Pre-coordination of structural and semantic conventions to capture breast cancer indicators in the context of the European University Hospital Alliance (EUHA)

Tijs Delabastita[b], Pieter Maertens[b], Jared Houghtaling[a], Pieter Stijnen[b], Guy Vanden Boer[b], Johan Van Eldere[b]

[a] edenceHealth NV

[b] University Hospitals Leuven (UZL)

LIMIT OF 1000 WORDS EXCLUDING REFERENCES, AUTHORS & AFFILIATIONS, AND FIGURE CAPTIONS – 21/04/2023 – 1019 words

Background (178 words)

University Hospitals Leuven (UZ Leuven) is an active participant in the European University Hospital Alliance (EUHA) and is spearheading a network-wide study investigating critical breast cancer quality indicators using data in the Observational Medical Outcomes Partnership (OMOP) common data model (CDM). One anticipated outcome of the study is an evaluation of the effects of the COVID pandemic on quality of care and patient reported outcomes for patients with breast cancer. Given the diversity in OMOP data source maturity across the consortium, UZ Leuven must accommodate existing mapping decisions while also defining clear conventions for capturing data with sufficient granularity to support the study. The intent of the work presented here is therefore three-fold: it aims (1) to describe the relevant breast cancer indicators and their underlying OMOP CDM tables, (2) to outline the process of pre-coordinating the structural and conceptual mapping approach across a network to suit the needs of a particular research question, and (3) to discuss the challenges and findings throughout this process, with an eye toward capturing community feedback regarding best practices.

Methods (90 words)

The preliminary step in establishing the particular study was to define the set of indicators. The quality indicators published by the European Society of Breast Cancer Specialists (EUSOMA) [1] were a solid starting point. During discussions with clinicians and data custodians, the set of indicators was reduced and modified based on the clinical relevance and the feasibility to capture and transform the necessary data behind the indicators for all participating partners. Later, discussions (mainly with data custodians) were organized to clarify the definitions and the data components of the indicators.

Results (523 words)

Eight out of 17 EUSOMA indicators (that often consist of multiple parts) were retained and 1 indicator on patient reported outcome measures was added (Table 1).

Table 1 Final set of indicators after reduction and modification of the original set of EUSOMA quality indicators

Indicator
Proportion of patients <= 70 yrs. with invasive breast cancer who received postoperative radiation therapy after surgical resection of the primary tumour in the framework of breast conserving therapy.
Proportion of patients with invasive cancer who underwent sentinel lymph-node biopsy with no more than 5 nodes excised.
Proportion of patients with HER2-positive invasive carcinoma treated with neoadjuvant chemotherapy who received neoadjuvant trastuzumab.
Time interval of <= 6 weeks, from the date of first diagnostic examination within the breast centre to the date of surgery or start of other treatment.
Proportion of cancer cases examined preoperatively by MRI (excluding patients treated with primary systemic treatment). Please specify subtype: lobular, ductal, mixed, uncertain.
Proportion of patients with invasive breast cancer who received a single operation for the primary tumour (excluding reconstruction).
Proportion of patients receiving immediate reconstruction at the same time of mastectomy for invasive breast cancer.
Proportion of patients (BRCA1 and BRCA2 patients excluded) with invasive breast cancer, including patients treated with neo-adjuvant systemic therapy, who underwent breast conserving therapy as primary treatment.
Patient reported outcome measures are available.

HER2 = human epidermal growth factor receptor 2; BRCA1/2 = Breast cancer gene 1/2;

From the set of indicators, 11 important data components and their underlying OMOP CDM tables were deduced. It was agreed to use the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) vocabulary for cancer diagnoses and the medical prescription normalized (RxNorm) vocabulary for systemic therapies (Table 2). It was intended to use SNOMED CT for coding surgical procedures. However, UZ Leuven's source tables contain procedures coded in the International Classification of Diseases Procedure Coding System 10th edition (ICD 10 PCS) vocabulary. Since there is no direct mapping available between two standard vocabularies, i.e. ICD 10 PCS and SNOMED CT, both the SNOMED CT and ICD 10 PCS vocabulary are allowed for coding surgical procedures.

The standard vocabularies and/or the underlying OMOP CDM tables for certain oncology specific components remain to be discussed (Table 2). Furthermore, it is still unclear how the OMOP CDM oncology extension will be implemented.

Table 2 Important data components, their underlying OMOP CDM tables and standard vocabularies

Important data components	OMOP CDM tables	Standard vocabularies
Diagnosis	Condition_occurrence	SNOMED CT
Tumour subtype	[1]	[1]
Tumour staging	Measurement	[1]
HER2-status	Measurement	[1]
BRCA1 & BRCA2-status	Measurement	[1]
Mammography – date	Procedure_occurrence	ICD 10 PCS/SNOMED CT [2]
MRI – date & type	Procedure_occurrence	ICD 10 PCS/SNOMED CT [2]
Surgery – date & type	Procedure_occurrence	ICD 10 PCS/SNOMED CT [2]
Radiotherapy – date	Procedure_occurrence	ICD 10 PCS/SNOMED CT [2]
Trastuzumab – date & type	Drug_exposure	RxNorm
Hormone therapy – date & type	Drug_exposure	RxNorm

[1] remains to be discussed; [2] we allow these vocabularies because both are standard vocabularies and standard to standard mappings are not available

SNOMED CT = Systematized Nomenclature of Medicine Clinical Terms; ICD 10 PCS = International Classification of Diseases Procedure Coding System 10th edition; RxNorm = medical prescription normalized;

Conclusion (198 words)

During this project, the main focus shifted from discussions to distil a proper set of indicators were the main focus to aligning all partners' data transformation to the OMOP CDM.

The process so far learned that organizing discussions to align the partners on every aspect of the project is essential. During these discussions, the SME contracted in the context of a Data Partner Call of the European Health Data and Evidence Network (EHDEN) often gives useful input. Furthermore, we feel that the clinical research questions behind the study also help to make deliberate choices. Still it is challenging to motivate partners to invest time in preparing these discussions.

In the near future, important challenges will be to clarify the transformation of oncology-specific data into the OMOP CDM and the use of the OMOP CDM oncology extension. Documentation on this topic seems to be immature. Later, designing and writing queries to calculate the indicators taking will be challenging due to the differences in data transformation between the partners. Many partner-specific adaptations might decrease the transparency of the study and the value of using and investing effort into a federated data model.

Bibliography

[1] “Quality indicators in breast cancer care: An update from the EUSOMA working group,” *European Journal of Cancer,* vol. 86, pp. 59-81, 2017.