

Mining Careflow Patterns in Data Warehouses of Breast Cancer Patients

Lucia Sacchi, Daniele Segagni, Arianna Dagliati,
Alberto Zambelli and Riccardo Bellazzi

Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy
IRCCS Fondazione Salvatore Mugerì, Pavia, Italy

Collecting a Data Warehouse of Oncology Patients

Collaboration between University of Pavia and the IRCCS Fondazione S. Maugeri (FSM) hospital (Pavia, Italy)

ONCO-i2b2

i2b2 DW infrastructure to integrate FSM hospital information system (HIS) and a cancer biobank that manages both plasma and cancer tissues

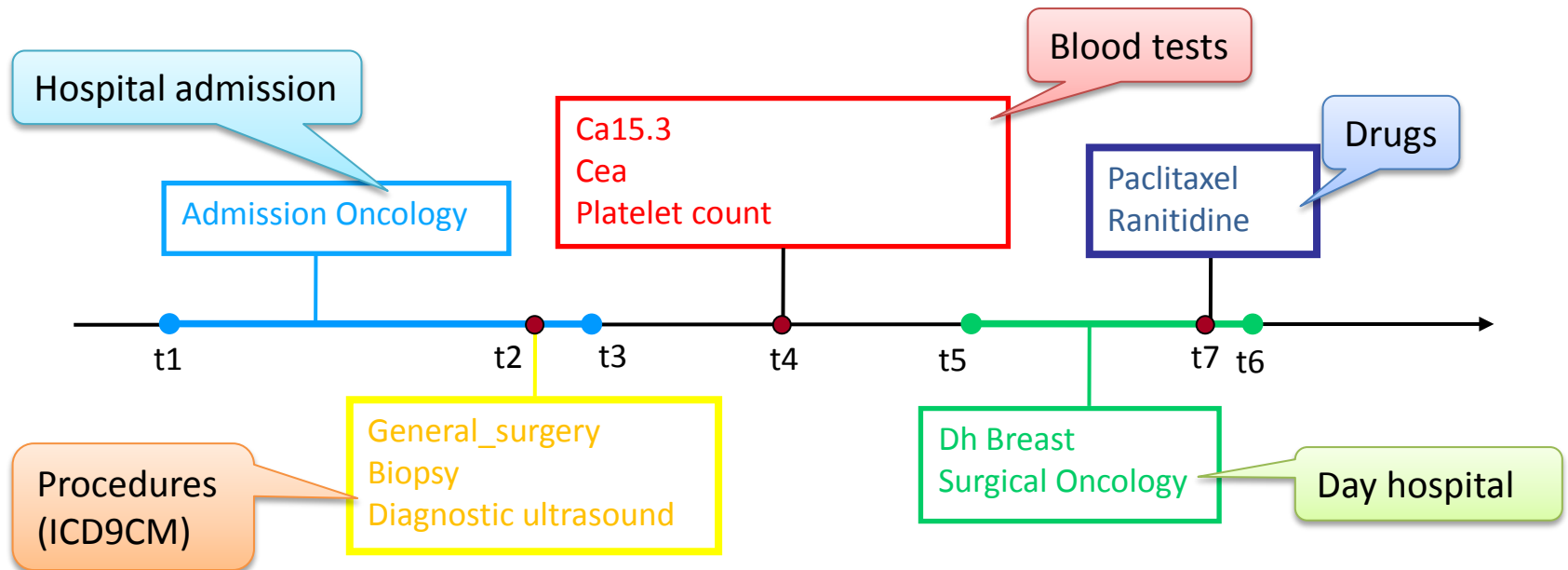
Currently data related to 28,838 oncology patients

Knowledge discovery to find frequent, meaningful paths and build structured scenarios

The Data Warehouse

Information on:

Admission, Day Hospital, Procedures, Drugs, Blood tests



For each patient: integrated temporal clinical history

How to Exploit this Data for Research?

Objective: extract the most frequent **careflows** from clinical histories

PROCESS MINING

- Mines workflows from event logs
- Doesn't take into account quantitative clinical data (only "process" data)

TEMPORAL DATA MINING

- Able to extract frequent patterns from clinical data with temporal features (e.g. TARs)
- Cannot generate complex temporal histories (e.g. chains of events)

How to Exploit this Data for Research?

Objective: extract the most frequent **careflows** from clinical histories

PROCESS MINING

- Mines workflows from event logs
- Doesn't take into account quantitative clinical data (only "process" data)

TEMPORAL DATA MINING

- ~~Able to extract frequent patterns from clinical data with temporal features (e.g. TARs)~~
- Cannot generate complex temporal histories (e.g. chains of events)

Methods

Use process mining techniques to extract frequent flows of care in the data warehouse

Enrich the mined careflows with clinical information

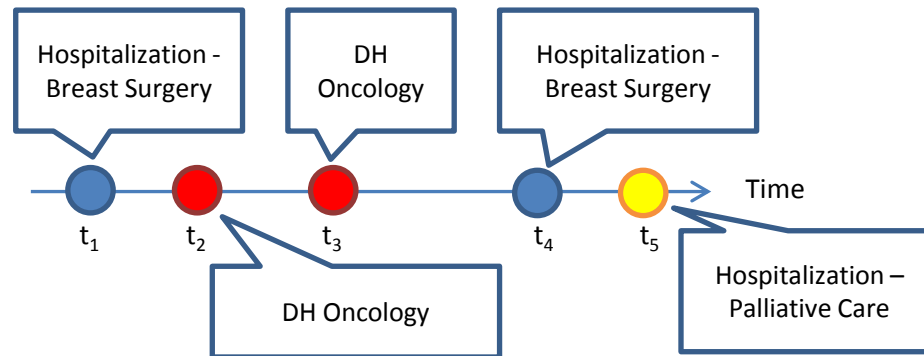
Methods

Use process mining techniques to extract frequent flows of care in the data warehouse

Enrich the mined careflows with clinical information

Our Process Mining Approach

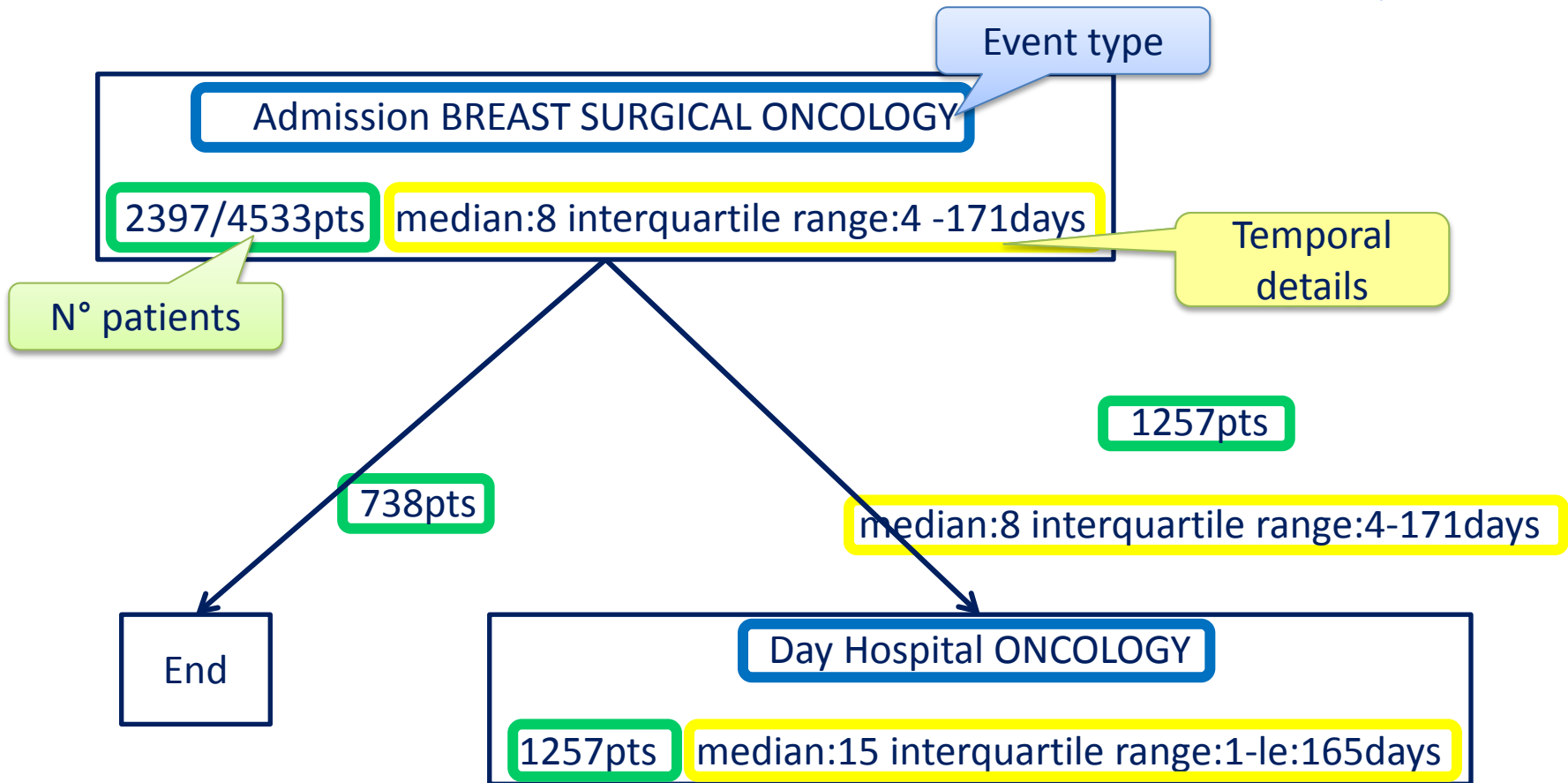
Currently works on sequences of in-hospital and day hospital (DH) accesses to the different FSM wards



Allows extracting the most frequent clinical paths on the basis of threshold parameters

Each step of a clinical pattern is enriched with temporal information

Clinical Histories Characterization

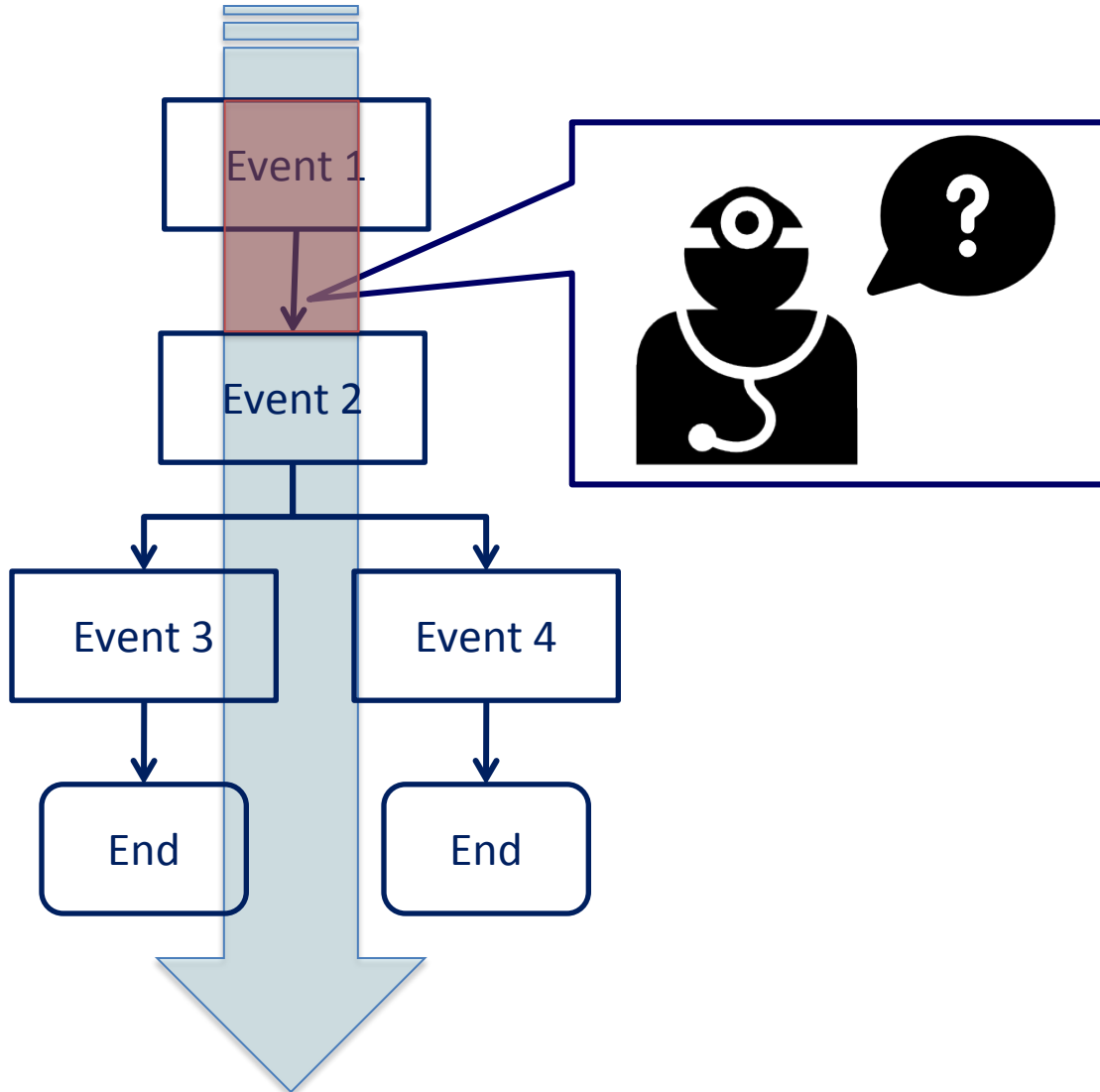


Methods

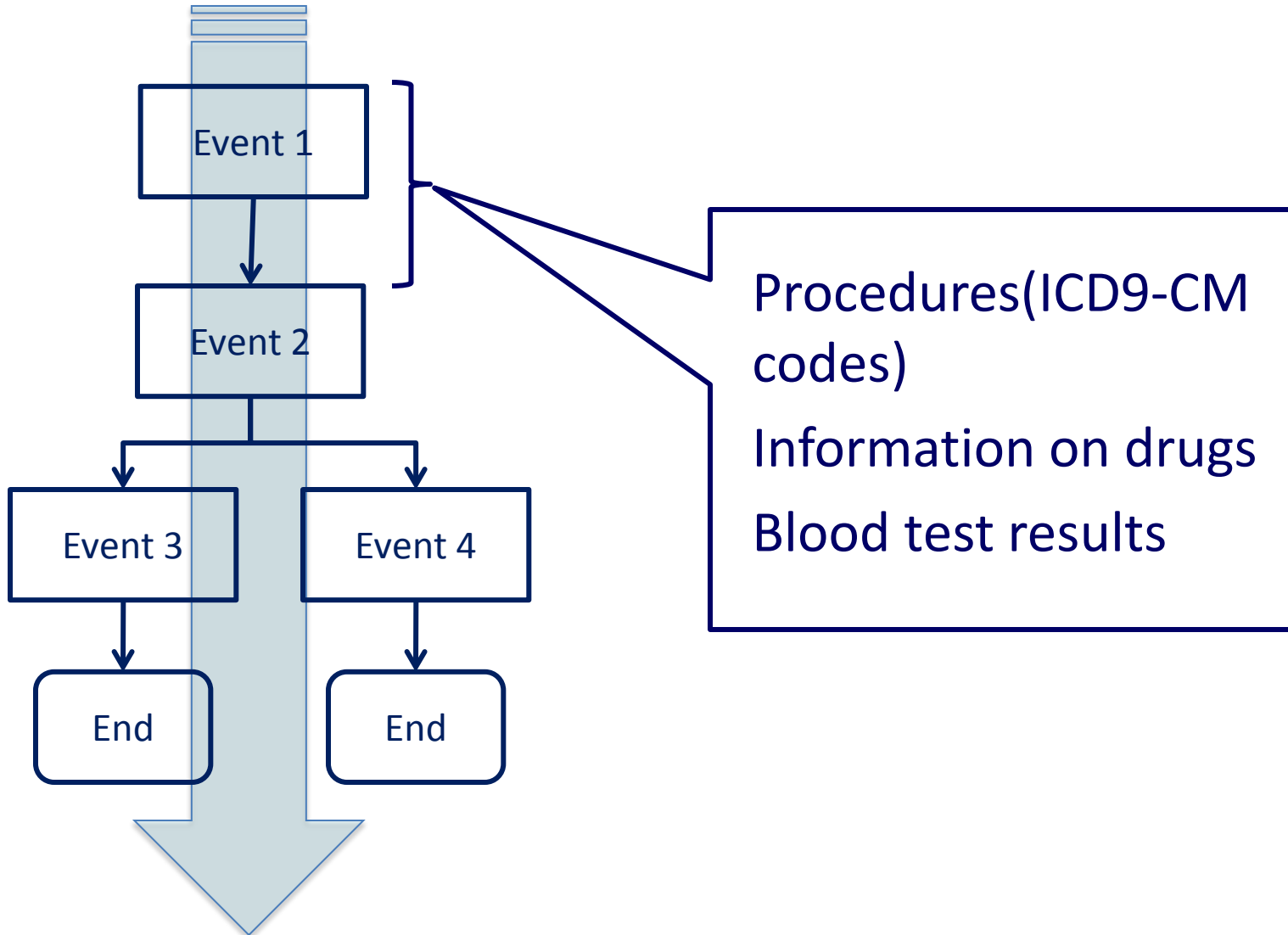
Use process mining techniques to extract frequent flows of care in the data warehouse

Enrich the mined careflows with clinical information

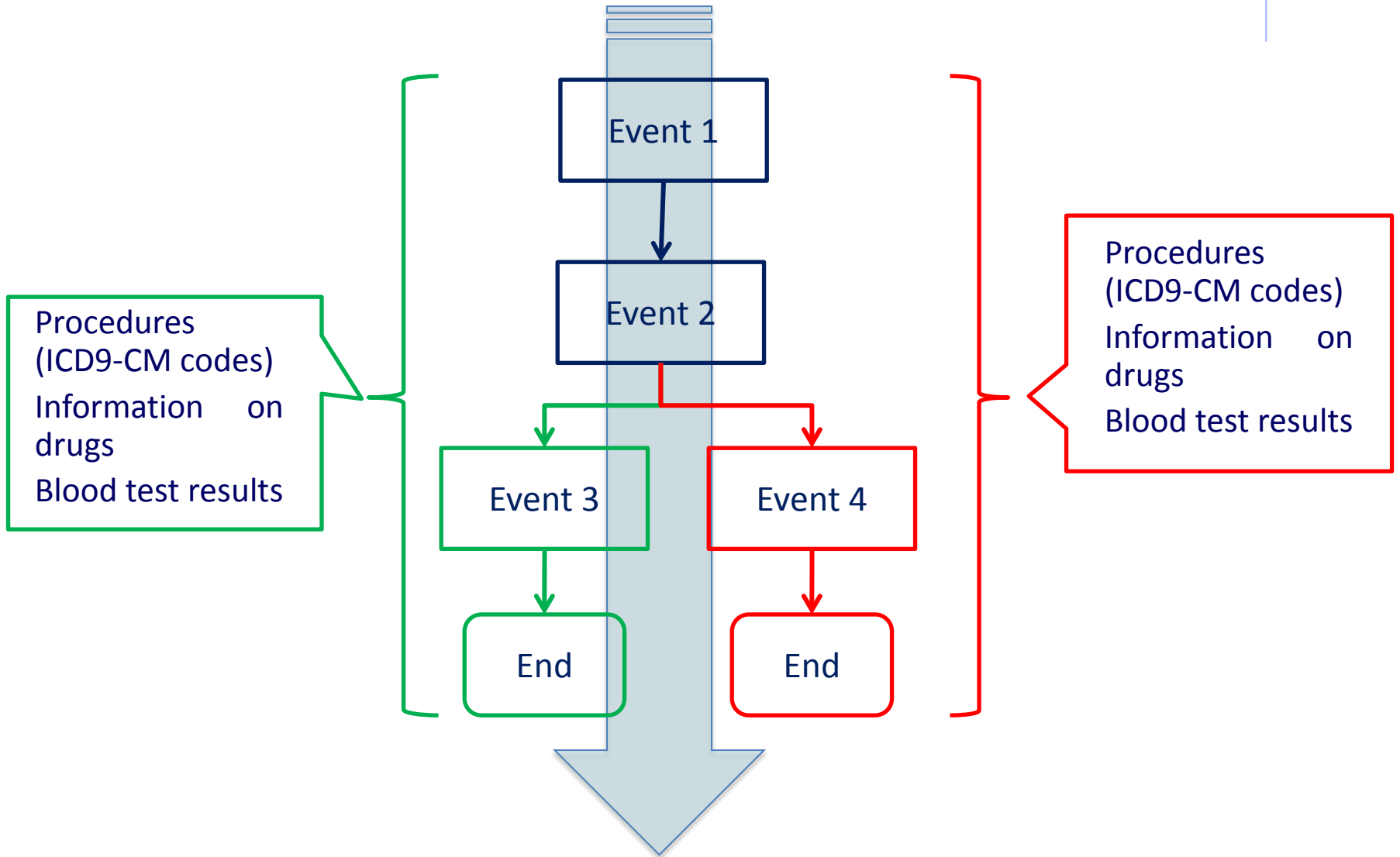
Enriching Clinical Histories



Enriching Clinical Histories



Enriching Clinical Histories



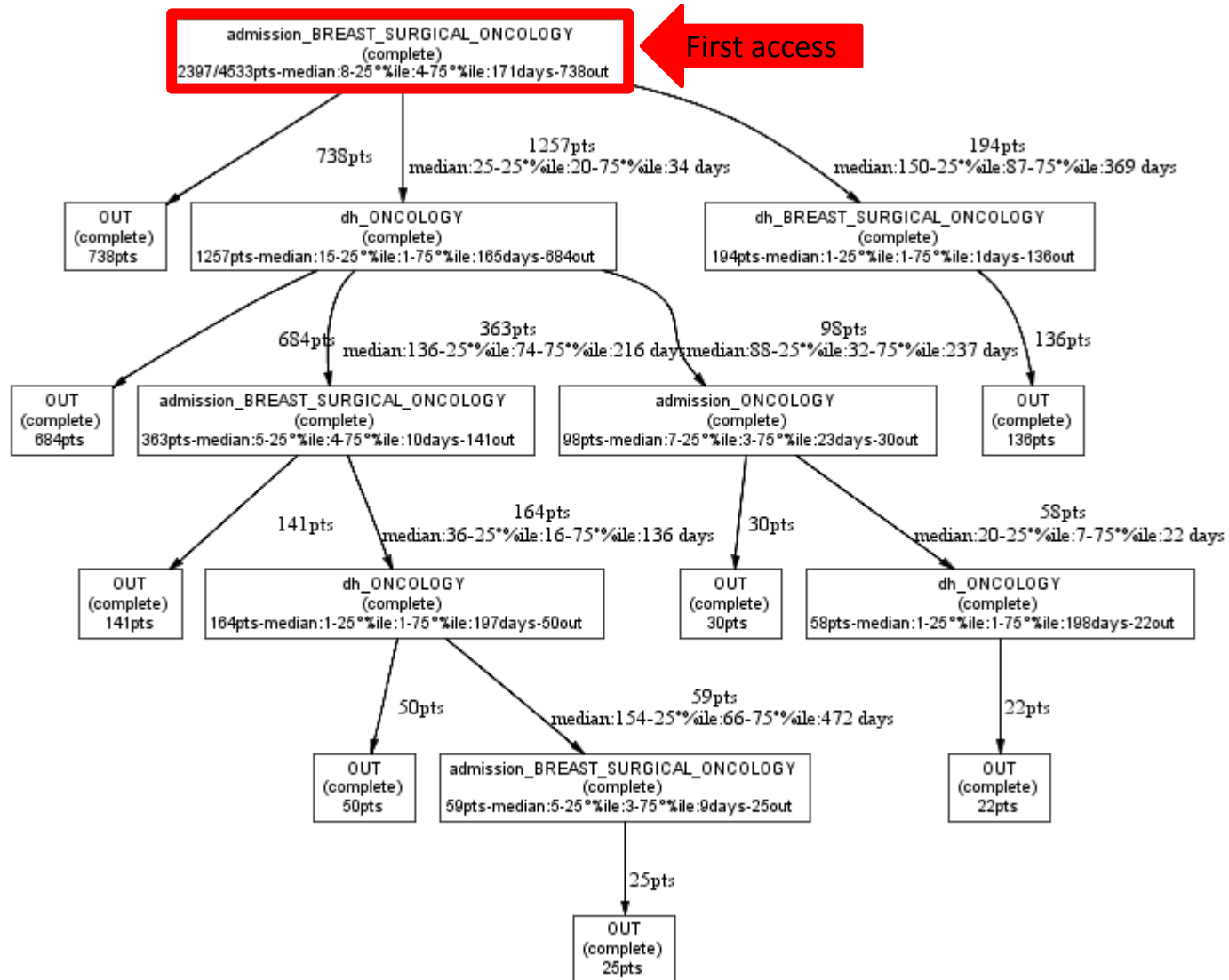
Results – A Case Study on Breast Cancer

Patients with a diagnosis of malignant neoplasm of female breast (ICD9-CM code 174.0-174.9) or personal history of malignant neoplasm (ICD9-CM V10.3)	8.727
Number of patients with a single hospital admission	2.713
Number of patients with a single procedure	420
Number of patients with a single DH	227
Number of patients with at least one event before 2000	1446

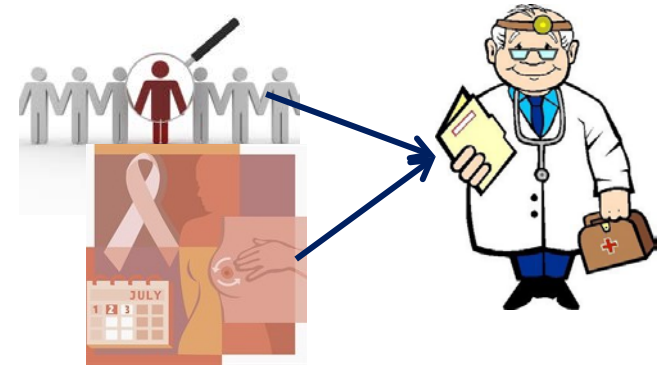
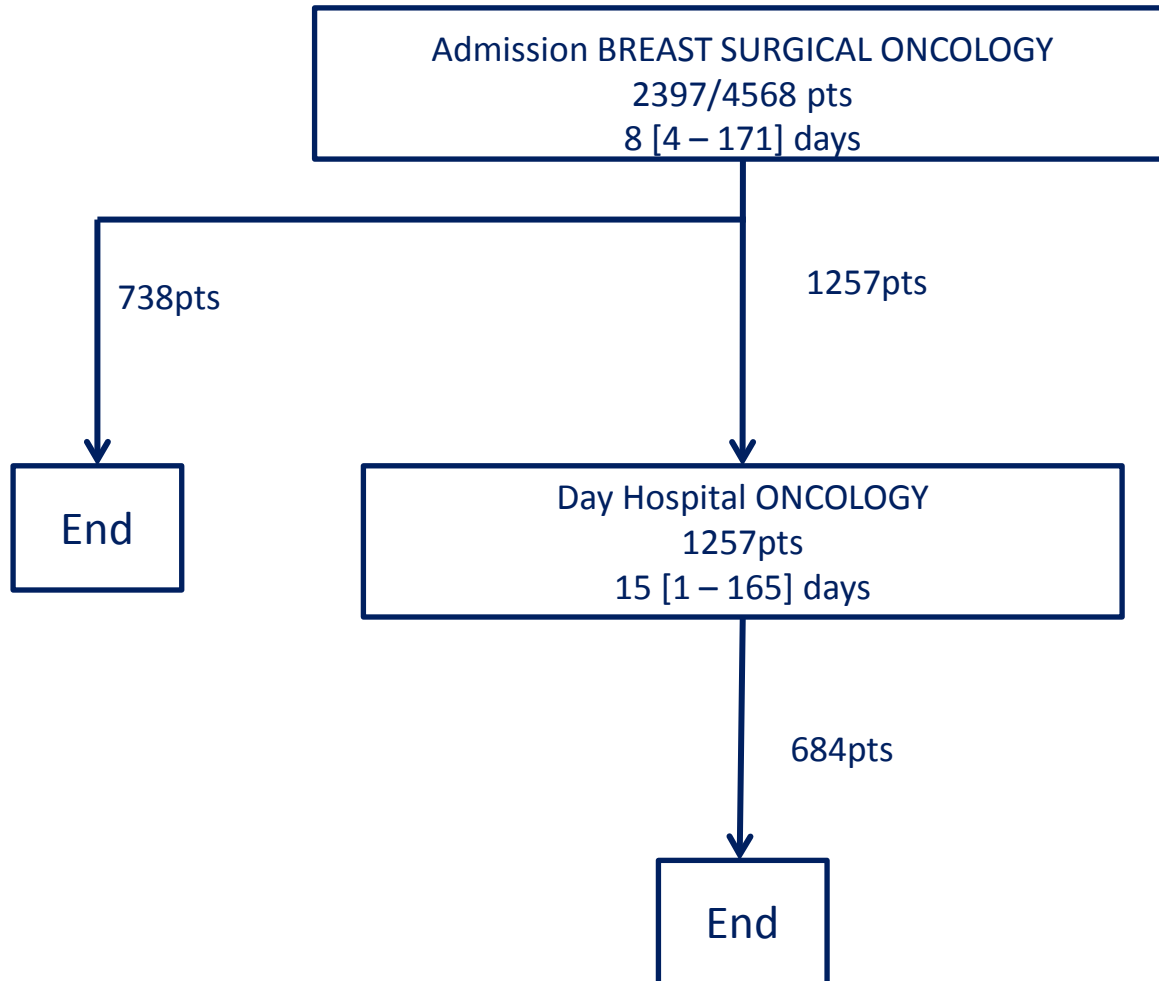


4.568 subjects

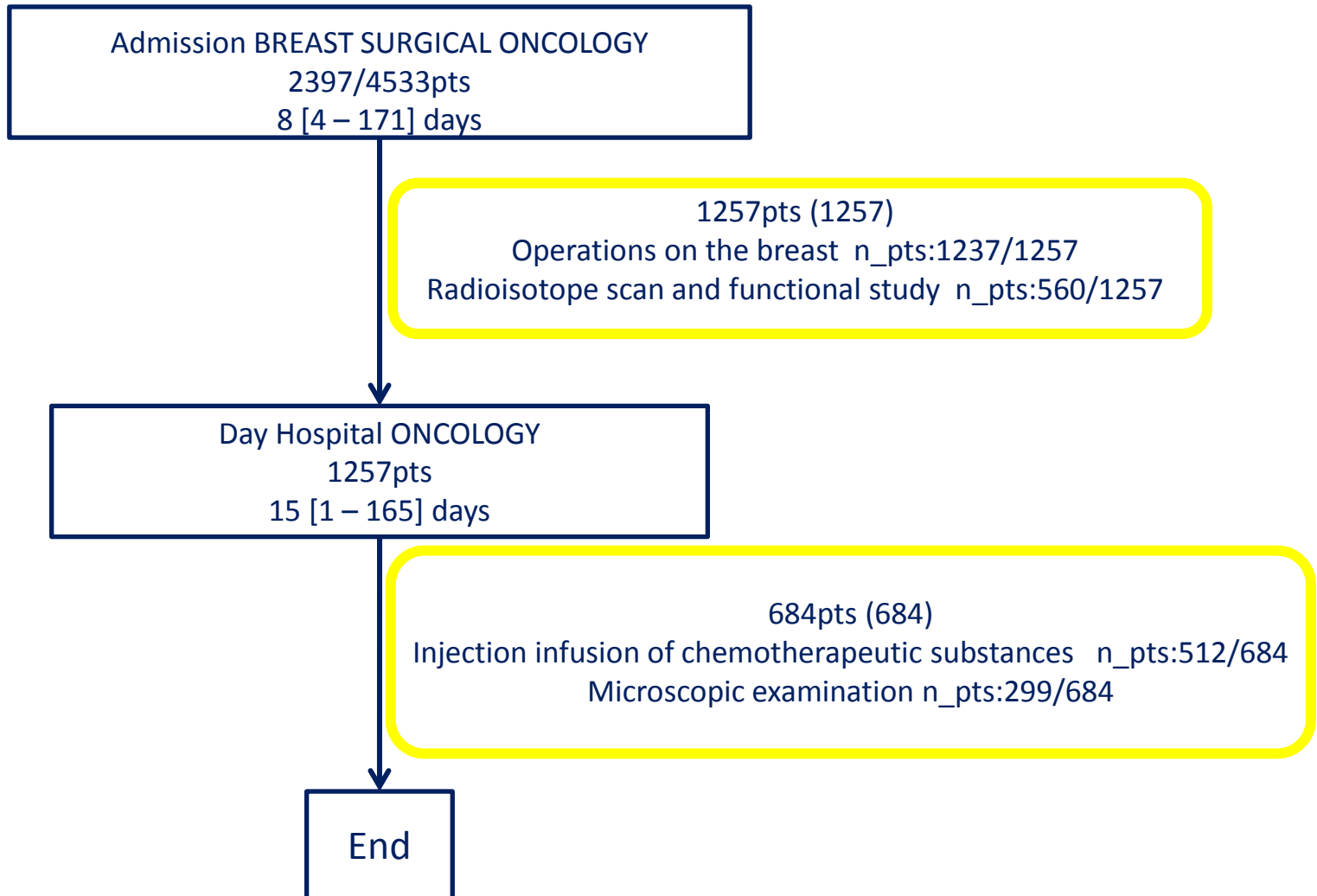
Results



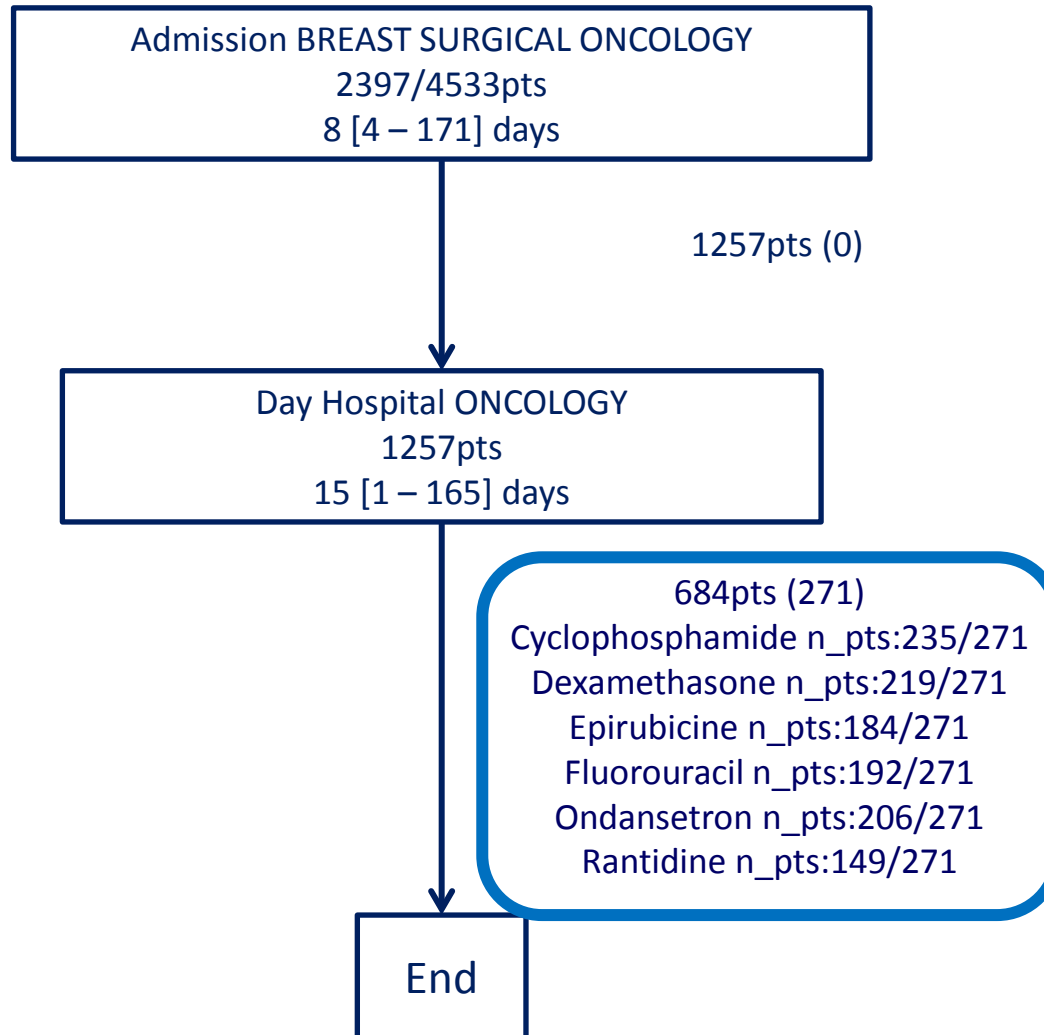
Most Frequent Clinical Pattern



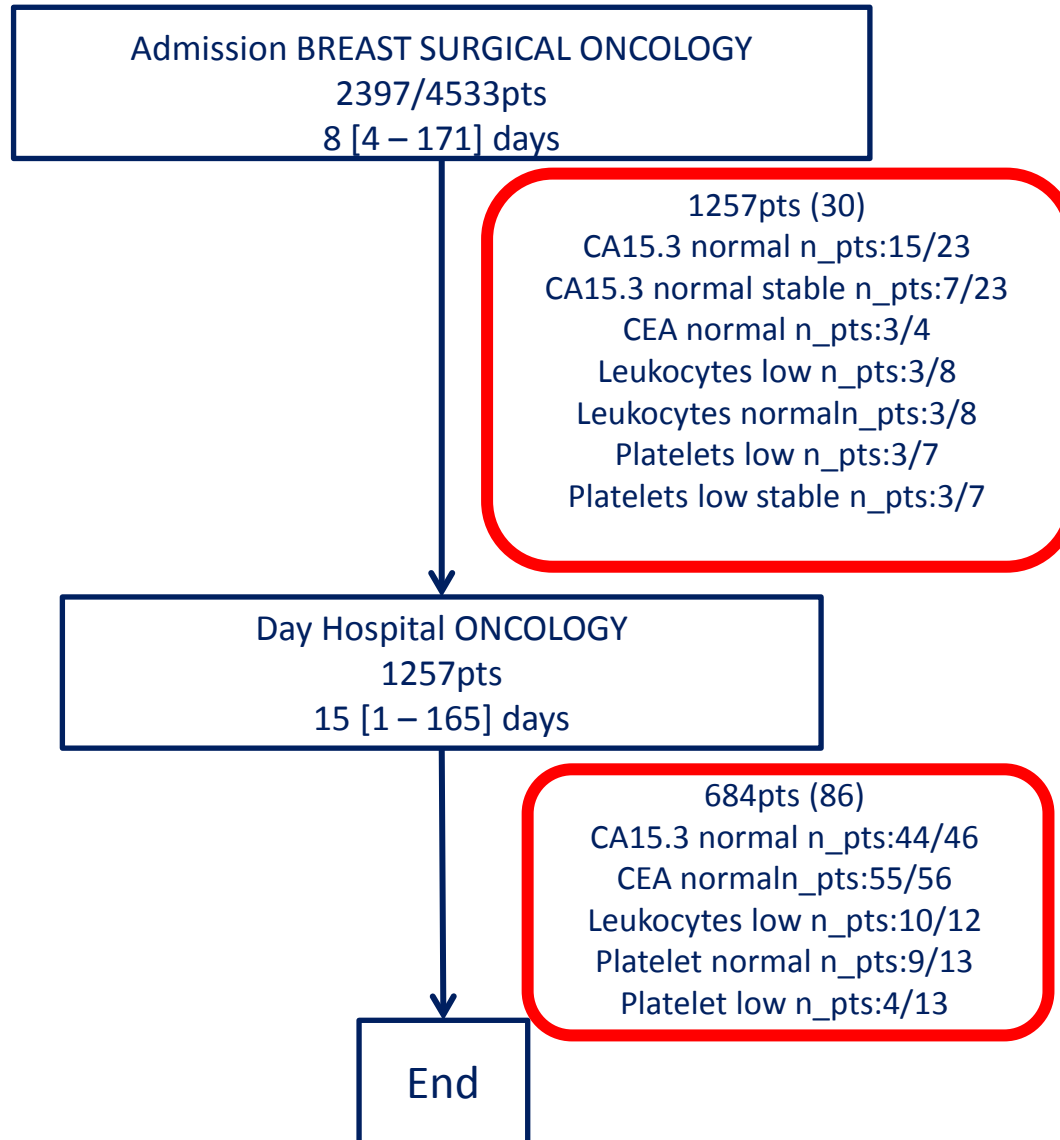
Enriching the History - Procedures



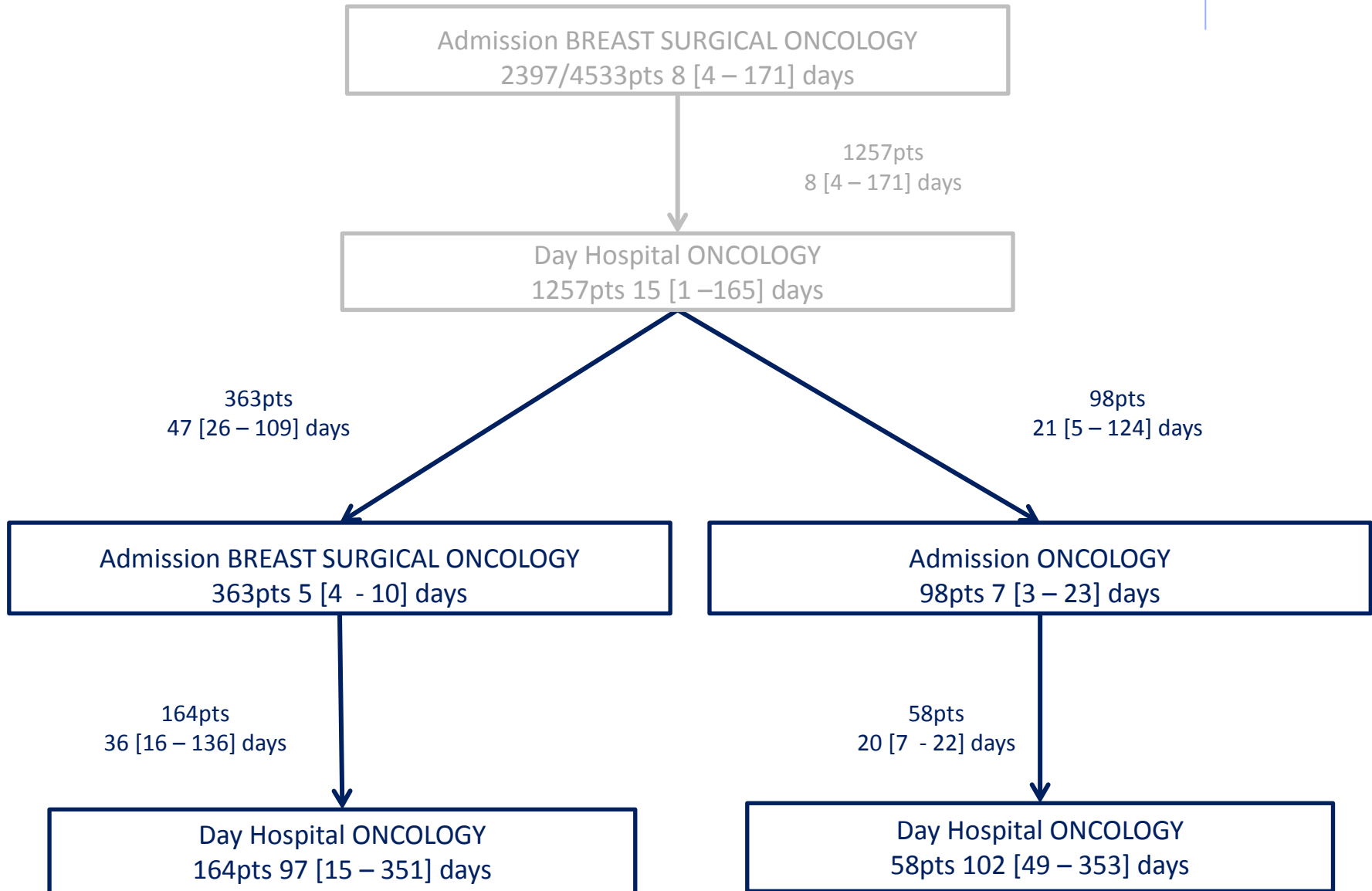
Enriching the History - Drugs



Enriching the History – Blood Tests



Patients with Complications



Conclusions

- Heterogeneous information collected in the ONCO-i2b2 Data Warehouse used for research.
- A methods able to extract frequent careflows composed by clinical and administrative events.
- Data Mining approaches exploitation to create an enriched careflow model.
- Add more drugs to the analyzed spectrum (only oncology drugs up to now)
- Stratify patients on the cancer morphological features to check if different types of tumors lead to different careflows.

Thank you