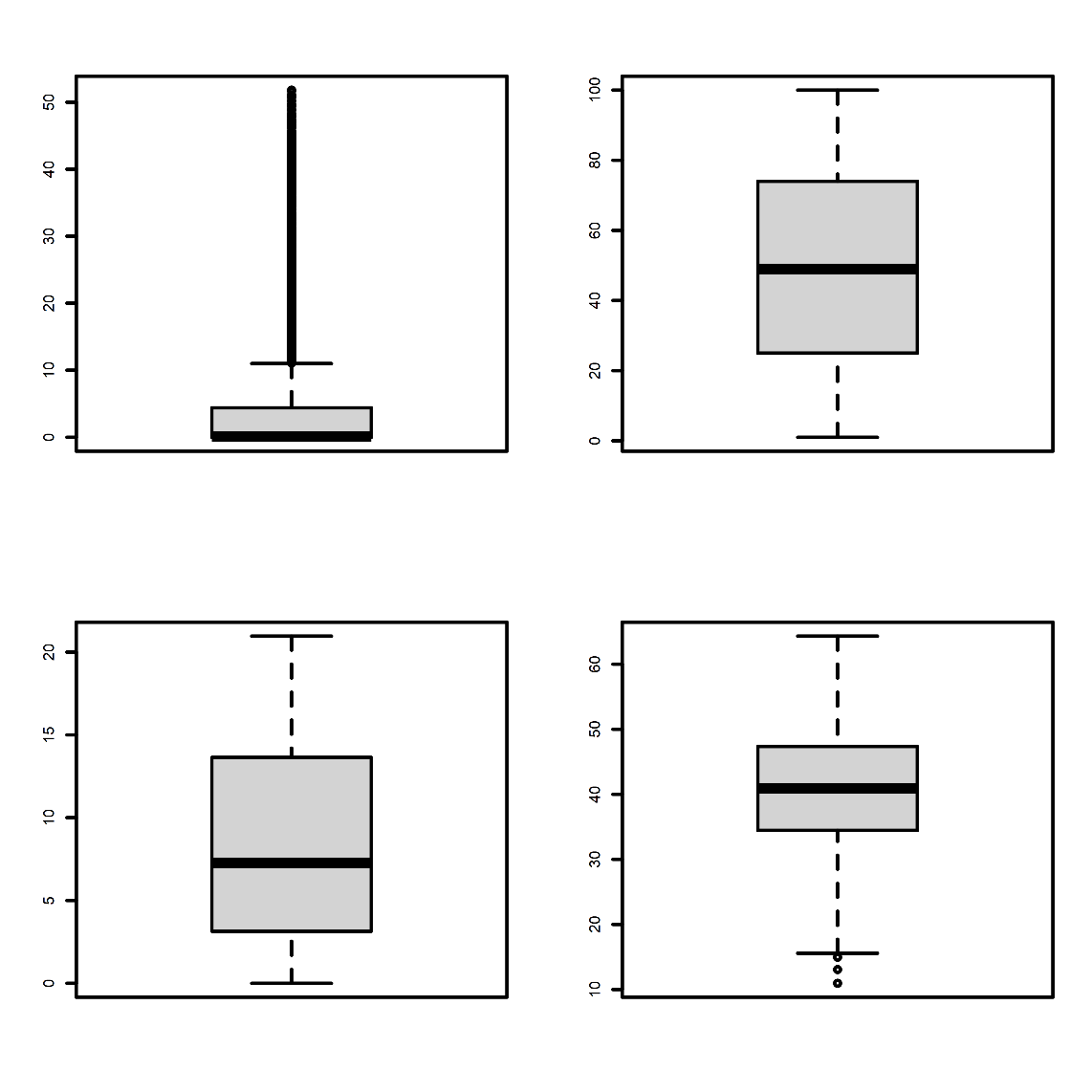
The file “Test data.xls” contains age, sex, heart2, and outcome variables for a group of individuals. Please write an R script to perform 2 regression analyses: (1) a logistic regression with “outcome” as the binary outcome, and (2) a linear regression with “outcome2” as the continuous outcome.

Please send us your R script, the analysis results, and a list of variables in each analysis that have a statistically significant association (P-value < 0.05) with the outcome. Please destroy the data file afterwards.

Ans: Please do find the attachment for analysis results along with code. I am also including results for variables with significant p values in here.

We tested for assumptions and checked for outliers below using boxplots. We removed individuals who are missing and have z score of heart2 >4 as outliers. We also tested other necessary assumptions before performing linear and logistic regression based on binary and continuous variables in R using glm(). We supplied family=binomial(link=”logit” to run logistic regression in R.

Box Plots to Determine Outliers

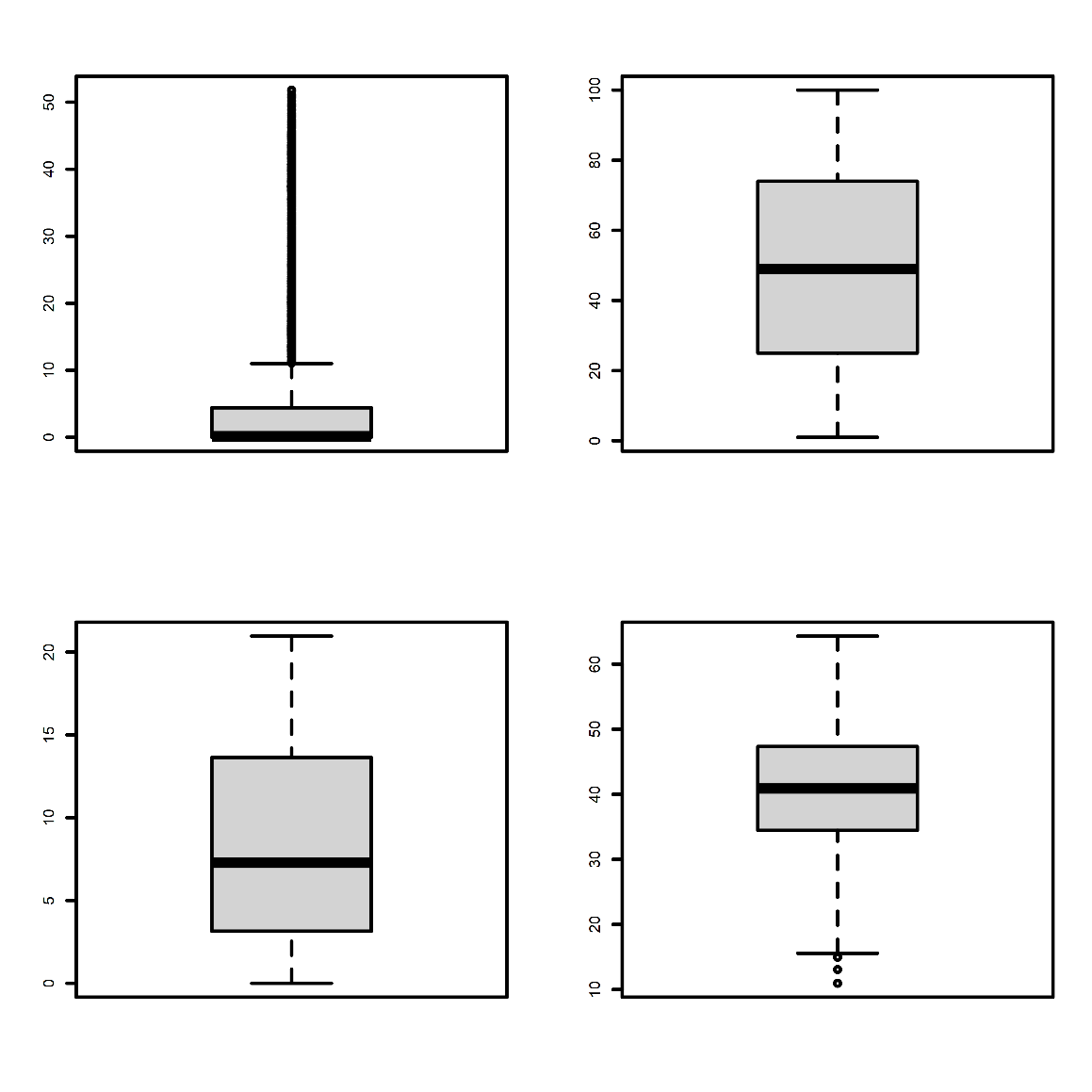


Age\_allevents

Agedx

Primary Outcome

Heart2

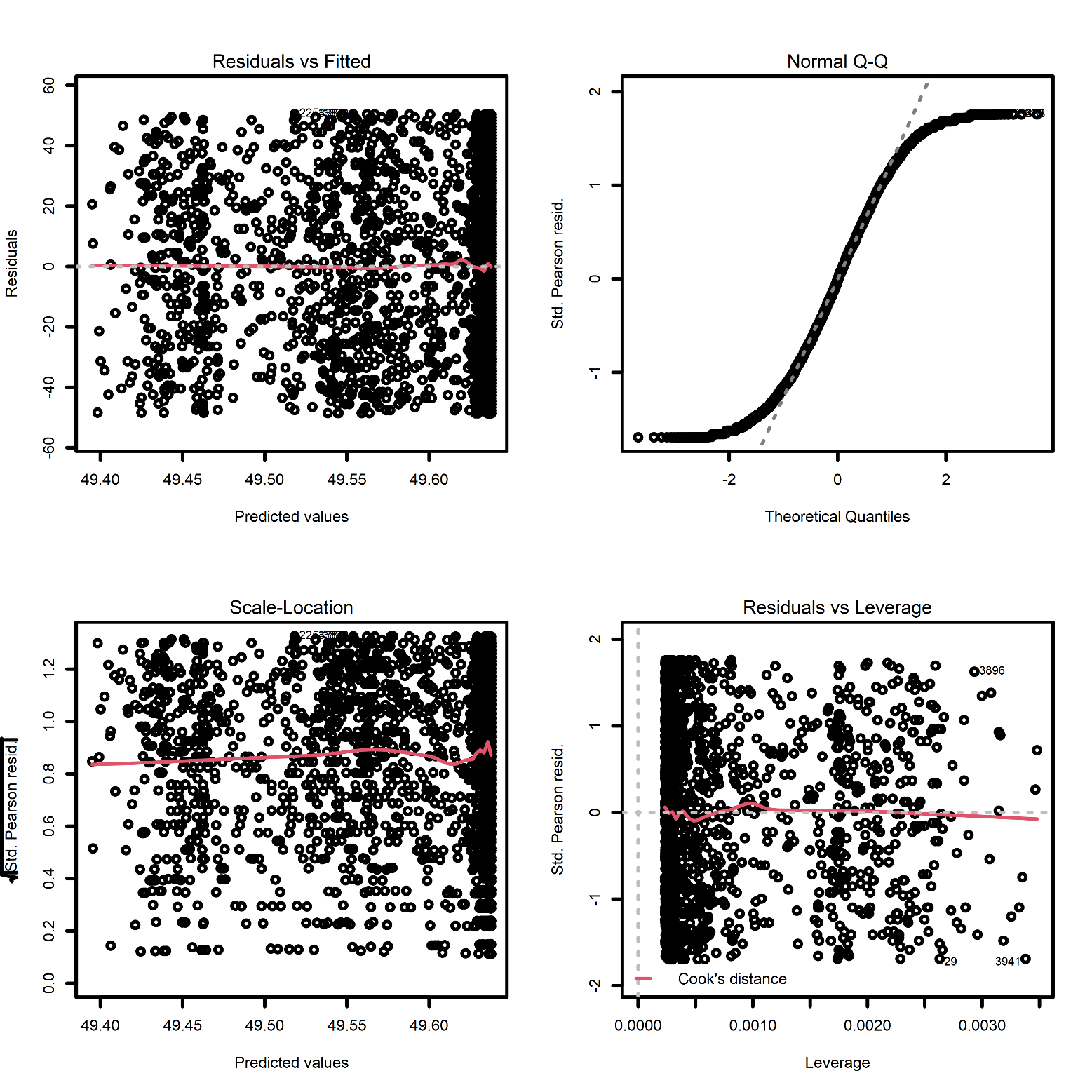


Age\_allevents

Agedx

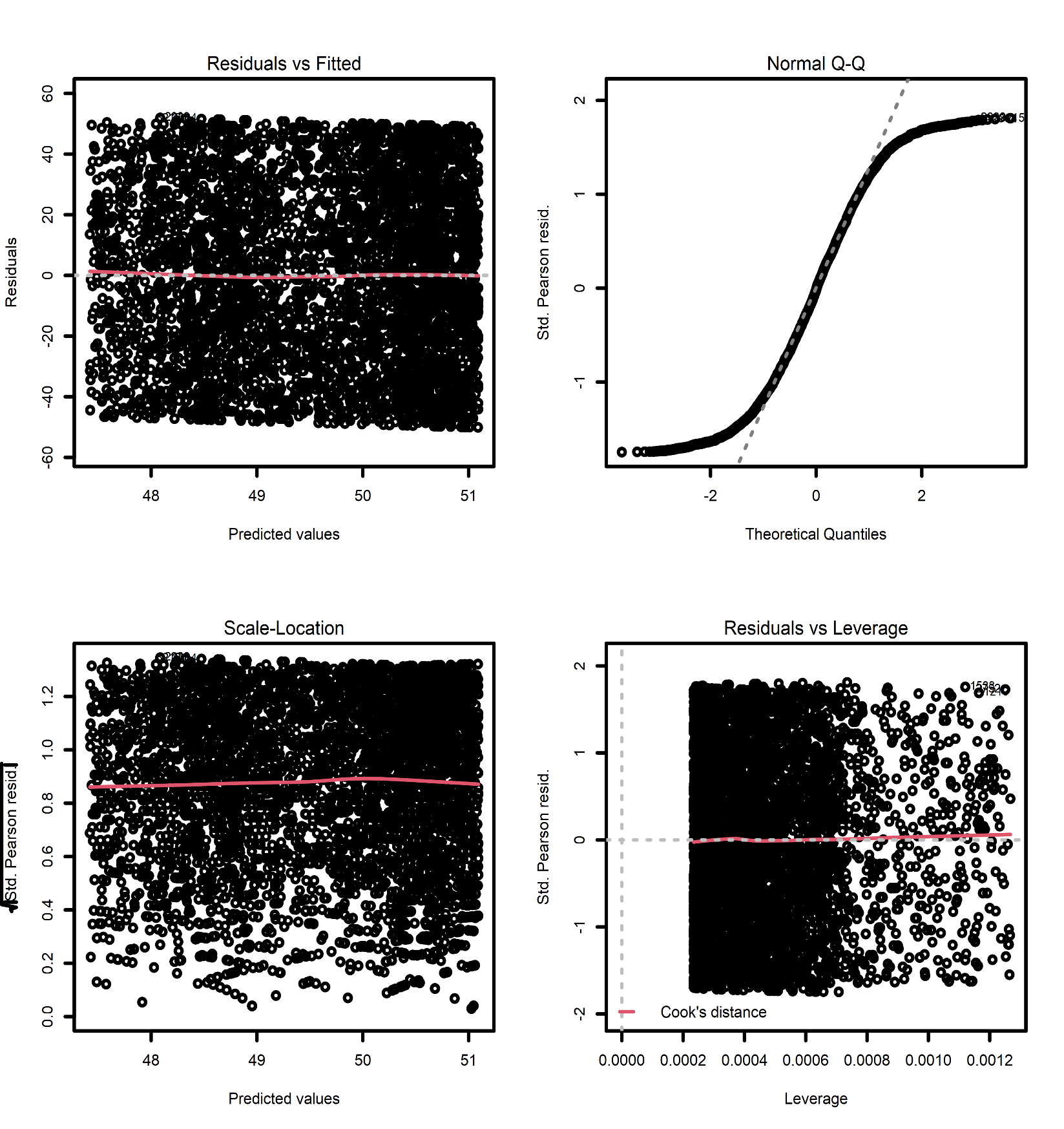
Primary Outcome

Heart2

We then plotted model statistics for each variable associating with outcome2 to determine linearity

Heart 2

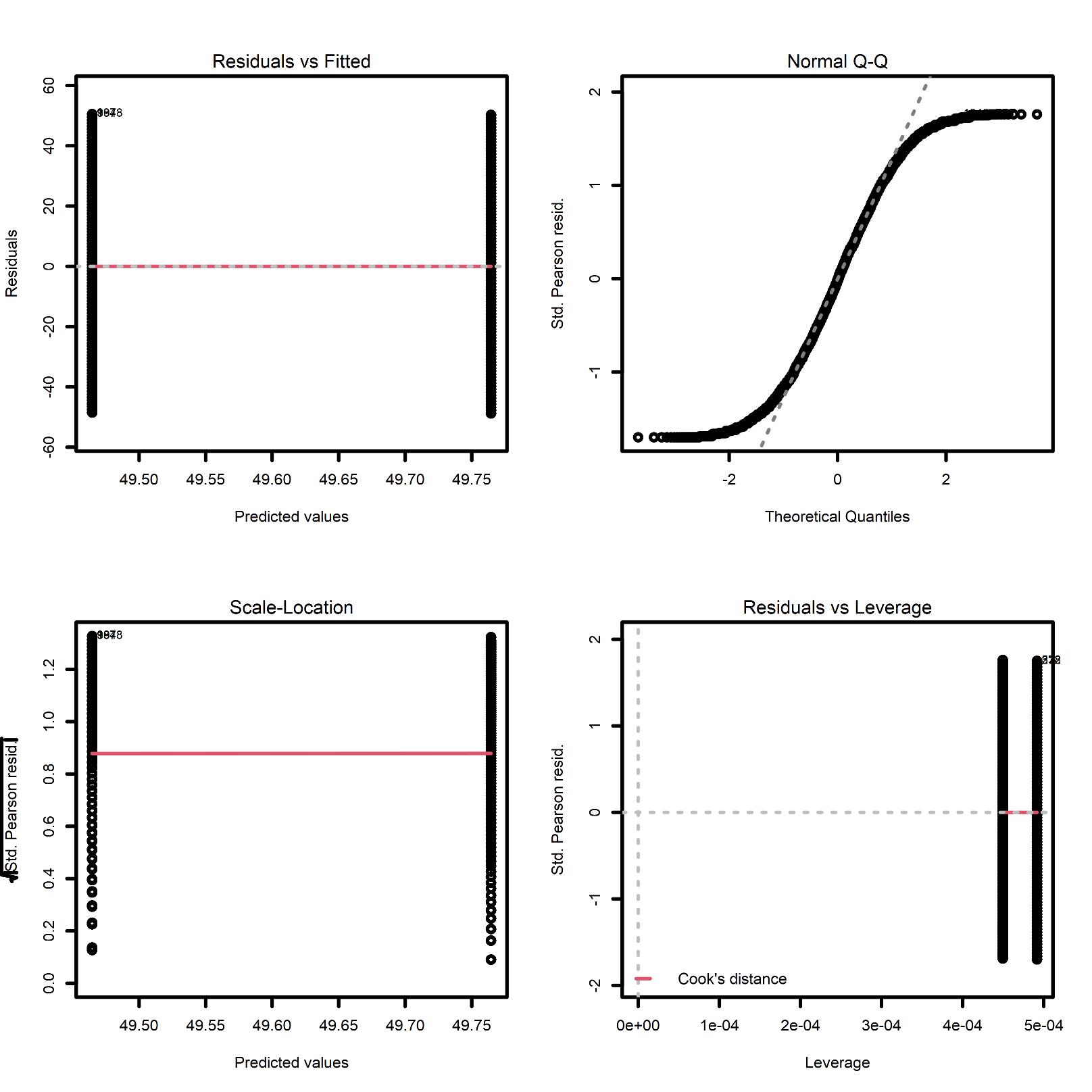
AgeDx



A picture containing diagram

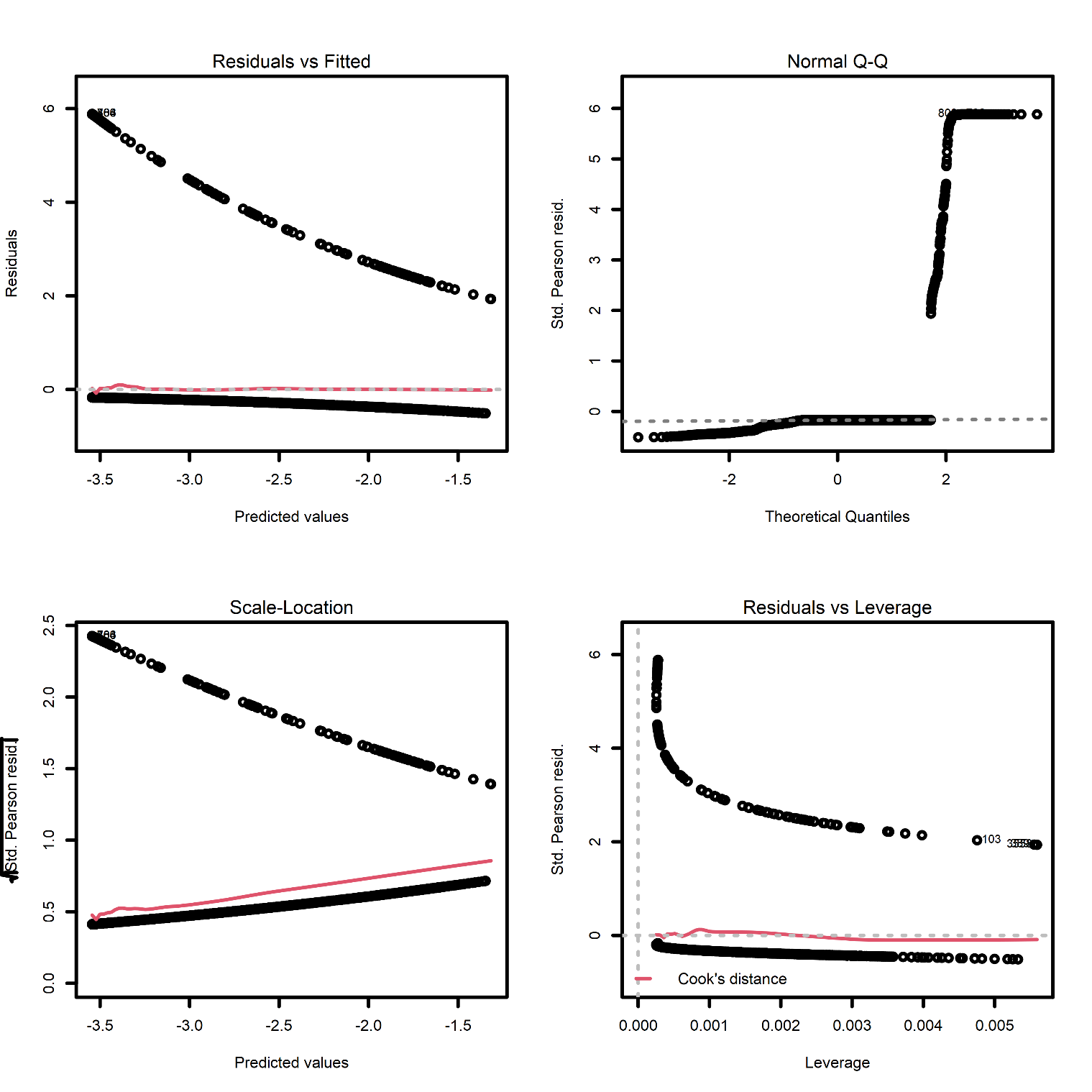
Description automatically generated

Age\_Allevents

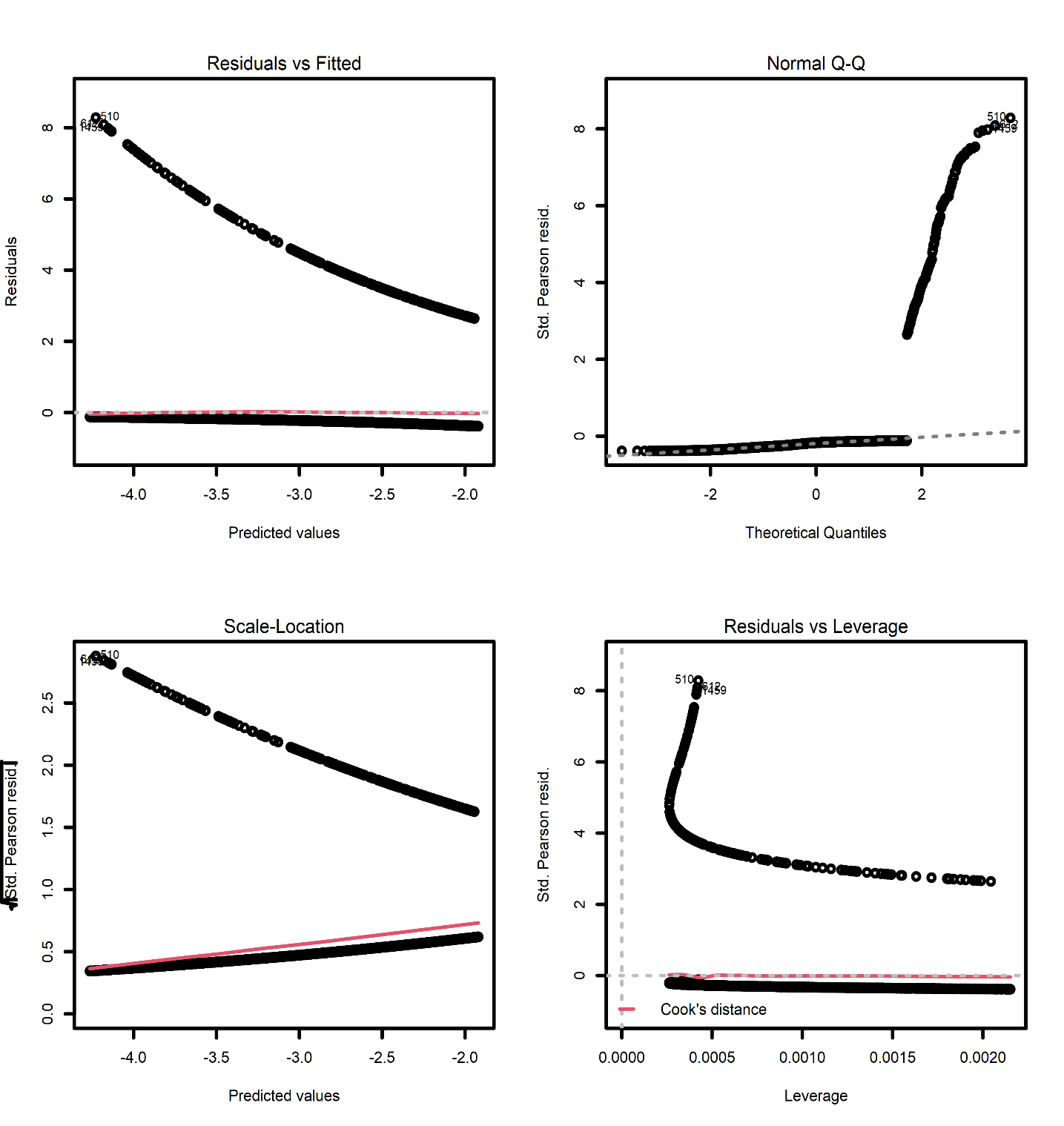


Sex

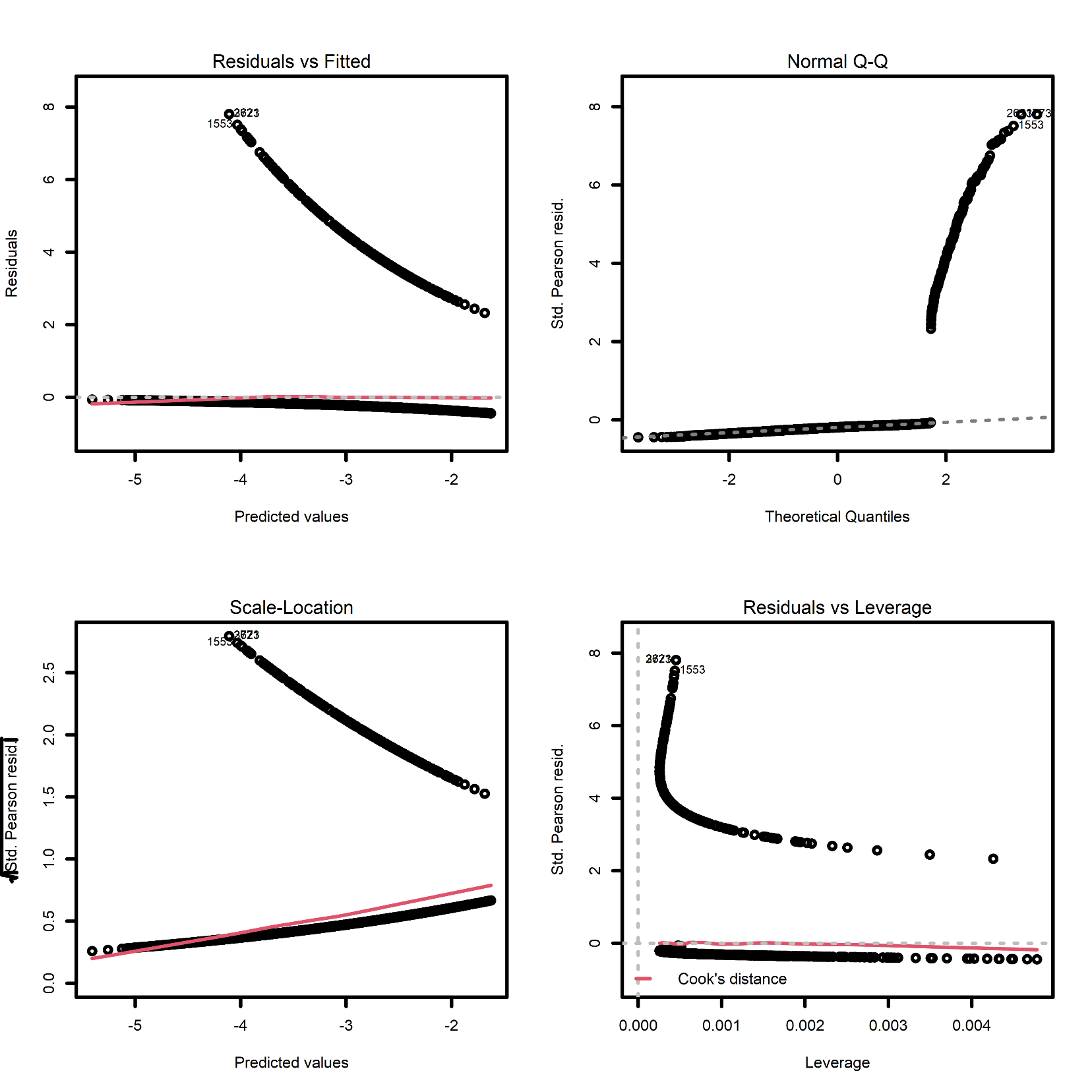
Logistic Regression models as above



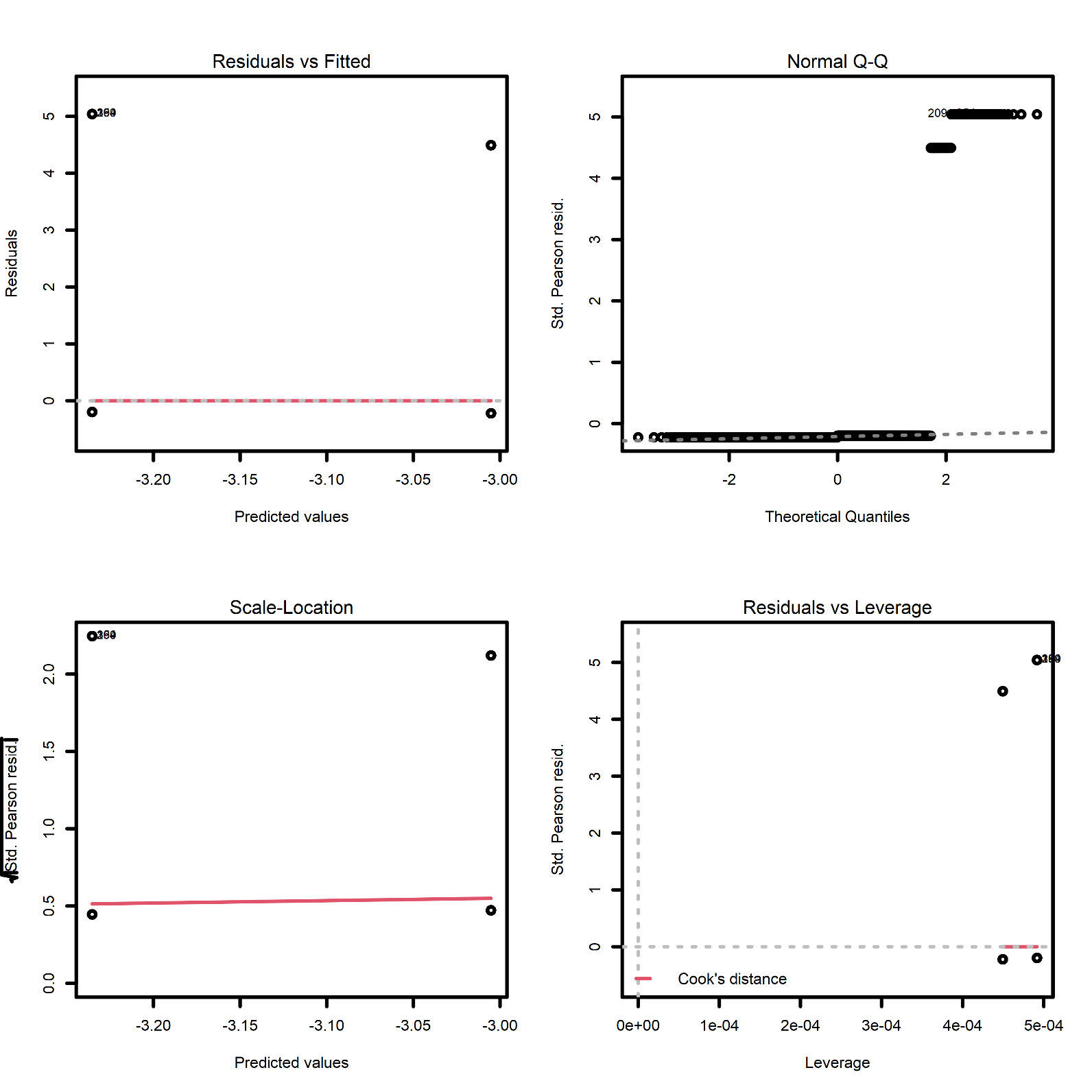
Heart2



Agedx



Age\_AllEvents



Sex

Association of Outcomes with covariates using linear regression

Text

Description automatically generated

Logistic regression results

Text

Description automatically generated

We have 3 variables crossing pvalue cutoff