The data (HEALTH.xlsx) are from a randomized, controlled trial among women of childbearing age to evaluate the effects of an educational intervention. One response variable of interest is the participants self-rating of health status as either good or poor. The researchers would like to assess the effect of the intervention on self-rated health across the follow-up period, as well as whether these effects are influenced by the mothers age. There are n=80 women enrolled in this trial. These data were measured at 4 points in time: randomization, 3 months, 6 months, and 12 months post-randomization. Each observation in the dataset contains values for the following variables (in this order):

ID: unique participant identification code
TIME: the visit number for this observation of this participant
     1. corresponds to a participants visit at the time of randomization
     2. corresponds to the 3 month post-randomization visit
     3. corresponds to the 6 month post-randomization visit
     4. corresponds to the 12 month post-randomization visit
TRT: the treatment group to which the participant has been randomized
     Control
     Intervention
HEALTH: participants self-rated level of health for this visit
     Good
     Poor
AGEGROUP: participants age group at time of randomization   15-24 (years old)
     25 to 34 (years old)
     35+ (years old)

(a) Evaluate the bivariate, cross-sectional relationship between randomized group assignment and participants health self-rating at the time of randomization. Interpret and discuss these findings.

(b) Perform a longitudinal data analysis across all study follow-up visits (but not at randomization) to describe the relationship of the participants self-ratings as a function of the effects of health self-rating at the baseline, treatment group, month post randomization, and age group as predictors. Fit a GEE with unstructured correlation structure. Interpret your results.

(c) Fit a generalized linear mixed effects model with subject-specific random intercepts. Interpret your estimates. How are the interpretations different from the GEE model?