Practical 6

A study recruited 100 heavy drinkers to participate in an educational program, who were then followed up every three months for 4 times. At each follow-up, the readiness to change (RTC) on alcohol behaviour was assessed through a questionnaire, which was then converted to a score. The number of heavy drinking days (HDD) in the past 30 days was also recorded. Data are stored in the file 'rtc.csv', which included:

- id indicates patient identity number
- males were recorded with male = 1
- age indicates age in years at baseline
- time indicates the occasion of follow-up
- rtc is the score on readiness to change on alcohol use (0: not ready to change drinking behaviour; 10: trying actively to change drinking behaviour)
- hdd is the outcome of interest, number of heavy drinking days (over six standard drinks) per 30 days

Perform the following analysis:

(A) Give a brief overview of the data including the sample size, age and sex distribution of the subjects, and the pattern of readiness to change score and heavy drinking days.

(B) Fit appropriate generalized estimating equation models (GEE) to assess the association between RTC and HDD, accounting for potential confounding effect of sex and age. It is expected that HDD at different follow-ups are correlated for the same person. Use at least two different correlation structures for your GEE model and select the best model (Should use log link/poisson family).

(C) For the selected model in (B), summarize your results in a table (including the 95% confidence intervals).