APAC Scientific Forum

2025.08.07

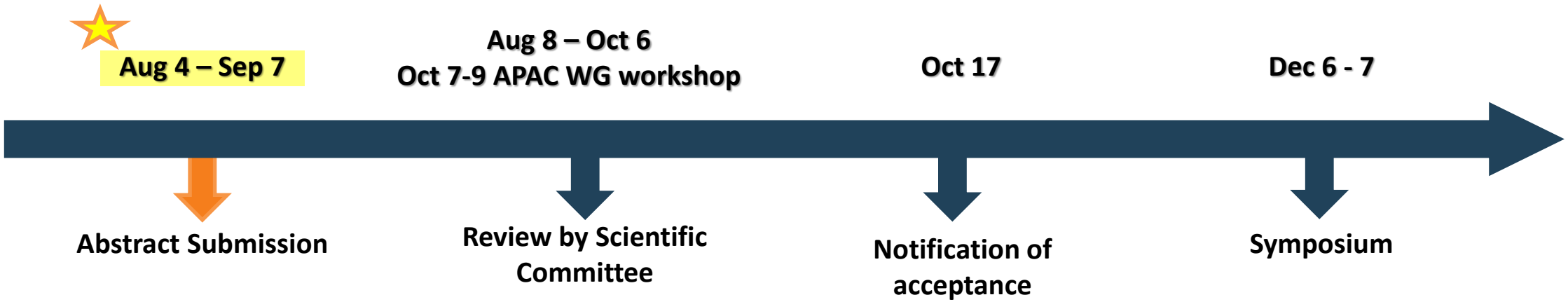
Agenda

- Promoting 2025 APAC Symposium
- Scientific Forum
  - CDM for oncology & AI-empowered clinical oncology data structure *by Subin Kim*

2025 OHDSI APAC Symposium

December 6-7 • Shanghai Jiao Tong University, China

Proposed Timeline for Collaborators Showcase



- Registration will be announced soon so stay tuned!

Brief Report Submission

Now Available!!

OHDSI APAC

2025 OHDSI APAC Collaborator Showcase Brief Report Submission Form

Thank you for your interest in the 2025 OHDSI APAC Collaborator Showcase! We are delighted that you are considering joining our research community and presenting your work at this year's symposium. The 2025 OHDSI APAC Symposium will be held in person **December 6-7** at the Shanghai Jiao Tong University in Shanghai, China.

Please take a few minutes to fill out this submission form to help the OHDSI APAC Scientific Review Committee better understand your work. The deadline to submit your brief report is **September 7**. You will receive a confirmation email of your responses upon completion. If the committee has selected your work to be presented at this year's symposium, you will be notified via email by **October 17**.

Should you need to change your responses to any of the questions on this form, please click on the "Edit response" button in the confirmation email you received. Should you need to revise your brief report, please email apacsymposium@ohdsi.org. Your submission will be removed, and you will need to submit again with the revised PDF.

Submissions are open!
All submissions are due **September 7**

2025 OHDSI APAC Symposium Agenda *(Tentative)*

Day 1 (Dec 6)

Tutorial Sessions 📚 - *tbc*

- Introduction of OHDSI/OMOP
- OMOP CDM and Vocabulary
- ETL
- Analytics

Day 2 (Dec 7)

Updates & Studies 💻

- OHDSI APAC and Regional Chapter updates
- 2025 APAC Studies: Overviews and Results
- Real-world Data Developments in China

Discussions & Presentations 💬

- Cross-community Panel Discussion
- Collaborator Showcase: Poster Presentations and Lightning Talks

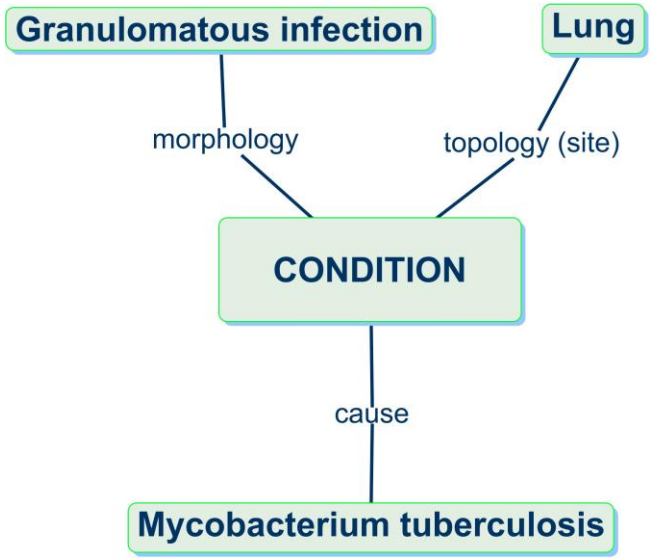
CDM for oncology & AI-empowered clinical oncology data structure

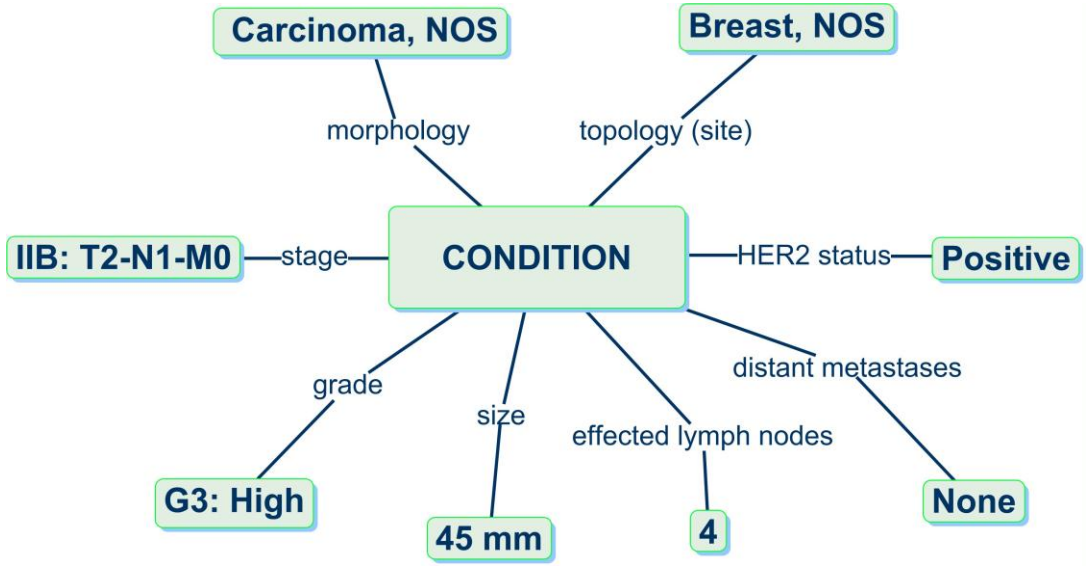
Subin Kim
Department of Biomedical Systems Informatics
Yonsei University College of Medicine

Challenges in standard OMOP CDM in oncology research

Problem 1: Cancer needs more detail

- **Granularity**





Normal Condition

Most normal conditions are defined by three main di mensions implicitly, plus some extra attributes

Cancer

- Cause is not known, but morphology and topology are detailed and explicit
- The many tumor attributes (modifiers) are also explicit and well defined

Problem 1: Cancer needs more detail

- Research question
  - “What is the progress-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 1: Cancer needs more detail

- Research question
  - “What is the progress-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
Lung	Anatomical site
Non-small cell	Histology
Metastatic disease	Tumor attribute
MET exon 14 skipping	Genomic variant
First line treatment	Treatment episode
Capmatinib	Regimen
Progression	Disease episode

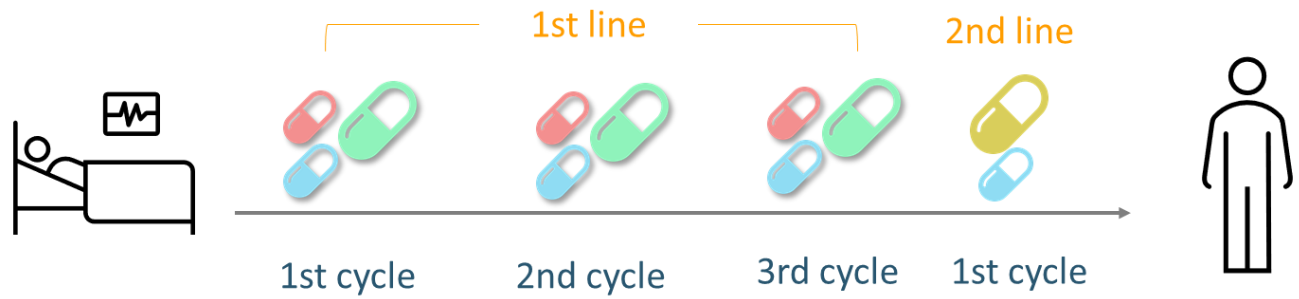
Mostly unavailable in the standard CDM

Problem 2: Chemotherapy is Not Standardized in CDM

- Chemotherapy is administered in complex, multi-cycle schedules
- Personalized regimens increase the complexity of standardization

Common treatment

Treatment on regimen

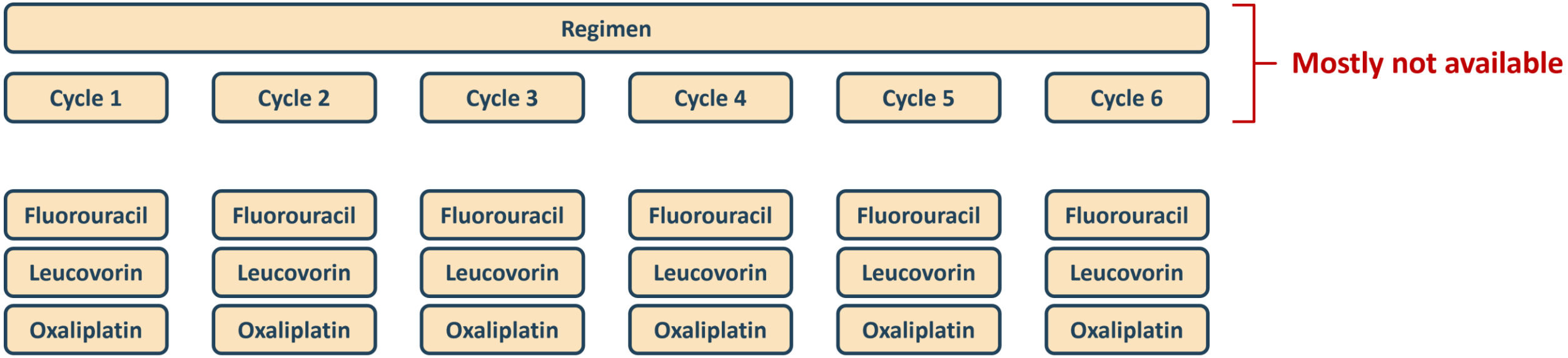


Benefits

1) Enhance clinical outcome
2) Reduced risk of complications
3) Reduced risk of side effects
4) Toxicity issues

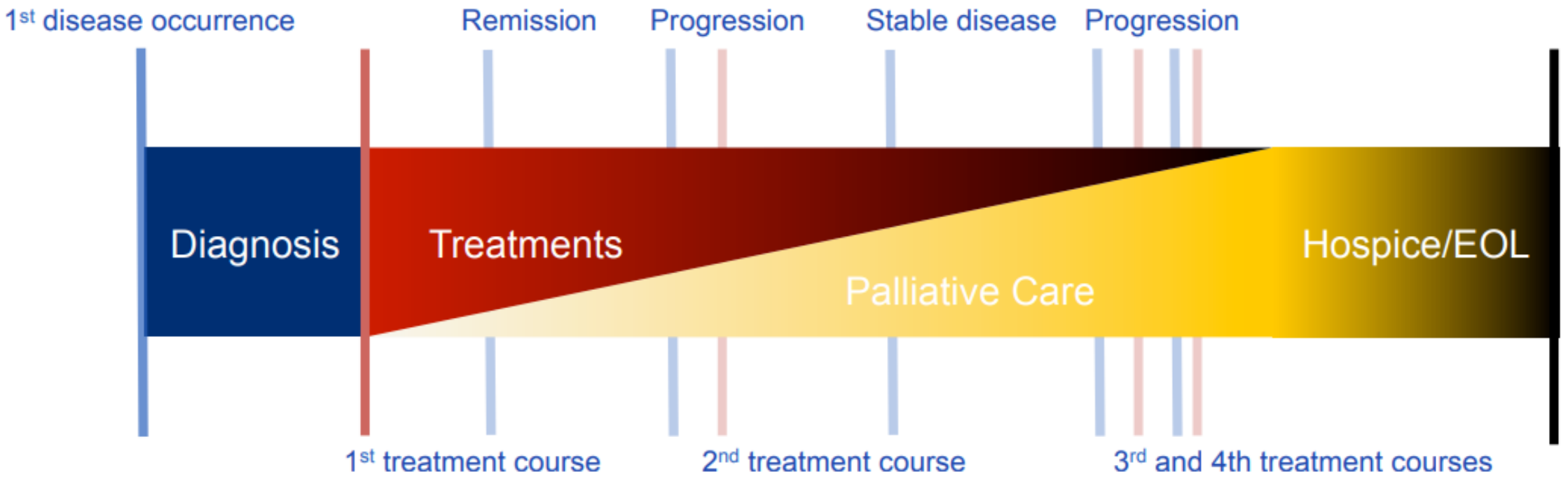
Problem 2: Chemotherapy is Not Standardized in CDM

- Example: FOLFOX (Fluorouracil, Leucovorin, Oxaliplatin) up to 6 cycles



Problem 3: Abstraction of episodes from clinical events

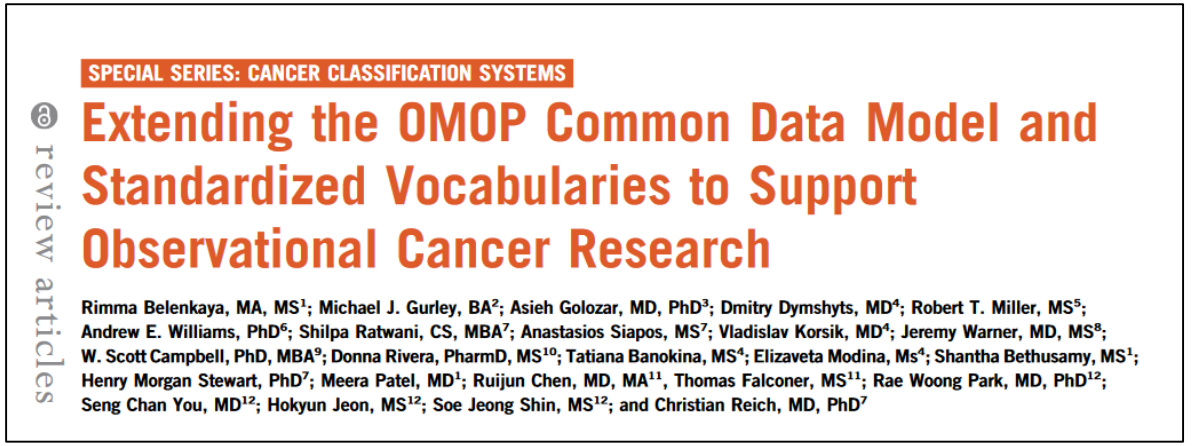
- A need to track **patients' trajectory**, but not supported in the source data
  - **Disease** episode (diagnosis to incidence of outcomes)
  - **Treatment** episode



Oncology CDM Extension

Oncology CDM Extension

- **Cancer Disease Model**
  - Cancer Diagnosis: Base Diagnosis + Diagnostic Modifiers
    (One-to-many connection between them)
- **Cancer Treatment Model**
  - Composite Level (Treatment Episodes) or Individual Level (standard OMOP)
- **Cancer Episode Model**
  - Continuous periods of disease or treatment with distinct clinical meaning
  - Composed of multiple events
  - Essential for conducting cancer research


SPECIAL SERIES: CANCER CLASSIFICATION SYSTEMS

review articles

Extending the OMOP Common Data Model and Standardized Vocabularies to Support Observational Cancer Research

Rimma Belenkaya, MA, MS[1]; Michael J. Gurley, BA[2]; Asieh Golozar, MD, PhD[3]; Dmitry Dymshyts, MD[4]; Robert T. Miller, MS[5]; Andrew E. Williams, PhD[6]; Shilpa Ratwani, CS, MBA[7]; Anastasios Siapos, MS[7]; Vladislav Korsik, MD[4]; Jeremy Warner, MD, MS[8]; W. Scott Campbell, PhD, MBA[9]; Donna Rivera, PharmD, MS[10]; Tatiana Banokina, MS[4]; Elizaveta Modina, Ms[4]; Shantha Bethusamy, MS[1]; Henry Morgan Stewart, PhD[7]; Meera Patel, MD[1]; Ruijun Chen, MD, MA[11], Thomas Falconer, MS[11]; Rae Woong Park, MD, PhD[12]; Seng Chan You, MD[12]; Hokyun Jeon, MS[12]; Soe Jeong Shin, MS[12]; and Christian Reich, MD, PhD[7]

Belenkaya, R., et al., *JCO Clinical Cancer Informatics* (2021).

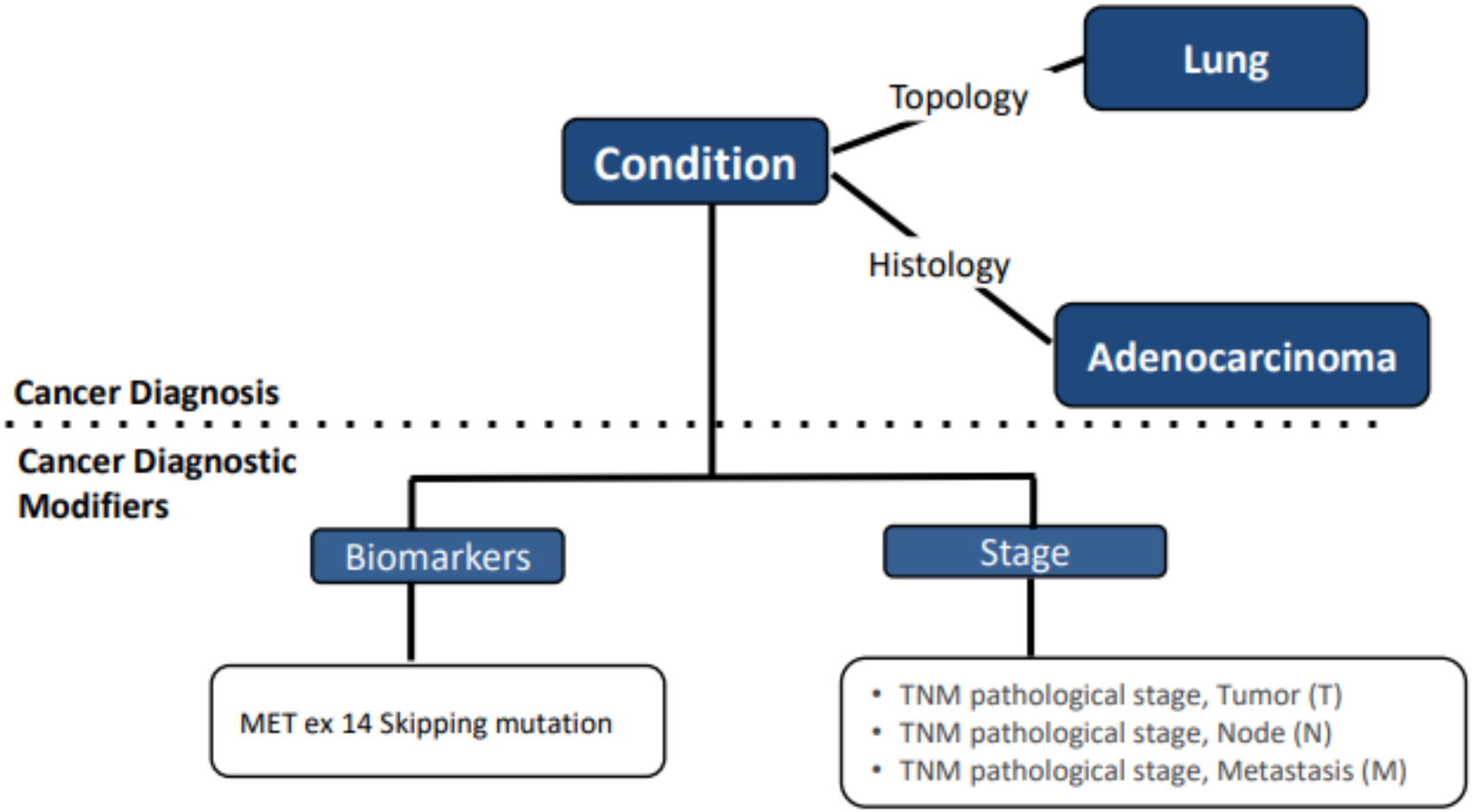
Oncology CDM Extension

- **Cancer Disease Model**
  - Cancer Diagnosis: Base Diagnosis + Diagnostic Modifiers
    (One-to-many connection between them)
- **Cancer Treatment Model**
  - Composite Level (Treatment Episodes) or Individual Level (standard OMOP)
- **Cancer Episode Model**
  - Continuous periods of disease or treatment with distinct clinical meaning
  - Composed of multiple events
  - Essential for conducting cancer research

Belenkaya, R., et al., *JCO Clinical Cancer Informatics* (2021).

Cancer Disease Model

- **Base Diagnosis**
  - Topography + Histology
- **Diagnostic Modifiers**
  - Staging/Grading
  - Topography
  - Histological pattern
  - Dimension
  - Extension/Invasion
  - Metastasis
  - Margin
  - Biomarker



Cancer Disease Model

- **Solution: ICD-O-3**

ICD-10

C18.7

Malignant neoplasm of sigmoid colon

ICD-O-3

8140/3-C18.7

Histology ICD-O + Location ICD-O
(Morphology + Topology)

Adenocarcinoma, NOS, of sigmoid colon

Cancer Disease Model

- Mapping base diagnosis into Condition_Occurrence table

Base diagnosis in Condition_Occurrence table

Field	Content	
condition_occurrence_id	123456789	
person_id	10001	
condition_concept_id	**44504380**	**OMOP Concept ID**
condition_start_datetime	2025-07-04	
condition_type_concept_id	32535	
condition_source_value	**8140/3-C18.7**	**ICD-O-3**
condition_source_concept_id	44504380	

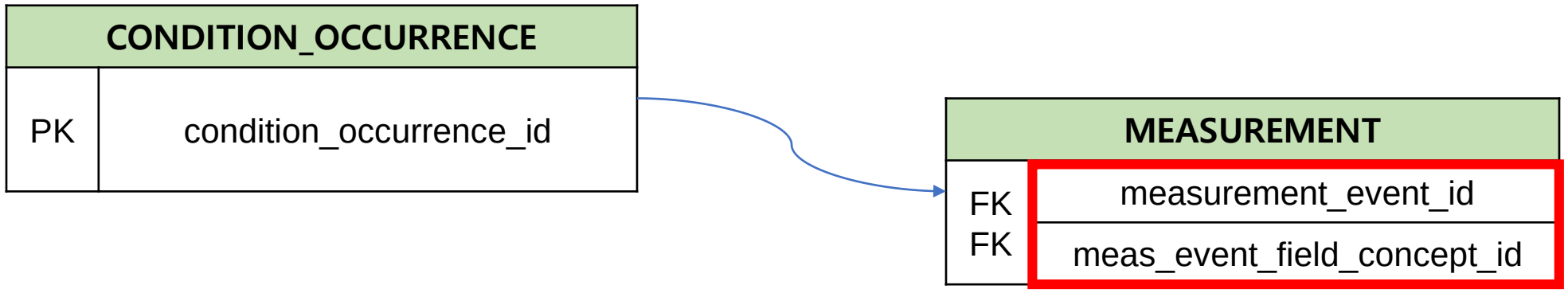
Cancer Disease Model

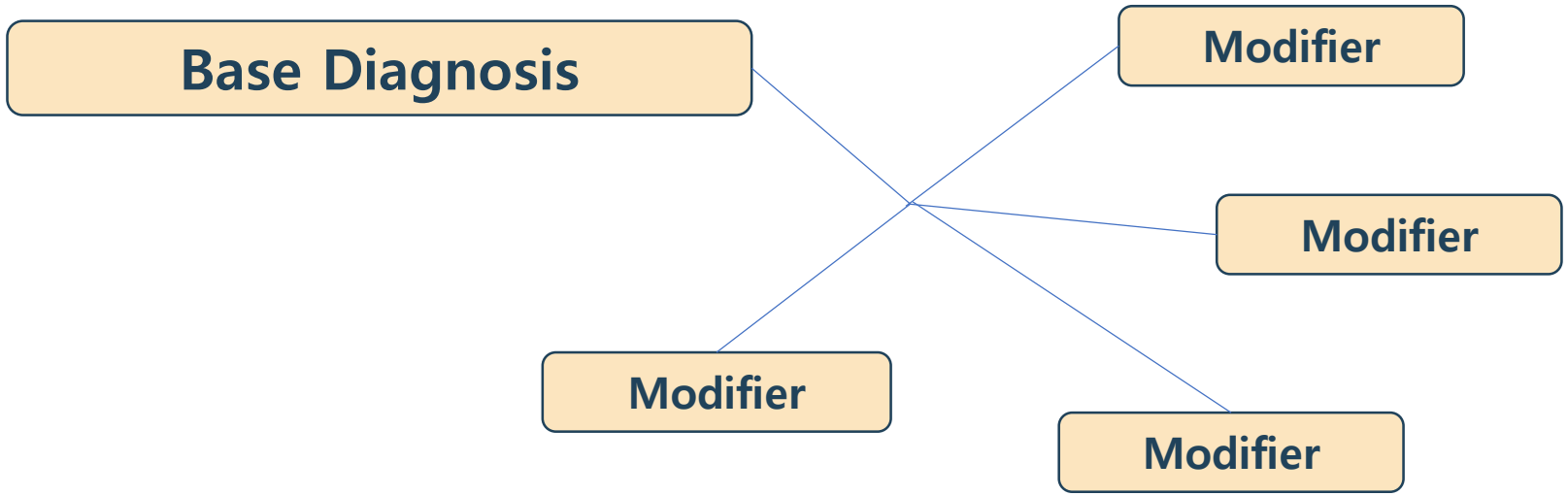
- Diagnostic Modifiers in CDM & Vocabulary
  - CDM: **MEASUREMENT**, Domain **"*Measurement*"**
  - Vocabulary
    - **Standard <u>Modifier</u> concepts** - from designated **"*<u>Cancer Modifier</u>*"** vocabulary

Modifier Type	Description	Vocabulary
Topography	More detail than ICD-O	Cancer Modifier, adopted by OMOP from CAP and NAACCR
Histology pattern	More detail than ICD-O	Cancer Modifier, adopted by OMOP from CAP and NAACCR
Staging/Grading	Externally defined	NCIt, Cancer Modifier, adopted by OMOP from CAP and NAACCR
Extension/Invasion	Local tumor growth	Cancer Modifier, adopted by OMOP from CAP and NAACCR
Nodes	Growth into lymphatic system	Cancer Modifier, adopted by OMOP from CAP and NAACCR
Metastasis	Distant growth (except lymph nodes	OMOP, adopted from CAP and NAACCR
Dimension	Tumor size	OMOP, adopted from CAP and NAACCR
Margin	Margin after surgery	OMOP, adopted from CAP and NAACCR
Biomarker	Genomic variants	HGNC, OMOP, adopted from CAP, CGI, CIViC, ClinVar, JAX, NAACCR, NCIt,

Cancer Disease Model

- Diagnostic Modifiers in CDM & Vocabulary
  - CDM: **New columns for modifiers** in measurement table to link with condition_occurrence table



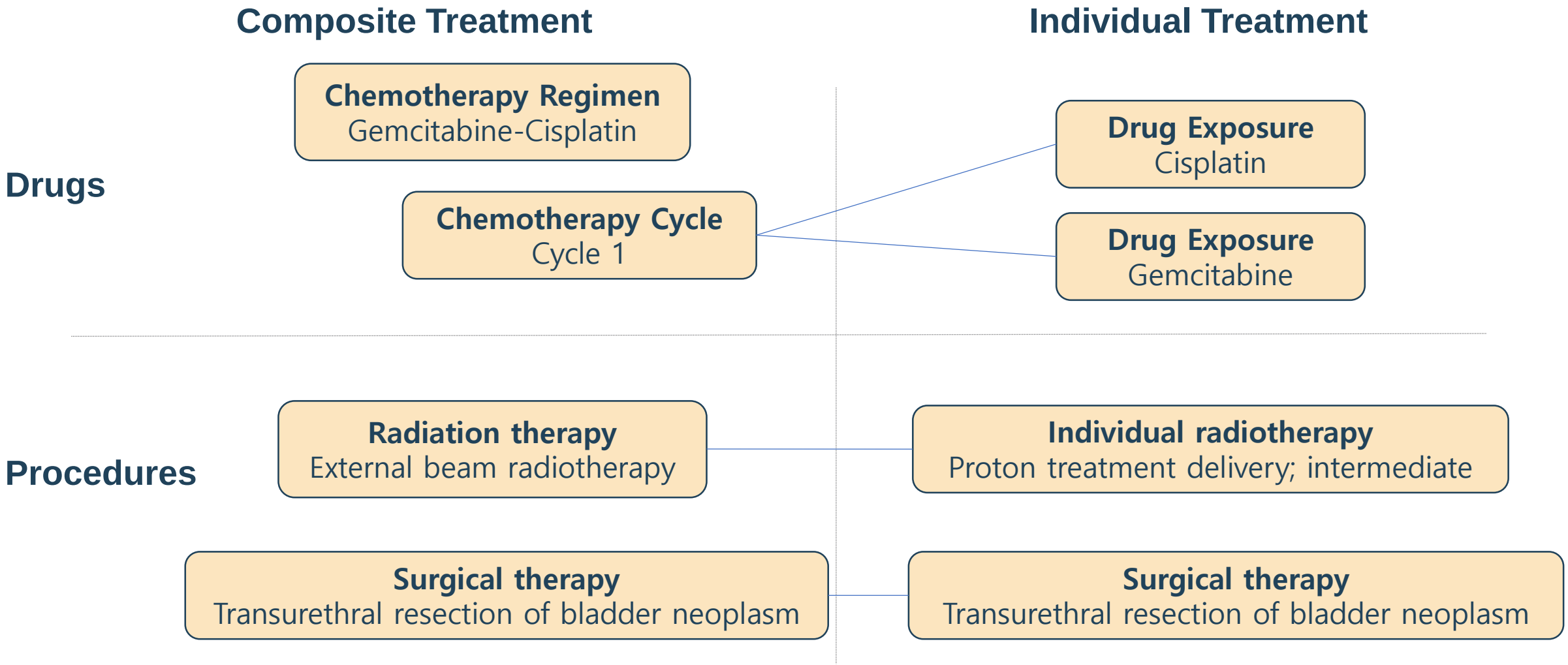


Oncology CDM Extension

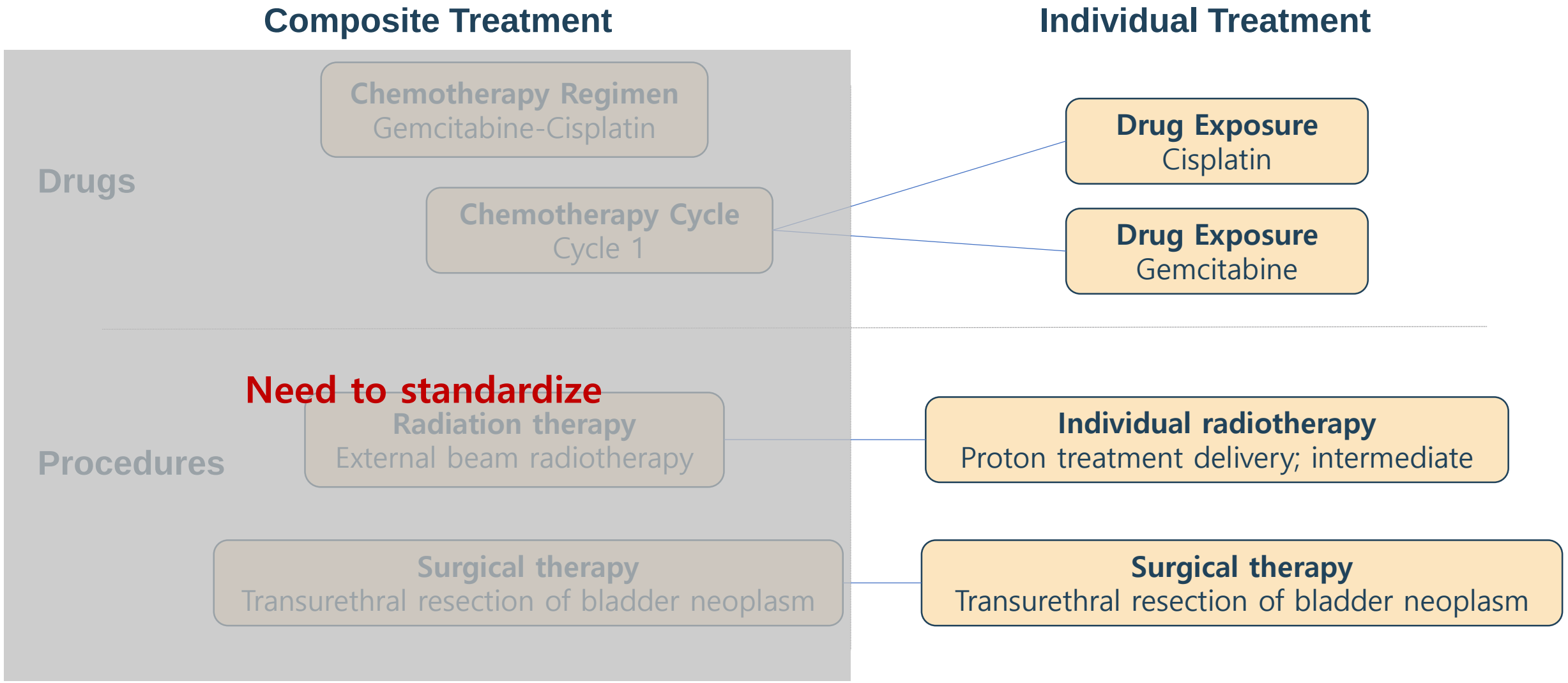
- **Cancer Disease Model**
  - Cancer Diagnosis: Base Diagnosis + Diagnostic Modifiers
    (One-to-many connection between them)
- **Cancer Treatment Model**
  - Composite Level (Treatment Episodes) or Individual Level (standard OMOP)
- **Cancer Episode Model**
  - Continuous periods of disease or treatment with distinct clinical meaning
  - Composed of multiple events
  - Essential for conducting cancer research

Belenkaya, R., et al., *JCO Clinical Cancer Informatics* (2021).

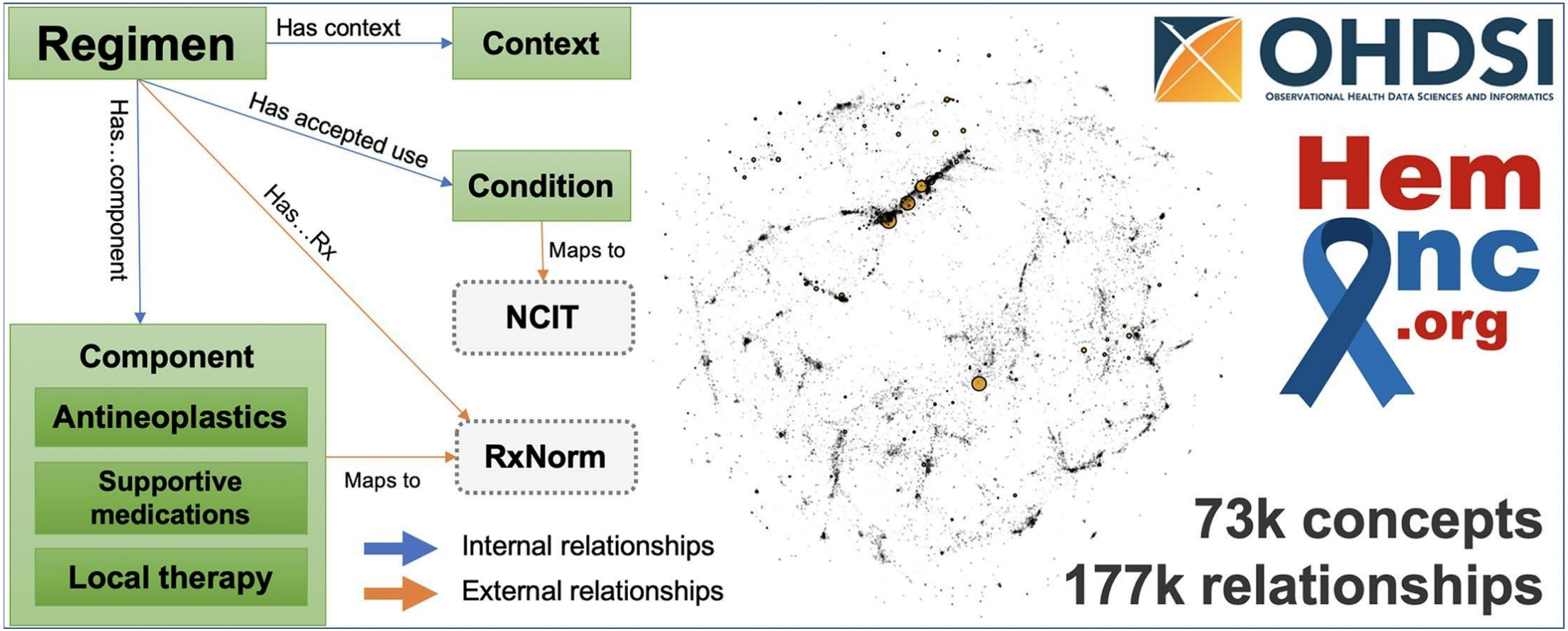
Cancer Treatment Model



Cancer Treatment Model



Cancer Treatment Model



Warner, J. L., et al., *Journal of biomedical informatics* (2019).

Cancer Treatment Model

FOLFOX

DETAILS

Domain ID	Regimen
Concept Class ID	Regimen
Vocabulary ID	HemOnc
Concept ID	35806596
Concept code	33193
Validity	Valid
Concept	Standard

LANGUAGE	SYNONYM CONCEPT
English	FOLinic acid, Fluorouracil, OXaliplatin

Valid start	27-May-2019
Valid end	31-Dec-2099

TERM CONNECTIONS (23)

HIERARCHY | RELATED CONCEPTS

RELATIONSHIP	RELATES TO	CONCEPT ID	VOCABULARY
Has accepted use (HemOnc)	Pancreatic cancer	42542217	HemOnc
Has context (HemOnc)	Non-curative therapy	35803588	HemOnc
Has cytotoxic chemotherapy (HemOnc)	Fluorouracil	35803077	HemOnc
	Leucovorin	35803081	HemOnc
	Oxaliplatin	35803227	HemOnc
Has cytotoxic chemotherapy - RxNorm (HemOnc)	fluorouracil	955632	RxNorm
	leucovorin	1388796	RxNorm
	oxaliplatin	1318011	RxNorm
Has modality (HemOnc)	Chemotherapy	35803401	HemOnc
Is a	Chemotherapy-containing regimen	37557736	HemOnc

Oncology CDM Extension

- **Cancer Disease Model**
  - Cancer Diagnosis: Base Diagnosis + Diagnostic Modifiers
    (One-to-many connection between them)
- **Cancer Treatment Model**
  - Composite Level (Treatment Episodes) or Individual Level (standard OMOP)
- **Cancer Episode Model**
  - Continuous periods of disease or treatment with distinct clinical meaning
  - Composed of multiple events
  - Essential for conducting cancer research

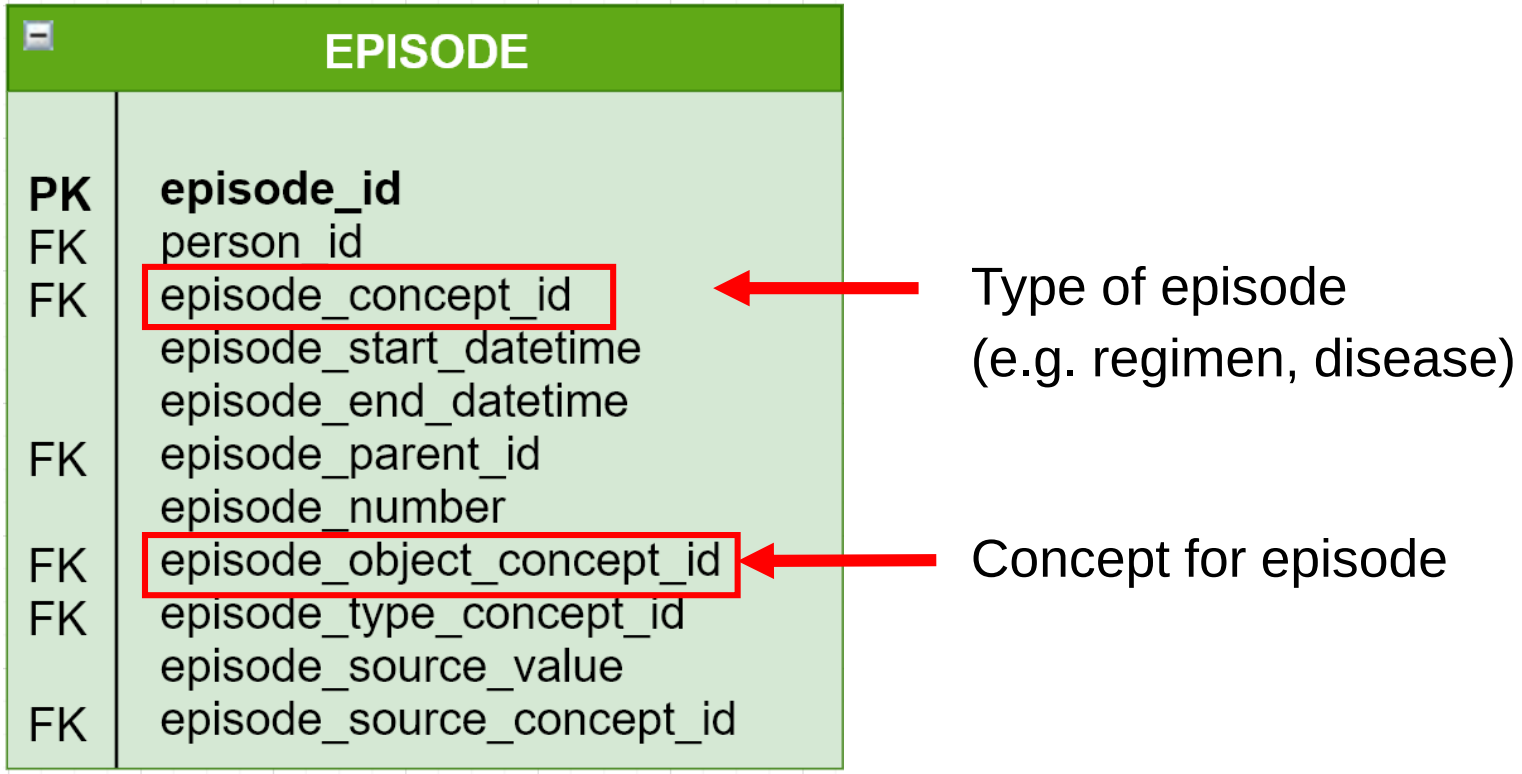
Belenkaya, R., et al., *JCO Clinical Cancer Informatics* (2021).

Cancer Episode Model

- **Episode Model**
  - **What are Episodes?**
    - Continuous periods of disease or treatment that have distinct **clinical meaning** and are composed of multiple events, e.g.
    - Progressive Disease Episode
    - Treatment Regimen
  - **Why do we need them?**
    - Overall and Progression-Free Survival (OS, PFS)
    - Time to progression

Cancer Episode Model

- **New table: EPISODE**



Cancer Episode Model

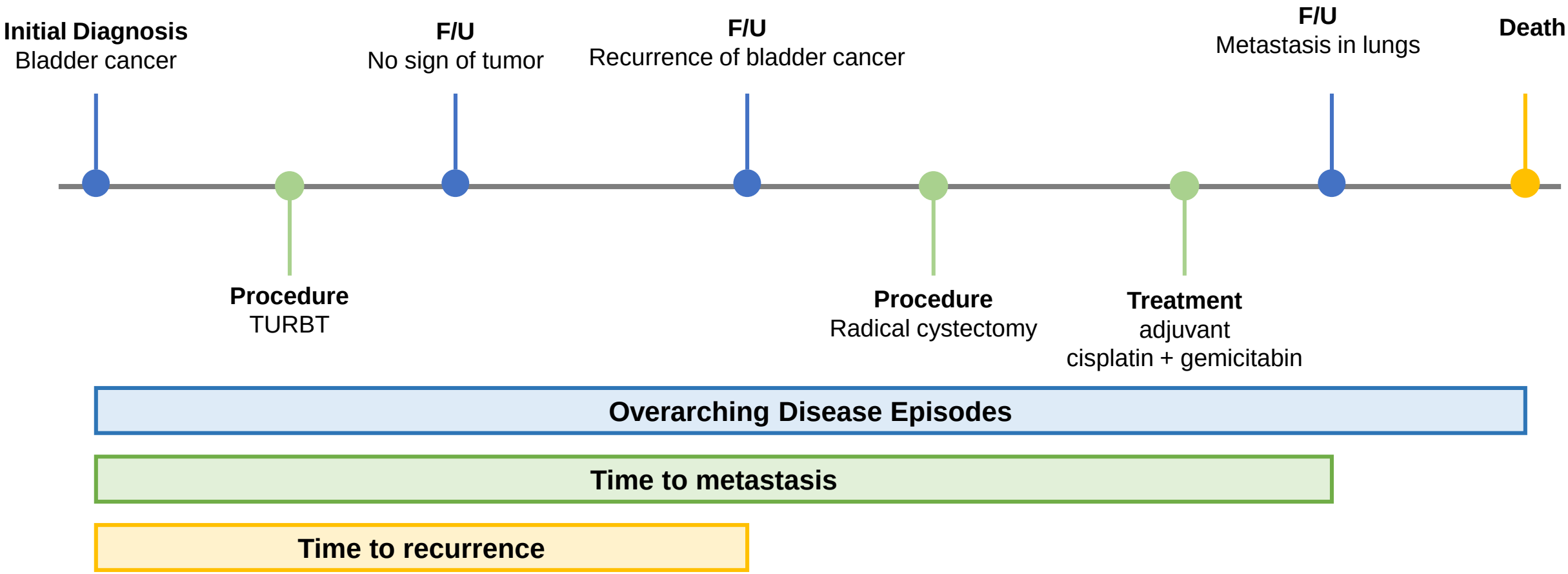
- **Disease Episodes in OMOP CDM**

Overarching	/	Extent	/	Dynamic

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

Cancer Episode Model

- **Disease Episodes in OMOP CDM**



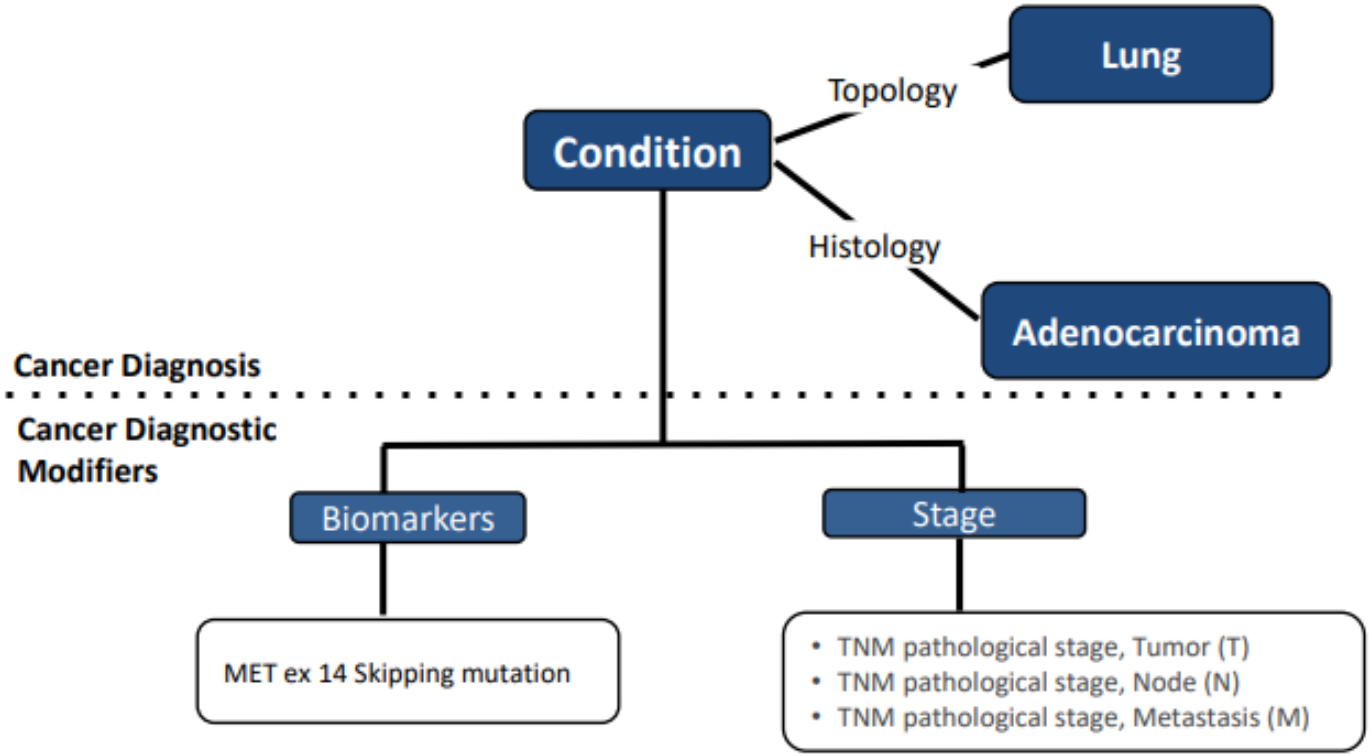
AI-empowered
clinical oncology data structure

Challenges of EHR for Oncology CDM

- Cancer-specific data is **unstructured** in EHR
  - Challenges to standardize clinical data into Oncology CDM

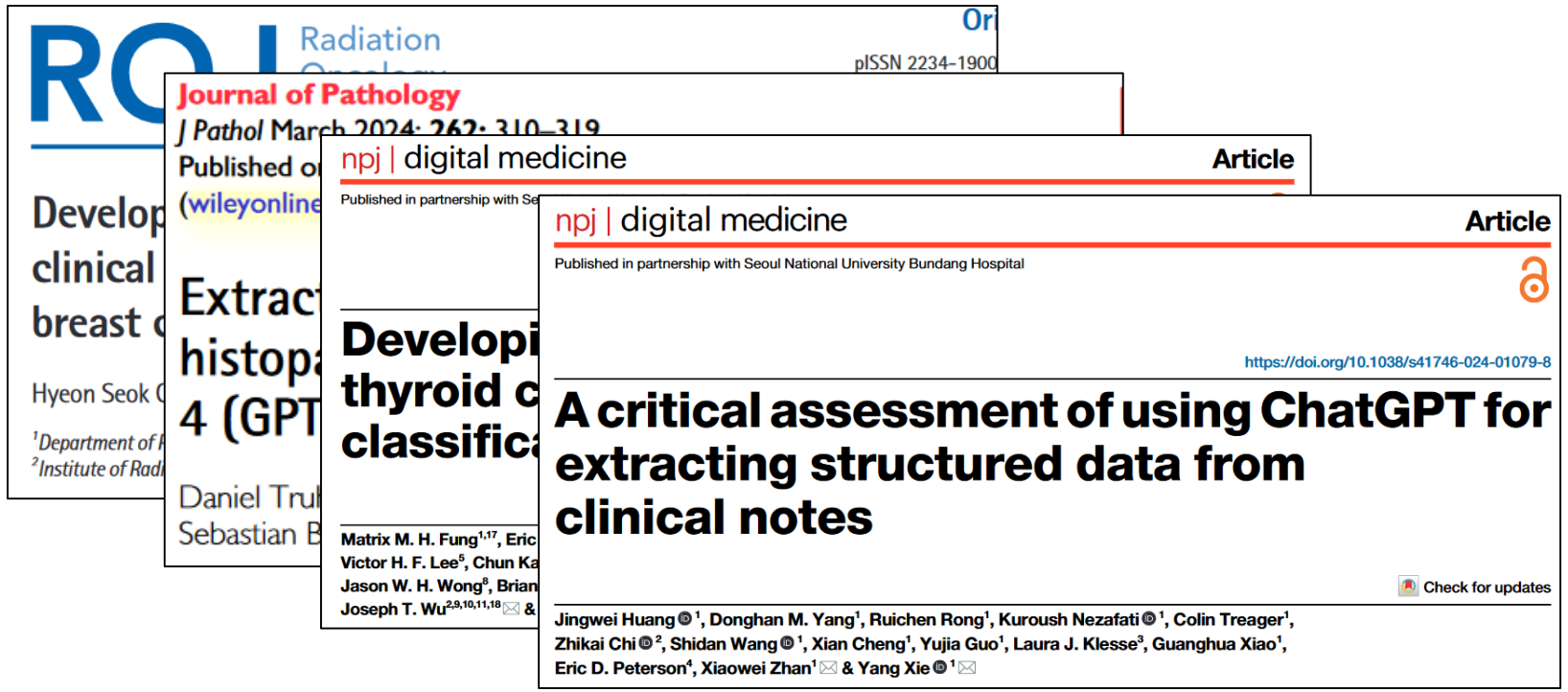
Cancer Modifiers

- Staging/Grading
- Topography
- Histological pattern
- Dimension
- Extension/Invasion
- Metastasis
- Margin
- Biomarker



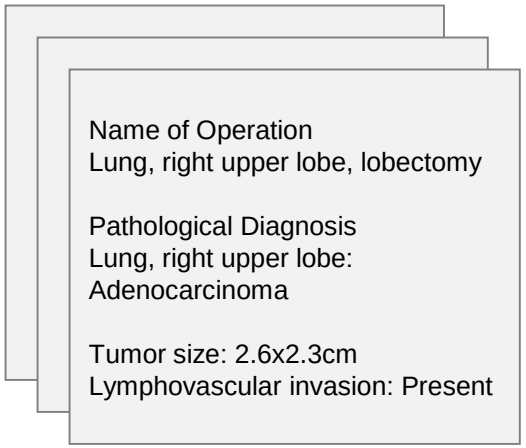
Generative LLM can extract cancer data

- Generative LLMs are being actively studied for their potential to extract cancer-related data

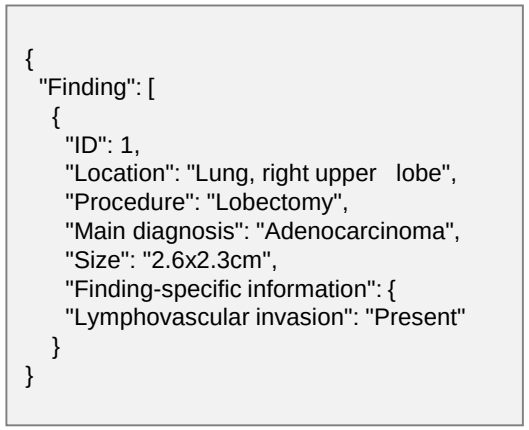


NLP pipeline

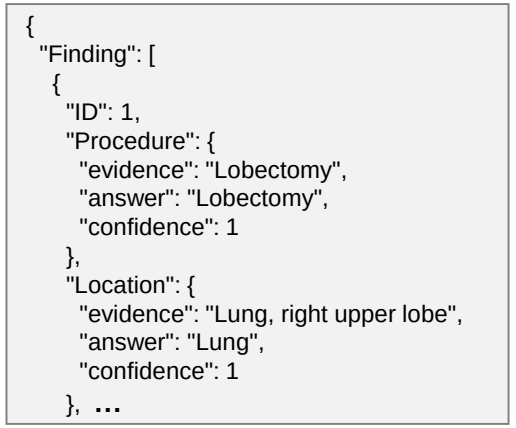
- Objective: To extract and structure cancer modifiers



Raw data



Parsing



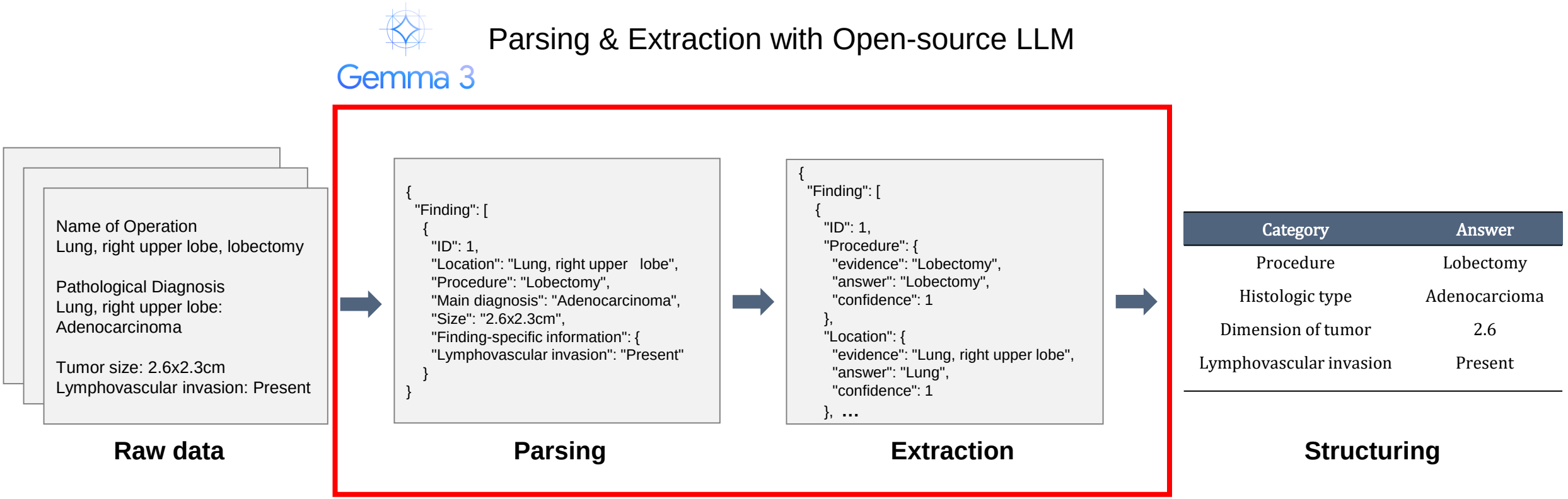
Extraction

Category	Answer
Procedure	Lobectomy
Histologic type	Adenocarcioma
Dimension of tumor	2.6
Lymphovascular invasion	Present

Structuring

NLP pipeline

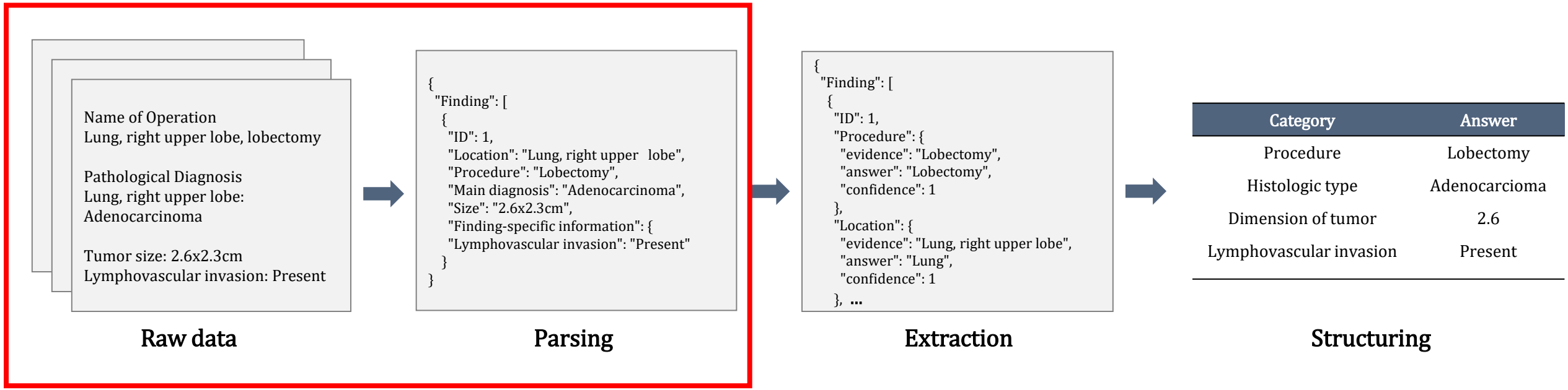
- Objective: To extract and structure cancer modifiers



Category	Answer
Procedure	Lobectomy
Histologic type	Adenocarcioma
Dimension of tumor	2.6
Lymphovascular invasion	Present

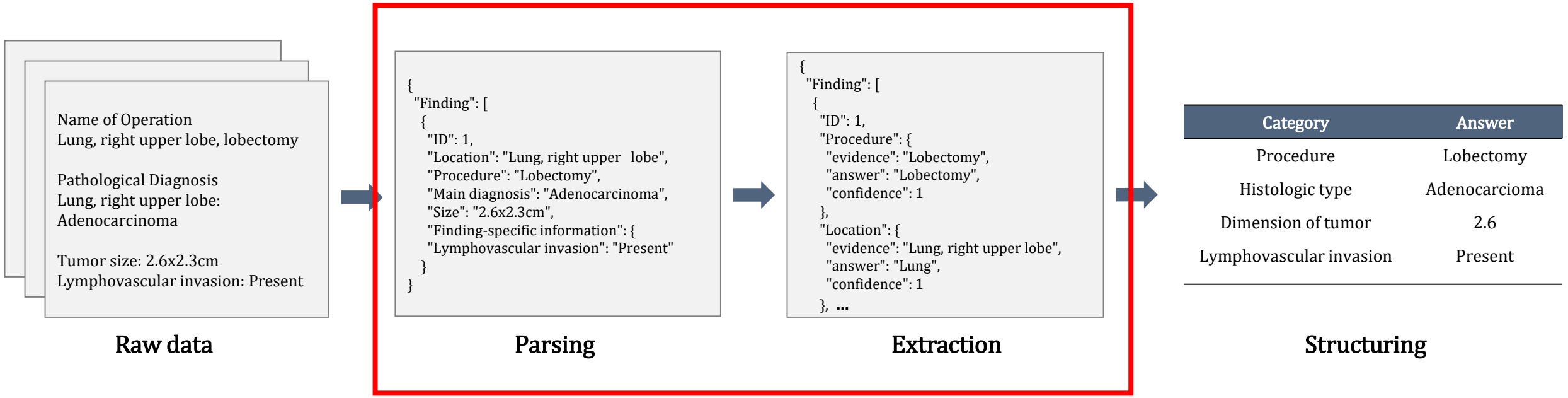
NLP pipeline

- **Step 1: Parsing**
  - Classify the entire clinical text into four domains: Finding, Lymph node, Biomarker, Others



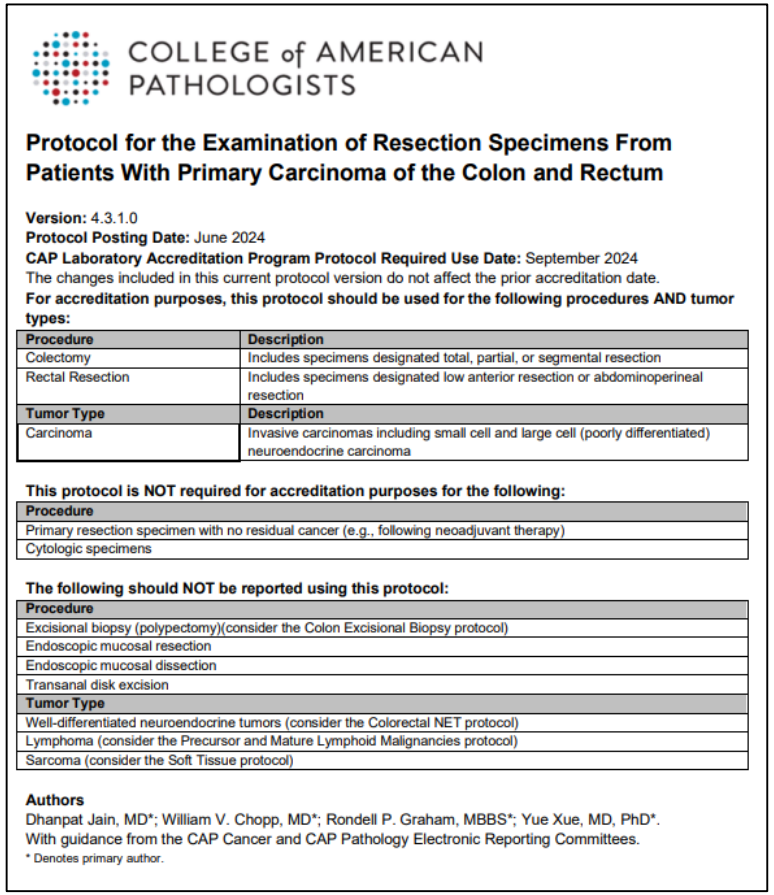
NLP pipeline

- **Step 2: Extraction**
  - Extract cancer-specific modifiers and convert into JSON format



What to extract?

- Target Cancer Types: Colorectal, Breast, and Lung Cancers



COLLEGE of AMERICAN PATHOLOGISTS

Protocol for the Examination of Resection Specimens From Patients With Primary Carcinoma of the Colon and Rectum

Version: 4.3.1.0
Protocol Posting Date: June 2024
CAP Laboratory Accreditation Program Protocol Required Use Date: September 2024
The changes included in this current protocol version do not affect the prior accreditation date.
For accreditation purposes, this protocol should be used for the following procedures AND tumor types:

Procedure	Description
Colectomy	Includes specimens designated total, partial, or segmental resection
Rectal Resection	Includes specimens designated low anterior resection or abdominoperineal resection
Tumor Type	**Description**
Carcinoma	Invasive carcinomas including small cell and large cell (poorly differentiated) neuroendocrine carcinoma

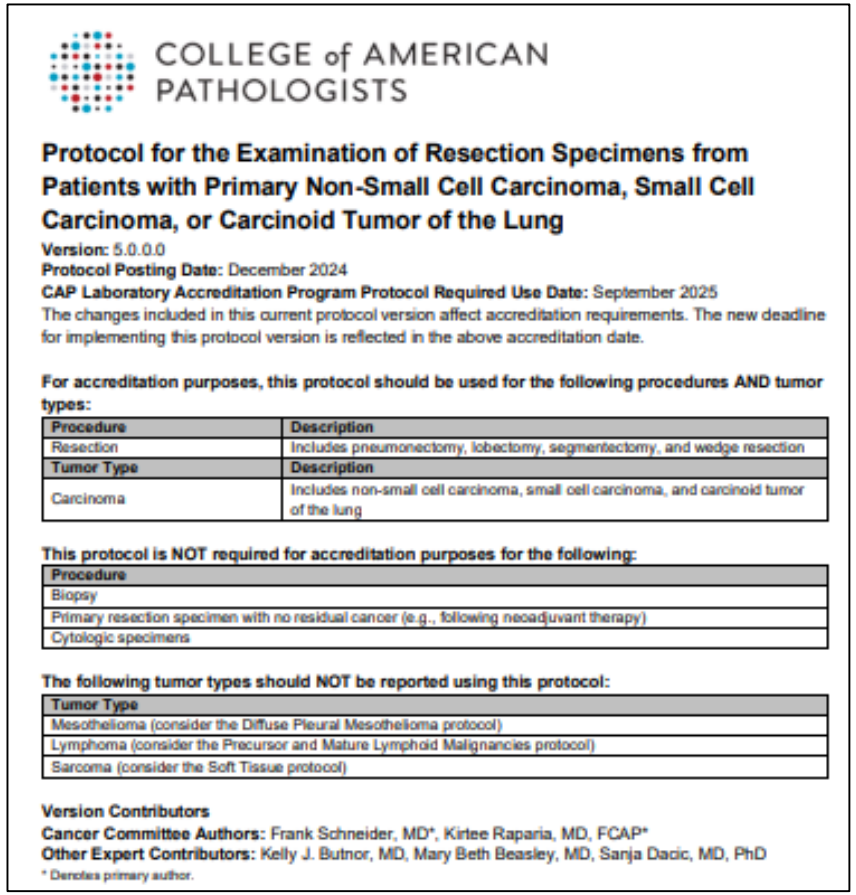
This protocol is NOT required for accreditation purposes for the following:

Procedure
Primary resection specimen with no residual cancer (e.g., following neoadjuvant therapy)
Cytologic specimens

The following should NOT be reported using this protocol:

Procedure
Excisional biopsy (polypectomy)(consider the Colon Excisional Biopsy protocol)
Endoscopic mucosal resection
Endoscopic mucosal dissection
Transanal disk excision
Tumor Type
Well-differentiated neuroendocrine tumors (consider the Colorectal NET protocol)
Lymphoma (consider the Precursor and Mature Lymphoid Malignancies protocol)
Sarcoma (consider the Soft Tissue protocol)

Authors
Dhanpat Jain, MD*; William V. Chopp, MD*; Rondell P. Graham, MBBS*; Yue Xue, MD, PhD*.
With guidance from the CAP Cancer and CAP Pathology Electronic Reporting Committees.
* Denotes primary author.



COLLEGE of AMERICAN PATHOLOGISTS

Protocol for the Examination of Resection Specimens from Patients with Primary Non-Small Cell Carcinoma, Small Cell Carcinoma, or Carcinoid Tumor of the Lung

Version: 5.0.0.0
Protocol Posting Date: December 2024
CAP Laboratory Accreditation Program Protocol Required Use Date: September 2025
The changes included in this current protocol version affect accreditation requirements. The new deadline for implementing this protocol version is reflected in the above accreditation date.

For accreditation purposes, this protocol should be used for the following procedures AND tumor types:

Procedure	Description
Resection	Includes pneumonectomy, lobectomy, segmentectomy, and wedge resection
Tumor Type	**Description**
Carcinoma	Includes non-small cell carcinoma, small cell carcinoma, and carcinoid tumor of the lung

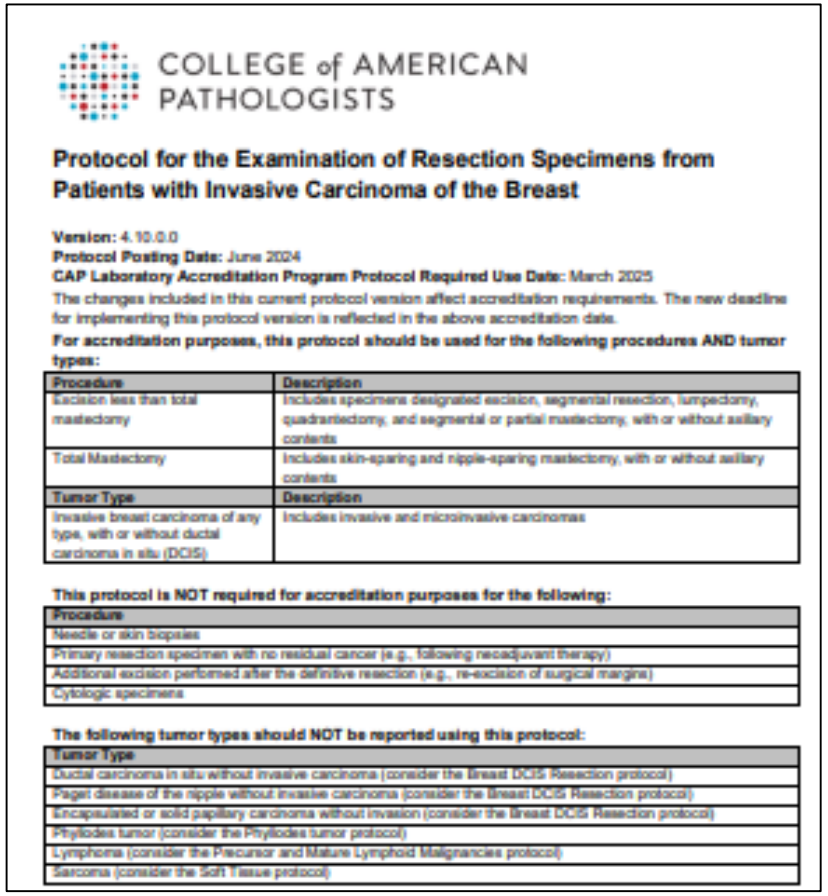
This protocol is NOT required for accreditation purposes for the following:

Procedure
Biopsy
Primary resection specimen with no residual cancer (e.g., following neoadjuvant therapy)
Cytologic specimens

The following tumor types should NOT be reported using this protocol:

Tumor Type
Mesothelioma (consider the Diffuse Pleural Mesothelioma protocol)
Lymphoma (consider the Precursor and Mature Lymphoid Malignancies protocol)
Sarcoma (consider the Soft Tissue protocol)

Version Contributors
Cancer Committee Authors: Frank Schneider, MD*, Kirtee Raparia, MD, FCAP*
Other Expert Contributors: Kelly J. Butnor, MD, Mary Beth Beasley, MD, Sanja Dacic, MD, PhD
* Denotes primary author.



COLLEGE of AMERICAN PATHOLOGISTS

Protocol for the Examination of Resection Specimens from Patients with Invasive Carcinoma of the Breast

Version: 4.10.0.0
Protocol Posting Date: June 2024
CAP Laboratory Accreditation Program Protocol Required Use Date: March 2025
The changes included in this current protocol version affect accreditation requirements. The new deadline for implementing this protocol version is reflected in the above accreditation date.
For accreditation purposes, this protocol should be used for the following procedures AND tumor types:

Procedure	Description
Excision less than total mastectomy	Includes specimens designated excision, segmental resection, lumpectomy, quadrantectomy, and segmental or partial mastectomy, with or without axillary contents
Total Mastectomy	Includes skin-sparing and nipple-sparing mastectomy, with or without axillary contents
Tumor Type	**Description**
Invasive breast carcinoma of any type, with or without ductal carcinoma in situ (DCIS)	Includes invasive and microinvasive carcinomas

This protocol is NOT required for accreditation purposes for the following:

Procedure
Needle or skin biopsies
Primary resection specimen with no residual cancer (e.g., following neoadjuvant therapy)
Additional excision performed after the definitive resection (e.g., re-excision of surgical margins)
Cytologic specimens

The following tumor types should NOT be reported using this protocol:

Tumor Type
Ductal carcinoma in situ without invasive carcinoma (consider the Breast DCIS Resection protocol)
Paget disease of the nipple without invasive carcinoma (consider the Breast DCIS Resection protocol)
Encapsulated or solid papillary carcinoma without invasion (consider the Breast DCIS Resection protocol)
Phyllodes tumor (consider the Phyllodes tumor protocol)
Lymphoma (consider the Precursor and Mature Lymphoid Malignancies protocol)
Sarcoma (consider the Soft Tissue protocol)

What to extract?

- Extracted variables for colorectal cancer patients

No.	Category
1	Procedure
2	Location
3	Histologic type
4	Histologic grade
5	Dimension
6	Depth of invasion
7	Resection Margin
8	Lymphovascular invasion
9	Perineural invasion
10	Lymph node stations
11	Number of lymph nodes with metastasis
12	Number of lymph nodes with examined
13	Tumor deposit

No.	Category
14	Tumor budding
15	Microsatellite instability
16	MLH1
17	MSH2
18	PMS2
19	MSH6
20	KRAS
21	NRAS
22	BRAF
23	EGFR
24	Mitotic count
25	Ki-67 index

NLP pipeline

- Location: ["Cecum", "Ileocecal valve", "Ascending colon", "Hepatic flexure", "Transverse colon", "Splenic flexure", "Descending colon", "Sigmoid colon", "Rectosigmoid junction", "Rectum", "Appendix", "Colon, not otherwise specified", "Other: <specify details>", "NA"]

- Histologic type: ["Adenocarcinoma, not otherwise specified","Mucinous adenocarcinoma","Signet-ring cell carcinoma","Medullary adenocarcinoma","Serrated adenocarcinoma","Micropapillary carcinoma","Adenoma-like adenocarcinoma","Adenosquamous carcinoma","Undifferentiated carcinoma","Carcinoma with sarcomatoid component","Neuroendocrine carcinoma, not otherwise specified","Large cell neuroendocrine carcinoma","Small cell neuroendocrine carcinoma", "Mixed neuroendocrine-non-neuroendocrine neoplasm","Neuroendocrine tumor","Tubular/Tubulovillous/Villous adenoma","Sessile serrated lesion","Traditional serrated adenoma","Other Adenoma/Polyp","Other: <specify details>","NA"]

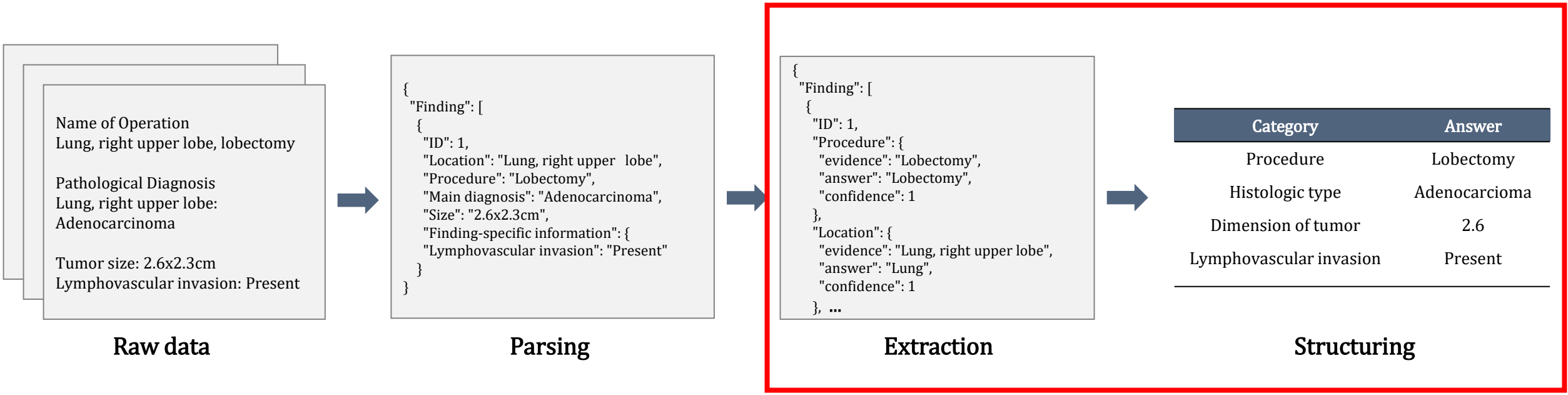
- Histologic grade: ["Well-differentiated","Moderately differentiated","Poorly differentiated","Undifferentiated","Low grade dysplasia","High grade dysplasia","Grade 1","Grade 2","Grade 3","NA"]

- Depth of invasion: ["Cannot be assessed","Intramucosal carcinoma (in situ)","Invades submucosa","Invades into muscularis propria","Invades through muscularis propria into the pericolonic or perirectal tissue/Invades subserosa","Invades through the visceral peritoneum","Directly invades or adheres to adjacent structures","NA"]

- Resection margin: ["Free from carcinoma", "Involved by carcinoma", "NA"]

NLP pipeline

- **Step 3: Structuring**
  - Convert data from JSON format to a structured tabular format



NLP pipeline

- Preliminary Result

Type	No. of reports	No. of category	Accuracy (%)	Precision (%)	Recall (%)
Colorectum	100	1,637	98.4	98.8	99.6
Breast	100	2,614	96.5	96.7	99.9
Lung	100	1,528	93.8	94.1	99.7

Mapping extracted values to OMOP CDM

Extracted data will be coverted to OMOP CDM

Category	Answer
Procedure	Lobectomy
Histologic type	Adenocarcioma
Dimension of tumor	2.6
Lymphovascular invasion	Present

Category	Answer	Concept ID
Procedure	Lobectomy	4054047
Histologic type	Adenocarcioma	37152526
Dimension of tumor	2.6	36768664
Lymphovascular invasion	Present	36768891

Structuring

Standardization

Mapping extracted values to OMOP CDM

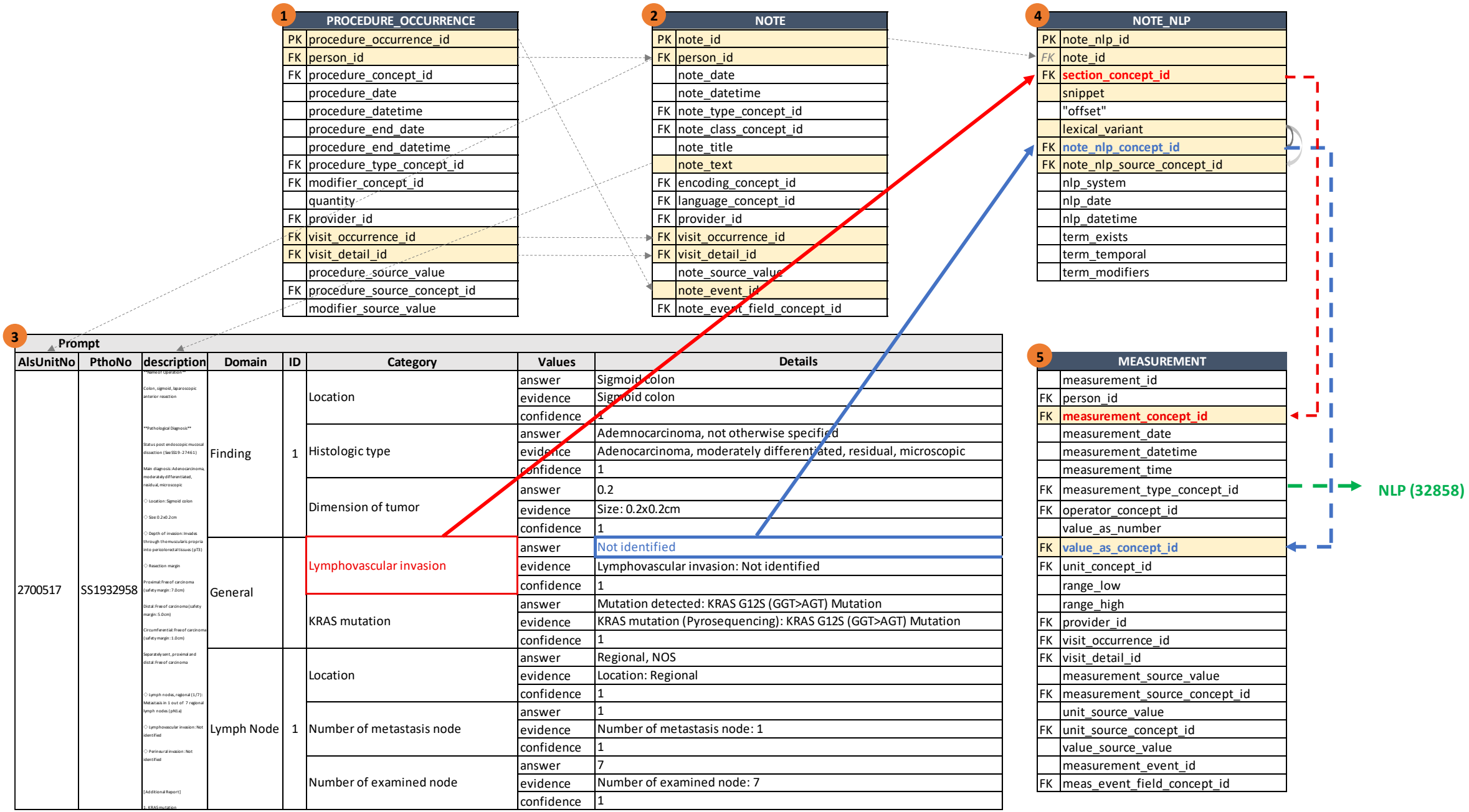
- Determine vocabulary for each cancer modifier

Category	Domain	Vocabulary	Category	Domain	Vocabulary
Diagnosis	Condition	ICD-O-3	Resection margin	Measurement	Cancer Modifier
Topography	Measurement	Cancer Modifier	Lymph node metastasis	Measurement	SNOMED
	Measurement	Cancer Modifier		Observation	SNOMED
Histology	Observation	SNOMED	Biomarker	Measurement	OMOP Genomic
Grade	Observation	SNOMED			LOINC
Dimension	Measurement	Cancer Modifier			NAACCR
Invasion	Measurement	Cancer Modifier	T stage	Measurement	Cancer Modifier
			N stage	Measurement	Cancer Modifier

Mapping extracted values to OMOP CDM

- Determine vocabulary for each cancer modifier

		measurement_concept_id			
Category	source_value_1	domain_id_	vocabulary_1	class_id_1	concept_id_
Dimension	<the greatest dimmension of tumor in centimenters (cm)>	Measurement	Cancer Modifier	Dimension	36768255
Depth of invasion	Intramucosal carcinoma (in situ)	Measurement	Cancer Modifier	Histopattern	36769623
Depth of invasion	Invades submucosa	Measurement	Cancer Modifier	Extension/Invasion	36768886
Depth of invasion	Invades into muscularis propria	Measurement	Cancer Modifier	Extension/Invasion	36769076
Depth of invasion	Invades through muscularis propria into the pericolonic or perirectal tissue	Measurement	Cancer Modifier	Extension/Invasion	36769648
Depth of invasion	Invades through the visceral peritoneum	Measurement	Cancer Modifier	Extension/Invasion	36769563
Depth of invasion	Directly invades or adheres to adjacent structures	Measurement	Cancer Modifier	Extension/Invasion	36770430
Resection margin	Free from carcinoma	Measurement	Cancer Modifier	Margin	36770153
Resection margin	Involved by carcinoma	Measurement	Cancer Modifier	Margin	36768316
Perineural invasion	Present	Measurement	Cancer Modifier	Extension/Invasion	36768846


1
PROCEDURE_OCCURRENCE
PK procedure_occurrence_id
FK person_id
FK procedure_concept_id
procedure_date
procedure_datetime
procedure_end_date
procedure_end_datetime
FK procedure_type_concept_id
FK modifier_concept_id
quantity
FK provider_id
FK visit_occurrence_id
FK visit_detail_id
procedure_source_value
FK procedure_source_concept_id
modifier_source_value
2
NOTE
PK note_id
FK person_id
note_date
note_datetime
FK note_type_concept_id
FK note_class_concept_id
note_title
note_text
FK encoding_concept_id
FK language_concept_id
FK provider_id
FK visit_occurrence_id
FK visit_detail_id
note_source_value
note_event_id
FK note_event_field_concept_id
4
NOTE_NLP
PK note_nlp_id
FK note_id
FK section_concept_id
snippet
"offset"
lexical_variant
FK note_nlp_concept_id
FK note_nlp_source_concept_id
nlp_system
nlp_date
nlp_datetime
term_exists
term_temporal
term_modifiers
3
Prompt
AlsUnitNo
PthoNo
description
Domain
ID
Category
Values
Details
2700517
SS1932958
Finding
1
Location
answer
Sigmoid colon
evidence
Sigmoid colon
confidence
1
Histologic type
answer
Ademnocarcinoma, not otherwise specified
evidence
Adenocarcinoma, moderately differentiated, residual, microscopic
confidence
1
Dimension of tumor
answer
0.2
evidence
Size: 0.2x0.2cm
confidence
1
General
Lymphovascular invasion
answer
Not identified
evidence
Lymphovascular invasion: Not identified
confidence
1
KRAS mutation
answer
Mutation detected: KRAS G12S (GGT>AGT) Mutation
evidence
KRAS mutation (Pyrosequencing): KRAS G12S (GGT>AGT) Mutation
confidence
1
Lymph Node
1
Location
answer
Regional, NOS
evidence
Location: Regional
confidence
1
Number of metastasis node
answer
1
evidence
Number of metastasis node: 1
confidence
1
Number of examined node
answer
7
evidence
Number of examined node: 7
confidence
1
5
MEASUREMENT
measurement_id
FK person_id
FK measurement_concept_id
measurement_date
measurement_datetime
measurement_time
FK measurement_type_concept_id
FK operator_concept_id
value_as_number
FK value_as_concept_id
FK unit_concept_id
range_low
range_high
FK provider_id
FK visit_occurrence_id
FK visit_detail_id
measurement_source_value
FK measurement_source_concept_id
unit_source_value
FK unit_source_concept_id
value_source_value
measurement_event_id
FK meas_event_field_concept_id
NLP (32858)

ETL Case

조직병리진단 결과

병리번호	SS2280154			
의뢰과/임상의사	대장항문외과/	병동	퇴실	
검체채취일	2022-12-13 18:51			
검사물 접수일	2022-12-14 14:12			
보고일	2022-12-27			

[Name Of Operation]

Ascending colon and abdominal wall, laparoscopic right hemicolectomy and en-bloc excision

[Pathological Diagnosis]

Adenocarcinoma, moderately differentiated
◇ Location: Ascending colon and cecum
◇ Size: 5.5 x 5.5 cm
◇ Depth of invasion: Tumor directly invades adjacent structures (abdominal wall) (pT4b)
◇ Lymphovascular invasion: Not identified
◇ Extramural large vessel (venous) invasion: Not identified
◇ Perineural invasion: Not identified
◇ Associated finding: 1. Tubulovillous adenoma, low grade 2. Extracellular mucin production
◇ Resection margin
- Proximal: Free from carcinoma (safety margin: 20.0 cm)
- Distal: Free from carcinoma (safety margin: 19.0 cm)
- Abdominal wall: Free from carcinoma (safety margin: 0.3 cm)

◇ Lymph node: Metastasis in 3 out of 21 regional lymph nodes (3/21) (pN1b)
- Tumor deposit: not identified

[Additional Report]
1. KRAS mutation (Pyrosequencing): KRAS G12D (GGT>GAT) Mutation
2. NRAS mutation (Pyrosequencing): Wild
3. Microsatellite instability : MSS
- NR-21: MSI(-)
- NR-24: MSI(-)
- NR-27: MSI(-)
- BAT-25: MSI(-)
- BAT-26: MSI(-)

[TNM Stage] pT4b/pN1b

ETL Case

조직병리진단 결과

병리번호	SS2280154			
의뢰과/임상의사	대장항문외과/	병동	퇴실	
검체채취일	2022-12-13 18:51			
검사물 접수일	2022-12-14 14:12			
보고일	2022-12-27			

[Name Of Operation]

Ascending colon and abdominal wall, laparoscopic right hemicolectomy and en-bloc excision

[Pathological Diagnosis]

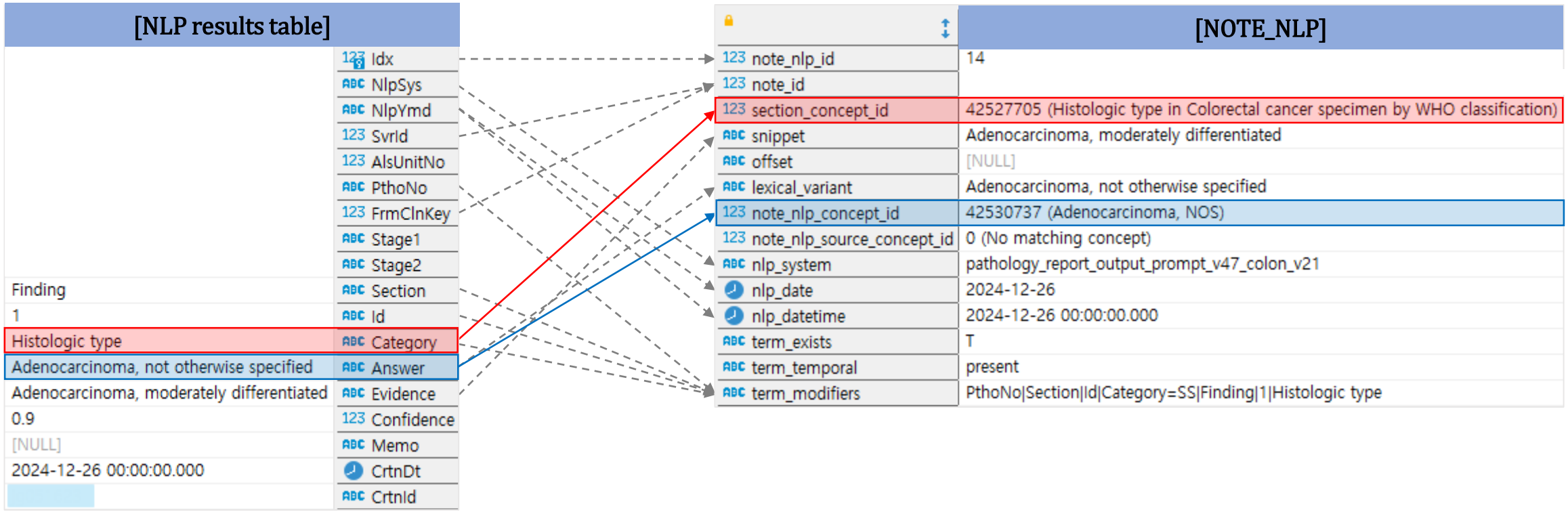
Adenocarcinoma, moderately differentiated
◇ Location: Ascending colon and cecum
◇ Size: 5.5 x 5.5 cm
◇ Depth of invasion: Tumor directly invades adjacent structures (abdominal wall) (pT4b)
◇ Lymphovascular invasion: Not identified
◇ Extramural large vessel (venous) invasion: Not identified
◇ Perineural invasion: Not identified
◇ Associated finding: 1. Tubulovillous adenoma, low grade 2. Extracellular mucin production
◇ Resection margin
- Proximal: Free from carcinoma (safety margin: 20.0 cm)
- Distal: Free from carcinoma (safety margin: 19.0 cm)
- Abdominal wall: Free from carcinoma (safety margin: 0.3 cm)

◇ Lymph node: Metastasis in 3 out of 21 regional lymph nodes (3/21) (pN1b)
- Tumor deposit: not identified

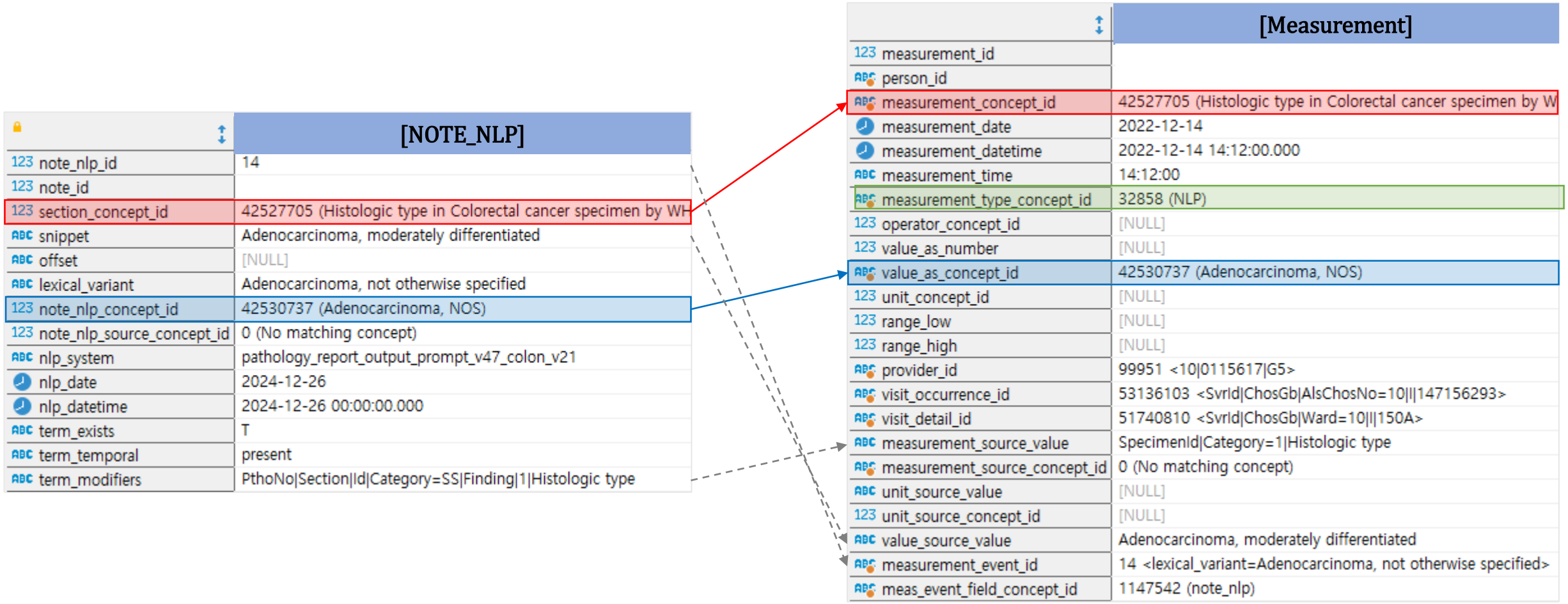
[Additional Report]
1. KRAS mutation (Pyrosequencing): KRAS G12D (GGT>GAT) Mutation
2. NRAS mutation (Pyrosequencing): Wild
3. Microsatellite instability : MSS
- NR-21: MSI(-)
- NR-24: MSI(-)
- NR-27: MSI(-)
- BAT-25: MSI(-)
- BAT-26: MSI(-)

[TNM Stage] pT4b/pN1b

Progress


[NLP results table]
Idx
NlpSys
NlpYmd
SvrId
AlsUnitNo
PthoNo
FrmClnKey
Stage1
Stage2
Finding
Section
1
Id
Histologic type
Category
Adenocarcinoma, not otherwise specified
Answer
Adenocarcinoma, moderately differentiated
Evidence
0.9
Confidence
[NULL]
Memo
2024-12-26 00:00:00.000
CrtnDt
CrtnId
[NOTE_NLP]
note_nlp_id
14
note_id
section_concept_id
42527705 (Histologic type in Colorectal cancer specimen by WHO classification)
snippet
Adenocarcinoma, moderately differentiated
offset
[NULL]
lexical_variant
Adenocarcinoma, not otherwise specified
note_nlp_concept_id
42530737 (Adenocarcinoma, NOS)
note_nlp_source_concept_id
0 (No matching concept)
nlp_system
pathology_report_output_prompt_v47_colon_v21
nlp_date
2024-12-26
nlp_datetime
2024-12-26 00:00:00.000
term_exists
T
term_temporal
present
term_modifiers
PthoNo|Section|Id|Category=SS|Finding|1|Histologic type

ETL Case


[NOTE_NLP]
note_nlp_id 14
note_id
section_concept_id 42527705 (Histologic type in Colorectal cancer specimen by WH
snippet Adenocarcinoma, moderately differentiated
offset [NULL]
lexical_variant Adenocarcinoma, not otherwise specified
note_nlp_concept_id 42530737 (Adenocarcinoma, NOS)
note_nlp_source_concept_id 0 (No matching concept)
nlp_system pathology_report_output_prompt_v47_colon_v21
nlp_date 2024-12-26
nlp_datetime 2024-12-26 00:00:00.000
term_exists T
term_temporal present
term_modifiers PthoNo|Section|Id|Category=SS|Finding|1|Histologic type
[Measurement]
measurement_id
person_id
measurement_concept_id 42527705 (Histologic type in Colorectal cancer specimen by W
measurement_date 2022-12-14
measurement_datetime 2022-12-14 14:12:00.000
measurement_time 14:12:00
measurement_type_concept_id 32858 (NLP)
operator_concept_id [NULL]
value_as_number [NULL]
value_as_concept_id 42530737 (Adenocarcinoma, NOS)
unit_concept_id [NULL]
range_low [NULL]
range_high [NULL]
provider_id 99951 <10|0115617|G5>
visit_occurrence_id 53136103 <SvrId|ChosGb|AlsChosNo=10||147156293>
visit_detail_id 51740810 <SvrId|ChosGb|Ward=10||150A>
measurement_source_value SpecimenId|Category=1|Histologic type
measurement_source_concept_id 0 (No matching concept)
unit_source_value [NULL]
unit_source_concept_id [NULL]
value_source_value Adenocarcinoma, moderately differentiated
measurement_event_id 14 <lexical_variant=Adenocarcinoma, not otherwise specified>
meas_event_field_concept_id 1147542 (note_nlp)

ETL Case

[NOTE_NLP]	
note_nlp_id	11
note_id	
section_concept_id	36768255 (Largest Dimension of Tumor)
snippet	Size: 5.5 x 5.5 cm
offset	[NULL]
lexical_variant	5.5
note_nlp_concept_id	[NULL]
note_nlp_source_concept_id	0 (Largest Dimension of Tumor)
nlp_system	pathology_report_output_prompt_v47_colon_v21
nlp_date	2024-12-26
nlp_datetime	2024-12-26 00:00:00.000
term_exists	T
term_temporal	present
term_modifiers	PthoNo|Section|Id|Category=SS|Finding|1|Dimension of tumor

[Measurement]	
measurement_id	
person_id	
measurement_concept_id	36768255 (Largest Dimension of Tumor)
measurement_date	2022-12-14
measurement_datetime	2022-12-14 14:12:00.000
measurement_time	14:12:00
measurement_type_concept_id	32858 (NLP)
operator_concept_id	[NULL]
value_as_number	5.5
value_as_concept_id	[NULL]
unit_concept_id	[NULL]
range_low	[NULL]
range_high	[NULL]
provider_id	99951 <10|0115617|G5>
visit_occurrence_id	53136103 <SvrId|ChosGb|AlsChosNo=10||147156293>
visit_detail_id	51740810 <SvrId|ChosGb|Ward=10||150A>
measurement_source_value	SpecimenId|Category=1|Dimension of tumor
measurement_source_concept_id	0 (No matching concept)
unit_source_value	[NULL]
unit_source_concept_id	[NULL]
value_source_value	Size: 5.5 x 5.5 cm
measurement_event_id	11 <lexical_variant=5.5>
meas_event_field_concept_id	1147542 (note_nlp)

Take-Home Message

- Oncology research requires a more granular and structured data
- The OMOP CDM extension introduces detailed models for diagnosis, treatment, and clinical episodes
- Generative LLMs enable scalable extraction of cancer data from unstructured clinical text
- Together, these innovations pave the way for efficient, reproducible, and collaborative cancer research

YONSEI UNIVERSITY
COLLEGE OF MEDICINE