**Project Proposal Group 8**

**Group Members:**

Xin Xie, Xingjian Xie, Yue Wang, Zoe Zhao

**Data set and source:**

We chose the NINDS dataset provided by professor Kate Crespi on Canvas.

Website to the original dataset: <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0050113#s2>

**Dataset Description:**

There are 100 separate variables recorded for the 624 patients enrolled in the NINDS study. Text S2 documents each individual variable in the order in which it is displayed in the Excel format.

Patient results can be divided into nine natural groupings: A, baseline characteristics; B, patient demographics; C, baseline vital signs and laboratory values; D, head computerized tomography results; E, treatment randomization and timing of drug delivery; F, deleterious effects (death and hemorrhage); G, neurological outcomes; H, dosing of rt-PA and use of heparin; and I, stroke type and clinical outcomes. Text S3 summarizes the data for each of the above groups and details the location of missing data, inconsistencies with previously reported data, and protocol violations not previously reported. Texts S4 through S12 present the results of every variable in the groups noted above.

**Dependent variable:**

Y: los (length of hospital stay from admission to discharge in days)

**Research question to be answered:**

Among all the patients in this dataset, is the length of hospital stay associated with the baseline NIH stroke scale after controlling for age, race, gender, history of hypertension, history of diabetes, and smoking years?

**Predictor variables (tentative list)**

X1: nihssb (nih stroke scale at baseline)

X2: age (age in years)

X3: brace (patient's race)

X4: bgender (patient's gender)

X5: bhyper (history of hypertension at baseline)

X6: bdiab (history of diabetes at baseline)

X7: cursmker (smoking (in year) prior to this stroke)