**DETAILED PROPOSAL**

 **Background**

1. Challenges related to representation of cancer diagnosis in source data.

Cancer diagnosis is usually recorded in two sources: Cancer/Tumor Registries and electronic medical records (EMR). In Tumor Registries, cancer diagnoses are abstracted from pathology reports and EMR. Pathology-based diagnosis is coded in ICD-O representing a combination of cancer histology and topography. ICD-O is considered a gold standard to annotate cancer diagnosis. EMR-based diagnosis is coded in ICD-9/10. Although diagnosis in Cancer Registries is most accurate and granular, in most (non-SEER) states it is only recorded for the first cancer occurrence and not recorded for recurrent cancer. In EMRs, cancer diagnosis is recorded as regular billing and problem list diagnoses using ICD-9/10 coding that is less granular than ICD-O. Therefore, tracking of cancer diagnosis at the same level of granularity throughout the course of disease is a serious challenge.

In addition to cancer histology and topography, there are other cancer features that define cancer diagnosis and determine outcomes and treatments. Similar to histology and topography, these features are abstracted from pathology reports into Cancer Registries for the first cancer occurrence (with the exception for SEER states where all cancer occurrences are recorded in Cancer Registries). Unlike histology and topography, these additional features are rarely available in EMR in a structured form. They may be present in pathology systems. Deriving and reconciling one “clean” set of cancer attributes for each cancer occurrence is critical and challenging.

Identifying cancer occurrences is another challenge because, with the exception of SEER states that explicitly record cancer recurrences, they are not available in a structured form in EMRs. Patterns of ICD9/10 diagnoses after initial diagnosis do not reliably track the recurrence status of a patient’s cancer diagnosis.

1. Challenges of cancer diagnosis representation in OMOP CDM and Vocabulary.

Since in the source data cancer diagnosis is represented in ICD-O for the first occurrence and in ICD-9/10 for all occurrences, the challenge is to connect these two representations. OHDSI’s standard for diagnosis representation is SNOMED. ICD-9 and ICD-10 have been successfully mapped to SNOMED. One challenge is to validate existing mappings between ICD-O and SNOMED CT and propose new SNOMED CT coding to cover all the existing ICD-O histology and topography combinations. The other challenge is to reconcile SNOMED diagnosis resulted from mappings from ICD-O and ICD-9/10.

Cancer histology-topography must be linked to other key diagnostic features like stage, and grade. Although there are a few common features for many cancer types (e.g. stage, grade), their values vary and other features (e.g. tumor size, laterality, biomarkers) are specific to each cancer type. There are presently two attribute-value tables that may house cancer diagnostic features, Observation and Measurement. However, none of them has an explicit linkage to Condition\_Occurrence. Implications on extending either of these tables to support the linkage and store modifier tables must be carefully evaluated.

From the vocabulary stand point, many cancer diagnosis features are covered in LOINC and SNOMED CT. The challenge is to create a set of these concepts for each cancer type (e.g. breast, lung, etc.). There is an ongoing initiative that intends to create and align these sets with the ICCR (International Collaboration on Cancer Reporting) data sets. However, at this point, it has only covered a few cancer types.

1. Treatment clinical event welter.

Oncology treatments often create a clinical event welter that thwarts many analytic use cases. Instead, we want a concept that aggregates lower-level clinical events into a higher-level abstraction. Some examples of clinical event welter:

* Beam IMRT to left breast 15 Fractions at 267 cGy Dose over 27 days.
	+ Spawns 72 entries in the PROCEDURE\_OCCURRENCE table across 11 CPT codes.
* Docetaxel + Carboplatin, 21-DAY cycle, 6 cycles over 115 days
	+ Spawns 22 entries in the PROCEDURE\_OCCURRENCE table across 5 CPT codes and 50 entries in the DRUG\_EXPOSURE table across 13 RxNorm codes.

In the oncology community, there is a widely shared *treatment* concept: tumor registries, practice guidelines, clinical trials databases and oncology analytic platforms all employ the concept of a *treatment*. The *treatment* concept aggregates lower-level clinical events into a higher-level abstraction. But we still want this higher-level abstraction *treatment* concept to connect to lower-level clinical events.

Can we have both the benefit of a higher-level oncology treatment abstraction and connection to lower-level clinical events? Can we find a place for pre-made oncology *treatment* abstractions connected to lower-level clinical events? Conversely, can we derive higher-level oncology *treatments* from lower-level clinical events ? Finally, can we attach unconnected pre-made oncology *treatment* abstractions to lower-level clinical events?

Adding support for the representation of *treatment* abstractions within the OMOP CDM will enable the following use cases:

* Classify each treatment at a level intuitive to oncology professionals/researchers.
* Enumerate how many oncology treatments have been performed on a patient.
* Characterize when each treatment begins and ends.
* Describe when a treatment is “switched”.
* Reuse the grouping of low-level clinical events present in source systems. Not lose *treatments* abstractions during conversion into the OMOP CDM.
* A target for the algorithmic derivation of treatment abstractions when not present in our source systems
* Harmonize EHR/claims database oncology treatment data and tumor registry oncology treatment data.
* Attribute properties to an oncology treatment as a whole.
1. Absence of abstraction layer representing clinician’s and researcher’s view

Clinicians and researches view cancer as a chronic disease with a series of disease episodes: first occurrence, remission, relapse, end of life event. Cancer is maintained with a treatment course comprised of one or more treatment modalities, multiple regimens, and cycles. Disease progression is monitored at prescribed intervals and reported as outcomes (e.g stable disease). These abstractions of disease, treatment, and outcomes are rarely available in the source data. Derivation and persistence of these abstractions are critical since they are key variables for prediction of disease progression, disease free and overall survival.

Currently, Condition\_Era and Drug\_Era are the structures that house derived condition and drug exposure episodes. However, neither their attributes nor algorithms of their derivation will support complex abstractions of cancer disease and treatment.

1. **Overall Approach**
* **Cancer diagnosis**
	+ In our current approach, we define cancer diagnosis as a combination of **histology** (morphology) + **topography** (anatomy)
* **Diagnosis modifiers**
	+ Diagnostic and treatment features that vary between different cancer diagnoses and treatments are represented as modifiers and explicitly linked to the respective diagnosis or treatment
	+ Examples of diagnosis modifiers are stage, grade, laterality, foci, tumor biomarkers. These diagnostic features are assessed when a patient is first diagnosed and also (possibly) for each cancer recurrence. Repeated measurements of the same modifier (lymph node invasion) may be recorded. Different modifiers may be recorded on different dates
	+ Examples of treatment modifiers are surgery laterality, radiotherapy dosage and frequency.
* **Disease and treatment abstraction layer**
	+ Disease and treatment abstractions will be modeled as episodes, a new CDM construct that can be used to represent other abstractions such as episode of care.
	+ These abstraction may be derived algorithmically pre- or post-ETL or extracted from the source data directly. In addition to the regular OMOP type\_concept\_ID, we propose to store references to the derivation algorithms in the vocabulary.
	+ Disease abstractions include first occurrence, remissions, relapses, and end of life event.
	+ There should be one set of “verified” cancer modifiers associated with each cancer occurrence and relapse.
	+ Treatment abstractions include treatment course, treatment regimen, and treatment cycle.
1. **Representing cancer diagnosis as histology and topography combination**

The International Classiﬁcation of Diseases for Oncology (ICD-0)1 is a dual-axis vocabulary used to identify cancer topography (anatomic site) and histology (morphology) to track and report cancer incidence, survival and mortality.

The topography code describes the anatomical site of origin of the neoplasm. The code always has a preﬁx of “C”, followed by a three digit number that indicates the site (two digits) and the subsite (one digit), separated by a decimal point. Example: C18.4: the C18 indicates that the site is the colon and the 4 indicates that the sub-site is the transverse colon.

The histology code describes the characteristics of the tumor itself, including its cell type and biological activity. The code is composed of four digits that indicate the cell type or histology and one digit that indicates the behavior. The ﬁrst four digits are separated from the last (behavior) digit by a forward slash (/). The behavior digit can be: 0 (benign), 1 (uncertain behavior), 2 (carcinoma in situ), 3 (malignant, primary site), 6 (malignant, metastatic site), 9 (malignant, uncertain whether primary or metastatic site), Examples: Squamous cell carcinoma in situ, NOS = 8070/2, Adenocarcinoma, NOS = 8140/3, Carcinoma, NOS = 8010/3.

Each combination of these two dimensions, histology and topography, rolls up to a unique cancer diagnosis. For example, Carcinoma, NOS (8010/3) and ‘Unspecified part of bronchus or lung’ (C34.9) rolls to ‘Carcinoma of the lung’. However, there is no single code to annotate this unique cancer diagnosis. National Cancer Institute (NCI) Surveillance, Epidemiology, and End Result Program (SEER) provide validation lists for “coherent” topography/morphology or site/histology combinations2.

To represent cancer diagnosis, the combination of histology and topography, in the OMOP CDM Condition domain (Condition\_Occurrence) without changes of the existing structure, we propose to perform a pre-coordinated collapse of the ICD-O axes, histology and topography, to a single OMOP originated concept representing unique cancer diagnosis and preserve linkages between these single codes and the ICD-O axes in the OMOP vocabulary3.

Our proposed approach will support adherence to the OMOP CDM/Vocabulary conventions:

* One required \_concept\_id field will be populated in the corresponding domain table, Condition\_Occurrence.
* Vocabulary-related attributes are stored in a vocabulary data model in a uniform way

If a new proper SNOMED code is created, the OMOP-originated concept can be easily replaced by it.

A similar mapping approach is used for representing LOINC/HL7 clinical note types and CDO ontology in the OMOP CDM NLP tables.

**Detailed implementation**

1. Create pre-coordinated source (non-standard) concepts in the Concept table representing unique cancer diagnosis by collapsing the two ICD-O axes, histology and topography into unique concepts using SEER validation lists <https://seer.cancer.gov/icd-o-3/>. For example, a new pre-coordinated concept derived by collapsing ‘*Adenocarcinoma, NOS’* (8140/3) and ‘*Sigmoid colon’* (C18.7) will be ‘*Adenocarcinoma of sigmoid colon’* (8140/3- C18.7):

|  |  |
| --- | --- |
| **Field** | **Record** |
| concept\_id | 36517865 |
| concept\_name | Adenocarcinoma of sigmoid colon |
| concept\_code | 8140/3- C18.7 |
| vocabulary\_id | ICDO3 |

1. Create mapping between the new pre-coordinated source concept and a standard SNOMED concept ‘*Adenocarcinoma of sigmoid colon’* (301756000) in Concept\_Relationship table

|  |  |  |
| --- | --- | --- |
| **Field** | **Record 1** | **Record 2** |
| concept\_id\_1 | 36517865 | 4200514 |
| concept\_id\_2 | 4200514 | 36517865 |
| relationship\_id | *Maps to* | *Is mapped to* |

1. Each pre-coordinated SNOMED concept is linked to morphology/histology (‘*Has associated morphology*’) and anatomic site/topography (‘*Has finding site’*) in the Concept\_Relationship table thus supporting analysis along those axes:

|  |  |  |
| --- | --- | --- |
| **concept\_id\_1** | **concept\_id\_2** | **relationship\_id** |
| 4200514 | 4290838 | *Has associated morphology* |
| 4200514 | 4244588 | *Has finding site* |

Where 4290838 represents a SNOMED concept of ‘*Malignant adenomatous neoplasm – category*’ and 4244588 represents a SNOMED concept of ‘*Sigmoid colon structure’*.

1. Detailed ETL instructions are provided in the Appendix.
2. **CDM Representation and Extension**
3. **Overview**

****

**Fig 1. High level ERD**

Existing tables are depicted in blue and new tables and relationships in red.

* Cancer diagnoses are stored in CONDITION\_OCCURRENCE as pre-coordinated concepts combining histology and topography.
* Cancer treatment clinical events are stored in the PROCEDURE\_OCCURRENCE and DRUG\_EXPOSURE tables.
* Disease and treatment abstractions (e.g. first cancer occurrence, treatment regimen hormonal therapy) are represented in the new EPISODE table.
* Links between the disease and treatment episodes and the underlying events (conditions, procedures, drugs) are stored in the new EPISODE\_EVENT table.
* Additional diagnostic and treatment features are stored in the Measurement table as modifiers of the respective condition, treatment, or episode. Measurement table is extended to include a reference to the condition, treatment, or episode record.
1. **Representing cancer disease and treatment abstractions as episodes**

**New EPISODE table**

Episode represents disease and treatment abstractions like first disease occurrence or treatment regimen derived algorithmically or extracted directly from the source data. This table can be also used to represent other abstractions such as episode of care.

| **Field** | **Required** | **Type** | **Description** |
| --- | --- | --- | --- |
| episode\_id | yes | integer | A unique identifier for each Episode. |
| person\_id | yes | integer | A foreign key identifier to the Person who is undergoing the Episode. The demographic details of that Person are stored in the PERSON table. |
| episode\_concept\_id | yes | integer | A foreign key that refers to a standard Episode Concept identifier in the Standardized Vocabularies. Examples of an Episode Concept can be: Treatment Regimen, Treatment Cycle, Disease First Occurrence, Remission, Relapse, Episode of Care |
| episode\_start\_datetime | yes | date | The date and time on which the Episode was started. |
| episode\_end\_datetime | yes | date | The date and time on which the Episode was ended. |
| episode\_parent\_id | no | integer | A foreign key that refers to a parent Episode entry representing an entire episode if the episode spans multiple cycles. |
| episode\_number | no | integer | An ordinal count for an Episode that spans multiple times |
| episode\_object\_concept\_id | yes | integer | A foreign key that refers to a concept identifier in the Standardized Vocabularies describing disease, treatment, or other abstraction that the episode describes. For example, ‘Breast Carcinoma’ or ‘Chemotherapy’. |
| episode\_type\_concept\_id | yes | integer | A foreign key that refers to a standard Episode Type Concept identifier in the Standardized Vocabularies reflecting the provenance of the episode derivation. It may reference a derivation algorithm, sources such as cancer registry, EMR, etc. |
| episode\_source\_value | no | varchar(50) | The source code for the Episode as it appears in the source data. This code is mapped to a standard episode Concept in the Standardized Vocabularies and the original code is, stored here for reference. |
| episode\_source\_concept\_id | no | integer | A foreign key to a Episode Concept that refers to the code used in the source. |

**Conventions**

* Valid Episode Concepts belong to the ‘Episode’ domain.
* Valid Episode Type Concepts belong to the ‘Episode Type' vocabulary in the 'Type Concept' domain.
* Valid Episode Object Concepts belong to different domains based on the corresponding concept class/vocabulary of the Episode Concept.
	+ ‘Disease Episode’:
		- ‘Condition’ domain
	+ ‘Treatment Episode’
		- ‘Procedure/Treatment’ domain
	+ ‘Episode of Care Episode’
		- ‘Episode of Care’ domain.

**Vocabulary Extensions to Represent Episodes**

1. **Add ‘Episode’ domain.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **concept\_name** | **domain\_id** | **vocabulary\_id** | **concept\_class\_id** | **standard\_concept** | **concept\_code** |
| Disease Episode | Episode | Episode | Disease Episode | S | OMOP generated |
| Treatment Regimen Episode | Episode | Episode | Treatment Episode | S | OMOP generated |
| Treatment Cycle Episode | Episode | Episode | Treatment Episode | S | OMOP generated |
| Episode of Care Episdoe | Episode | Episode | Episode | S | OMOP generated |

1. **Add ‘Episode Type’ vocabulary.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **concept\_name** | **domain\_id** | **vocabulary\_id** | **concept\_class\_id** | **standard\_concept** | **concept\_code** |
| Pre-made episode in source system | Type Concept | Episode Type | Episode Type | S | OMOP generated |
| Algorithmically-derived episode pre-ETL  | Type Concept | Episode Type | Episode Type | S | OMOP generated |
| Algorithmically-derived episode post-ETL | Type Concept | Episode Type | Episode Type | S | OMOP generated |

1. **Add concepts to new ‘Procedure/Treatment’ Domain. Based the entries on NAACCR/SEER treatment variables**

| **concept\_name** | **domain\_id** | **vocabulary\_id** | **concept\_class\_id** | **standard\_concept** | **concept\_code** |
| --- | --- | --- | --- | --- | --- |
| Chemotherapy | Procedure/Treatment | Procedure/Treatment | Drug Treatment | S | OMOP generated |
| Homonal Therapy | Procedure/Treatment | Procedure/Treatment | Drug Treatment | S | OMOP generated |
| Immunotherapy | Procedure/Treatment | Procedure/Treatment | Drug Treatment | S | OMOP generated |
| External beam, NOS | Procedure/Treatment | Procedure/Treatment | Radiation Therapy Treatment | S | OMOP generated |
| External beam, photons | Procedure/Treatment | Procedure/Treatment | Radiation Therapy Treatment | S | OMOP generated |
| External beam, protons | Procedure/Treatment | Procedure/Treatment | Radiation Therapy Treatment | S | OMOP generated |
| External beam, electrons | Procedure/Treatment | Procedure/Treatment | Radiation Therapy Treatment | S | OMOP generated |
| External beam, neutrons | Procedure/Treatment | Procedure/Treatment | Radiation Therapy Treatment | S | OMOP generated |
| External beam, carbon ions | Procedure/Treatment | Procedure/Treatment | Radiation Therapy Treatment | S | OMOP generated |
| Brachytherapy, NOS | Procedure/Treatment | Procedure/Treatment | Radiation Therapy Treatment | S | OMOP generated |
| Brachytherapy, intracavitary, LDR | Procedure/Treatment | Procedure/Treatment | Radiation Therapy Treatment | S | OMOP generated |
| Brachytherapy, intracavitary, HDR | Procedure/Treatment | Procedure/Treatment | Radiation Therapy Treatment | S | OMOP generated |
| Brachytherapy, Interstitial, LDR | Procedure/Treatment | Procedure/Treatment | Radiation Therapy Treatment | S | OMOP generated |
| Brachytherapy, Interstitial, HDR | Procedure/Treatment | Procedure/Treatment | Radiation Therapy Treatment | S | OMOP generated |
| Brachytherapy, electronic | Procedure/Treatment | Procedure/Treatment | Radiation Therapy Treatment | S | OMOP generated |
| Radioisotopes, NOS | Procedure/Treatment | Procedure/Treatment | Radiation Therapy Treatment | S | OMOP generated |
| Radioisotopes, Radium-232 | Procedure/Treatment | Procedure/Treatment | Radiation Therapy Treatment | S | OMOP generated |
| Radioisotopes, Strontium-89 | Procedure/Treatment | Procedure/Treatment | Radiation Therapy Treatment | S | OMOP generated |
| Radioisotopes, Strontium-90 | Procedure/Treatment | Procedure/Treatment | Radiation Therapy Treatment | S | OMOP generated |

1. **Placeholder for NAACCR treatment data dictionary items.**
2. **Placeholder for Observational Research in Oncology Toolbox (OROT) classification vocabulary.**
3. **Placeholder for entries in CONCEPT\_RELATIONSHIP between ‘Procedure/Treatment’ domain, NAACCR data dictionary items and OROT classification vocabulary.**

**New EPISODE\_EVENT table**

Episode\_Event stores links between cancer disease and treatment episodes and the underlying events (conditions, procedures, drugs, etc. ).

| **Field** | **Required** | **Type** | **Description** |
| --- | --- | --- | --- |
| episode\_id | yes | integer | A foreign key identifier to the Episode that the Episode Event belongs to. |
| event\_id | Yes | integer | A foreign key identifier to the underlying event (condition, procedure, measurement, etc.) record in a respective table for which an episode is recorded. |
| event\_table\_concept\_id | yes | integer | A foreign key identifier to the standardized concept corresponding to the table (condition\_occurrence, procedure\_occurrence, etc.) where the underlying event is stored. |

**Conventions**

* Some episodes may not have links to the underlying events. For such episodes, EPISODE\_EVENT table is not populated.

**Example of a query**

The query below retrieves all underlying condition and procedure records related to an episode:

SELECT episode.\*, condition\_occurrence\_start\_datetime, condition\_occurrence\_concept\_id as underlying\_event\_concept\_id

FROM episode

JOIN episode\_event ON episode.episode\_id = episode\_event.episode\_id

JOIN condition\_occurrence ON event\_id = condition\_occurrence\_id

AND event\_table\_concept\_id = ‘condition\_occurrence’

UNION ALL

SELECT episode.\*, procedure\_occurrence\_start\_datetime, procedure\_occurrence\_concept\_id as underlying\_event\_concept\_id

FROM episode

JOIN episode\_event ON episode.episode\_id = episode\_event.epsiode\_id

JOIN procedure\_occurrence ON event\_id = procedure\_occurrence\_id

AND event\_table\_concept\_id = ‘procedure\_occurrence’

1. **Representing cancer diagnosis features as diagnosis modifiers and treatment features as treatment modifiers.**

**Extending MEASUREMENT table**

The Measurement table may contain cancer diagnosis, treatment, and episode modifiers such as cancer stage, grade, lymph node involvement, tumor size, tumor biomarkers, radiotherapy total dose and others (numeric or categorical) obtained through laboratory tests, imaging, and pathology reports.

To explicitly link cancer diagnosis, treatment, or episode record to its modifier, we propose to add the following fields to the Measurement table:

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Required** | **Type** | **Description** |
| modifier\_of\_event\_id | No | integer | A foreign key identifier to the event (e.g. condition, procedure, episode) record for which the modifier is recorded. |
| modifier\_of\_field\_concept\_id | No | integer | The concept representing the table field concept that contains the value of the event id for which the modifier is recorded (e.g. Condition\_Occurrence.condition\_occurrence\_id). |

**Conventions**

* Modifier records are similar to regular measurement records in that they require a standardized test or some other activity to generate a quantitative or qualitative result. However, modifiers are not independent measurements but modifiers which add specificity to cancer diagnosis, treatment, or episode. For example, LOINC 44648-4 'Histologic grade' may modify cancer diagnosis of “Tubular carcinoma” recorded in CONDITION\_OCCURRENCE. Therefore, although modifier\_of\_event\_id and modifier\_of\_table\_concept\_id are not required fields, they must be populated for modifiers.
* Repeated modifier records (lymph node invasion) may be associated with one or multiple condition occurrence records.
* Modifiers for the same condition record may be recorded on different dates.
* One set of “verified” cancer modifiers must be associated with a disease or treatment episode.
* Valid Concepts for the value\_as\_concept field normally belong to the ‘Meas Value’ domain.

**Vocabulary Extensions to Represent Cancer Diagnosis Modifiers and Treatment Modifiers**

1. **Standardized diagnosis modifier terminology from Nebraska Lexicon Project:**

We recommend leveraging standard terminology developed by Nebraska Lexicon Project 4. This initiative intends to implement CAP (College of American Pathologists) Protocol Templates5 by providing terminology binding between LOINC and SNOMED CT**.** The majority of the associated terminology development is modeled within SNOMED Observable entity hierarchy. Coded LOINC observables are linked to SNOMED value sets. Terminology sets have been completed by clinical and genomic biomarkers for breast and colorectal cancers.

* 1. For those cancer types not yet covered in Nebraska Lexicon, we will be using North American Association of Central Cancer Registries (NAACCR) data dictionary concepts to represent cancer diagnostic features.
	2. Mappings should be created between NAACCR and Nebraska Lexicon concepts to support ETL for standardized concepts.
1. **Placeholder for Standardized treatment modifier terminology from NAACCR data dictionary.**
2. **Add concepts to the ‘Meas Type’ vocabulary in the ‘Type Concept’ domain;**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **concept\_name** | **domain\_id** | **vocabulary\_id** | **concept\_class\_id** | **standard\_concept** | **concept\_code** |
| Cance Registry | Type Concept | Meas Type | Meas Type | S | OMOP generated |

**EXAMPLES OF CANCER DATA IN OMOP CDM**

1. **Cancer diagnosis record**



1. **Diagnosis modifier record**



1. **Diagnosis episode and related event records**



1. **Treatment episode and related event records**



**References**

1. ICDO-3 vocabulary

<http://codes.iarc.fr/usingicdo.php>

1. SEER histology and topography combinations

<https://seer.cancer.gov/icd-o-3/>.

1. Proposed approach for mapping ICDO-3 to SNOMED

<http://www.ohdsi.org/web/wiki/lib/exe/fetch.php?media=documentation:oncology:poster2018-improvement_of_cancer_diagnosis_representation_in_omop_cdm3_1_.pdf>

1. Nebraska Lexicon

<https://www.unmc.edu/pathology/informatics/tdc.html>

1. CAP Protocol Templates

[http://www.cap.org/web/oracle/webcenter/portalapp/pagehierarchy/cancer\_protocol\_templates.jspx](http://www.cap.org/web/oracle/webcenter/portalapp/pagehierarchy/cancer_protocol_templates.jspx?_afrLoop=208489263945329)

1. FORDS

<https://www.facs.org/~/media/files/quality%20programs/cancer/ncdb/fords%202016.ashx>

**Appendix A**

**ETL instructions for mapping ICD-O to SNOMED**

**COMPLETE ICD-O SOURCE CODES**

Cancer diagnoses are usually represented by a combination of ICD-O-3 histology and topography codes. To map this combination to SNOMED follow these steps:

1. Transform diagnosis SOURCE VALUE
	1. Histology code. In the source, it is normally formatted like this: 8070/3, where 8070 is histology type and 3 is tumor behavior. If histology type and behavior are stored separately, concatenate them to get one histology concept, e.g. 8070/3.
	2. Topography code. the source, it is normally formatted like this: C50.2. Be aware of the dot. if the source doesn't have the dot, insert it after the 3d character: C502 -> C50.2. If the source code contains only 3 characters, the dot is not required: C50 -> C50.
	3. Source value. Concatenate histology code and topography code using hyphen: 8070/3-C50.2. This value will be stored in the CONDITION\_OCCURRENCE.CONDITION\_SOURCE\_VALUE field.
2. Extract value of diagnosis SOURCE CONCEPT ID

Concept ID for the combined histology/topography code is stored in the CONCEPT table. The following SQL shows how to extract its value for the above example:

SELECT CONCEPT\_ID

FROM CONCEPT

WHERE CONCEPT\_CODE = ‘8070/3-C50.2’

 AND VOCABULARY\_ID = ‘ICDO3’

The resulting value 36517865 will be stored in the CONDITION\_OCCURRENCE.CONDITION\_SOURCE\_CONCEPT\_ID field and will be used in mapping to a standard SNOMED code (next section).

1. Extract value of STANDARD CONCEPT ID

Source concept ID of the combined histology/topography code is mapped to a standard concept ID in the CONCEPT\_RELATIONSHIP table. The following SQL shows how to extract its value for the above example:

SELECT CONCEPT\_ID\_2

FROM CONCEPT\_RELATIONSHIP

WHERE CONCEPT\_ID\_1 = 36517865

 AND RELATIONSHIP\_ID = 'Maps to'

The resulting value [36517865] will be stored in the CONDITION\_OCCURRENCE.CONDITION\_ CONCEPT\_ID field.

**INCOMPLETE ICD-O SOURCE CODES**

In some cases when the source data are incomplete, apply the following approach.

1. Tumor behavior is not known

Use 1 (uncertain behavior) to making your code complete: 8070 -> 8070/1

1. Topography is unknown.

Use mappings from this file <https://seer.cancer.gov/tools/conversion/ICD03toICD9CM-ICD10-ICD10CM.xls> (last 3 tabs of this file) to obtain topography if you have ICD-10 code for this diagnosis. Note, if you have long ICD-10CM code, you need to cut it off to have only 5 symbols (including dot): C50.211 -> C50.2

In case when a patient has several cancer diagnoses, use ICD-10 from the date closest to the ICD-O histology date.

**Appendix B**

**ETL instructions for mapping Treatment abstractions.**

The varying levels of grouping/abstraction of lower-level clinical events into TREATMENTS available within source systems will require different ETL strategies.

1. Oncology EHR contains TREATMENT groupings/abstractions natively.
	* No algorithmic derivation necessary. Use OROT to map clinical event codes to TREATMENT concepts. Insert low-level clinical events, the grouping/abstraction structures and the connections between them.
2. EHR records administrations/prescriptions of the drugs in a chemotherapy regimen or each fraction of a radiation therapy treatment. No grouping/abstractions natively.
	* Insert the low-level clinical events. Algorithmically derive TREATMENT abstractions/groupings and connections between them. Use OROT to map clinical event codes to TREATMENT concepts.
3. Tumor Registry records that a chemotherapy regimen or a radiation therapy treatment occurred and an EHR records administrations/prescriptions of the drugs in a chemotherapy regimen or each fraction of a radiation therapy treatment. No grouping/abstractions natively.
	* Insert low-level clinical events and the grouping/abstraction structures. Use OROT to algorithmically derive connections between TREATMENT abstractions/groupings and low-level clinical events.
4. Tumor Registry records that a chemotherapy regimen or a radiation therapy treatment occurred.
	* Insert only into the TREATMENT grouping/abstraction structure.