Patient utilization profile

I referred to the data description provided in *Table 1* of research paper located here [(http://downloads.hindawi.com/journals/bmri/2014/781670.pdf)]. Here are few factors we can consider to create the utilization profile.

num_procedures

These are number of procedures performed during given encounter. If patient has muliple encounters, we will use average number of procedures `avg_num_procedures`.

num_lab_procedures

These are number of lab test performed during given encounter. If patient has multiple encounters, we will use average number of lab test `avg_num_lab_procedures`.

num_medications

These are number of distinct generic names administered during the encounter. If patient has multiple encounters, we will use average number of medications administered `avg_num_meds`.

number_outpatient

This shows number of outpatient visits of the patient in the year preceding the encounter. As we don't have a date of encounter in the data, it will be difficult to determine which is the most recent encounter if a patient has more than one encounter. We could use the `max(encounter_id)` for given patient assuming the encounter with largest `encounter_id` is the most recent. Other option, for simplicity, we can take average of `num_outpatient` for patients with multiple encounters (`avg_num_outpatient`).

number_emergency

This shows number of emergency visits of the patient in the year preceding the encounter. If a patient has more than one encounter as comment as `number_outpatient` applies to this feature.

number_inpatient

This shows number of inpatient visits of the patient in the year preceding the encounter. If a patient has more than one encounter as comment as `number_outpatient` applies to this feature.

A1Cresult

This is the result for HbA1c test. This data field has one of the following values {'>8','>7','Norm','None'}. For each encounter, we can create new variable `A1C_elevated_flag` with values `0` if the `A1Cresult` is in {'None','Norm'} or `1` if the `A1Cresult` is in {'>8','>7'}. For the patients with multiple encounters, we can calculate total number of `A1C_elevated_flag` (`tot_a1c_elevated`).

readmitted

This dat field shows if the patient was readmitted to hospital and number of days to the readmission. It has one of the following values {'<30','>30','NO'} which indicates *readmission within 30 days of previous inpatient hospitalization, readmission more than 30 days after previous inpatient hospitalization, no readmission* respectively. We can create new variable for each encounter `readmission_flag` which is `1` if `readmitted` is in {'NO'} or `0` if `readmitted` is in {'<30','>30'}. For the patients with mutiple encounters, we can calcualted total number of `readmission_flag` (`tot_readmissions`).

age

This is patient's age grouped in 10-year intervals [0, 10), [10, 20), ..., [90, 100). For the simplicity in the calcualtion, we can create new variable `age_category` with values from `1` to `10` mapped to each of the age grouping. e.g. `1` will be mapped to `[0,10)`, `2` will be mapped to `[10,20)` .. so on.

Finally, for the *PatientClass* we will have 9 values {`age_category`, `avg_num_procedures`, `avg_num_lab_procedures`, `avg_num_meds`, `avg_num_outpatient`, `avg_num_emergency`, `avg_num_inpatient`, `tot_a1c_elevated`, `tot_readmissions`}