New concept proposal

Oncology Diagnosis

Author	**Sylvain Pradervand, Nicolas Freundler**	**Date last updated**	**09.11.2023**
Project	SPO	**Contact**	DCC
Dataset Release	2024.1	**Consulted expert**	-

1 Rationale

Oncology Diagnosis carries information about the topography, the morphology, the observation date and the type of the cancers and is therefore the central concept to construct both cancer categorization and temporal features for any kind of analyses.

2 Comparison to other standards/data models

2.1 ICD-10-GM

ICD-10-GM terminology aims at being comprehensive to all diagnoses and oncology diagnoses are therefore only a part of it, namely Chapter II. ICD-10-GM classifies cancers according to different criteria across the different levels of a single hierarchy: malignancy, then primality, then the body site. The ICD-10-GM code can be added to the code attribute of proposed Oncology Diagnosis

2.2 ICD-O-3

ICD-O-3 is similar to ICD-10-GM, but is specific to oncology. It categorizes cancer through two classifications: the topography (body site) and the morphology (adenocarcinoma, squamous cells, etc …). It also comprehends more types than ICD-10-GM. Both ICD-O-3 topography and morphology codes can be added to the code attribute of proposed Oncology Diagnosis, and alongside ICD-10-GM code.

2.3 OncoTree

OncoTree describes itself as a dynamic and flexible classification of tumors for precision oncology. The first level of classification is by body site, then each sub-class has specific classification criteria. OncoTree codes can be a standard for the proposed Oncology Diagnosis, and alongside ICD-10-GM and ICD-O-3 codes.

2.4 NCI Thesaurus

NCI Thesaurus has three classifications: by site, by morphology and by so-called special classification. The latter covers diverse classification criteria related to aetiology, malignancy, metastases and primality. OncoTree by itself contains a mapping, if available, to NCI codes and Unified Medical Language System (UMLS) codes.

2.5 NICER

The National Institute for Cancer Epidemiology and Registration (NICER) promotes cancer registration and cancer epidemiological studies. NICER provides a list of variables covering diagnosis, first treatment and course of the disease. NICER diagnosis-related fields are date of informing the patient, date and type of cancer, type and characteristics of tumor, tumor spread at the time of diagnosis, stage of disease, diagnostic method and method of first detection (for instance diagnosis basis, *Diagnosegrundlage*, has codes for Clinical, Imaging, Histology of primary tumor, etc), tumor-related prognostic factor. NICER's data dictionary uses ICD-O-3.

2.6 ENCR

The European Network of Cancer Registries (ENCR) is, from our perspective, the European equivalent of NICER. ENCR's ontology uses ICD-O-3.

3 Concept information

Concept or concept compositions or inherited	General concept name	General description	Contextualized concept name	Contextualized description	Type	Standard	Value set or subset	Meaning binding	Cardinality of ComposedOf	Additional Information
concept	Oncology Diagnosis	determination of the presence of an oncological disease, from expressed signs and symptoms and assessments such as biopsy, tumor marker test, imaging, or the like	Oncology Diagnosis	determination of the presence of an oncological disease, from expressed signs and symptoms and assessments such as biopsy, tumor marker test, imaging, or the like	Diagnosis			SNOMED CT: 363346000 | Malignant neoplastic disease (disorder) |		
inherited	code	coded information specifying the concept	code	coded information describing the oncology diagnosis	Code	ICD-10-GM, ICD-O-3 Morphology, ICD-O-3 Topography, OncoTree			1:n	
inherited	record datetime	datetime the concept was recorded	record datetime	datetime the oncology	temporal				0:1	

				diagnosis was recorded						
inherited	subject age	age of the individual at the time of the event	subject age	age of the individual at the time of the oncology diagnosis	Age				0:1	
composedOf	incidence datetime	datetime of incidence of the concept	incidence datetime	datetime of incidence of the oncology diagnosis	temporal				1:1	corresponds to date of incidence in Swiss national cancer data dictionary
composedOf	body site	anatomical site or structure associated to the concept	primary tumor location	origin site of malignant disease	Body Site				0:1	

General concept	**Cardinality for concept to Administrative Case**	**Cardinality for concept to Provider Institute**	**Cardinality for concept to Subject Pseudo Identifier**	**Cardinality for concept to Source System**
Oncology Diagnosis	0:1	1:1	1:1	1:n

*The date of the first event (of the 7 listed below) to occur chronologically should be chosen as incidence date. If an event of higher priority occurs within three months of the date initially chosen, the date of the higher priority event should take precedence.
Order of declining priority:
1. Date of first histological or cytological (including flow cytometry, liquid biopsy) confirmation of this malignancy (with the exception of histology or cytology at autopsy). This date should be, in the following
order:
a) date when the specimen was taken
b) date of receipt by the pathologist
c) date of the pathology report.
2. Date of first positive genomic/molecular test diagnostic of this malignancy (see examples)
3. Date of admission to the hospital because of this malignancy.
4. When evaluated at an outpatient clinic only: date of first consultation at the outpatient clinic because of this malignancy.
5. Date of diagnosis, other than 1, 2, 3 or 4, for example:
a) date of first positive tumor marker test diagnostic for this malignancy
b) date of first imaging (includes PET, CT or MRI) diagnostic for this malignancy
c) date of multidisciplinary team meeting (MDT) for this malignancy.
6. Date of death, if no other information is available other than the fact that the patient has died because of a malignancy.
7. Date of death, if the malignancy is discovered at autopsy.
Whichever date is selected, the date of incidence should not be later than the date of the start of the treatment, or decision not to treat, or date of death. The choice of the date of incidence does not determine the coding of the item "basis of
diagnosis".

2. Date of first positive genomic/molecular test diagnostic of this malignancy
Examples of molecular tests that could be used to define incidence date
- T-cell receptor rearrangement – T-cell lymphoma
- BCR-ABL fusion gene (Philadelphia chromosome) – Chronic myeloid leukaemia, acute lymphoblastic leukaemia, and acute myelogenous leukaemia
- JAK2 gene mutation – myeloproliferative neoplasms
- PML/RARα fusion gene – Acute promyelocytic leukaemia
- Circulating tumor DNA (ctDNA) – as part of diagnosis and cancer screening in future
Source ENCR 2022
https://encr.eu/sites/default/files/Recommendations/ENCR%20Recommendation%20DOI_Mar2022.pdf

4 Impact on the SPHN Dataset

ICD-O-3 Diagnosis is not required as it is fungible in Oncology Diagnosis through the multiplicity of the code composedOf.

5 Discussion

5.1 ENCR date of incidence

The used definition of incidence date comes from a document by NICER in 2017. For informational purposes only, here is reported a guideline published by ENCR in 2022. It is worth noticing that the indications are almost literally the same, but this version has a new item in priority, its point 2. The examples they provided for these new points are also reported here.

The date of the first event (of the 7 listed below) to occur chronologically should be chosen as incidence date. If an event of higher priority occurs within three months of the date initially chosen, the date of the higher priority event should take precedence.
Order of declining priority:
1. Date of first histological or cytological (including flow cytometry, liquid biopsy) confirmation of this malignancy (with the exception of histology or cytology at autopsy).
This date should be, in the following order:
a) date when the specimen was taken
b) date specimen received by pathologist
c) date of the pathology report.
2. Date of first positive genomic/molecular test diagnostic of this malignancy (see examples)
3. Date of admission to the hospital because of this malignancy.
4. When evaluated at an outpatient clinic only: date of first consultation at the outpatient clinic because of this malignancy.
5. Date of diagnosis, other than 1, 2, 3 or 4, for example:
a) date of first positive tumor marker test diagnostic for this malignancy
b) date of first imaging (includes PET, CT or MRI) diagnostic for this malignancy
c) date of multidisciplinary team meeting (MDT) for this malignancy.
6. Date of death, if no other information is available other than the fact that the patient has died because of a malignancy.
7. Date of death, if the malignancy is discovered at autopsy.
Whichever date is selected, the date of incidence should not be later than the date of the start of the treatment, or the decision not to treat, or the date of death. The choice of the date of incidence does not determine the coding of the item "basis of diagnosis".

2. Date of first positive genomic/molecular test diagnostic of this malignancy
Examples of molecular tests that could be used to define incidence date
• T-cell receptor rearrangement – T-cell lymphoma
• BCR-ABL fusion gene (Philadelphia chromosome) – Chronic myeloid leukaemia, acute lymphoblastic leukaemia, and acute myelogenous leukaemia
• JAK2 gene mutation – myeloproliferative neoplasms
• PML/RARα fusion gene – Acute promyelocytic leukaemia
• Circulating tumor DNA (ctDNA) – as part of diagnosis and cancer screening in future

Source ENCR 2022
https://encr.eu/sites/default/files/Recommendations/ENCR%20Recommendation%20DOI_Mar2022.pdf

UPDATE: the most recent definition will be used, namely the one of ENCR, March 2022. The previous one is reported below.
The date of the first event (of the six listed below) to occur chronologically should be chosen as incidence date. If an event of higher priority occurs within three months of the date initially chosen, the date of the higher priority event should take precedence.
Order of declining priority:
1. Date of first histological or cytological confirmation of this malignancy (with the

exception of histology or cytology at autopsy). This date should be, in the following order:
a) date when the specimen was taken (biopsy)
b) date of receipt by the pathologist
c) date of the pathology report.
2. Date of admission to the hospital because of this malignancy.
3. When evaluated at an outpatient clinic only: date of first consultation at the outpatient clinic because of this malignancy.
4. Date of diagnosis, other than 1, 2 or 3.
5. Date of death, if no information is available other than the fact that the patient has died because of a malignancy.
6. Date of death, if the malignancy is discovered at autopsy.
Whichever date is selected, the date of incidence should not be later than the date of the start of the treatment, or decision not to treat, or date of death.
The choice of the date of incidence does not determine the coding of the item "basis of diagnosis".
ENCR clarification:
Incidence is the "date of diagnosis i.e. the date of confirmation of the invasive cancer". Only this cancer may be counted as "incidence".
However, there are cases where "in situ" or "highly suspicious" is reported first (e.g. breast) and later on this changes to invasive cancer (e.g. during the operation invasive parts are found or as a result of the first cytology/second biopsy). Some colleagues prefer to use the date of "in situ" or "highly suspicious" diagnosis and some prefer the date of invasive diagnosis. It is the view of the ENCR Steering Committee that only the latter is in compliance with ENCR/IARC.
source NICER Core Dataset 2017
(https://www.nicer.org/archive/assets/files/data/ncd_4.1_abbrev_version_201706.pdf)

5.2 ICD-O-3 topography

It has been discussed if the topography codes of ICD-O-3 terminology should be put into the body site composedOf because, semantically, the topography (within the human body) is a body site and the other way around. However, it has been decided to not move the topography coded information to the body site composedOf as:
1) it is a hierarchical terminology for diagnosis codification on an equal footing with ICD-O-3 morphology, ICD-10 and OncoTree, and
2) to a certain extent, some levels of ICD-10 and OncoTree are topographical as well.

5.3 Free text diagnosis

A free text diagnosis information from a medical report or electronic health record is desired by the project for example in case only a high level Oncology Diagnosis code can be provided. This kind of information (e.g. adenocarcinoma of left lung) can be captured in the new concept Source Data (string value).

6 Example

In the following example, there are two codes from terminology ICD-O-3 and they have different values. This is expected and they represent two different pieces of information: the first one concerns the tumor's topology and the second one its histology and behaviour. Note that the ICD-10-GM and the ICD-O-3

topological codes are the same here, but it is not the case in general (for instance malignant melanoma of the skin is under C43._ codes in the former, C44._ in the latter)

Oncology Diagnosis:
- code:
  - identifier: LUAD
  - name: Lung Adenocarcinoma
  - coding system and version : ONCO-TREE-2021-11-02
- code:
  - identifier: C34.9
  - name: Bronchus or lung, unspecified
  - coding system and version: ICD-10-GM-2019
- code:
  - identifier: C34.9
  - name: Lung, NOS
  - coding system and version: ICD-O-3.1
- code:
  - identifier: 8140/3
  - name: Adenocarcinoma, NOS
  - coding system and version: ICD-O-3.1
- record datetime: 2015-01-01
- coding datetime: 2015-01-01
- subject age: 35
- incidence datetime: 2015-01-01
- tumor location:
  - code:
    - identifier: 44029006
    - name: Left lung structure (body structure)
    - coding system and version: SNOMED-CT-2023-03-10
  - laterality:
    - code:
      - identifier: 7771000
      - name: Left (qualifier value)
      - coding system and version: SNOMED-CT-2023-03-10