**Sequential data analysis tool**

The aim of this project is to develop a standalone tool in java to analyze the self-management behavior of renal transplant patients collected over a period of year.

List of example questions the tool should help to answer are:

* what time of the day and on what day do people measure themselves?
* what time of the day and on what day do they enter measure measurement?
* What is the difference in time between measurement device and entering data in Mijnnierinzicht?
* Is there a difference between StatSensor measurement and what patients enter into Mijnnierinzicht
* How often do patients measure themselves before they enter data into Mijnnierinzicht?
* If a patient did measure multiple time, what measure do he/she eventually enter into Mijnnierinzicht?
* Do patient follow up advice given my Mijnnierinzicht?
* What are the conditions under which people start deviating from their normal measurement routine?
* What are the conditions under which people overwrite their initial data entered in Mijnnierinzicht?
* Find cases where there is a difference between device measurement and what is entered in Mijnnierinzicht
* find cases where Mijnnierinzicht advice to contact the hospital
* How well do patients follow up advice of Mijnnierinzicht to re-measure again?
* Do external factors, such as holidays, have an effect on the patient’s measurement routine?
* Is the value of creatine, blood pressure, HR effecting the patient’s measurement routine?

Key components of the tool with should be supported with graphical user interface.

**Input module (must have)**

* Editor for user to specify the structure of data files (XML). User should able to store this specification file and read in again.
* Based on data specification file (XML) created by the user, this module able read raw data files from:
  + creatine data statSensor
  + blood pressure, hand held device
  + hard rate, hand held device
  + Mijnnierinzicht website: creatine level, advice, blood pressure, HR, weight
  + hospital appointments data

**Analysis specification module (must have)**

* editor in which user can specify sequential data operations that needs to be carried out on the data
* editor to specify codes that could be linked with observed patterns
* interpreter of data operation files created in editor by the user

**Analysis module**

This module consisted of the parser functions the user can use on data or processed data, and functions to store the results in that can be read by statistical application such as R or SPSS. (must have)

* Chunking algorithm (must have)
  + Automatic breaking up that event sequence for specific time period to analyze frequency specific codes in separate blocks (e.g. breaking data up in months and each month record frequency of specific codes)
* Coding algorithms (must have)
  + automatic coding of behavior patterns (e.g. single device measurement – single web entry; multiple device measurement – single web entry)
  + automatic recoding event sequence from high frequency events domain (low level events: measurement device, entry into website) to low frequency event domain (more abstract: normal monitoring, additional monitoring, unscheduled monitoring)
* Conversion algorithms (should have)
  + automatic generating behavior of web site response (e.g. advice to do nothing, advice to re-measure, advice to contact hospital) in order to creating new input stream for analysis
* Connection algorithms (must have)
  + automatic data synchronization whereby events recorded in different data files (e.g. measure device, and web site, or national holidays) are merged into single data file with right event order
* Constrains algorithms (could have)
  + Automatic filtering data for certain event codes, or time periods
* Comparison algorithms
  + Lag sequential analysis (LSA) for identifying dependency (temporal relationship) between events (e.g. A->B, or B->A) (must have)
  + State Transition matrix/markov chain (e.g. count how often did patterns AB, AC, AA, occurs), higher order State transition matrix (e.g. count how often did patterns A??C, or A??B happened) (should have)
  + Time between events (e.g. time between measurement with device and data entered into web site) (must have)

**Visualization module**

* User supported in exploring the data visually. This could be raw data, or data that has been passed.
* User can specify what data need to be visualized (must have)
* Besides showing visualization (must have)
* graph should also be store (could have)
* Required visualizations:
  + Frequency Bar of specific events (must have)
  + Frequency of time between measurements:
    - stem and leave (should have)
    - histogram (could have)
    - box plot (must have)
  + Temporal relationships:
    - markov chain graph (could have)
    - state transition matrix (must have)
  + Time series: 2D Line graph, frequencies of occurrence of specific events (Y) set against months (X) (could have)