Querying the PCORnet Tumor Table

Brad McDowell

June 9, 2026

https://redcap.link/tt_reference

Advantages of the Tumor Table

- Date of diagnosis: allows incidence to be determined (compared to prevalence)
- Stage at diagnosis
- Site specific variables (e.g., hormone receptor status for breast cancer, KRAS status for colorectal cancer, Ki-67 for neuroendocrine tumor)

Tumor Table Records

- Based on NAACCR 2025 specification
- One row per tumor (in principle!)
  - Multiple records for same tumor could represent updated records, ascertainment from >1 source, errors
- Each tumor corresponds to a unique combination of PATID and hospital sequence number (some sites use central sequence number).
  - It is possible that records for the same tumor is found in multiple facilities (for registries that cover >1 facility). Each tumor record is defined as a unique combination of PATID, sequence number, and FACILITY.

Creating Cancer Cohorts: ICD-O-3

- Cancer site (primarySite, NAACCR #400): ICD-O-3 topology codes (not ICD-10 codes!)

ICD-10 Code	Description
C34	Malignant neoplasm of bronchus and lung
C34.0	Malignant neoplasm of main bronchus
C34.00	Malignant neoplasm of unspecified main bronchus
C34.01	Malignant neoplasm of right main bronchus
C34.02	Malignant neoplasm of left main bronchus
C34.1	Malignant neoplasm of upper lobe, bronchus or lung
C34.10	Malignant neoplasm of upper lobe, unspecified bronchus or lung
C34.11	Malignant neoplasm of upper lobe, right bronchus or lung
C34.12	Malignant neoplasm of upper lobe, left bronchus or lung
C34.2	Malignant neoplasm of middle lobe, bronchus or lung
C34.3	Malignant neoplasm of lower lobe, bronchus or lung
C34.30	Malignant neoplasm of lower lobe, unspecified bronchus or lung
C34.31	Malignant neoplasm of lower lobe, right bronchus or lung
C34.32	Malignant neoplasm of lower lobe, left bronchus or lung
C34.8	Malignant neoplasm of overlapping sites of bronchus and lung
C34.80	Malignant neoplasm of overlapping sites of unspecified bronchus and lung
C34.81	Malignant neoplasm of overlapping sites of right bronchus and lung
C34.82	Malignant neoplasm of overlapping sites of left bronchus and lung
C34.9	Malignant neoplasm of unspecified part of bronchus or lung
C34.90	Malignant neoplasm of unspecified part of unspecified bronchus or lung
C34.91	Malignant neoplasm of unspecified part of right bronchus or lung
C34.92	Malignant neoplasm of unspecified part of left bronchus or lung

Does not include carcinoid tumors of the lung

ICD-O-3 lung topology

C34 BRONCHUS AND LUNG

C34.0 Main bronchus
Carina
Hilus of lung

C34.1 Upper lobe, lung
Lingula of lung
Upper lobe, bronchus

C34.2 Middle lobe, lung
Middle lobe, bronchus

C34.3 Lower lobe, lung
Lower lobe, bronchus

C34.8 Overlapping lesion of lung *(see section 4.2.6)*

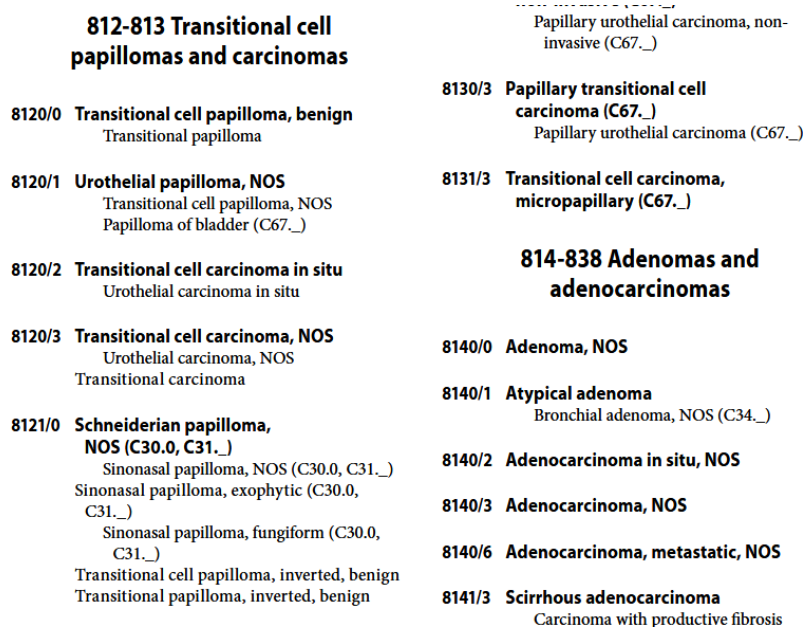
C34.9 Lung, NOS
Bronchus, NOS
Bronchiole
Bronchogenic
Pulmonary, NOS

Creating Cancer Cohorts: ICD-O-3

- Cancer site (primarySite, NAACCR #400): ICD-O-3 topology codes (they are NOT ICD-10 codes!!)
  - ICD-O-3 codes are stored without the decimal in the NAACCR standard
  - Primary tumors only, not metastatic sites or recurrent disease

Creating Cancer Cohorts: ICD-O-3

- Cancer histology (histologicTypeIcdO3, NAACCR #522): ICD-O-3 histology. Refers to microscopic characteristics of the tumor cells.



812-813 Transitional cell papillomas and carcinomas

8120/0 Transitional cell papilloma, benign
Transitional papilloma

8120/1 Urothelial papilloma, NOS
Transitional cell papilloma, NOS
Papilloma of bladder (C67._)

8120/2 Transitional cell carcinoma in situ
Urothelial carcinoma in situ

8120/3 Transitional cell carcinoma, NOS
Urothelial carcinoma, NOS
Transitional carcinoma

8121/0 Schneiderian papilloma, NOS (C30.0, C31._)
Sinonasal papilloma, NOS (C30.0, C31._)
Sinonasal papilloma, exophytic (C30.0, C31._)
Sinonasal papilloma, fungiform (C30.0, C31._)
Transitional cell papilloma, inverted, benign
Transitional papilloma, inverted, benign
Papillary urothelial carcinoma, non-invasive (C67._)

8130/3 Papillary transitional cell carcinoma (C67._)
Papillary urothelial carcinoma (C67._)

8131/3 Transitional cell carcinoma, micropapillary (C67._)

814-838 Adenomas and adenocarcinomas

8140/0 Adenoma, NOS

8140/1 Atypical adenoma
Bronchial adenoma, NOS (C34._)

8140/2 Adenocarcinoma in situ, NOS

8140/3 Adenocarcinoma, NOS

8140/6 Adenocarcinoma, metastatic, NOS

8141/3 Scirrhous adenocarcinoma
Carcinoma with productive fibrosis

Creating Cancer Cohorts: ICD-O-3

- Cancer histology (histologicTypeIcdO3, NAACCR #522): ICD-O-3 histology. Refers to microscopic characteristics of the tumor cells.
  - It is important to put some thought into which codes to include.
    - ICD-O-3 manual organizes the codes into basic groups (e.g., 805X-808X correspond to squamous cell neoplasms)
    - Let previous lit be a guide.
    - Empirical approach: Create a list histology codes associated with tumor site of interest along with their counts and ask PI to decide which they want to include.
  - Histology combined with behavior (e.g., 8140/3) is called “morphology”
- Cancer behavior (behaviorCodeIcdO3, NAACCR #523): Usually interested in malignant behavior (behaviorCodeIcdO3 = ‘3’)

Stage

- There are many stage variables (clinical/pathological versions of TNM components, e.g., pT2pN1pM0, stage groups, localized/regional/distant)
- There are many changes over time to these variables, especially in 2018. Make sure to examine changes in the data!

Stage group variables (Stage 1-4) are most often used:

	Dx prior to 2018	Dx 2018 and after
Clinical stage	tnmClinStageGroup (#970)	ajccTnmClinStageGroup (#1004)
Pathological Stage	tnmPathStageGroup (#910)	ajccTnmPathStageGroup (#1014)

Creating Cancer Cohorts: other important variables

- Date of diagnosis (dateOfDiagnosis)
- Sequence number (sequenceNumberHospital): (00 = solitary, 01 = first of many, etc.)
- Class of case (classOfCase): 00-22 = “analytic” and likely to have more complete data available

First course of treatment

- All treatment variable names start with “rx”.
- Will only capture treatments that occur ~6 months after dx
- Variables for several kinds of cancer therapy. The best data is for surgery (rxSummSurgPrimSite)
- Surgery codes are site specific (see resources)

Potential pitfalls

- NAACCR spec changes over time – big change in 2018
  - Always look at how your variables change over time (including whether they're even populated/null)
- Hospital vs Central versions of variables – mainly affects treatment variables and sequence number
- Ascertainment lag
- Completeness of record: just because there is a variable doesn't mean it's well-populated (e.g., chemotherapy). Possible differences between hospitals vs health systems (e.g., Allina, Intermountain)

Resources

- Tumor table specification: https://github.com/gpcnetwork/Tumor-Table
- NAACCR data dictionary: https://apps.naaccr.org/data-dictionary/data-dictionary/version=25/chapter-view/
- ICD-O-3 manual: https://iris.who.int/server/api/core/bitstreams/f63d7652-1254-47f4-8ab7-3c65ad60d57d/content
- SEER Registrar Staging Assistant: https://staging.seer.cancer.gov/
  - Helpful for finding site-specific variables mentioned above
- Surgery codes: https://www.naaccr.org/crosswalks-interoperability/
- Histology codes: See “histology codes.csv”

Example

```
data tumor00; set indata.tumor_v2025;
    where substr(primarysite,1,3) = 'C50' and behaviorCodeIcdO3 = '3'
        and 2008 <= input(substr(dateOfDiagnosis,1,4),8.) <= 2022;

    dxyr = input(substr(dateOfDiagnosis,1,4),8.);
run;

data tumor01; set tumor00;
    if histologicTypeIcdO3 in ('8500','8520','8480','8470','8503','8504','8507','8509','8522','8523','8524');
run;

titlel 'Receptor status, 2008-2017';
proc freq data = tumor01;
    where 2008 <= dxyr <= 2017;
    format csSiteSpecificFactor16 $recep_f.;
    tables csSiteSpecificFactor16;
run;

titlel 'ER, HER2 status, 2018+'; *PR not assessed because it can be either +/-;
proc freq data = tumor01;
    where dxyr >= 2018;
    format estrogenReceptorSummary $er_f. her2OverallSummary $her2_f.;
    tables estrogenReceptorSummary*her2OverallSummary / norow nocol nopercent;
run;
titlel;
```

Malignant breast cancer diagnosed between 2008-2022

Eligible histology codes determined with PI. Used “histology codes.csv” for that conversation.

Receptor status is a site-specific factor captured differently before/after 2018. See Registrar Staging Assistant

```
data tumor02; set tumor01;
    if 2008 <= dxyr <= 2017 then
        do;
            if csSiteSpecificFactor16 in ('100','110') then erpos_her2neg = 1;
            else erpos_her2neg = 0;
        end;
    else if dxyr >= 2018 then
        do;
            if estrogenReceptorSummary = '1' and her2OverallSummary = '0' then erpos_her2neg = 1;
            else erpos_her2neg = 0;
        end;
run;

title1 'ER+,HER2-';
proc freq data = tumor02;
    format dxyr era.;
    tables erpos_her2neg*dxyr / norow nopercent;
run;

data tumor03; set tumor02;
    if erpos_her2neg;
run;
```

Examining whether receptor status looks different across different eras.

Table of erpos_her2neg by dxyr			
erpos_her2neg	dxyr		
Frequency Col Pct	CS Era	EOD Era	Total
0			
1			
Total			

```
titlel 'Surgery codes - consider 20-80 to indicate lumpectomy/mastectomy';
proc freq data = tumor03;
    tables rxSummSurgPrimSite*dxyr / norow nocol nopercent;
run;
titlel;

data tumor04; set tumor03;
    if '20' <= rxSummSurgPrimSite <= '80';
run;

data tumor05; set tumor04;
    if 2008 <= dxyr <= 2017 then
        do;
            if substr(tnmPathStageGroup,1,1) in ('1','2') then stagelor2 = 1;
            else stagelor2 = 0;
        end;
    else if dxyr >= 2018 then
        do;
            if substr(ajccTnmPathStageGroup,1,1) in ('1','2') then stagelor2 = 1;
            else stagelor2 = 0;
        end;
run;
```

See surgery codes resource. Be careful about recent changes.

Different stage variables depending on dx year. Make sure to examine stage distributions for different era – it really matters for some cancer sites.

Possible next steps for this query

- Select one record for each patient (possibly first or most complete tumor record)
- Link to other CDM tables to assess procedures, lab values/other clinical measurements, identify therapies (Carnahan et al., https://pubmed.ncbi.nlm.nih.gov/32795185/), etc.
- Deidentify and recruit for prospective data collection (e.g., NET-PRO, Share Thoughts on Breast Cancer: https://redcap.link/LEC2026_cancer_research)

Whew!

- Any questions?

https://redcap.link/tt_reference

Related Datasets

- Surveillance, Epidemiology, and End Results (SEER)
- Cancer in North America Data Products (CiNA)
- National Cancer Database (NCDB)