THE FUTURE OF CANCER DATA: UNLOCKING INSIGHTS WITH PATHOLOGY REPORTING

Data to Discovery: Novel Applications of Pathology Data

Michael Gurley, Brigette Rabel,
Ross Simpson, MD,
Peter Paul Yu, MD, FACP, FASCO

OCTOBER 6 | 2:15–3:00 PM CT

COLLEGE of AMERICAN PATHOLOGISTS
Laboratory Quality Solutions

CAP23 | CHICAGO

#PATHDATA

Pathology Synoptic Data

Local Usages

Ross W. Simpson, MD

Park Nicollet / Methodist Hospital

St. Louis Park, MN

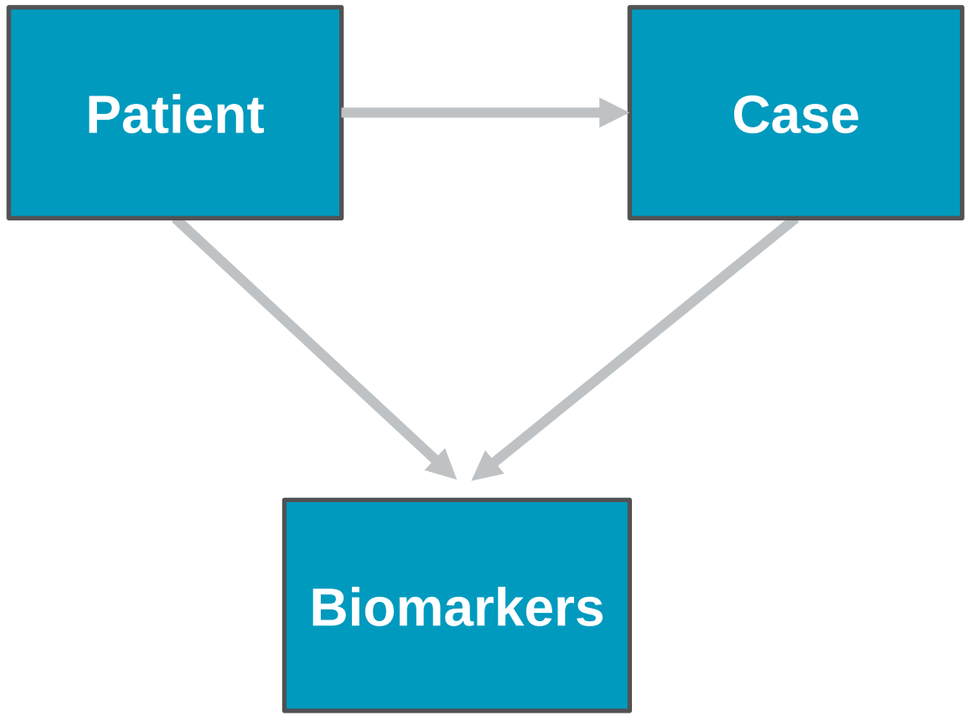
Disclosures

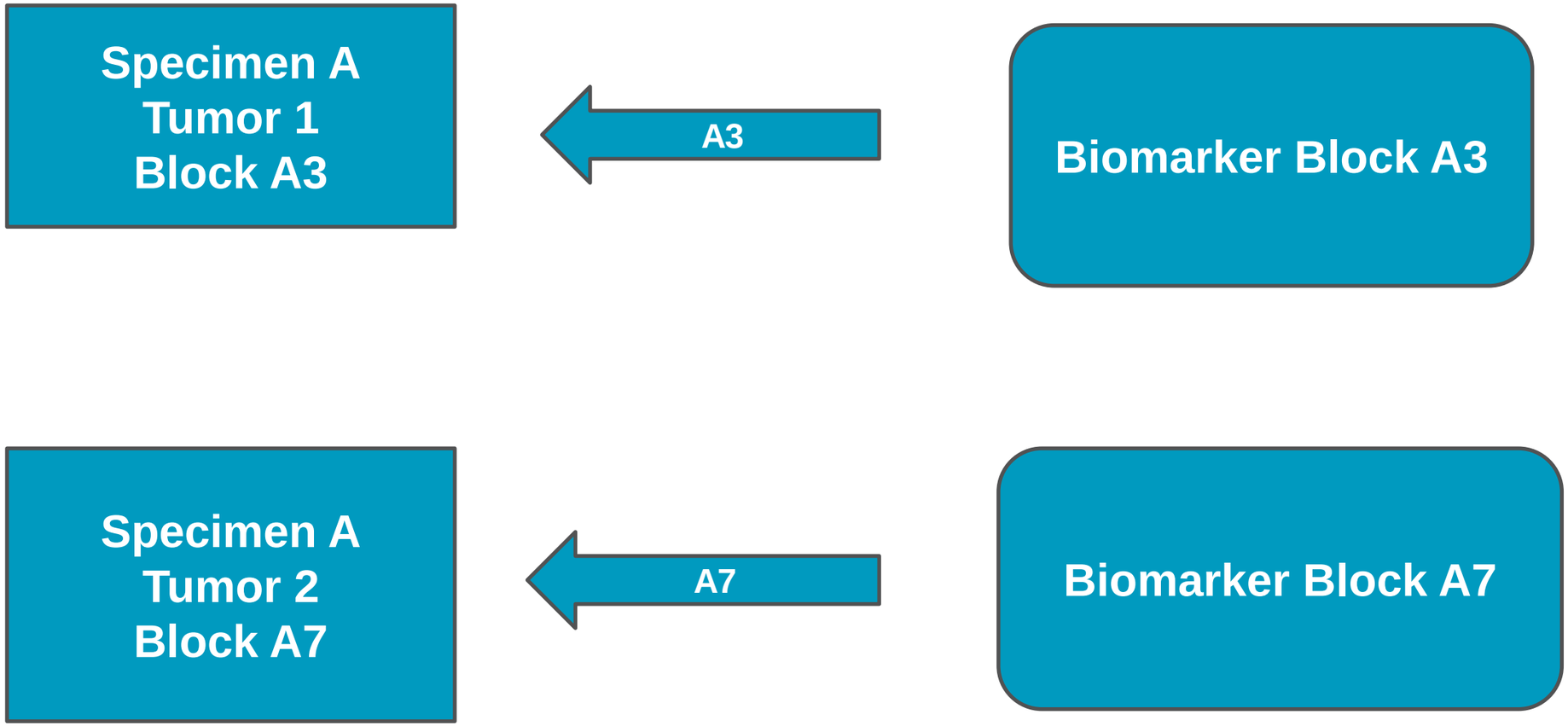
- **No relevant disclosures**

Validation/Curation

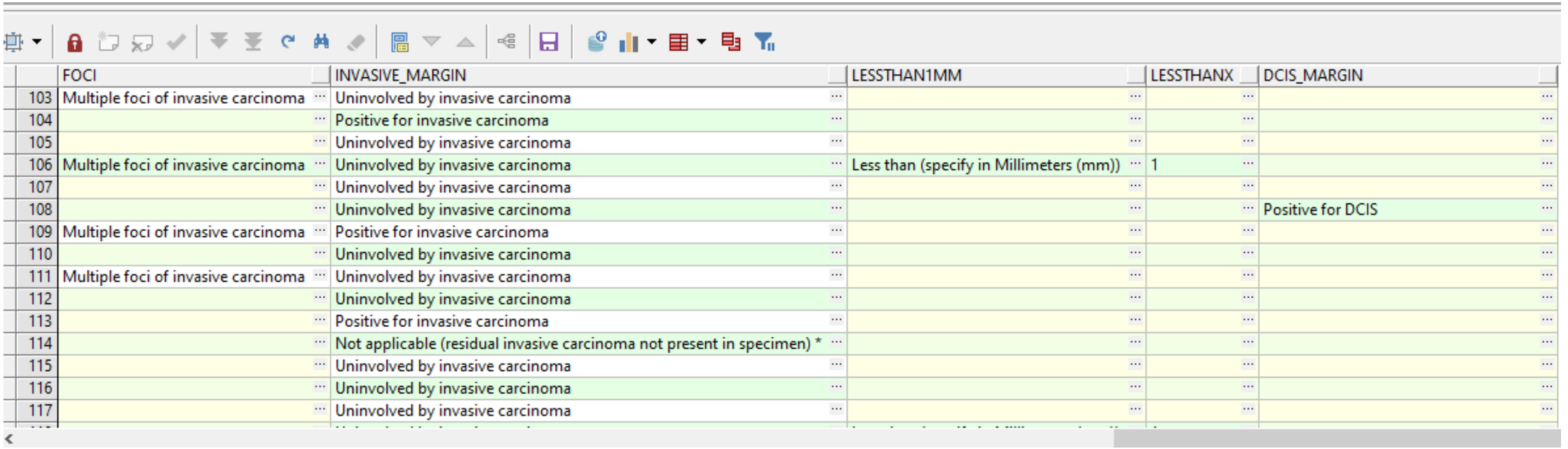
- Are the forms being filled out consistently?
- Is the dataset complete?
- Is the data accurate – especially in complicated cases with multiple tumors?
- Are Biomarkers correctly mapped?

SQL





Example: Breast Cancer Margins



	FOCI	INVASIVE_MARGIN	LESSTHAN1MM	LESSTHANX	DCIS_MARGIN
103	Multiple foci of invasive carcinoma	Uninvolved by invasive carcinoma			
104		Positive for invasive carcinoma			
105		Uninvolved by invasive carcinoma			
106	Multiple foci of invasive carcinoma	Uninvolved by invasive carcinoma	Less than (specify in Millimeters (mm))	1	
107		Uninvolved by invasive carcinoma			
108		Uninvolved by invasive carcinoma			Positive for DCIS
109	Multiple foci of invasive carcinoma	Positive for invasive carcinoma			
110		Uninvolved by invasive carcinoma			
111	Multiple foci of invasive carcinoma	Uninvolved by invasive carcinoma			
112		Uninvolved by invasive carcinoma			
113		Positive for invasive carcinoma			
114		Not applicable (residual invasive carcinoma not present in specimen) *			
115		Uninvolved by invasive carcinoma			
116		Uninvolved by invasive carcinoma			
117		Uninvolved by invasive carcinoma			

Variation by: Surgeon, PA (grossing), Pathologist, tumor type, tumor size, tumor grade

Sample Size

CAP / AP-LIS Vendor Goal

- **To make use of the data easy**

Provincial Health Services Authority

Provincial AP Synoptic Reporting Program

Brigette Rabel

Provincial AP Strategic Lead and Synoptic Reporting Program Coordinator

Provincial Health Services Authority

Disclosure Statement

I have no financial disclosure or conflicts of interest with the material in this presentation.

Provincial Health Services Authority

Synoptic Reporting Pathologist Dashboard 2022

Legend	
▲	Observed value may deviate from the target
●	As expected
	No target or < 5 cases

Tissue	Metric Description	Suggested/ Published Target	P160	Site	Health Authority	Province
Breast	Number of checklists submitted		30	251	693	3438
	% Cases with invasive ductal carcinoma	70-90%	● 70%	● 71%	● 74%	● 72%
	% Cases with invasive lobular carcinoma	10-15%	● 10%	● 15%	● 11%	● 12%
	% Cases where biomarkers are resulted as performed	> 95%	● 100%	● 99%	● 99%	● 99%
Colorectal	Number of checklists submitted		9	132	334	1715
	% Colon resections with at least 12 lymph nodes examined	> 90%	● 100%	▲ 89%	● 91%	● 93%
	% Cases with serosal penetration (pT4a and pT4b)	> 20%	▲ 0%	▲ 20%	▲ 20%	● 21%
	% Cases with lymphovascular invasion	> 30%	● 33%	● 49%	● 33%	● 33%
Endometrium	Number of checklists submitted		1	26	94	739
	% Cases with lymphovascular invasion		0%	15%	17%	23%
Lung	Number of checklists submitted		31	154	154	581
	% Cases with visceral pleural invasion		13%	24%	24%	24%
Prostate	Number of checklists submitted		4	33	152	1196
	% Cases with extra prostatic extension		0%	36%	45%	50%
Quality	% all tissues 14 day TAT met	> 85%	▲ 60%	▲ 49%	▲ 61%	▲ 72%
	% all tissues 6 working day TAT met	> 85%	▲ 13%	▲ 16%	▲ 34%	▲ 54%

Fictitious Data for Demonstration Purposes

Provincial Health Services Authority

Synoptic Reporting Surgeon Dashboard 2022

Legend	
▲	Observed value may deviate from the target
●	As expected
	No target or < 5 cases

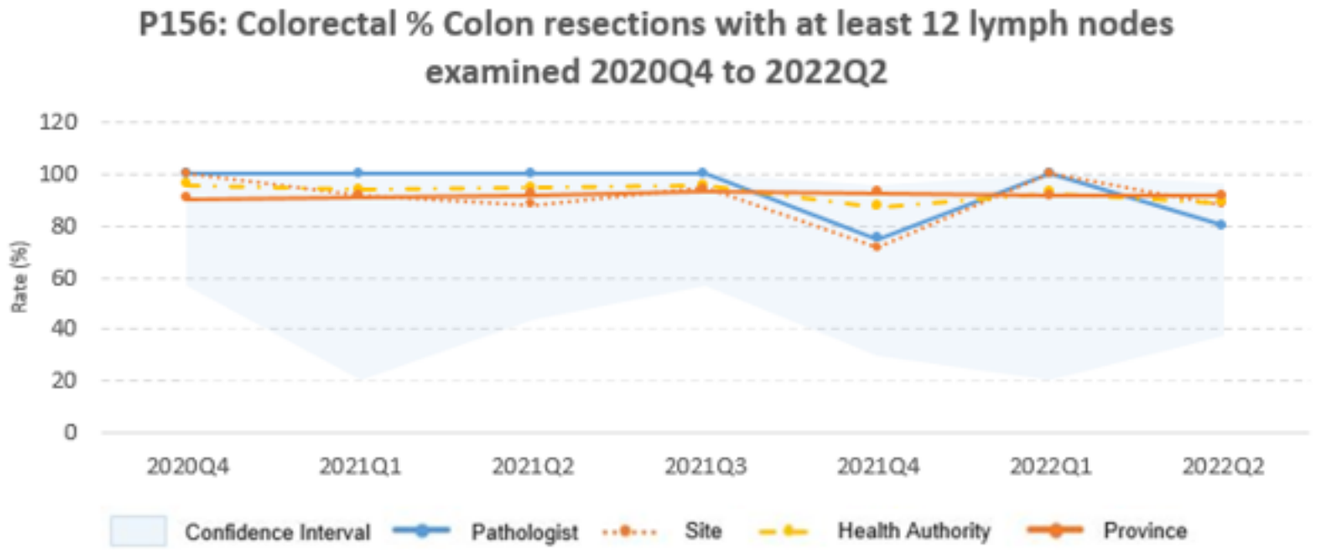
Tissue	Metric Description	Target	G7403	Hospital(s)	Health Authority	Province
Breast	Number of checklists submitted		18	61	823	3438
	% Cases with margins positive for invasive carcinoma (excisions)		20%	9%	11%	15%
	% Cases with margins positive for in situ carcinoma (excisions)		10%	9%	6%	7%
Colorectal	Number of checklists submitted		8	20	408	1715
	% Checklists - colon		75%	90%	79%	77%
	% Checklists - rectal		25%	10%	21%	23%
	% Colon resections with at least 12 lymph nodes examined	> 90%	▲ 83%	● 94%	● 96%	● 93%
	% Rectal resections (with y) with at least 12 lymph nodes examined		50%	50%	95%	86%
	% Cases with macroscopic intactness of mesorectum incomplete	< 10 %	0%	0%	● 3%	● 8%
	% Cases with macroscopic intactness of mesorectum complete or near complete	> 90 %	100%	100%	● 96%	▲ 90%
	% Cases with radial/mesenteric margin positive for tumour		0%	5%	5%	5%
Prostate	Number of checklists submitted		1	1	248	1196
	% cases pT2		100%	100%	51%	47%
	% cases pT3a		0%	0%	36%	35%
	% cases pT3b		0%	0%	13%	18%
	% pT2 cases with positive margins	< 20 %	0%	0%	● 17%	▲ 24%
	% Cases with lymph nodes submitted or found		0%	0%	96%	91%

Fictitious Data for Demonstration Purposes

Synoptic Reporting Dashboard Deep Dive

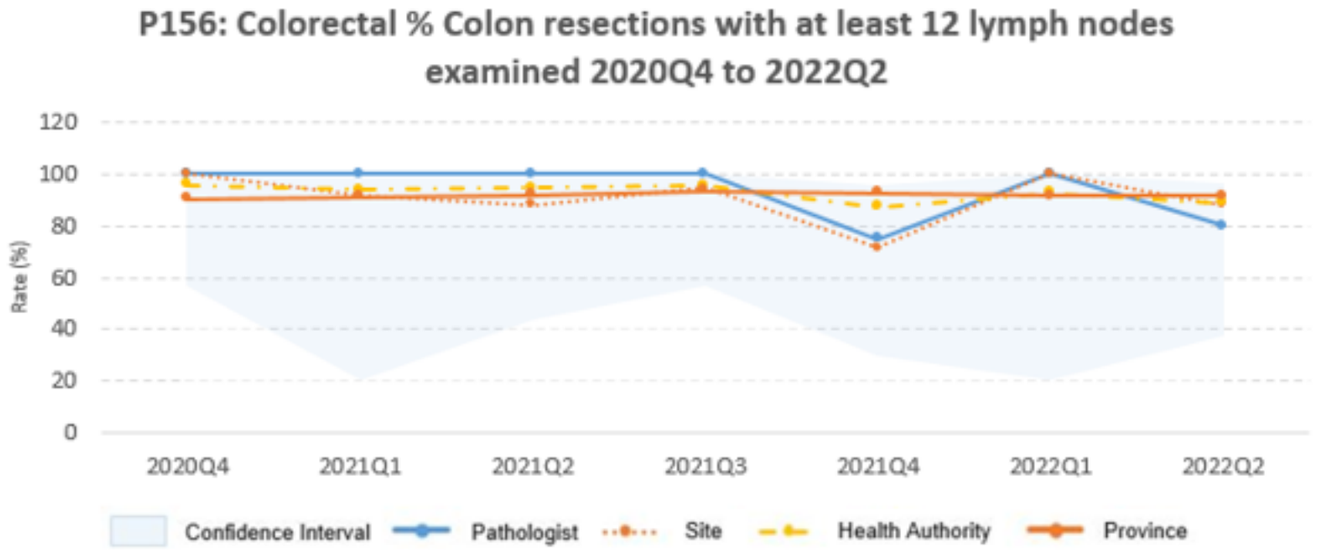
Provincial Health Services Authority

Choose a tissue	Choose your ID	Choose one metric
Colorectal	P156	% Cases with macroscopic intactness of mesorectum in...
Breast	P157	% Cases with macroscopic intactness of mesorectum n...
Endometrium	P158	% Cases with radial/mesenteric margin positive for tu...
Lung	P160	% Cases with serosal penetration (pT4a and pT4b)
Prostate	P162	% Colon resections with at least 12 lymph nodes exami...
Quality	P163	% Rectal resections (with y) with at least 12 lymph nod...
	P164	% Rectal resections (without y) with at least 12 lymph n...
	P168	Number of checklists submitted



Synoptic Reporting Dashboard Deep Dive

Choose a tissue	Choose your ID	Choose one metric
Colorectal	P156	% Cases with macroscopic intactness of mesorectum in...
Breast	P157	% Cases with macroscopic intactness of mesorectum n...
Endometrium	P158	% Cases with radial/mesenteric margin positive for tu...
Lung	P160	% Cases with serosal penetration (pT4a and pT4b)
Prostate	P162	% Colon resections with at least 12 lymph nodes exami...
Quality	P163	% Rectal resections (with y) with at least 12 lymph nod...
	P164	% Rectal resections (without y) with at least 12 lymph n...
	P168	Number of checklists submitted



What we are measuring

We are counting the number of Colon and Rectum Resection checklists submitted where the number of lymph nodes examined is 12 or more. This number is divided by the number of Colon and Rectum Resection checklists with lymph nodes received.
- Excludes cases where "Rectum" is listed as a primary tumour site

Why it matters

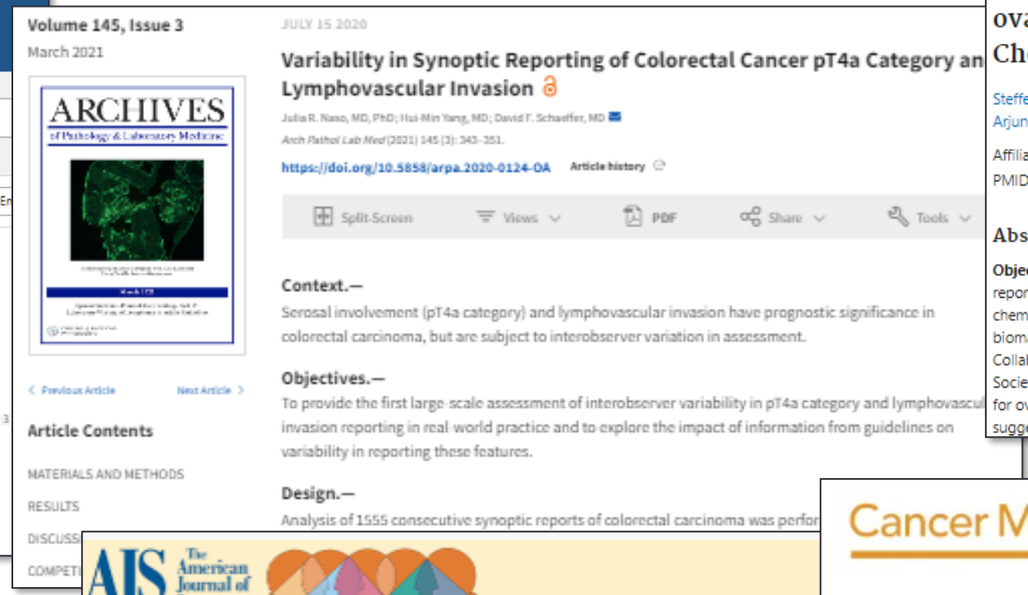
The accuracy and predictive value of stage II assignment are directly proportional to the thoroughness of the surgical technique in removing all regional nodes and the pathologic examination of the resection specimen in identifying and harvesting all regional lymph nodes for microscopic assessment. The National Quality Forum lists the presence of at least 12 lymph nodes in a surgical resection among the key quality measures for colon cancer care. The likelihood of detecting metastasis increases with the number of lymph nodes examined; hence 12 lymph nodes should be considered the minimum target, but all possible lymph nodes should be retrieved and examined. The clinical outcome is linked to lymph node harvest in stage II disease, indicating a positive effect of optimal mesenteric resection by the surgeon, optimal lymph node harvest from the resection specimen by the pathologist, or both.

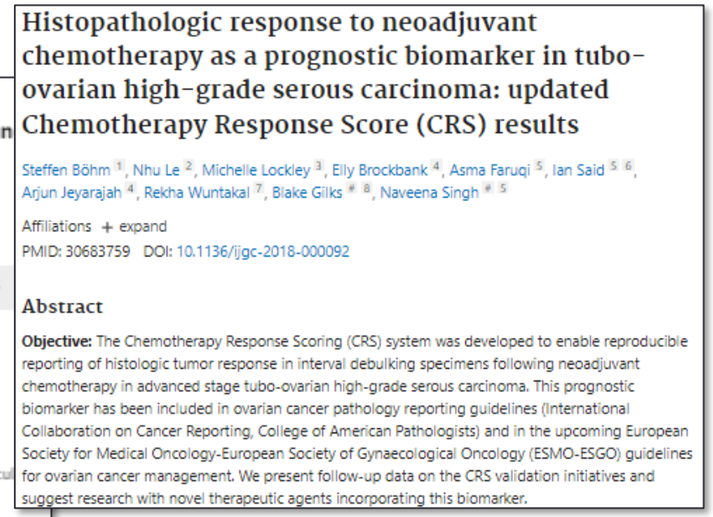
What you can do

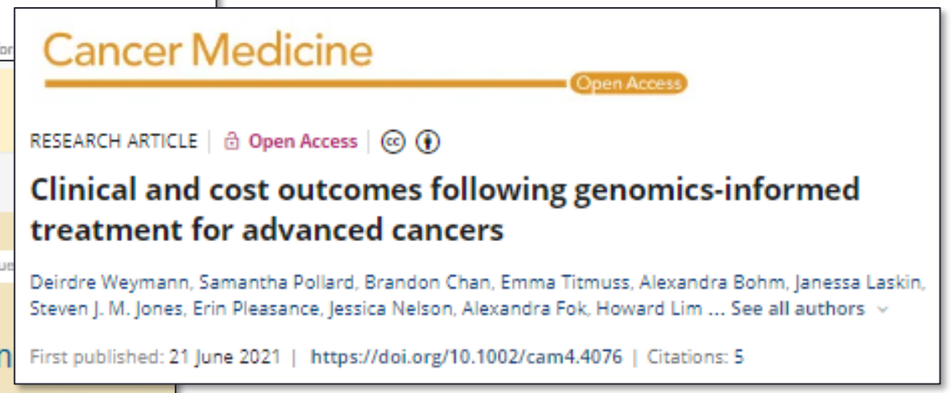
The number of lymph nodes recovered from a resection specimen is dependent on several factors. Surgical technique, surgery volume, and patient factors (e.g., age and anatomic variation) alter the actual number of nodes in a resection specimen, but the diligence and skill of the pathologist in identifying and harvesting lymph nodes in the resection specimen also are major factors. Lymph nodes may be more difficult to identify in specimens from patients who are obese or elderly, or after neoadjuvant therapy. Because it has been shown that nodal metastasis in colorectal cancer is often found in small lymph nodes (<5 mm in diameter), diligent search for lymph nodes is required on gross examination of resection specimens. If fewer than 12 lymph nodes are found, re-examining the specimen for additional lymph nodes, with or without visual enhancement techniques, should be considered. The pathology report should clearly state the total number of lymph nodes examined and the total number involved by metastases. Data are

Synoptic Reporting Data is used to Inform Research Projects and Quality Initiatives

Provincial Health Services Authority







BC CANCER
Provincial Health Services Authority

GENOVA (Genetic testing of ENdometrial and OVArian cancer), a new initiative from BC Cancer's Gynecologic Oncology Tumour Group and Hereditary Cancer Program (HCP)

ARCHIVES
of Pathology & Laboratory Medicine

Using Pathology Synoptic Reporting Data to Create Individual Dashboards for Pathologists and Surgeons

Gurpal Bisra, MMOR, MSc, BASc; Brigette Rabel, MLT; Nick van der Westhuizen, MB, FRCPC

• *Context.*—Electronic synoptic pathology reporting using xPert from mTuitive is available to all pathologists in British Columbia, Canada. Comparative feedback reports for pathologists and surgeons were created by using the synoptic reporting software.

Objective.—To use data stored in a single central data repository to provide nonpunitive confidential comparative feedback reports (dashboards) to individual pathologists and surgeons for reflection on their practice and to use aggregate data for quality improvement initiatives.

Design.—Integration of mTuitive middleware in 5 different laboratory information systems to have 1 software solution (xPert) sending discrete data elements to the central data repository. Microsoft Office products were used to build comparative feedback reports and made the infrastructure sustainable. Two different types of reports were developed: individual confidential feedback reports (dashboards) and aggregated data reports.

Results.—Pathologists have access to an individual confidential live feedback report for the 5 major cancer sites. Surgeons get an annual confidential emailed PDF report. Several quality improvement initiatives were identified from the aggregate data.

Conclusions.—We present 2 novel dashboards: a live pathologist dashboard and a static surgeon dashboard. Individual confidential dashboards incentivize use of nonmandated electronic synoptic pathology reporting tools and have increased adoption rates. Use of dashboards has also led to discussions about how patient care may be improved.

(*Arch Pathol Lab Med.* doi: 10.5858/arpa.2021-0542-OA)

Cancer is the leading cause of death in Canada and is responsible for 30% of all deaths.[1] Lung, breast, colorectal, and prostate cancer are the most commonly diagnosed types of cancer in Canada (excluding nonmelanoma skin cancer).[1] These cancers account for about half (48%) of all new cancer cases[1] so these were the cancer types we focused on in this study.

The pathology report for resected cancer specimens forms the basis for evaluating the need for adjuvant therapy and for patient prognostication.[2] The pathology reporting of cancer specimens needs to be standardized, complete, and structured so that therapeutic dilemmas and delays do not occur because of incomplete pathology reports.[2,3] The use of standardized structured datasets for pathology cancer reporting has been shown to improve patient care and clinical outcomes.[4,5] In particular, standardized electronic synoptic reporting using the College of American Pathologists (CAP)–approved cancer checklists ensures complete reports, reduces clinical errors, and provides reliable aggregate data as discrete data elements to analyze for laboratory quality assurance, research, cancer registry surveillance, and other secondary uses.[6–8] The electronic version of the CAP Cancer Protocols (eCCs) is currently used by 35% to 40% of all practicing anatomic pathologists in the United States and Canada.[4] CAP Cancer Protocols and Checklists have been endorsed by the Canadian Association of Pathologists (CAP-ACP) and are now a pan-Canadian content standard for pathology reporting in Canada.[9]

Health care organizations and professional groups are being asked to develop clear quality measures and metrics that can lead to performance improvements at provider and

Accepted for publication January 27, 2023.

Supplemental digital content is available for this article. See text for hyperlink.

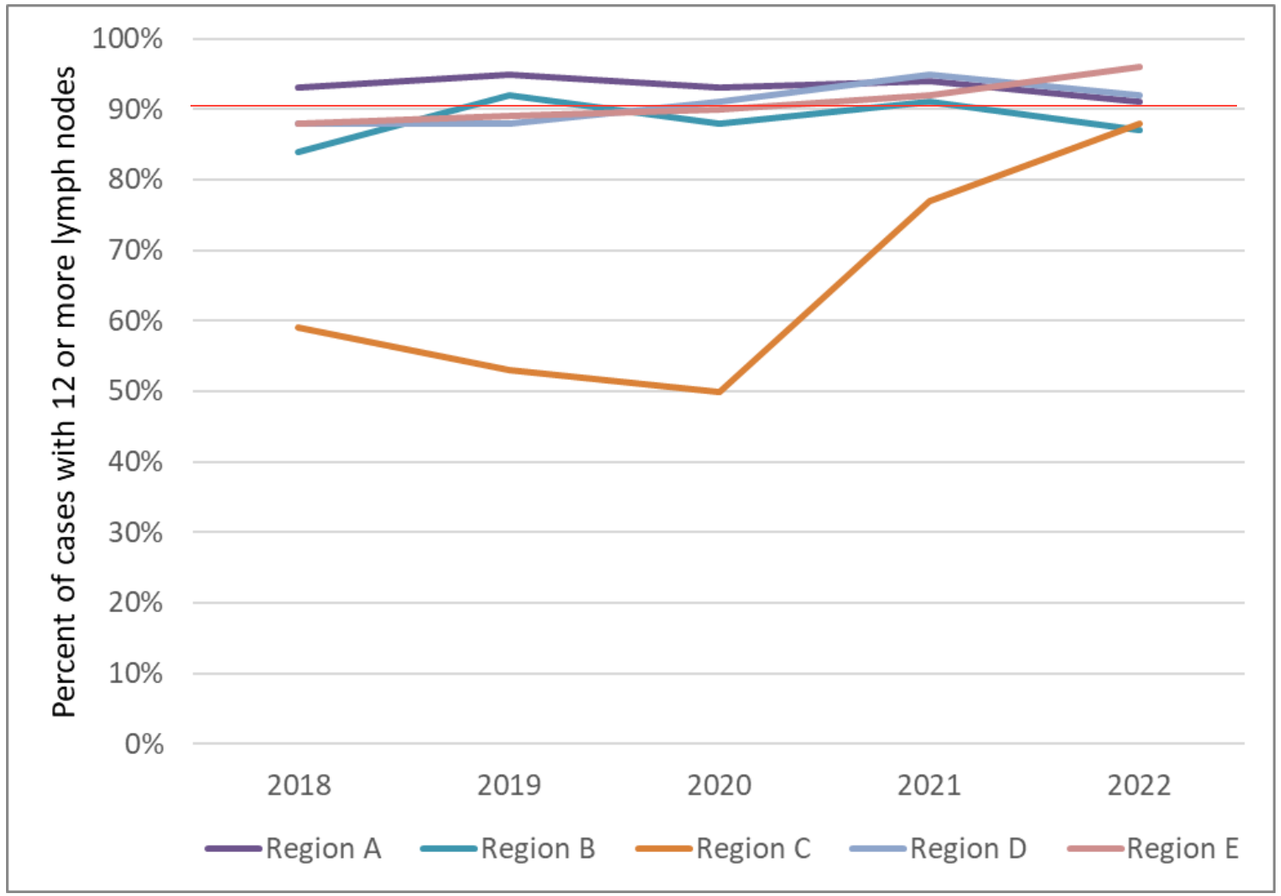
From Data Science, Bisrey Analytics Ltd, Delta, British Columbia, Canada (Bisra); Anatomical Pathology, Provincial Health Services Authority, Vancouver, British Columbia, Canada (Rabel); the Department of Pathology and Laboratory Medicine, Royal Jubilee

Provincial Health Services Authority

https://meridian.allenpress.com/aplm/article/doi/10.5858/arpa.2021-0542-OA/492384/Using-Pathology-Synoptic-Reporting-Data-to-Create

Colorectal Resection Lymph Node Harvest Rates in British Columbia

Provincial Health Services Authority



Target: 90%

Harvesting a minimum of 12 lymph nodes is a key quality measure for colon cancer care

This measures both:

√ Optimal mesenteric resection by surgeon

√ Optimal lymph node harvest from the resection specimen by the pathologist

British Columbia's Renal Pathology Service

Provincial Health Services Authority

BC's Medical Kidney Biopsy (Non-Cancer) Synoptic Report

Statistics

- # Transplant biopsies
- # Native biopsies
- # Biopsies from community hospitals

Quality

- Biopsy adequacy by community site
  → Support with education/training as needed
- Diagnoses by Nephrologist
  → Assist nephrologists with their QA

Research

- PROMIS codes linked to clinical databases
- Various histologic features captured– easy to identify cases for research

Provincial Health Services Authority

Northwestern Medicine®

Robert H. Lurie Comprehensive Cancer Center of Northwestern University

Future of Cancer Data Summit

Unlocking Insights with Pathology Reporting

Michael Gurley | m-gurley@northwestern.edu
Administrative Director of Cancer Informatics
Robert H. Lurie Comprehensive Cancer Center of Northwestern University
10/6/2023

Disclosures

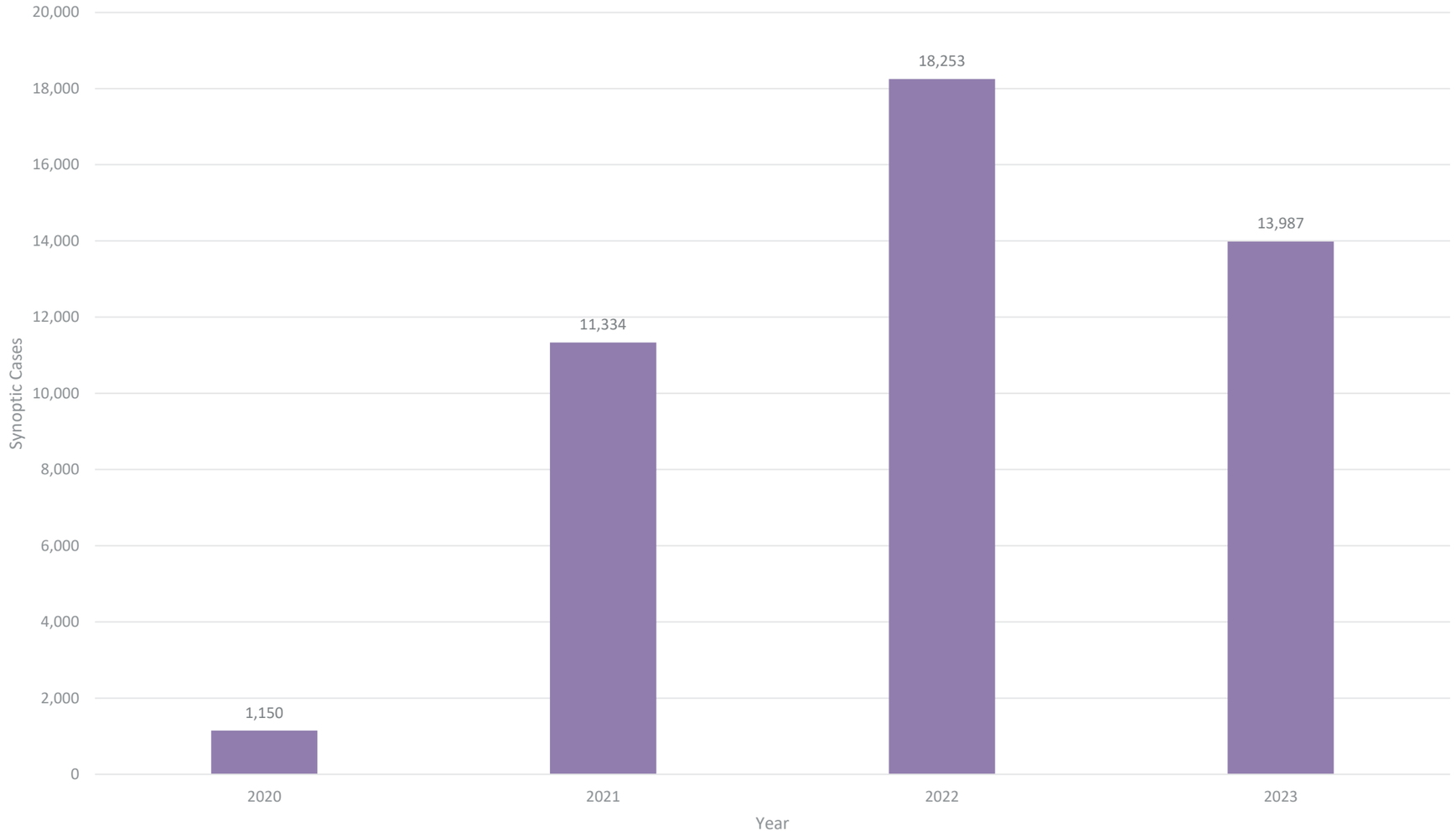
- Ownership Interest in Textractor, Inc.

Synoptic Pathology at Northwestern Medicine

- This discrete capture of CAP cancer protocol data as synoptic reports has been ongoing since 2020 at Northwestern Medicine (NM).

- The synoptic reports contain discrete representations of desired fundamental oncology data points.

Synoptic Case Growth



Northwestern Medicine®

Robert H. Lurie Comprehensive Cancer Center of Northwestern University

Finding the Data

- NM research data analysts noticed ‘Synoptic Reports’ sections appearing in many pathology cases.
- First at tempts at using ‘synoptic’ data involved regular expressions and NLP to extract this new consistently formatted, tableized data.
- In late 2022, NM Research Analytics investigated if and how this new nicely formatted data might be discretely stored within the Epic Clarity data model.
- NM research analytics discovered that the CAP cancer protocols are implemented as synoptic reports that leverage Epic Smart Data Elements within the Epic Beaker module.
- NM Research Analytics developed boilerplate SQL that could extract the discrete fundamental oncology data points for incorporation into analytic data sets.

Organizing the Data

- The native Epic Smart Data Elements data structures containing synoptic data can be queried in Clarity but the retrieval is not intuitive or performant.

- To optimize the use of discrete synoptic data, NM Research Analytics has built an extension data model and ETL to its existing pathology data mart that represents discrete synoptic data in a simple entity attribute value (EAV) structure.

- Share our pathology data mart ETL that incorporates the synoptic pathology data from Clarity Smart Data Elements with other Epic hostages (customers).

Organizing the Data

- The pathology finding extension tables are integrated into the existing NM pathology data mart by adding a new fact PATHOLOGY_FINDING table hanging off the existing PATHOLOGY_CASE_REPORT_SECTION table.

- Synoptic reports are represented in Epic Beaker as a textual report section and discrete Smart Form concepts. The data mart transforms the discrete Smart Form concepts into a compact discrete representation in the PATHOLOGY_FINDING table.

- The connection of discrete findings to the textual representation in the pathology case report section enables querying a pathology case based on discrete values.

Organizing the Data

- Each finding is implemented as a question and answer. The data type of the answer depends on the data type of the question. Supported question data types include: ‘category’, and ‘string’.

- CAP cancer protocols often allow for answers to be supplemented with free text. So, a categorical answer will specify if it can detour to free text.

- The data mart contains dimension tables that set up the expectations for each CAP cancer protocol: PATHOLOGY_FINDING_FORM, PATHOLOGY_FINDING_QUESTION, and PATHOLOGY_FINDING_ANSWER. These tables function like REDCap instruments.

Organizing the Data

- The tables define each CAP cancer protocol's questions, answers, and dependencies. Each CAP cancer protocol has multiple published versions. Each version has a begin and end date of clinical applicability. For the most part, questions, answers and dependencies remain constant across published versions, but additions and removals can occur.

- Epic Smart forms use stable identifiers for questions and answer concepts across versions that are sourced from the CAP eCP Ckeys. So, querying for data across multiple published versions of the same CAP cancer protocol is possible.

- The data mart supports the tracking of published versions of a CAP cancer protocol in PATHOLOGY_FINDING_FORM table.

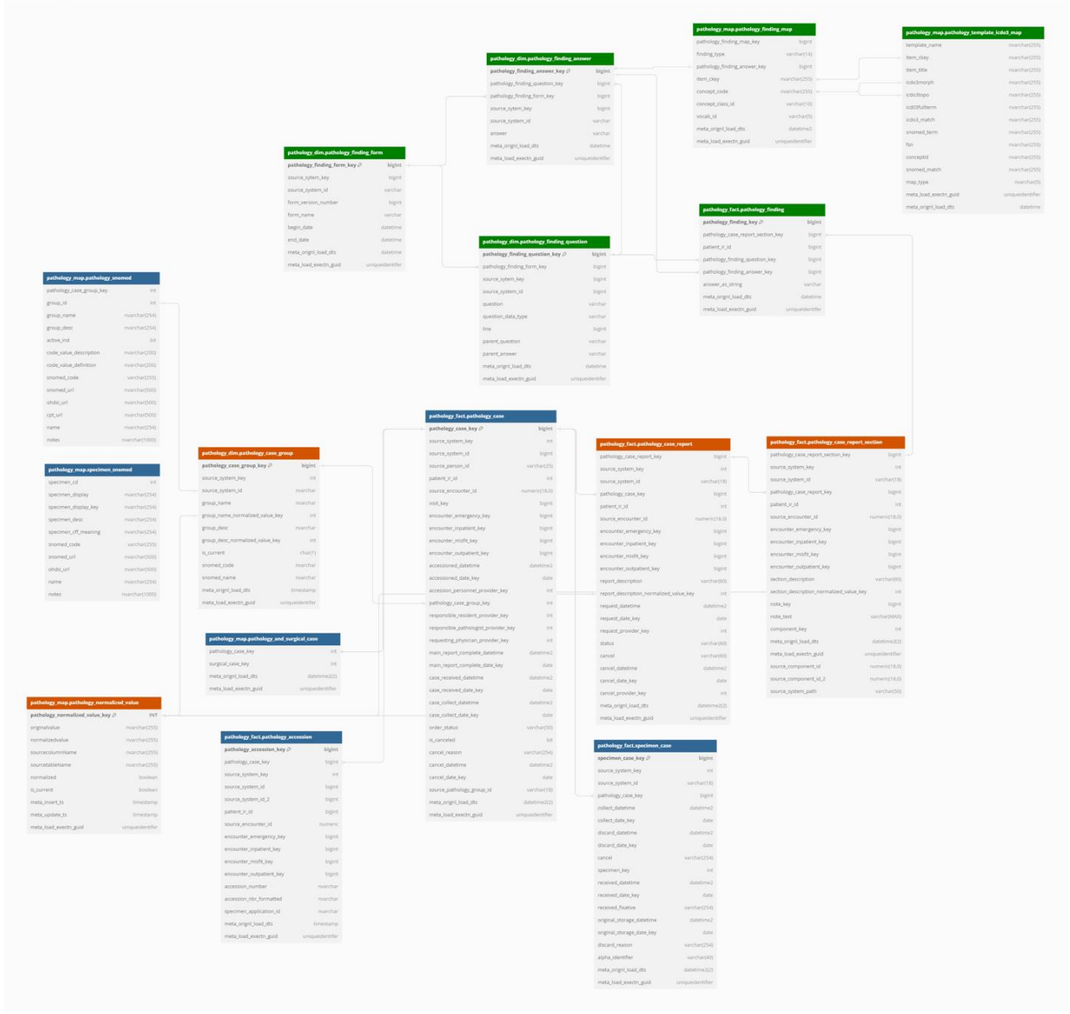
Organizing the Data

- Often It is desirable to map CAP cancer protocol data to standardized healthcare vocabularies. Both the PATHOLOGY_FINDING_QUESTION and PATHOLOGY_FINDING_ANSWER can be related to a polymorphic mapping table that allows for specifying a map to concepts in external, standardized vocabularies.

- CAP itself provides maps of answers to ICD-O-3 site and histology in their CAP eCC distribution. NM has incorporated the mappings into its pathology data mart.

- Future plans include mappings to all Ckey concepts to SNOMED via the Nebraska Lexicon.

Organizing the Data



Using the Data

- Surgical Volume Based on Histology.
  - ICD-10 diagnosis codes provide no support for partitioning diagnoses by histology.
  - Cancer programs want to know how many surgeries spawned what types of pathology-confirmed histologies.
  - Using NM pathology data mart's connection of pathology cases to synoptic data and to surgeries enables the reporting of surgical volume by histology.

Using the Data

- Clinical Trials Recruitment.
  - Many of the fundamental oncology data points that are essential in matching patients to clinical trial eligibility criteria can be sourced from discrete synoptic pathology data. For example: ICD-O-3.2 site/histology; AJJCC TNM staging; grading; and biomarkers
  - NM Research Analytics has incorporated the querying of discrete synoptic pathology data into its delivery of clinical trial recruitment reports.
  - Connecting prior treatment Epic Beacon episodes and MOSAIQ treatment plans via problem list diagnoses joined to synoptic ICD-O-3.2 sites.

Using the Data

- Tumor Registry Case Finding and Case Abstraction.

  - Tumor registry casefinding and case abstraction can be sourced from discrete synoptic pathology data. For example: ICD-O-3.2 site/histology; AJJCC TNM staging; grading; and biomarkers

  - NM IT has set up a direct feed of synoptic pathology results from Epic to our tumor registry vendor NeuralFrame.

Using the Data

- Clinical Specimen Archive Annotation.
    - Many of the attributes requested by investigators for identifying cohorts of patients with clinical specimens residing in the NM clinical specimen archive can be sourced from discrete synoptic pathology data. For example: ICD-O-3.2 site/histology; AJJCC TNM staging; grading; and biomarkers
    - NM Research Analytics has incorporated the querying of discrete synoptic pathology data into its delivery of accession numbers for submission to the NM Clinical Specimen Release Committee.

Using the Data

- Automating Imports into REDCap Data Repositories.
    - Many of the attributes that are manually abstracted by data coordinators for input into REDCap data repositories (for example, SPORE data repositories) can be sourced from discrete synoptic pathology data. For example: ICD-O-3.2 site/histology; AJJCC TNM staging; grading; and biomarkers
    - NM Research Analytics has incorporated the querying of discrete synoptic pathology data for export into REDCap via the REDCap API.

Using the Data

- Automating Imports into REDCap Data Repositories.

Using the Data

- Automating Imports into Investigator-Initiated Clinical Trial Case Report Forms.
    - Many of the attributes that are manually abstracted by data coordinators for input into Northwestern's (NU) eCRF system to support investigator-initiated clinical trials can be sourced from discrete synoptic pathology data. For example: ICD-O-3.2 site/histology; AJJCC TNM staging; grading; and biomarkers
    - NM Research Analytics is in the process of developing the querying of discrete synoptic pathology data for import into the NU's eCRF system.

Challenges

- CAP eCC is Not an Ontology
  - The lack of ontology in the synoptic pathology concept Smart Form hierarchy as imported from CAP eCC makes querying across protocol templates a hardcoded affair.
  - Example: the concept ‘Acinar adenocarcinoma’ does not have the same internal identifier (cKey) across all templates or templates for the same cancer type. For example:

Template Name	Item Title	Item Ckey
Prostate Gland: Needle Biopsy (Case Level)	Acinar adenocarcinoma	330696.1
Prostate Gland: Needle Biopsy (Specimen Level)	Acinar adenocarcinoma	330451.1
Prostate Gland: Radical Prostatectomy	Acinar adenocarcinoma	56746.1
Prostate Gland: Transurethral Prostatic Resection (TURP), Enucleation Specimen (Simple or Subtotal Prostatectomy)	Acinar adenocarcinoma	40264.1

Future Plans

- Incorporating ino the NM pathology data mart the Nebraska Lexicon ontology mappings of Ckeys to SNOMED concepts To improve flexibility in querying. The missing ontology!

- Harmonizing NLP-derived pathology findings from pre-synoptic pathology reports in our data mart to be able to deliver pathology findings across time. The hope is CAP eCC will serve as a knowledge graph to aid LLMs with retrieval augmented generation (RAG).

- Mapping CAP Ckeys to SNOMED for AJCC staging
  - https://build.fhir.org/ig/HL7/fhir-mCODE-ig/ValueSet-mcode-cancer-stage-value-vs.html

Takeaway

- For the longest time, data has been missing from the EHR.
  - No histology!
  - No pathological staging!
  - No grading!
  - No biomarkers!

- The data is finally there in the EHR! We need to start delivering the data from synoptic reports to accelerate research, clinical care, and operations.

- Using the data will motivate the expansion of synoptic reporting to biopsies and resistant disease types.

Northwestern Medicine®

Robert H. Lurie Comprehensive Cancer Center
OF NORTHWESTERN UNIVERSITY

The Future of Cancer Data:
Unlocking Insights With Pathology Reporting Summit

October 6, 2023