# PHP 2516: HW #2

*Blain Morin*

*October 15, 2018*

## Question 1: Using the 'calcium_allL' dataset please answer the following questions assuming an unstructured covariance pattern:

### a. Describe the missing patterns you see in the data (if any).

We used the xtdescribe command to show the pattern of missingness in the data:

| Freq. | Percent | Cum. | Pattern |
|-------|---------|--------|---------|
| 91 | 81.25 | 81.25 | 11111 |
| 7 | 6.25 | 87.50 | 1.... |
| 6 | 5.36 | 92.86 | 11... |
| 5 | 4.46 | 97.32 | 111.. |
| 3 | 2.68 | 100.00 | 1111. |
| 112 | 100.00 | | XXXXX |

We see from the table above that 91 people had all five observations, 7 people had only the first observation, 6 people had only the first and second observation, and etc. It appears that people that drop out of the study do not return at a later time.

**b. Plot the observed data (means and spaghetti plots). What do you observe?**



Trajectories for the First 8 Individuals
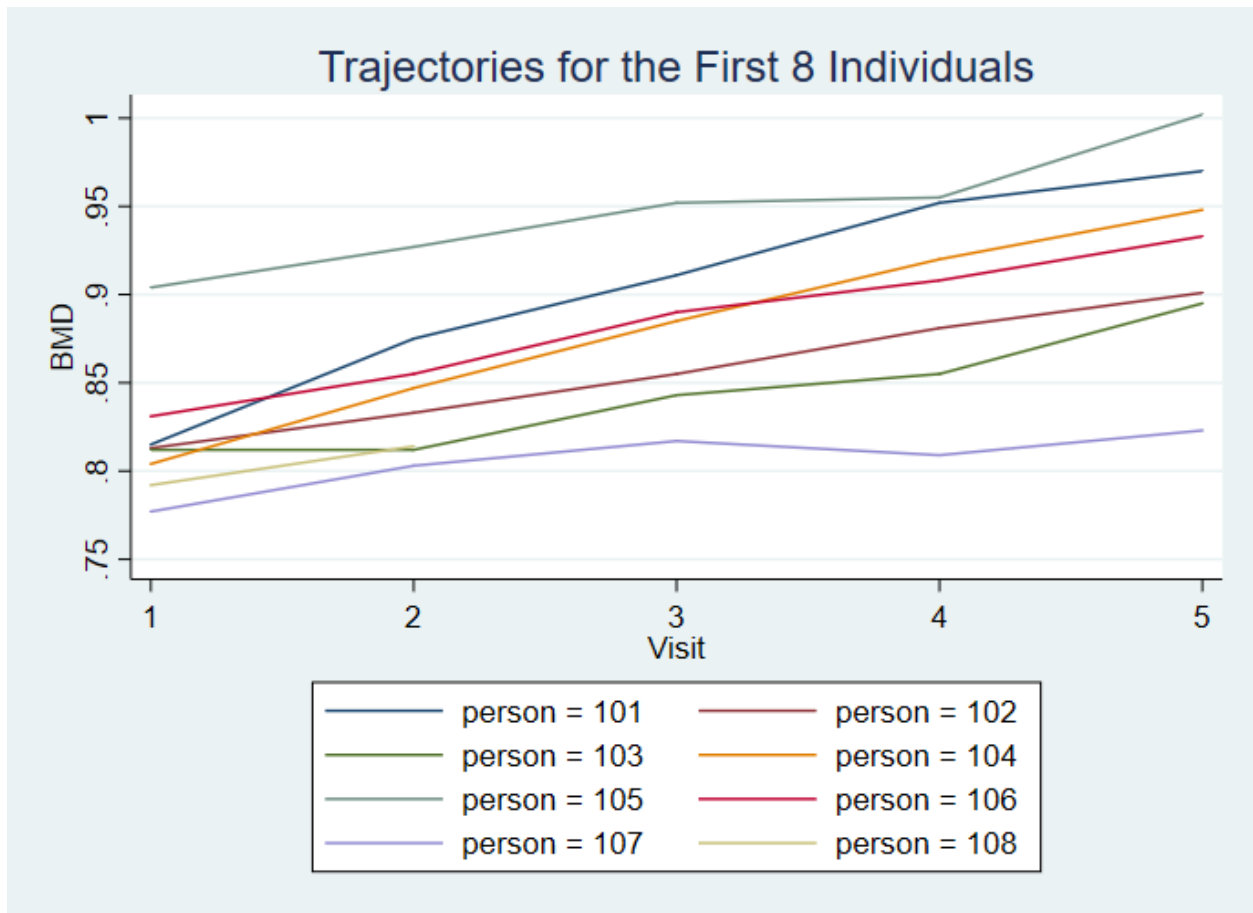
The spaghetti plot above shows the trajectories for the first 8 individuals. We see that there appears to be an upward trend of bone mineral density over time.

Average BMD Over Time By Group
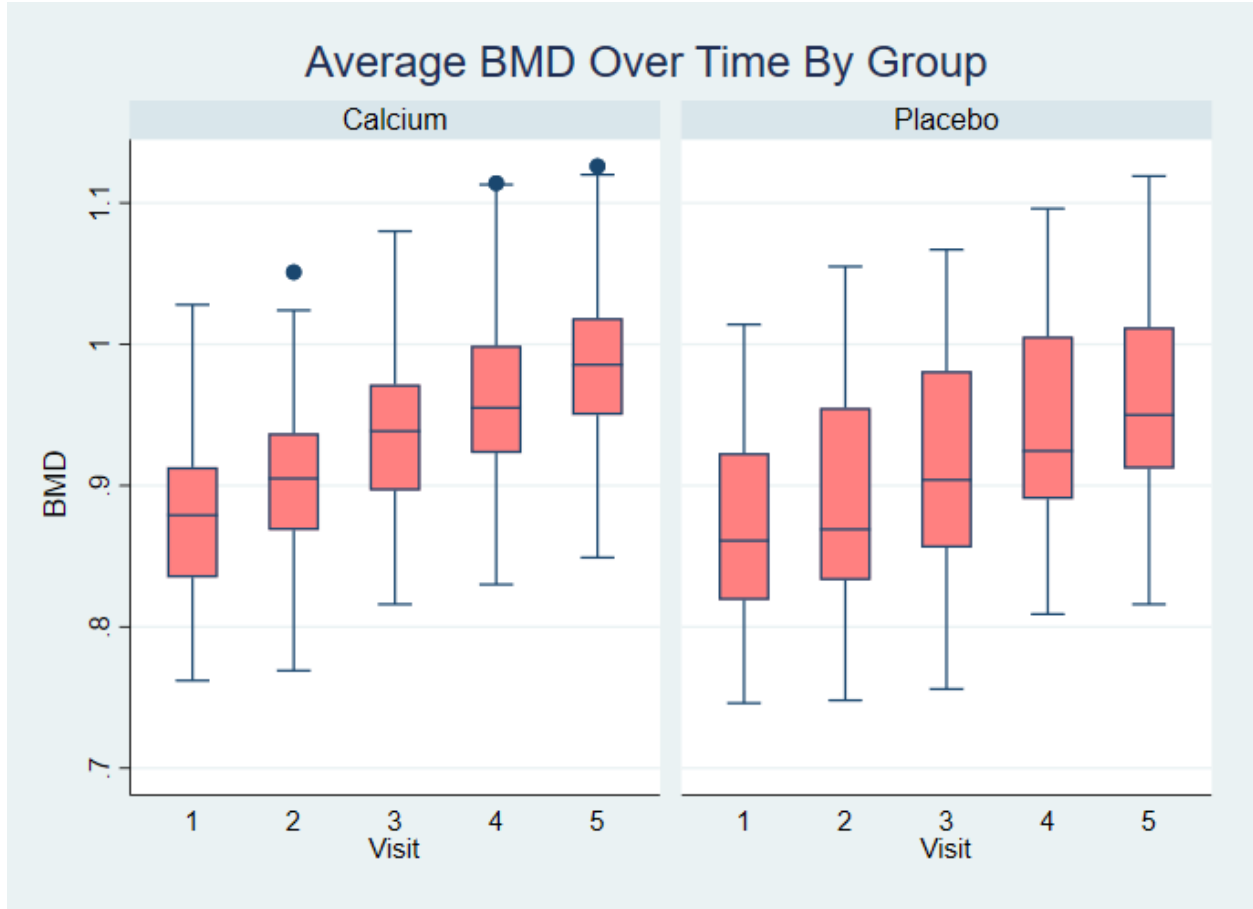
The boxplots also display an upward trend in bone mineral density. Visually, the calcium group and control group seem to have the same rate of increase. The overall means in the calcium group appear to be higher than the means in the placebo group.

**c. Conduct a Mean Response Profiles analysis (Model 1) with only time and treatment group. What is your overall conclusion about the changes in the mean response over time and the effect of the treatment group on those changes?**

The mean response profile uses all variables as categorical and takes the form:

$$(1) : BMD_i = \beta_0 + \sum_{n=1}^{4} \beta_n visit + \beta_5 tx + \sum_{m=6}^{10} \beta_m visit * tx + \epsilon_i$$

The regression result from model (1) is:

| bmd | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] |
|---|---|---|---|---|---|---|
| **visit** | | | | | | |
| 2 | .0200641 | .0023589 | 8.51 | 0.000 | .0154408 | .0246875 |
| 3 | .0441663 | .0033297 | 13.26 | 0.000 | .0376402 | .0506924 |
| 4 | .0705428 | .0039355 | 17.92 | 0.000 | .0628294 | .0782563 |
| 5 | .0870585 | .0043331 | 20.09 | 0.000 | .0785657 | .0955513 |
| | | | | | | |
| **tx** | | | | | | |
| Treatment | .0103844 | .011881 | 0.87 | 0.382 | -.012902 | .0336707 |
| | | | | | | |
| **visit#tx** | | | | | | |
| 2#Treatment | .0069691 | .0033527 | 2.08 | 0.038 | .000398 | .0135402 |
| 3#Treatment | .0123047 | .0047558 | 2.59 | 0.010 | .0029836 | .0216259 |
| 4#Treatment | .0124964 | .0056195 | 2.22 | 0.026 | .0014824 | .0235103 |
| 5#Treatment | .0189692 | .0061952 | 3.06 | 0.002 | .0068268 | .0311115 |
| | | | | | | |
| _cons | .8700702 | .0083258 | 104.50 | 0.000 | .8537519 | .8863884 |

The regression coefficients are interpreted in detail in Question 4.

**d. Suppose that you are mainly interested in describing the trends in the mean responses over time adjusting for age. Find a model for the mean that best fits your data.**

We start the model selection process with a full marginal model (using time as a continuous predictor) (model2):

$$(2): BMD_i = \beta_0 + \beta_1 visit + \beta_2 visit^2 + \beta_3 tx + \beta_4 age + \beta_5 tx * visit + \beta_6 tx * visit^2 + \beta_7 tx * age + \epsilon$$

| bmd | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] |
|---|---|---|---|---|---|---|
| visit | .0017705 | .0096601 | 0.18 | 0.855 | -.017163 | .0207041 |
| visit2 | -.0001578 | .0005065 | -0.31 | 0.755 | -.0011505 | .000835 |
| tx | .0951248 | .2652323 | 0.36 | 0.720 | -.424721 | .6149706 |
| age | .0431855 | .0183667 | 2.35 | 0.019 | .0071875 | .0791835 |
| txvisit | .0117916 | .0133344 | 0.88 | 0.377 | -.0143433 | .0379265 |
| txvisit2 | -.0005234 | .0007238 | -0.72 | 0.470 | -.001942 | .0008952 |
| txage | -.0084595 | .0251906 | -0.34 | 0.737 | -.0578321 | .0409131 |
| _cons | .3900983 | .1933508 | 2.02 | 0.044 | .0111378 | .7690589 |

We see that no predictors are statistically significant in this full model. We then run the model without the interaction between treatment and visit^2 (dropped because of complexity):

$$(3): BMD_i = \beta_0 + \beta_1 visit + \beta_2 visit^2 + \beta_3 tx + \beta_4 age + \beta_5 tx * visit + \beta_6 tx * age + \epsilon$$

| bmd | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| visit | .0035449 | .0094089 | 0.38 | 0.706 | -.0148963 | .021986 |
| visit2 | -.0004217 | .0003609 | -1.17 | 0.243 | -.001129 | .0002856 |
| tx | .0938985 | .2650567 | 0.35 | 0.723 | -.425603 | .6134 |
| age | .0429718 | .0183593 | 2.34 | 0.019 | .0069883 | .0789553 |
| txvisit | .008424 | .0125525 | 0.67 | 0.502 | -.0161785 | .0330265 |
| txage | -.0083415 | .0251736 | -0.33 | 0.740 | -.0576809 | .0409979 |
| _cons | .3923219 | .1932742 | 2.03 | 0.042 | .0135114 | .7711323 |

Again, there are no statistically significant predictors. We decide again to drop the more complicated term, visit^2, and rerun the regression (model4):

$$(4): BMD_i = \beta_0 + \beta_1 visit + \beta_2 tx + \beta_3 age + \beta_4 tx * visit + \beta_5 tx * age + \epsilon$$

| bmd | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| visit | .0008886 | .0091447 | 0.10 | 0.923 | -.0170346 | .0188119 |
| tx | .0951723 | .2654472 | 0.36 | 0.720 | -.4250945 | .6154392 |
| age | .0429505 | .0183775 | 2.34 | 0.019 | .0069312 | .0789698 |
| txvisit | .0084772 | .0125714 | 0.67 | 0.500 | -.0161623 | .0331167 |
| txage | -.0084588 | .0252109 | -0.34 | 0.737 | -.0578713 | .0409537 |
| _cons | .3925635 | .193466 | 2.03 | 0.042 | .0133772 | .7717499 |

We see again that non of the predictors are significant at the .05 level. We then decide to drop the least significant interaction term (tx*age) (model5):

$$(5): BMD_i = \beta_0 + \beta_1 visit + \beta_2 tx + \beta_3 age + \beta_4 tx * visit + \epsilon$$

| bmd | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| visit | .0031635 | .0062964 | 0.50 | 0.615 | -.0091772 | .0155041 |
| tx | .0062377 | .0112506 | 0.55 | 0.579 | -.0158131 | .0282884 |
| age | .0383463 | .0125607 | 3.05 | 0.002 | .0137278 | .0629648 |
| txvisit | .0042865 | .0014821 | 2.89 | 0.004 | .0013815 | .0071914 |
| _cons | .4409402 | .1323538 | 3.33 | 0.001 | .1815314 | .7003489 |

In model 5, we see that the interaction between treatment and visit is significant as well as the coefficient on age. Since interaction terms lose their interpretability when the individual predictors of the interaction are

dropped, we elect to keep visit and treatment in the model, despite the lack of significance. We thus choose model 5 as our best model.

We also check the information criteria for all 5 models:

| Model | Obs | ll(null) | ll(model) | df | AIC | BIC |
|-------|-----|----------|-----------|-----|-----|-----|
| model1 | 501 | . | 1173.17 | 25 | -2296.34 | -2190.925 |
| model2 | 501 | . | 1181.694 | 23 | -2317.388 | -2220.406 |
| model3 | 501 | . | 1187.751 | 22 | -2331.502 | -2238.737 |
| model4 | 501 | . | 1194.156 | 21 | -2346.313 | -2257.764 |
| model5 | 501 | . | 1196.864 | 20 | -2353.728 | -2269.396 |

As expected we see that model 5 has the lowest AIC and BIC.

# Question 2: : Using the best model from q1 part d try the following covariance patterns:

We rerun our model 5 using seven different covariance structures. Here is the regression output for each:

## a. Unstructured

| bmd | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
|-----|-------|-----------|---|-------|------|---|
| visit | .0031635 | .0062964 | 0.50 | 0.615 | -.0091772 | .0155041 |
| tx | .0062377 | .0112506 | 0.55 | 0.579 | -.0158131 | .0282884 |
| age | .0383463 | .0125607 | 3.05 | 0.002 | .0137278 | .0629648 |
| txvisit | .0042865 | .0014821 | 2.89 | 0.004 | .0013815 | .0071914 |
| _cons | .4409402 | .1323538 | 3.33 | 0.001 | .1815314 | .7003489 |

## b. Exchangable

| bmd | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
|-----|-------|-----------|---|-------|------|---|
| visit | -.0052147 | .0073783 | -0.71 | 0.480 | -.0196759 | .0092465 |
| tx | .0072468 | .0130863 | 0.55 | 0.580 | -.0184019 | .0328955 |
| age | .0564111 | .0148839 | 3.79 | 0.000 | .0272392 | .0855829 |
| txvisit | .0044043 | .0009836 | 4.48 | 0.000 | .0024765 | .0063321 |
| _cons | .2503899 | .1576196 | 1.59 | 0.112 | -.0585387 | .5593186 |

## c. Autoregressive (1)

| bmd | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| visit | .007089 | .0064485 | 1.10 | 0.272 | -.0055498 | .0197278 |
| tx | .0057492 | .0130241 | 0.44 | 0.659 | -.0197774 | .0312759 |
| age | .0296696 | .0128456 | 2.31 | 0.021 | .0044926 | .0548466 |
| txvisit | .0047171 | .0015962 | 2.96 | 0.003 | .0015886 | .0078456 |
| _cons | .5348164 | .1360546 | 3.93 | 0.000 | .2681543 | .8014786 |

## d. Autoregressive (3)

| bmd | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| visit | .0068361 | .0064825 | 1.05 | 0.292 | -.0058694 | .0195416 |
| tx | .0058168 | .0130241 | 0.45 | 0.655 | -.0197099 | .0313435 |
| age | .0302843 | .0129249 | 2.34 | 0.019 | .0049519 | .0556166 |
| txvisit | .0047016 | .0015644 | 3.01 | 0.003 | .0016353 | .0077678 |
| _cons | .5282109 | .1368967 | 3.86 | 0.000 | .2598983 | .7965234 |

## e. Toeplitz (1)

| bmd | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| visit | .0138843 | .0104262 | 1.33 | 0.183 | -.0065506 | .0343192 |
| tx | .0067486 | .0129569 | 0.52 | 0.602 | -.0186464 | .0321435 |
| age | .0161105 | .0203566 | 0.79 | 0.429 | -.0237877 | .0560087 |
| txvisit | .0048153 | .0039542 | 1.22 | 0.223 | -.0029348 | .0125654 |
| _cons | .6778402 | .2145347 | 3.16 | 0.002 | .25736 | 1.09832 |

## f. Toeplitz (3)

| bmd | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| visit | .0072544 | .0073514 | 0.99 | 0.324 | -.0071539 | .0216628 |
| tx | .0024347 | .0136204 | 0.18 | 0.858 | -.0242608 | .0291303 |
| age | .0250645 | .0136917 | 1.83 | 0.067 | -.0017706 | .0518997 |
| txvisit | .0054828 | .0039831 | 1.38 | 0.169 | -.0023238 | .0132895 |
| _cons | .5882832 | .1447269 | 4.06 | 0.000 | .3046237 | .8719427 |

## g. Toeplitz (4)

| bmd | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| visit | .0072544 | .0073514 | 0.99 | 0.324 | -.0071539 | .0216628 |
| tx | .0024347 | .0136204 | 0.18 | 0.858 | -.0242608 | .0291303 |
| age | .0250645 | .0136917 | 1.83 | 0.067 | -.0017706 | .0518997 |
| txvisit | .0054828 | .0039831 | 1.38 | 0.169 | -.0023238 | .0132895 |
| _cons | .5882832 | .1447269 | 4.06 | 0.000 | .3046237 | .8719427 |

To choose the best fit from the seven different covariance structures, we examine the information criteria:

| Model | Obs | ll(null) | ll(model) | df | AIC | BIC |
|---|---|---|---|---|---|---|
| unstr | 501 | . | 1196.864 | 20 | -2353.728 | -2269.396 |
| exch | 501 | . | 1121.253 | 7 | -2228.506 | -2198.99 |
| auto1 | 501 | . | 1180.889 | 7 | -2347.777 | -2318.261 |
| auto3 | 501 | . | 1180.966 | 9 | -2343.931 | -2305.982 |
| toe1 | 501 | . | 852.9992 | 7 | -1691.998 | -1662.482 |
| toe3 | 501 | . | 1078.897 | 9 | -2139.795 | -2101.845 |
| toe4 | 501 | . | 1078.897 | 9 | -2139.795 | -2101.845 |

Because it has the lowest BIC, we believe that the model with a 1 period autoregressive covariance structure is the best fit.

# Question 3: Start with a simple model assuming only a linear trend over time, the main effect of treatment, no interaction between time and treatment, and no adjustment for other covariates.

**a,b,c.** Select the model that best fits the covariance in your data among the seven options given in question 2. Then fit the model that best describes the trends in the mean responses over time. Is the final "best" model for your data the same with the one resulted from the process followed in Question 2?

The simplified model is :

$$BMD_i = \beta_0 + \beta_1 visit + \beta_2 tx + \epsilon$$

Using the simplified model, we run the regression with the seven different covariance structures from question 2 and examine the information criteria for each regression:

| Model | Obs | ll(null) | ll(model) | df | AIC | BIC |
|-------|-----|----------|-----------|-----|-----------|-----------|
| unstr | 501 | . | 1197.707 | 18 | -2359.413 | -2283.514 |
| exch | 501 | . | 1113.933 | 5 | -2217.865 | -2196.782 |
| auto1 | 501 | . | 1182.795 | 5 | -2355.59 | -2334.507 |
| auto3 | 501 | . | 1182.822 | 7 | -2351.644 | -2322.128 |
| toe1 | 501 | . | 859.5332 | 5 | -1709.066 | -1687.983 |
| toe3 | 501 | . | 1084.229 | 7 | -2154.458 | -2124.941 |
| toe4 | 501 | . | 1084.229 | 7 | -2154.458 | -2124.941 |

Again, the model with a 1 period autoregressive covariance structure is the best model, in terms of BIC. Because this matches the conclusion in question 2, this may be evidence that the best covariance structure remains constant despite different predictor choices for nested models, although this point may be coincidence and warrants further investigation.

# Question 4: Interpret the regression coefficients from Model 1 and the best model(s) for the mean resulted from Q1.d and Q3. What is your final conclusion about the changes in the response over time between the two treatment groups?

*Interpretation for our mean response model (model 1):*

For model 1, the constant, .87, is the mean BMD for the control group at the first visit.

The beta coefficients for visit are the average increase BMD measurements for the placebo group in reference to the first visit. For example, the average BMD was .04 g/cm^2 higher on visit three for the placebo group compared to their first visit.

The coefficient on treatment is the average difference in BMD between the calcium and placebo group at the first visit (not significant).

The interaction coefficients represent the average difference between the calcium group and the placebo group at each of the follow up visits. For example, the average BMD measurement was .012 g/cm^2 higher for the calcium group at visit 4 than the placebo group.

***Interpretation for our best marginal model from Q1 (model 5):***

Since the data was not centered, the intercept, .44, represents the mean BMD for a 0 year old person in the placebo group.

The coefficient on visit means that the BMD for each additional visit increased by .003 on average for the placebo group.

The coefficient on treatment means that the average BMD for the treatment group at time zero was .006 g/cm^2 higher than the average in the placebo group (not significant).

The coefficient on age means that for a 1 year increase in age, we expect a .038 increase in BMD measurement on average.

The coefficient on the interaction means that the average marginal increase in BMD is .004 g/cm^2 higher for the calcium group compared to the placebo group.

***Interpretation for our best marginal model from Q3:***

The intercept, .841, is the average BMD for visit 1 in the placebo group.

The coefficient on visit means that for each additional visit, we expect the BMD level to be .024 g/cm^2 higher.

The coefficient on treatment means that we expect the BMD measurement to be .012 g/cm^2 higher for those in the calcium group compared to the placebo group, all else equal.

After reviewing the model results, it is clear that the mean BMD measurements are increasing over time for both the calcium group and the control group. Because the interaction is significant in model (5), there is evidence to suggest that the rate of increase in BMD is higher for those taking calcium compared to placebo.

# Question 5: Using the 'cd4' dataset please answer the following questions:

## a. Create spaghetti plots to explore the patterns of the individual trajectories over time, by treatment group. What do you observe?

Here are spaghetti plots for a few individuals, by treatment group:



10
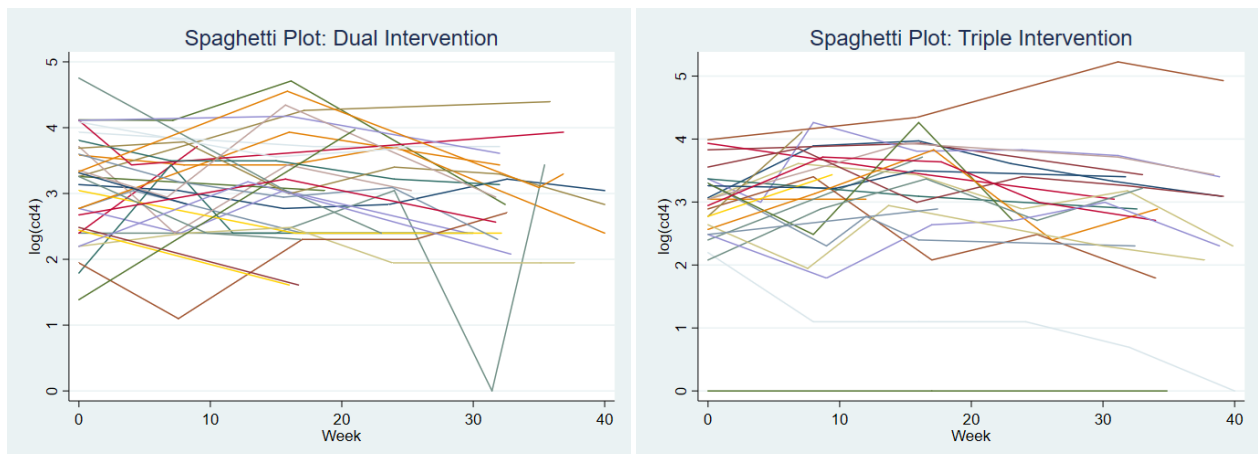
We observe that in both the dual intervention group and the triple intervention group, the individual trajectories vary greatly. Visually, it is difficult to identify a trend.

## b. Based on the plots of the observed data what model would you prefer to fit to express the changes in the primary outcome (log of CD4 counts) over time and any differences between the two treatment groups? Explain.

Because there is a lot of variation in initial log CD4 levels and individual trends over time, we may prefer a mixed effects model that allows for each individual to have their own intercept and slope.

# Question 6: Fit a marginal model to describe the trends in the mean responses over time by treatment group. For the mean model consider only a linear trend, and test whether the changes over time are different among the treatment groups without adjusting for any other covariates in the model. For the covariance pattern assume:

To run these regressions, we changed the week variable into an integer variable (intweek) by rounding all the weeks up.

## a. Exchangeable

| logcd4 | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| trt | .0193174 | .0642195 | 0.30 | 0.764 | -.1065505 | .1451852 |
| intweek | -.0131325 | .0008496 | -15.46 | 0.000 | -.0147977 | -.0114673 |
| trtintweek | .0119722 | .0016608 | 7.21 | 0.000 | .0087171 | .0152273 |
| _cons | 2.992349 | .03231 | 92.61 | 0.000 | 2.929022 | 3.055675 |

## b. Unstructured

There were too many parameters to estimate for the unstructured covariance matrix.

## c. Auto-regressive (1)

| logcd4 | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| trt | -.0467184 | .0663259 | -0.70 | 0.481 | -.1767149 | .083278 |
| intweek | -.01113 | .0013615 | -8.17 | 0.000 | -.0137984 | -.0084615 |
| trtintweek | .0141925 | .0026772 | 5.30 | 0.000 | .0089453 | .0194397 |
| _cons | 2.958933 | .03337 | 88.67 | 0.000 | 2.893529 | 3.024337 |

## d. Toeplitz(1)

| logcd4 | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| trt | .0153594 | .0549684 | 0.28 | 0.780 | -.0923766 | .1230954 |
| intweek | -.0105528 | .0013807 | -7.64 | 0.000 | -.0132589 | -.0078467 |
| trtintweek | .0116561 | .0027064 | 4.31 | 0.000 | .0063517 | .0169606 |
| _cons | 2.985846 | .0277626 | 107.55 | 0.000 | 2.931432 | 3.040259 |

# What do you observe? Do you think that a marginal model is appropriate for analyzing these data? Explain

The three models that run have statistically significant positive interactions between triple intervention and time, meaning that the marginal increase in log CD4 levels is higher for the triple intervention group than the dual intervention group. Although the conclusion is consistent among all the marginal models, a mixed model may be more appropriate to fit. We see in the spaghetti plots above that many individuals have a negative slope in the treatment group. If we are interested in predictions at the individual level, a mixed model that accounts for individual level characteristics would be a better choice.

# Question 7: Fit four different models for the covariance in the data, only assuming the following:

## a. A random intercept (model1)

| logcd4 | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| week | -.0045968 | .0027167 | -1.69 | 0.091 | -.0099213 | .0007278 |
| week2 | -.0002531 | .0000756 | -3.35 | 0.001 | -.0004013 | -.0001049 |
| trt | .423783 | .3108695 | 1.36 | 0.173 | -.18551 | 1.033076 |
| age | .0121742 | .0036784 | 3.31 | 0.001 | .0049647 | .0193838 |
| gender | -.0019041 | .0920833 | -0.02 | 0.984 | -.1823842 | .1785759 |
| weektrt | .0371581 | .0053368 | 6.96 | 0.000 | .0266982 | .0476181 |
| week2trt | -.0007291 | .0001475 | -4.94 | 0.000 | -.0010182 | -.0004399 |
| agetrt | -.0035082 | .0072268 | -0.49 | 0.627 | -.0176725 | .010656 |
| gendertrt | -.417861 | .181678 | -2.30 | 0.021 | -.7739433 | -.0617787 |
| _cons | 2.50258 | .1559242 | 16.05 | 0.000 | 2.196974 | 2.808186 |

## b. Random intercept and random slope for time t (model2)

| logcd4 | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| week | -.0037738 | .0026288 | -1.44 | 0.151 | -.0089262 | .0013785 |
| week2 | -.0002972 | .0000727 | -4.09 | 0.000 | -.0004397 | -.0001546 |
| trt | .3630766 | .3025003 | 1.20 | 0.230 | -.2298132 | .9559664 |
| age | .0115266 | .00358 | 3.22 | 0.001 | .0045098 | .0185433 |
| gender | .0093256 | .0895165 | 0.10 | 0.917 | -.1661236 | .1847748 |
| weektrt | .037052 | .0051613 | 7.18 | 0.000 | .026936 | .047168 |
| week2trt | -.0007078 | .0001418 | -4.99 | 0.000 | -.0009857 | -.0004299 |
| agetrt | -.0028666 | .0070498 | -0.41 | 0.684 | -.016684 | .0109508 |
| gendertrt | -.3779569 | .1766732 | -2.14 | 0.032 | -.72423 | -.0316838 |
| _cons | 2.5154 | .1515171 | 16.60 | 0.000 | 2.218432 | 2.812368 |

## c. Random intercept and random slopes for time t, and t^2 (model3)

| logcd4 | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| week | -.0037738 | .0026288 | -1.44 | 0.151 | -.0089262 | .0013785 |
| week2 | -.0002972 | .0000727 | -4.09 | 0.000 | -.0004397 | -.0001546 |
| trt | .3630764 | .3025003 | 1.20 | 0.230 | -.2298132 | .9559661 |
| age | .0115266 | .00358 | 3.22 | 0.001 | .0045098 | .0185433 |
| gender | .0093256 | .0895165 | 0.10 | 0.917 | -.1661235 | .1847748 |
| weektrt | .037052 | .0051613 | 7.18 | 0.000 | .026936 | .047168 |
| week2trt | -.0007078 | .0001418 | -4.99 | 0.000 | -.0009857 | -.0004299 |
| agetrt | -.0028666 | .0070498 | -0.41 | 0.684 | -.016684 | .0109508 |
| gendertrt | -.3779568 | .1766732 | -2.14 | 0.032 | -.7242298 | -.0316837 |
| _cons | 2.5154 | .1515171 | 16.60 | 0.000 | 2.218432 | 2.812368 |

## d. random intercept and random slopes for time t, t^2 and age (model4)

| logcd4 | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| week | -.0037738 | .0026288 | -1.44 | 0.151 | -.0089262 | .0013785 |
| week2 | -.0002972 | .0000727 | -4.09 | 0.000 | -.0004397 | -.0001546 |
| trt | .3630767 | .3025005 | 1.20 | 0.230 | -.2298135 | .9559669 |
| age | .0115266 | .00358 | 3.22 | 0.001 | .0045098 | .0185433 |
| gender | .0093256 | .0895166 | 0.10 | 0.917 | -.1661237 | .1847749 |
| weektrt | .037052 | .0051613 | 7.18 | 0.000 | .026936 | .047168 |
| week2trt | -.0007078 | .0001418 | -4.99 | 0.000 | -.0009857 | -.0004299 |
| agetrt | -.0028666 | .0070499 | -0.41 | 0.684 | -.0166841 | .0109509 |
| gendertrt | -.3779569 | .1766733 | -2.14 | 0.032 | -.7242302 | -.0316835 |
| _cons | 2.5154 | .1515172 | 16.60 | 0.000 | 2.218432 | 2.812368 |

## e. Can you assume a random slope for gender? Explain.

You can not assume a random slope for gender because gender is constant at all time points. Only time varying covariates can be included in the random part of the model.

## f. Plot the observed trajectories of four randomly selected individuals; two from the dual and two from the triple therapy. Compare these observed trajectories with the respective model predictions and the linear model mean estimates from each of the four previous models.
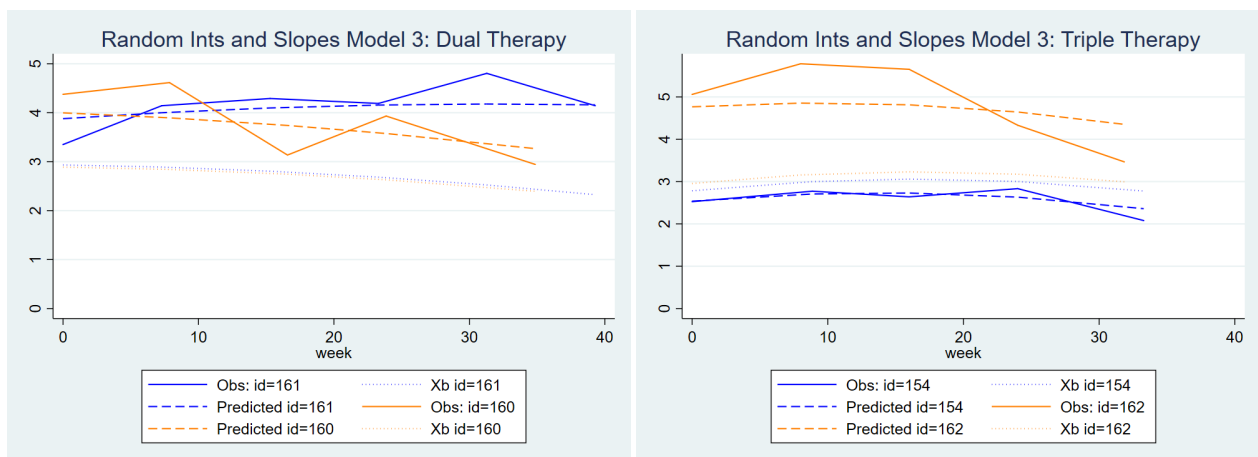
Here are the observed vs individual prediction vs marginal prediction trajectories for model 1, by group:
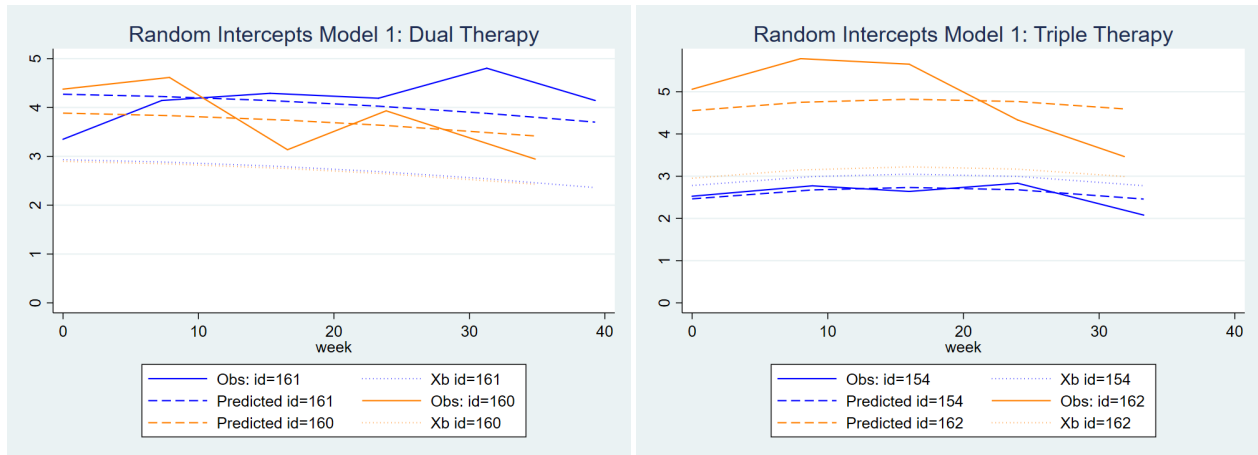
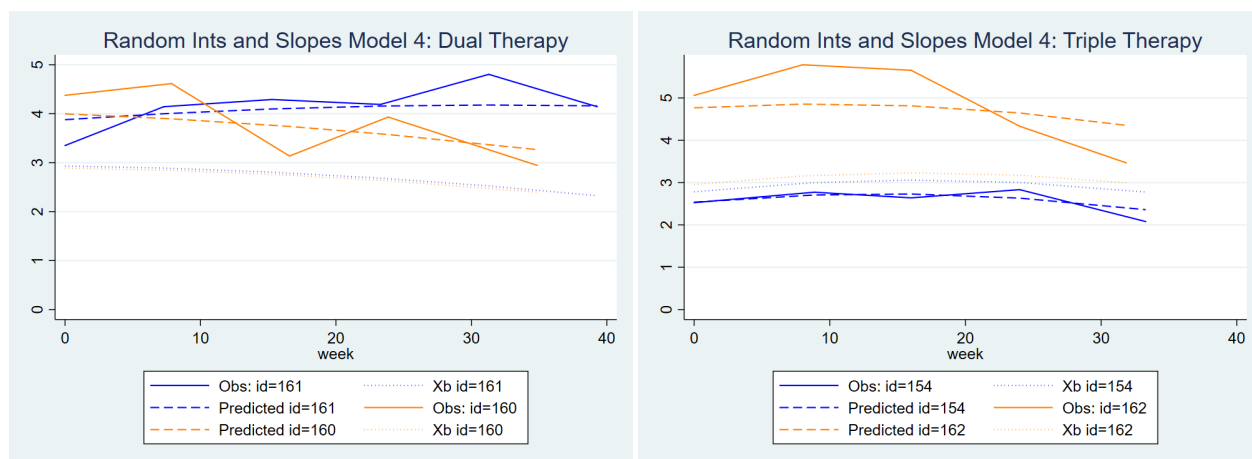The observed vs individual prediction vs marginal prediction trajectories for model 2, by group:
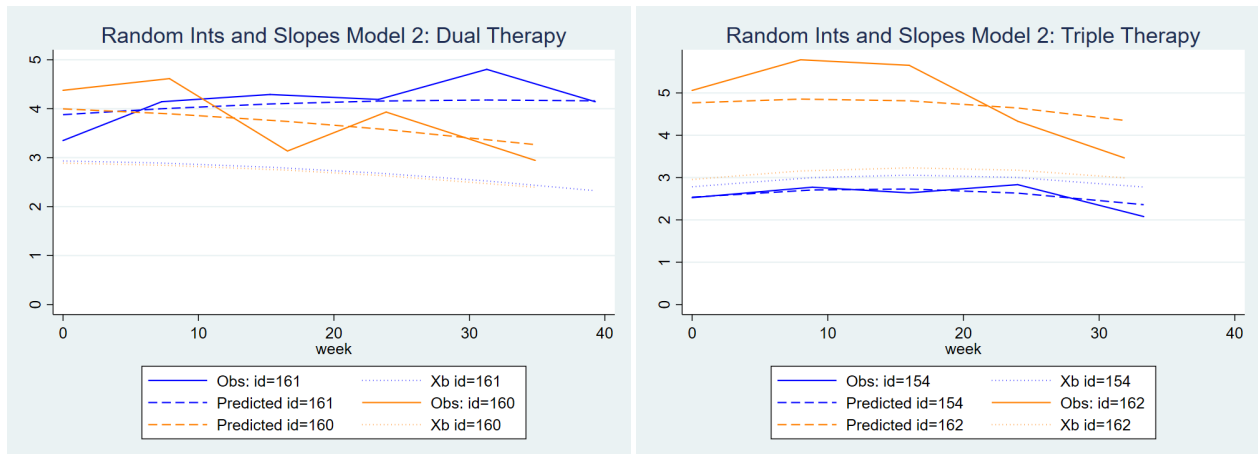


The observed vs individual prediction vs marginal prediction trajectories for model 3, by group:



Lastly, the observed vs individual prediction vs marginal prediction trajectories for model 4, by group:

| Random-effects Parameters | Estimate | Std. Err. | [95% Conf. Interval] | |
|---|---|---|---|---|
| **id: Independent** | | | | |
| var(week) | .0003098 | . | . | . |
| var(week2) | 2.90e-23 | . | . | . |
| var(age) | 1.06e-14 | . | . | . |
| var(_cons) | .6809629 | . | . | . |
| var(Residual) | .3238026 | . | . | . |

Figure 1:



In all instances, the mixed effects models predictions are closer to the observed predictions for the individual than the marginal model. We also see that including a random slope seems to improve the fit. Among the models with random slopes, it is difficult to determine visually which fits better (they look similar). The plots for the models with random slopes look similar because most of the variation is captured by the random intercepts. For example, here are the random parameters for model 4:

We see that the variance absorbed by the constant is .681, which is much larger than the variance in the slope terms.

## g. Which of the four models best fits the data? Explain.

We also examine the information criteria for each of the four models:

| Model | Obs | ll(null) | ll(model) | df | AIC | BIC |
|---|---|---|---|---|---|---|
| randint | 5,036 | . | -6092.666 | 12 | 12209.33 | 12287.62 |
| rand1 | 5,036 | . | -6022.316 | 13 | 12070.63 | 12155.45 |
| rand2 | 5,036 | . | -6022.316 | 14 | 12072.63 | 12163.97 |
| rand3 | 5,036 | . | -6022.316 | 10 | 12064.63 | 12129.88 |

The more complex model, which includes a random intercept and random slopes on time, time^2, and age has the lowest BIC and AIC. We thus believe that this model is best.

# Question 8: Based on the best model from the previous question what is your overall conclusion about the effect of the two treatments on the change of the log(CD4) over time?

To recap, here are the results of our best model (model 4):

| logcd4 | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| week | -.0037738 | .0026288 | -1.44 | 0.151 | -.0089262 | .0013785 |
| week2 | -.0002972 | .0000727 | -4.09 | 0.000 | -.0004397 | -.0001546 |
| trt | .3630767 | .3025005 | 1.20 | 0.230 | -.2298135 | .9559669 |
| age | .0115266 | .00358 | 3.22 | 0.001 | .0045098 | .0185433 |
| gender | .0093256 | .0895166 | 0.10 | 0.917 | -.1661237 | .1847749 |
| weektrt | .037052 | .0051613 | 7.18 | 0.000 | .026936 | .047168 |
| week2trt | -.0007078 | .0001418 | -4.99 | 0.000 | -.0009857 | -.0004299 |
| agetrt | -.0028666 | .0070499 | -0.41 | 0.684 | -.0166841 | .0109509 |
| gendertrt | -.3779569 | .1766733 | -2.14 | 0.032 | -.7242302 | -.0316835 |
| _cons | 2.5154 | .1515172 | 16.60 | 0.000 | 2.218432 | 2.812368 |

| Random-effects Parameters | Estimate | Std. Err. | [95% Conf. Interval] | |
|---|---|---|---|---|
| **id: Independent** | | | | |
| var(week) | .0003098 | . | . | . |
| var(week2) | 2.90e-23 | . | . | . |
| var(age) | 1.06e-14 | . | . | . |
| var(_cons) | .6809629 | . | . | . |
| var(Residual) | .3238026 | . | . | . |

The trend of log CD4 levels over time for the dual intervention group is given by the coefficients on week and week^2. The week coefficient means that log CD4 decreases initially at a rate of .0038. The week^2 coefficient means that the marginal rate of decrease in log CD4 levels is increasing. In other words, the slope starts negative and becomes increasingly more negative over time for the dual intervention group.

The trend of log CD4 levels over time for the triple intervention group is given by the coefficients on week, week x trt, week^2 and week^2 x trt. Initially, the average log CD4 levels are increasing at a rate of .0334 (week + week x trt). The rate of increase is decreasing over time by .001 (week^2 + week^2 x trt). In other words, the slope for the triple intervention group starts positive but becomes less positive, and eventually negative over time.