race: 4 binary variables with African American as reference

gender: removed unknown genders and turned into binary with female as ref

age: used midpoint of each interval as the approximation of patient age

weight: dropped

admissions type ID: dummy variable for each, ref = 1

discharge disposition ID: dummy variables for each ID, ref = 1 (discharged home), removed 11, 13, 14, 19, 20, 21 (end of life categories)

time in hospital:

payer code: ? converged to unknowns, ref = ?

medical speciality: top 10 as binary, rest = others, ref = ?,

diag1, diag2, diag3: top 10 as binary, rest = others, ref = others

max glu serum: binary. ref = none,

A1Cresult: binary, ref = none

medications: ref = no

dropped medications: acetohexamide, tolbutamide, troglitazone, examide, citoglipton, glipizide.metformin, glimepiride.pioglitazone metformin.rosiglitazone

, metformin.pioglitazone