**Methods**

This is a summary of what I have done so far.

We aim to predict, from a current snapshot of patients in the ED, the number of admissions to anticipate from this group. This will be done room by room. The indicators used to train the model are demographic information (age and sex), flow sheet readings (temperature, heart.rate), and lab readings. The model will produce a probabilistic prediction of admission for each patient. These predictions can be evaluated by the hosmer-lemeshow test and the MADCAP. The predictions can be converted into a probability distribution function for the number of admissions from this group and evaluated by the distribution evaluation (see Model Evaluation)

Focusing on the movements of patients that have visited the ED in EMAP, I have organized a bed moves matrix in which each stop a patient visited is stamped with an admission and discharge date and time. Attached to these rows is their ultimate fate of admission as a TRUE or FALSE. However, it is possible to label them with a finer classification, such as the ward in which the patient is admitted. Also, the patient’s age and sex are calculated and added as an indicator. If desired, one can add trajectory information, that is, the order of previously visited rooms and/or length of stay information, that is, the amount of time since arrival at ED.

The flow sheet readings are added to the matrix by adding a new column for every possible reading and attaching to each row of data the most recent reading with respect to the discharge time of that particular row. For each reading, NA is placed if the patient does not have the reading.

Similarly, lab readings are added to the matrix by adding a new column for each possible lab. While there are many possible labs, when using the historical data during the learning step, most labs are not performed for any patients and hence can be removed from possible indicators at this later step.

With the design matrix containing demographical, flow, and lab information, I have removed the columns which holds only one value and hence not valuable in training. One can also add new features to the matrix such as number of flow sheet readings. Running through the xgboost model will yield predictions of admission along with the probability predicted based on the model. The historical data are chosen to be those patients that are admitted in a set time period before the time in which the prediction takes place. The period currently set is one month.

These probabilities can be used to generate a distribution for the probability that patients will be admitted. This is done through the use of generating functions (for details, see admission\_generatingfunction.docx). To evaluate how well these distributions are, these distributions are generated over a period of time. By setting a threshold between 0 and 1, the distributions give us an expected number of times the number of admissions should be bounded by a parameter determined by the threshold and we compare this to the actual number of times this happens. For the details of this method, refer to Model Evaluation.docx.

The prediction probabilities obtained from the model are also enough for us to do the Hosmer-Lemeshow test and MADCAP. Refer to Model Evaluation.docx for a more detailed description.