

## ST 790, Homework 2 Solutions Spring 2018

### 1. *Finite-sample properties of estimators in a population-averaged model.*

(a) I did my simulation in R using the `gls()` function. I calculated the correct model-based standard errors and the robust standard errors using the homegrown function `robust.cov()`.

(b) Here are summaries of the results I got:

```
## > beta.est (Monte Carlo average of estimates)
## Method beta_0G beta_0B beta_1G beta_1B
## 1 TRUE 17.302856 16.283489 0.479162 0.78031
## 2 OLS 17.307295 16.308135 0.47925 0.777127
## 3 AR1 17.296802 16.30377 0.479771 0.777527
## 4 CS 17.307295 16.308135 0.47925 0.777127
## 5 UN 17.326025 16.287962 0.477398 0.780001

## > beta.sd (Monte Carlo SD of estimates)
## Method beta_0G beta_0B beta_1G beta_1B
## 1 TRUE 1.696502 1.364295 0.139641 0.113135
## 2 OLS 1.815164 1.460908 0.163254 0.132193
## 3 AR1 1.80458 1.437862 0.15995 0.128864
## 4 CS 1.815164 1.460908 0.163254 0.132193
## 5 UN 1.751252 1.421178 0.144219 0.117911

## > beta.se.model (Monte Carlo average of model-based SE estimates)
## Method beta_0G beta_0B beta_1G beta_1B
## 1 TRUE 1.66438 1.38004 0.13884 0.11512
## 2 OLS 1.63316 1.35414 0.14549 0.12064
## 3 AR1 1.54913 1.28447 0.1333 0.11053
## 4 CS 1.35391 1.12261 0.11523 0.09554
## 5 UN 1.59245 1.32039 0.1323 0.1097

## > beta.se.robust (Monte Carlo average of robust SE estimates)
## Method beta_0G beta_0B beta_1G beta_1B
## 1 TRUE 1.56275 1.34613 0.13066 0.11231
## 2 OLS 1.66273 1.43212 0.15066 0.1297
## 3 AR1 1.64602 1.4201 0.14742 0.12701
## 4 CS 1.66273 1.43212 0.15066 0.1297
## 5 UN 1.51996 1.3141 0.12679 0.10937

## > beta.sdse.model (Ratio of MC SDs to MC average of model-based SEs)
## Method beta_0G beta_0B beta_1G beta_1B
## 1 TRUE 1.019 0.989 1.006 0.983
## 2 OLS 1.111 1.079 1.122 1.096
## 3 AR1 1.165 1.119 1.2 1.166
## 4 CS 1.341 1.301 1.417 1.384
```

```
## 5      UN      1.1    1.076    1.09    1.075
```

```
## > beta.sdse.robust (Ratio of MC SDs to MC average of robust SEs)
```

```
## Method beta_0G beta_0B beta_1G beta_1B
## 1   TRUE    1.086    1.013    1.069    1.007
## 2   OLS     1.092     1.02    1.084    1.019
## 3   AR1     1.096    1.013    1.085    1.015
## 4   CS      1.092     1.02    1.084    1.019
## 5   UN      1.152    1.081    1.137    1.078
```

```
## > beta.mse (Ratio MSE(TRUE)/MSE(this estimator))
```

```
## Method beta_0G beta_0B beta_1G beta_1B
## 1   OLS     0.874    0.872    0.732    0.732
## 2   AR1     0.884     0.9    0.762    0.771
## 3   CS      0.874    0.872    0.732    0.732
## 4   UN      0.938    0.922    0.937    0.921
```

(c) The MC averages of the estimates of beta are all very close to the true values, which probably reflects that all of these estimators are indeed all consistent estimators for the true value of  $\beta$ , even with the covariance structure misspecified. In fact, it appears that the estimators are all approximately unbiased for this finite sample size.

(d) Under the true covariance model with  $\alpha$  known, the model-based standard errors do a pretty good job of approximating the standard deviation of the true sampling distribution of the estimator for beta (the ratio of the MC SD to the average of SEs is close to 1 for all four components of beta. When  $\alpha$  is not known, we still have the correct covariance structure, but the model-based standard errors seem to underestimate the true sampling variation somewhat, as all of the ratios are greater than 1. This might be because the first order asymptotic theory does not adequately account for the additional variation due to estimating alpha. However, when the covariance model used is incorrect (OLS, AR1, CS), the model-based standard errors understate the sampling variation even more dramatically, reflecting not only estimation of correlation parameters but also the fact that the covariance model is incorrect.

The robust standard errors really aren't needed for the ideal case where the true correlation matrix is known, and in fact they tend to understate the true sampling variation, whereas the model-based SEs get it approximately correct. In the more realistic case where the correlation structure is correctly specified but  $\alpha$  is estimated, the robust standard errors again aren't really needed, and they perform similarly to the model-based SEs, with perhaps even some degradation of performance. In both of these cases, the extra variation due to estimating the "meat" and "bread" of the robust covariance matrix is actually detrimental. In the cases of OLS, AR1, and CS, where the correlation model is incorrectly specified, the robust standard errors do lead to substantially better assessments of the true sampling variation relative to the model-based SEs, as advertised. However, in all cases, the robust SEs still do understate the true sampling variation somewhat; this is probably because of the fact that the first order theory is tenuous for this relatively small sample size.

(e) The MSE ratios here are all relative to the TRUE model, so that values less than 1 reflect inefficiency relative to that model. In all cases, the ratios are less than 1. The values in the last line are the MSE ratios comparing the estimator with the correlation structure correctly

specified as unstructured but with alpha estimated to that with the correlation structure correct and known (alpha known). While these values are all greater than 0.9, they do reflect some degradation of precision due to the need to estimate alpha rather than knowing it. The ratios for the other three cases with incorrect correlation models (OLS, AR1, CS) are pretty convincingly less than 1, particularly for the slope parameters, reflecting some pretty significant degradation of precision. Part of this is undoubtedly due to the fact that the correlation model is misspecified, as the first order theory predicts. The 1st order theory says that, at least for larger  $m$ , there should be no penalty for estimation of correlation parameters; however, the case where the correlation model is correct shows that there is a penalty, so it is reasonable to surmise that part of the reason the estimators for beta under incorrect correlation models show substantial loss of efficiency relative to the ideal case is due to the fact that the optimistic theory may not be completely relevant in this smallish sample size (small  $m$ ).

## 2. Another perspective on REML.

(a) (i) We show this by brute force; there are slicker ways to do this. We want to find

$$E\{(\mathbf{Y}_i - \mathbf{X}_i\hat{\beta})^T \mathbf{V}_i^{-1} \mathbf{V}_{\xi_{ki}} \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\hat{\beta}) \mid \tilde{\mathbf{x}}\}. \quad (1)$$

With  $\hat{\beta} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y}$ , it is straightforward to observe by substitution that  $E(\mathbf{Y}_i - \mathbf{X}_i\hat{\beta}) = \mathbf{0}$ . So from the result for the expectation of a quadratic form in Appendix A, we want to find

$$\begin{aligned} & \text{tr}[E\{(\mathbf{Y}_i - \mathbf{X}_i\hat{\beta})(\mathbf{Y}_i - \mathbf{X}_i\hat{\beta})^T \mid \tilde{\mathbf{x}}\} \mathbf{V}_i^{-1} \mathbf{V}_{\xi_{ki}} \mathbf{V}_i^{-1}] \\ &= \text{tr}[\mathbf{V}_i^{-1/2} E\{(\mathbf{Y}_i - \mathbf{X}_i\hat{\beta})(\mathbf{Y}_i - \mathbf{X}_i\hat{\beta})^T \mid \tilde{\mathbf{x}}\} \mathbf{V}_i^{-1/2} \mathbf{V}_{\xi_{ki}} \mathbf{V}_i^{-1/2}], \end{aligned} \quad (2)$$

using the usual properties of the trace. So we find  $E\{(\mathbf{Y}_i - \mathbf{X}_i\hat{\beta})(\mathbf{Y}_i - \mathbf{X}_i\hat{\beta})^T \mid \tilde{\mathbf{x}}\}$ . Adding and subtracting  $\mathbf{X}_i\beta$  in  $(\mathbf{Y}_i - \mathbf{X}_i\hat{\beta})$ ,

$$\begin{aligned} & (\mathbf{Y}_i - \mathbf{X}_i\hat{\beta})(\mathbf{Y}_i - \mathbf{X}_i\hat{\beta})^T \\ &= \{(\mathbf{Y}_i - \mathbf{X}_i\beta) - \mathbf{X}_i(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\beta)\} \{(\mathbf{Y}_i - \mathbf{X}_i\beta) - \mathbf{X}_i(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\beta)\}^T \\ &= (\mathbf{Y}_i - \mathbf{X}_i\beta)(\mathbf{Y}_i - \mathbf{X}_i\beta)^T - (\mathbf{Y}_i - \mathbf{X}_i\beta)(\mathbf{Y} - \mathbf{X}\beta) \mathbf{V}^{-1} \mathbf{X}(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}_i^T \\ &\quad - \mathbf{X}_i(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\beta)(\mathbf{Y}_i - \mathbf{X}_i\beta)^T \\ &\quad + \mathbf{X}_i(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\beta)(\mathbf{Y} - \mathbf{X}\beta)^T \mathbf{V}^{-1} \mathbf{X}(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}_i^T \end{aligned} \quad (3)$$

The expectation of the first term in (3) is  $\mathbf{V}_i$ , and the expectation of the last term can be seen to be

$$\mathbf{X}_i(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}_i^T. \quad (4)$$

For first middle term in (3), observe that  $(\mathbf{Y}_i - \mathbf{X}_i\beta)(\mathbf{Y} - \mathbf{X}\beta)^T$  is a  $(n_i \times N)$  matrix of the form

$$\{(\mathbf{Y}_i - \mathbf{X}_i\beta)(\mathbf{Y}_1 - \mathbf{X}_1\beta)^T, \dots, (\mathbf{Y}_i - \mathbf{X}_i\beta)(\mathbf{Y}_i - \mathbf{X}_i\beta)^T, \dots, (\mathbf{Y}_i - \mathbf{X}_i\beta)(\mathbf{Y}_m - \mathbf{X}_m\beta)^T\}.$$

Clearly, the conditional expectation of this matrix is, by independence across  $i$ ,

$$(\mathbf{0}, \dots, \mathbf{V}_i, \dots, \mathbf{0}). \quad (5)$$

Now  $\mathbf{V}^{-1}$  is a block diagonal matrix with  $\mathbf{V}_i^{-1}$ ,  $i = 1, \dots, m$ , on the diagonal, and  $\mathbf{X}$  is the  $(N \times p)$  matrix with the  $\mathbf{X}_i$ ,  $i = 1, \dots, m$ , stacked. It can then be seen that premultiplying  $\mathbf{V}^{-1} \mathbf{X}$

by (5) yields  $\mathbf{X}_i$  ( $n_i \times p$ ), so that the first middle term in (3) is equal to (4). Likewise, as the second middle term in (3) is just the transpose of the second, its expectation is also equal to (4). Combining, we thus obtain

$$E\{(\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})(\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})^T | \tilde{\mathbf{x}}\} = \mathbf{V}_i - \mathbf{X}_i(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}_i^T. \quad (6)$$

Thus, (2) becomes

$$\begin{aligned} & \text{tr}[\mathbf{V}_i^{-1/2} \{ \mathbf{V}_i - \mathbf{X}_i(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}_i^T \} \mathbf{V}_i^{-1/2} \mathbf{V}_{\xi_{ki}}^{-1/2} \mathbf{V}_{\xi_{ki}} \mathbf{V}_i^{-1/2}] \\ &= \text{tr}\{(\mathbf{I} - \mathbf{H}_i) \mathbf{V}_i^{-1/2} \mathbf{V}_{\xi_{ki}} \mathbf{V}_i^{-1/2}\} \\ &= \text{tr}\{(\mathbf{I} - \mathbf{H}_i) \mathbf{V}_i^{-1} \mathbf{V}_{\xi_{ki}}\}. \end{aligned}$$

where the last equality follows because all of these matrices are symmetric.

(ii) This term can be written using properties of the trace as

$$\begin{aligned} & \sum_{i=1}^m \text{tr} \left\{ (\mathbf{X}^T \mathbf{V} \mathbf{X})^{-1} \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{V}_{\xi_{ki}} \mathbf{V}_i^{-1} \mathbf{X}_i \right\} \\ &= \sum_{i=1}^m \text{tr} \left\{ \mathbf{V}_i^{-1/2} \mathbf{X}_i (\mathbf{X}^T \mathbf{V} \mathbf{X})^{-1} \mathbf{X}_i^T \mathbf{V}_i^{-1/2} \mathbf{V}_i^{-1} \mathbf{V}_{\xi_{ki}} \right\} \\ &= \sum_{i=1}^m \text{tr}(\mathbf{H}_i \mathbf{V}_i^{-1} \mathbf{V}_{\xi_{ki}}). \end{aligned} \quad (7)$$

(b) A summand of the ML equations is

$$(\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})^T \mathbf{V}_i^{-1} \mathbf{V}_{\xi_{ki}} \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}) - \text{tr}(\mathbf{V}_i^{-1} \mathbf{V}_{\xi_{ki}}).$$

Thus, it is immediate from (a)(i) that the expectation of this quantity given  $\tilde{\mathbf{x}}$  is

$$\text{tr}\{(\mathbf{I} - \mathbf{H}_i) \mathbf{V}_i^{-1} \mathbf{V}_{\xi_{ki}}\} - \text{tr}(\mathbf{V}_i^{-1} \mathbf{V}_{\xi_{ki}}) = -\text{tr}(\mathbf{H}_i \mathbf{V}_i^{-1} \mathbf{V}_{\xi_{ki}}),$$

which is not equal to zero. Likewise, using (7), a summand of the REML equation is

$$\begin{aligned} & (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})^T \mathbf{V}_i^{-1} \mathbf{V}_{\xi_{ki}} \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}) - \text{tr}(\mathbf{V}_i^{-1} \mathbf{V}_{\xi_{ki}}) + \text{tr}(\mathbf{H}_i \mathbf{V}_i^{-1} \mathbf{V}_{\xi_{ki}}) \\ &= (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})^T \mathbf{V}_i^{-1} \mathbf{V}_{\xi_{ki}} \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}) - \text{tr}\{(\mathbf{I} - \mathbf{H}_i) \mathbf{V}_i^{-1} \mathbf{V}_{\xi_{ki}}\}. \end{aligned}$$

It follows from (a)(i) that the conditional expectation of this quantity is equal to zero.

(c) (i) Under these conditions,  $\xi = \xi = \sigma^2$  and  $\mathbf{V}_i = \sigma^2 \mathbf{I}$ . Moreover, the  $i$ th diagonal element of  $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$  is  $H_{ii} = h_{ii} = \mathbf{X}_i(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}_i$ , a scalar. Thus, from (6), it is immediate that

$$E(r_i^2 | \tilde{\mathbf{x}}) = \sigma^2 - \sigma^2 \mathbf{X}_i(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}_i = \sigma^2(1 - h_{ii}).$$

(ii) Here,  $\mathbf{V}_{\xi_i} = 1$ , so, using this and the above, the REML estimating equation becomes

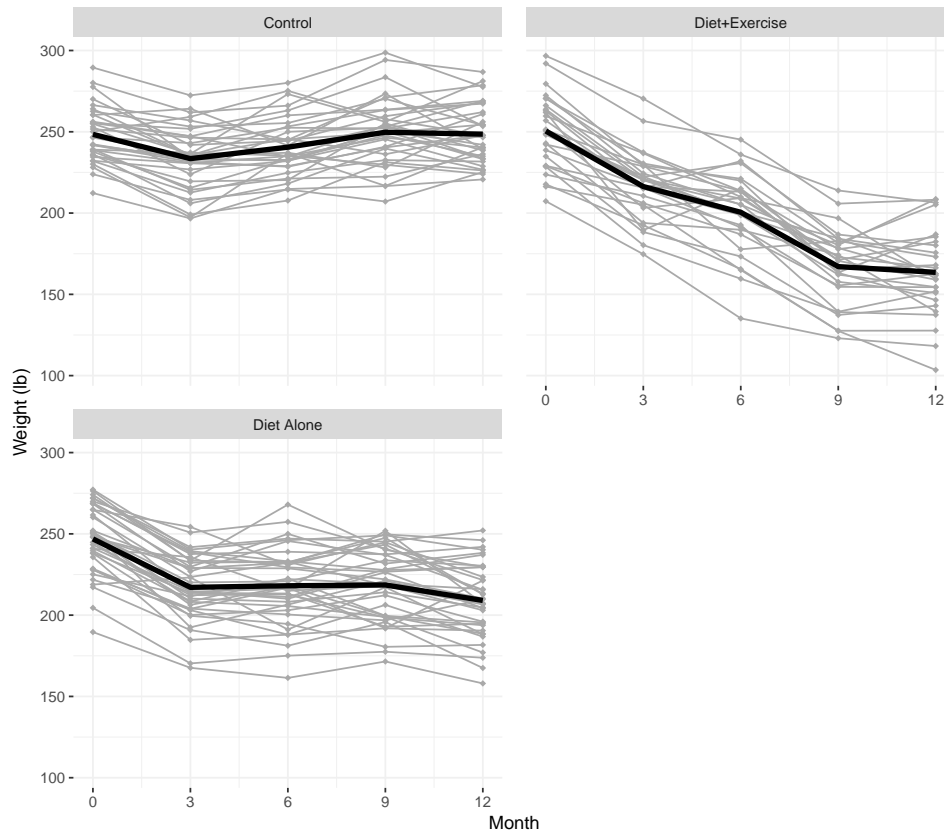
$$\sum_{i=1}^m \left\{ \frac{(\mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})^2}{\sigma^4} - \frac{(1 - h_{ii})}{\sigma^2} \right\} = 0,$$

which reduces by algebra to

$$\hat{\sigma}^2 = \left\{ \sum_{i=1}^m (1 - h_{ii}) \right\}^{-1} \sum_{i=1}^m (Y_i - \mathbf{X}_i \hat{\beta})^2 = (m - p)^{-1} \sum_{i=1}^m (Y_i - \mathbf{X}_i \hat{\beta})^2,$$

which is the usual bias-corrected estimator for the assumed constant variance in the classical linear model, where the last equality follows because  $\mathbf{H}$  is symmetric, idempotent matrix, so that  $\sum_{i=1}^m h_{ii} = \text{tr}(\mathbf{H}) = p$ .

3. *Effectiveness of weightloss programs, continued.* Here is a plot of the data again.



Because we are taking a population-averaged point of view, we interpret the investigators' questions as pertaining to the overall mean trajectories for each group. Thus, (i) can be interpreted as asking if the overall mean trajectories for the weight loss programs (Diet+Exercise and Diet Alone) decrease over the study period; (ii) can be addressed by assessing and comparing the patterns of change of the mean trajectories; (iii) can be interpreted as referring to the rate of change of the mean trajectories, and (iv) can be addressed by estimating and comparing the overall population means for each group.

The plot suggests that the overall mean trajectory for each group could be reasonably represented by a straight line, although there is hint of some curvature in each panel: for the Control condition, there seems to be an initial decline (maybe a "placebo effect?") followed by an eventual rise back to where things started; for the Diet+Exercise condition, the mean declines but then seems to plateau at the later times; and for the Diet Alone condition, there

is again an initial decline followed by what appears to be a plateau at a lower weight value. The investigators' questions (iii) seems to be directed to the issue of whether or not these patterns are "real" or if the overall mean trends are basically straight lines.

Accordingly, taking a PA perspective, a natural first step is to adopt a quadratic model for the population mean trajectory in each group, which can approximate these behaviors (although a nonlinear model that approaches an asymptote might be better!), and see if there is sufficient evidence suggesting there is curvature (in which case the rate of change is not constant; specifically, if the true population mean trajectory for any of the groups is quadratic, namely of the form

$$\beta_0 + \beta_1 t + \beta_2 t^2,$$

then, taking derivatives with respect to  $t$ , the rate of change at time  $t$  is

$$\beta_1 + 2\beta_2 t. \tag{8}$$

If the evidence supports the need for quadratic models over straight lines as representations of the mean trajectories for at least one group, then the rate of change for that group would not be constant. In that case, a reasonable thing to do would be to provide estimates of the rate of change at selected time points using (8), accompanied by standard errors (of course).

In the attached programs in SAS and R we fit a quadratic model with different intercepts, linear, and quadratic effects for each group with different posited structures for the overall covariance; because we have already investigated overall covariance structure in Homework 1, we stick to unstructured and compound symmetric models. SAS and R allow or don't allow the analyst to consider different versions of these structures – as we have emphasized, the `gls()` function in R does not support having different correlation structures in different groups. As it turns out, I found that a common unstructured covariance model is best supported; SAS does not allow the variances to be the same over time, while R does, and the evidence in R seems to support this latter model. So in SAS I used the common unstructured model with different variances over time, and in R used the common unstructured model with the same variance over time. In both cases, I didn't bother with robust standard errors, as with an unstructured correlation model it is a bit of overkill, but you may have used robust standard errors, and there is nothing wrong with that.

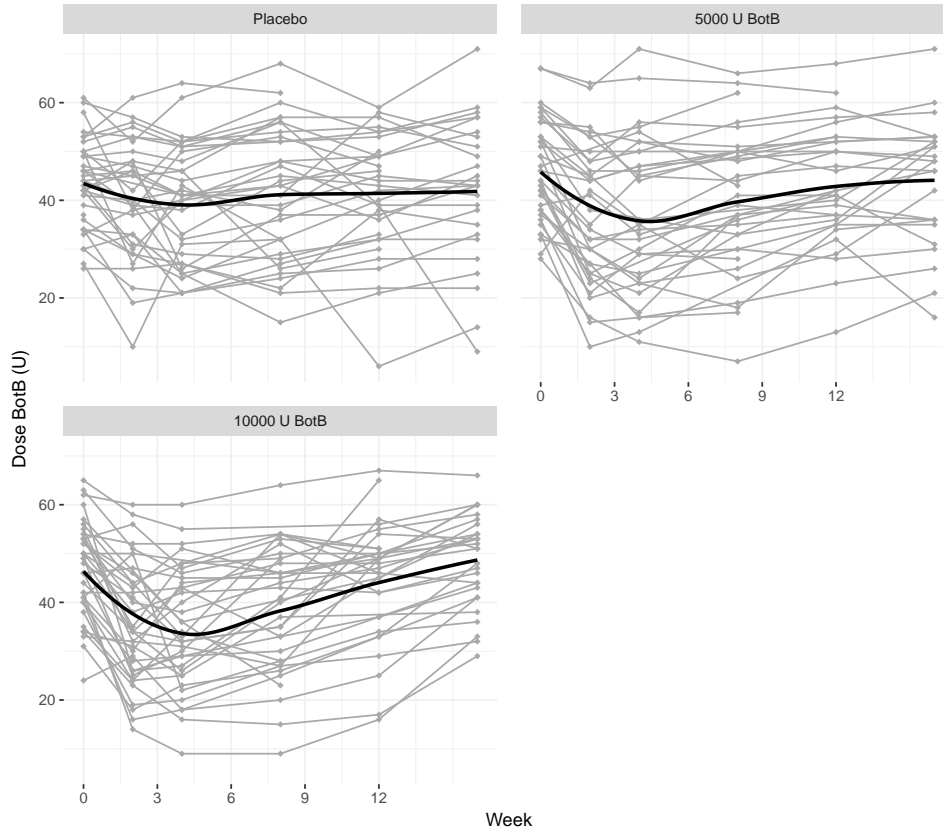
I did not simplify further to have a common intercept for all three groups on the basis of randomization, but you may have. Of course, it is also possible to work with an alternative parameterization of the model, treating one of the groups as the reference group.

I constructed a test of whether or not there is a quadratic effect in at least one group, and the evidence is very strong that there is; in fact, all of the estimates of the quadratic terms are very large relative to their standard errors. There does not seem to be strong evidence that these effects differ in a formal test, though, and the estimates are very similar. You might have chosen to reduce the model to have a common quadratic effect. See the programs.

I went ahead and used this same model to estimate the mean weight at 12 months for each group. Because the mean trajectories are not straight lines, it would be reasonable to estimate the rate of change for each group at several different time points using (8) above to get a sense of how change is different for each group; I did this for 8 months as an example.

You should have written out a formal model and stated your assumptions about the model, discussed the steps you took to settle on an appropriate working covariance structure assumption, formalized and addressed the investigators' questions in terms of your model, and provided standard errors for any estimates you provided to the investigators.

4. *Cervical dystonia clinical trial.* Here are spaghetti plots of the data with loess smooths superimposed.



In the 5000 U and 10000 U groups, the two “phases” are clearly visible; there is what looks to be a decline at what appears to be an approximately constant rate in each of these groups in phase 1, until week 4, when treatment was withdrawn, followed by an increase at an approximately constant rate in phase 2, from week 4 to week 16. In the placebo group, there is a hint of a decline in phase 1 while phase 2 shows a slight rebound and looks fairly flat over weeks 4 to 16.

Given that the investigators expect two phases, with a “change point” at 4 weeks, and the visual evidence supports two phases for which the mean trajectory for each group appears to follow an approximate straight line, a natural PA model is a linear spline model for the population mean for each group similar to that used for the guinea pig diet study in (5.21) of the notes, except that here, the slope of the first phase (and possibly the intercept, too), should be different for the three groups, as the treatments were started at baseline (time 0). So a reasonable first shot at a model for the overall mean at time  $t$  for individual  $i$  is

$$\sum_{\ell=1}^3 d_{i\ell} \{ \beta_{0\ell} + \beta_{1\ell}t + \beta_{2\ell}(t-4)_+ \},$$

where  $d_{i\ell} = 1$  if individual  $i$  was randomized to group  $\ell$ ,  $\ell = 1, 2, 3$ , in which case the constant rate of change in phase 1 for group  $\ell$  is  $\beta_{1\ell}$  and that for phase 2 is  $\beta_{1\ell} + \beta_{2\ell}$ . You might have parameterized your model differently. You could also assume a common intercept, given

randomization, or test this explicitly to investigate the integrity of the randomization. In the attached programs, I used 3 separate intercepts.

It is critical to recognize that here, because there are missing observations, we must be willing to believe that these data are missing according to a **missing at random** mechanism and that multivariate normality of the TWSTRS measures given covariates holds. In practice, we would have the opportunity to discuss this with the investigators and based this assumption on the substantive issues. From the discussion in Section 5.6 of the notes, we use maximum likelihood (ML) methods to carry out the analyses and model-based standard errors.

See the attached programs for initial fits of this model. I first fit a separate unstructured covariance model in each group and inspected the resulting group-specific estimates, which seem to suggest possible compound symmetry possibly different across groups, although there is some suggestion that correlations might damp out over time. For illustration, I just stuck to unstructured and compound symmetric structures in the attached; you may have also investigated an exponential structure (which is needed because the times are not equally spaced). I did not include age and gender in the initial model. Using `gls()` in R, from the information criteria, a common compound symmetric structure with common constant variance across groups and time was preferred based on BIC, while using SAS a common compound symmetric structure with common variances different over time was preferred based on BIC; recall R and SAS define BIC differently. I adopted the preferred model in each of R and SAS in my programs. You may have investigated other models and may have made a different decision, which is fine.

In the attached SAS and R programs, using the preferred models, there is very strong evidence that the steepness of the decline in phase 1 and rebound in phase 2 (slopes) is different from zero in each group and furthermore differs across groups. Estimates of the phase 1 slope are immediate from the fit and can be constructed from the fit for phase 2. The mean TWSTRS score can be estimated for each group; there is not strong evidence that these means differ. There is no evidence that the baseline means differ, as would be expected for a randomized study, so you may have chosen to reduce your model to have a common intercept.

I fit modified models allowing the baseline mean and phase 1 slope to depend on age and gender either in different or the same way for each group. There does not seem to be compelling evidence that mean TWSTRS at baseline or the slope of the mean trajectory at phase 1 is associated with either of these factors/

You may have considered other models and addressed things differently. You should have given a rationale for the model you adopted and then stated that model and how you formalized and addressed the questions of interest within it.