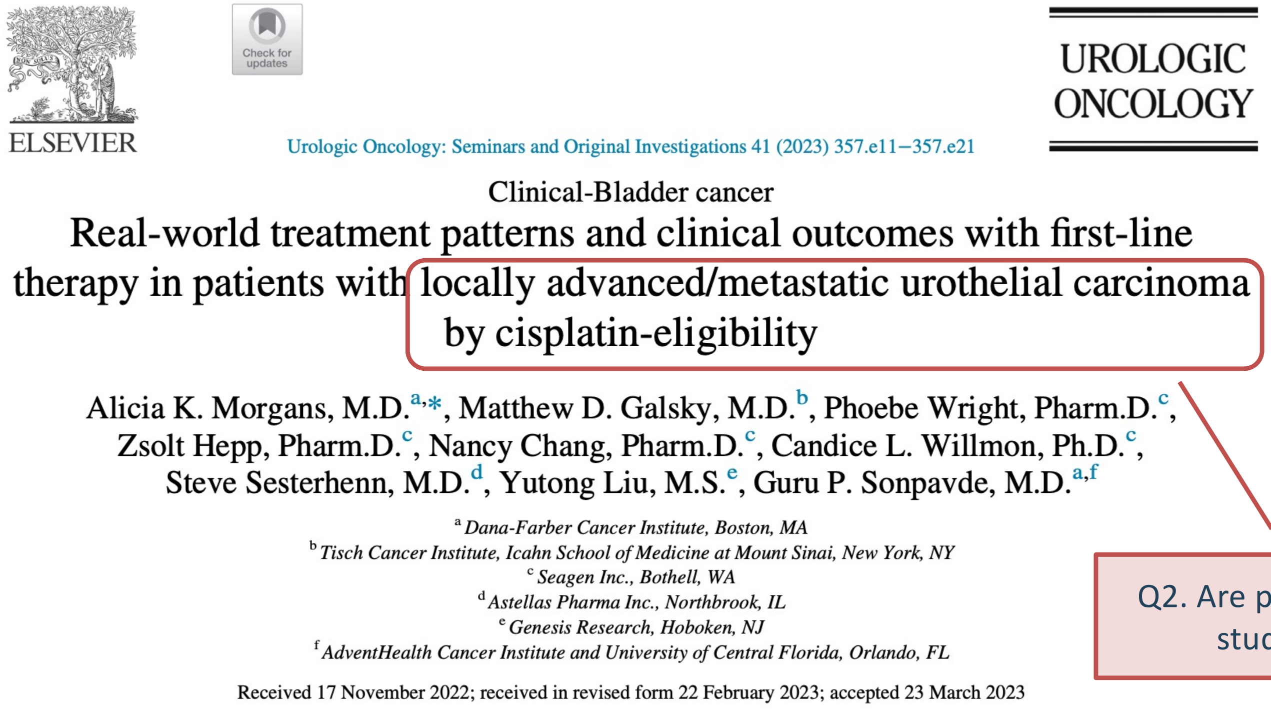
Adding RWE to guideline development

We need:

1. **Framework** for extracting populations and treatment recommendations from guideline
2. Process for Systematic RWE **Evaluation**
3. **Education** for guideline developers on RWD/E
4. Systematic approach to **develop** de-novo RWE for guideline integration

→ Generate RWE only if they can use it

Example Study



Urologic Oncology: Seminars and Original Investigations 41 (2023) 357.e11−357.e21

UROLOGIC
ONCOLOGY

Clinical-Bladder cancer

Real-world treatment patterns and clinical outcomes with first-line therapy in patients with locally advanced/metastatic urothelial carcinoma by cisplatin-eligibility

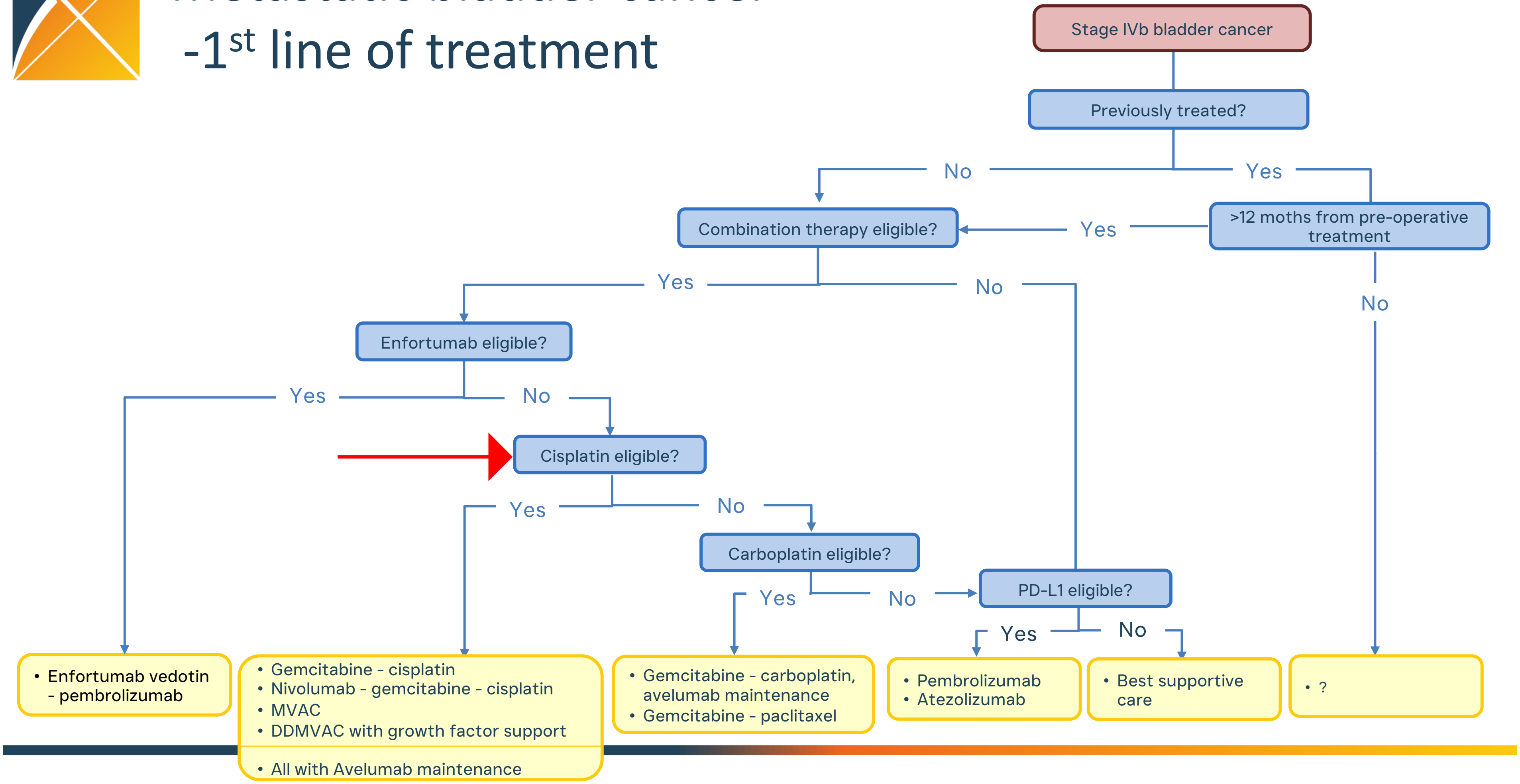
Alicia K. Morgans, M.D.[a,*], Matthew D. Galsky, M.D.[b], Phoebe Wright, Pharm.D.[c], Zsolt Hepp, Pharm.D.[c], Nancy Chang, Pharm.D.[c], Candice L. Willmon, Ph.D.[c], Steve Sesterhenn, M.D.[d], Yutong Liu, M.S.[e], Guru P. Sonpavde, M.D.[a,f]

[a] *Dana-Farber Cancer Institute, Boston, MA*
[b] *Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY*
[c] *Seagen Inc., Bothell, WA*
[d] *Astellas Pharma Inc., Northbrook, IL*
[e] *Genesis Research, Hoboken, NJ*
[f] *AdventHealth Cancer Institute and University of Central Florida, Orlando, FL*

Received 17 November 2022; received in revised form 22 February 2023; accepted 23 March 2023

Q2. Are participants in the study relevant?

Metastatic bladder cancer
-1st line of treatment



Stage IVb bladder cancer

Previously treated?

No

Yes

Combination therapy eligible?

Yes

>12 moths from pre-operative treatment

Yes

No

No

Enfortumab eligible?

Yes

No

Cisplatin eligible?

Yes

No

Carboplatin eligible?

Yes

No

PD-L1 eligible?

Yes

No

- Enfortumab vedotin - pembrolizumab

- Gemcitabine - cisplatin
- Nivolumab - gemcitabine - cisplatin
- MVAC
- DDMVAC with growth factor support

- All with Avelumab maintenance

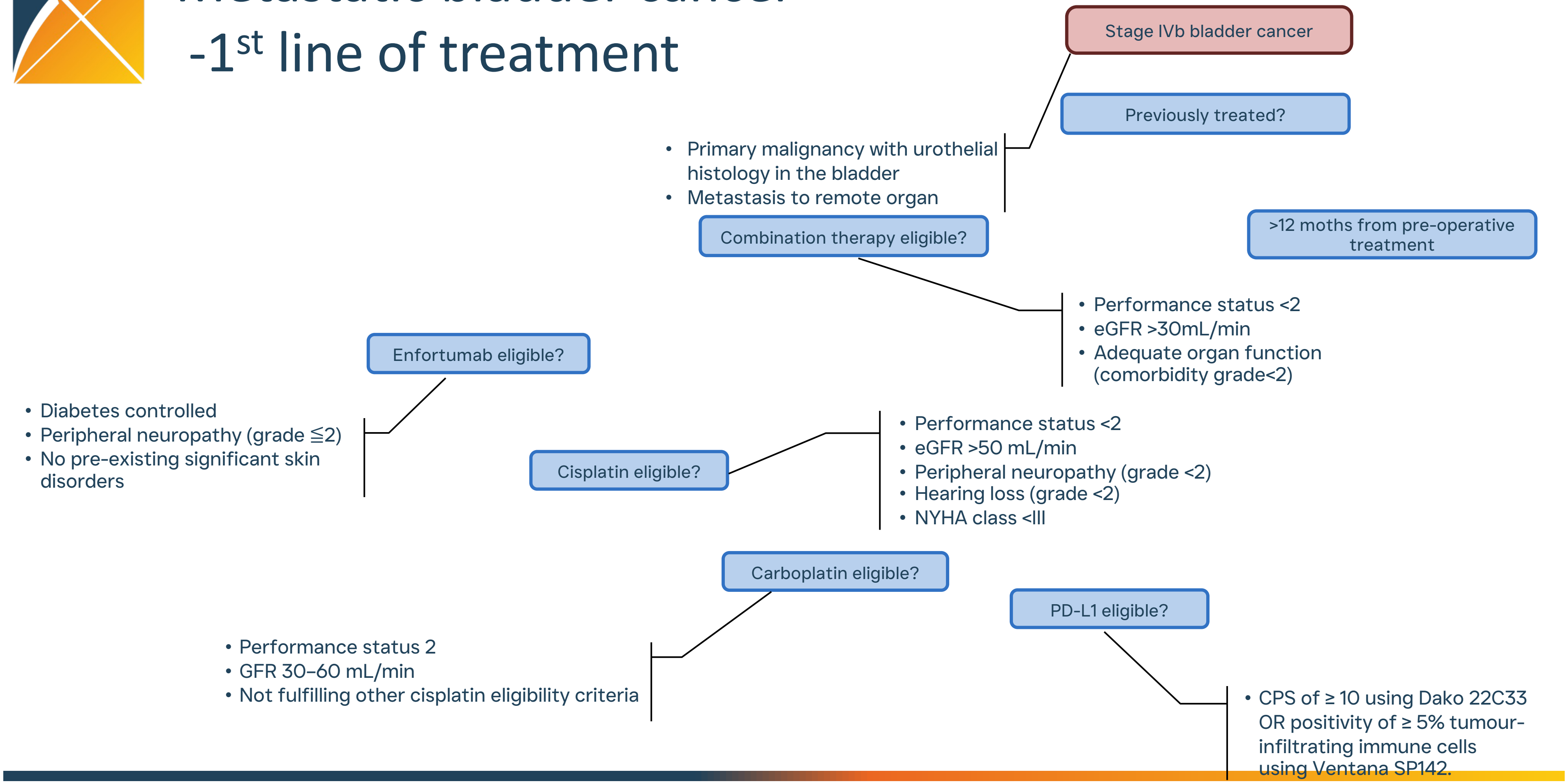
- Gemcitabine - carboplatin, avelumab maintenance
- Gemcitabine - paclitaxel

- Pembrolizumab
- Atezolizumab

- Best supportive care

- ?

Metastatic bladder cancer -1st line of treatment



Stage IVb bladder cancer

- Primary malignancy with urothelial histology in the bladder
- Metastasis to remote organ

Previously treated?

>12 moths from pre-operative treatment

Combination therapy eligible?

- Performance status <2
- eGFR >30mL/min
- Adequate organ function (comorbidity grade<2)

Enfortumab eligible?

- Diabetes controlled
- Peripheral neuropathy (grade ≦2)
- No pre-existing significant skin disorders

Cisplatin eligible?

- Performance status <2
- eGFR >50 mL/min
- Peripheral neuropathy (grade <2)
- Hearing loss (grade <2)
- NYHA class <III

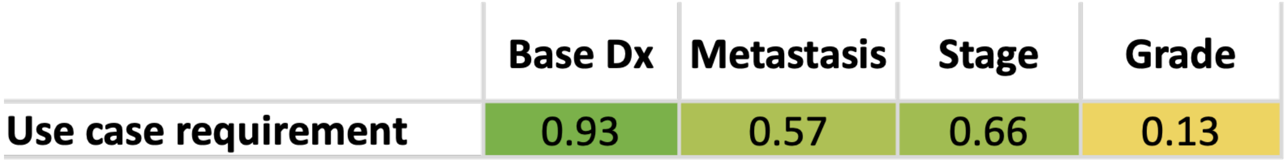
Carboplatin eligible?

- Performance status 2
- GFR 30–60 mL/min
- Not fulfilling other cisplatin eligibility criteria

PD-L1 eligible?

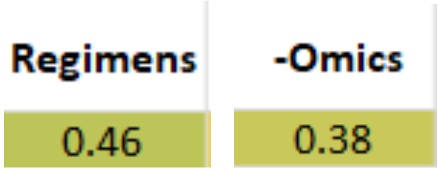
- CPS of ≥ 10 using Dako 22C33 OR positivity of ≥ 5% tumour-infiltrating immune cells using Ventana SP142.

Guidelinathon Data Readiness



	Base Dx	Metastasis	Stage	Grade
Use case requirement	0.93	0.57	0.66	0.13

Plus:



Regimens	-Omics
0.46	0.38

→ New round of iteration

Summary

- Cancer is more than vanilla OMOP
  - … if we want to do meaningful RWE
- Data need to be assessed
- Data often need to be fixed
- Oncology WG is innovating these
  - They need to become standard OHDSI

Join us at https://oncology.ohdsi.org