Brief solutions for Biostat 209 Repeated Measures Homework, Part 2

Q4. Here is a model run using age (in decades) and bmi (in units of 3) to get more reasonable sized effects.

```
xtgee htnm exer age10 bmi3 avgdrpwk, i(pptid) ef family(bino) robust

Iteration 1: tolerance = .22056762
Iteration 2: tolerance = .01441381
Iteration 3: tolerance = .00021058
Iteration 4: tolerance = 1.234e-06
Iteration 5: tolerance = 1.667e-07

GEE population-averaged model                   Number of obs      =      9094
Group variable:                       pptid     Number of groups   =      2031
Link:                                 logit     Obs per group: min =         1
Family:                            binomial                    avg =       4.5
Correlation:                   exchangeable                    max =         6
                                               Wald chi2(4)       =     37.26
Scale parameter:                          1    Prob > chi2        =    0.0000

                                  (Std. Err. adjusted for clustering on pptid)
------------------------------------------------------------------------------
             |             Semi-robust
     htnmeds | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    exercise |     .99181   .0488737    -0.17   0.867     .9004997    1.092379
       age10 |   1.399882   .1021289     4.61   0.000     1.213366    1.615069
        bmi3 |    1.13705   .0317067     4.61   0.000     1.076573    1.200923
    avgdrpwk |   .9956904   .0083791    -0.51   0.608     .9794024    1.012249
------------------------------------------------------------------------------
```

One example: The odds of being on HTN meds is about 40% higher for each decade older the participant is on entry into the study.

Q5. A day squared variable (gen day2=day*day) is needed in both the fixed and random portion of the model:

A Stata command is:

```
mixed logwght i.group*day i.group*day2 || mouseid: day day2, cov(un)
```

This fits much better than the model without day2 in the random statement (huge change in log likelihood and estimate of SD(day2) much larger than its SE). So it is needed in the model. It makes a big difference in the tests of the fixed effects as well. For example, testing for a curvature by group interaction (`testparm _IgroXday2*`) gives a p-value of about 0 with the day2 out of the random statement and a p-value of 0.04 with it in. Moral: if you are testing a particular term (like the group by quadratic interaction) and you leave out the variability in that term from the model it makes a big difference. Looks like final model needs all the terms above.

Q6. Need a xtgee model with a log link and the predictors of visit, BMI, sex and the interaction of sex and visit. Visit is not linear, so need to model it or treat it as categorical. The study exhibits what is sometimes called entry bias. Those with severe

pain are more likely to enroll and, afterwards, for some, the pain resolves. So odds of severe pain drop by 32% at year 1 (for the reference group = males) and then start climbing.

BMI: Associated with a one unit increase in BMI is a 9.8% higher chance of severe pain.
Sex: Males (the reference group) showed a 32% drop (100%*(1-.6800761)). But females have a coefficient of .680076*1.58377 = .89345 (for comparing visits 1 and 2), so about an 11% drop. Other coefficients have similar interpretations.