

**Regional cancer registry data**  
(source datasets)

**Sample of 159 breast cancer patients**

**Data collected from medical files**  
(medical reports)

**Algorithm**

**Representation of sequences**  
(chronological aggregation of events into states  
according to the level of granularity  
derived from clinical practice guidelines)

**Manually**

**Algorithm sequences**

Simple form  
Extended form

**Observed sequences**

Simple form  
Extended form

**Evaluation by similarity measure**  
(Levenshtein Edit Distance)