

ST 790, Homework 3 Solutions
Spring 2018

1. *Connection with individual estimation.*

(a) Because $\mathbf{X}_i = \mathbf{Z}_i \mathbf{A}_i$, we have immediately that

$$\hat{\beta} = \left\{ \sum_{i=1}^m \mathbf{A}_i^T (\mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i) \mathbf{A}_i \right\}^{-1} \sum_{i=1}^m \mathbf{A}_i^T (\mathbf{Z}_i \mathbf{V}_i^{-1} \mathbf{Y}_i).$$

So consider each of the components of this expression in turn. First using a standard matrix inversion formula in Appendix A,

$$\begin{aligned} \mathbf{V}_i^{-1} &= (\mathbf{R}_i + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T)^{-1} \\ &= \mathbf{R}_i^{-1} - \mathbf{R}_i^{-1} \mathbf{Z}_i (\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T \mathbf{R}_i^{-1}. \end{aligned}$$

Thus, substituting this into the leftmost component of $\hat{\beta}$, and using standard properties of the inverse of the product of matrices, by brute force,

$$\begin{aligned} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Y}_i &= \mathbf{Z}_i^T \{ \mathbf{R}_i^{-1} - \mathbf{R}_i^{-1} \mathbf{Z}_i (\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T \mathbf{R}_i^{-1} \} \mathbf{Y}_i \\ &= \{ \mathbf{I} - \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i (\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i)^{-1} \} \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Y}_i \\ &= \{ \mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i - \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i \} (\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Y}_i \\ &= \mathbf{D}^{-1} (\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Y}_i \\ &= \{ (\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i) \mathbf{D} \}^{-1} \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Y}_i \\ &= [(\mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i) \{ (\mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i)^{-1} + \mathbf{D} \}]^{-1} \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Y}_i \\ &= \{ (\mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i)^{-1} + \mathbf{D} \}^{-1} (\mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Y}_i \\ &= \{ \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i \}^{-1} + \mathbf{D} \}^{-1} \hat{\beta}_i. \end{aligned}$$

For the other component of $\hat{\beta}$, we have

$$\begin{aligned} \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i &= \mathbf{Z}_i^T \{ \mathbf{R}_i^{-1} - \mathbf{R}_i^{-1} \mathbf{Z}_i (\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T \mathbf{R}_i^{-1} \} \mathbf{Z}_i \\ &= \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i - \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i (\mathbf{D}^{-1} + \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i \\ &= \{ \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i \}^{-1} + \mathbf{D} \}^{-1} \end{aligned}$$

where the last equality follows from the standard matrix inversion formula in Appendix A. These calculations and (4) show that

$$\mathbf{W}_i = (\mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i)^{-1} + \mathbf{D}.$$

(b) Substituting the expression for \mathbf{Y}_i and using $\mathbf{X}_i = \mathbf{Z}_i \mathbf{A}_i$, we have

$$\begin{aligned} \hat{\beta}_i &= (\mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T \mathbf{R}_i^{-1} (\mathbf{Z}_i \mathbf{A}_i \beta + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i) \\ &= \mathbf{A}_i \beta + \mathbf{b}_i + (\mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{e}_i. \end{aligned}$$

It follows immediately using the fact that the conditional means of \mathbf{b}_i and \mathbf{e}_i are zero and that \mathbf{b}_i and \mathbf{e}_i are independent that

$$E(\hat{\beta}_i | \mathbf{x}_i) = \mathbf{A}_i \beta,$$

and

$$\begin{aligned}\text{var}(\hat{\beta}_i|\mathbf{x}_i) &= \mathbf{D} + (\mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{R}_i \mathbf{R}_i^{-1} \mathbf{Z}_i (\mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i)^{-1} \\ &= \mathbf{D} + (\mathbf{Z}_i^T \mathbf{R}_i^{-1} \mathbf{Z}_i)^{-1} = \mathbf{W}_i.\end{aligned}$$

(c) From (b) and independence across i , it follows immediately that

$$E(\hat{\beta}|\tilde{\mathbf{x}}) = \left(\sum_{i=1}^m \mathbf{A}_i^T \mathbf{W}_i^{-1} \mathbf{A}_i \right)^{-1} \sum_{i=1}^m \mathbf{A}_i^T \mathbf{W}_i^{-1} \mathbf{A}_i \beta = \beta,$$

so that $\hat{\beta}$ is conditionally unbiased. We find $\text{var}(\hat{\beta}|\tilde{\mathbf{x}})$ by brute force. We suppress the conditioning for brevity in the following argument:

$$\begin{aligned}\text{var}(\hat{\beta}|\tilde{\mathbf{x}}) &= E\{(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T\} \\ &= E \left[\left(\sum_{i=1}^m \mathbf{A}_i^T \mathbf{W}_i^{-1} \mathbf{A}_i \right)^{-1} \sum_{i=1}^m \mathbf{A}_i^T \mathbf{W}_i^{-1} (\hat{\beta}_i - \mathbf{A}_i \beta) \sum_{i=1}^m (\hat{\beta}_i - \mathbf{A}_i \beta)^T \mathbf{W}_i \mathbf{A}_i \left(\sum_{i=1}^m \mathbf{A}_i^T \mathbf{W}_i^{-1} \mathbf{A}_i \right)^{-1} \right] \\ &= \left(\sum_{i=1}^m \mathbf{A}_i^T \mathbf{W}_i^{-1} \mathbf{A}_i \right)^{-1} \sum_{i=1}^m \mathbf{A}_i^T \mathbf{W}_i^{-1} E \{ (\hat{\beta}_i - \mathbf{A}_i \beta)(\hat{\beta}_i - \mathbf{A}_i \beta)^T \} \mathbf{W}_i \mathbf{A}_i \left(\sum_{i=1}^m \mathbf{A}_i^T \mathbf{W}_i^{-1} \mathbf{A}_i \right)^{-1} \\ &= \left(\sum_{i=1}^m \mathbf{A}_i^T \mathbf{W}_i^{-1} \mathbf{A}_i \right)^{-1} \sum_{i=1}^m \mathbf{A}_i^T \mathbf{W}_i^{-1} \mathbf{W}_i \mathbf{W}_i^{-1} \mathbf{A}_i \left(\sum_{i=1}^m \mathbf{A}_i^T \mathbf{W}_i^{-1} \mathbf{A}_i \right)^{-1} \\ &= \left(\sum_{i=1}^m \mathbf{A}_i^T \mathbf{W}_i^{-1} \mathbf{A}_i \right)^{-1},\end{aligned}$$

where the third inequality follows by independence across i .

(d) Under these conditions, directly from (3), $\hat{\beta}$ reduces to

$$\hat{\beta} = \left(\sum_{i=1}^m \mathbf{x}_i^T \mathbf{x}_i \right)^{-1} \sum_{i=1}^m \mathbf{x}_i^T \mathbf{y}_i;$$

alternatively, this can be obtained from (4) by noting that

$$\mathbf{W}_i = \sigma^2 (\mathbf{Z}_i^T \mathbf{Z}_i)^{-1}$$

and

$$\hat{\beta}_i = \left(\sum_{i=1}^m \mathbf{Z}_i^T \mathbf{Z}_i \right)^{-1} \sum_{i=1}^m \mathbf{Z}_i^T \mathbf{y}_i$$

and substituting into (4) and simplifying. Thus, $\hat{\beta}$ reduces the OLS estimator treating the observations across $i = 1, \dots, m$ and $j = 1, \dots, n_i$ as mutually independent.

When $\mathbf{D} = \mathbf{0}$, this implies that there is no variation at all across individuals, so that, effectively, all observations can be viewed as arising from the same underlying inherent trajectory $\mathbf{A}_i \beta$. Moreover, with $\mathbf{R}_i = \sigma^2 \mathbf{I}_{n_i}$, the components of \mathbf{e}_i are mutually independent for $j = 1, \dots, n_i$, so that all N observations can be viewed as independent observations on this trajectory.

The obvious estimator is thus the OLS estimator, which is based on the assumption that all observations are independent. More formally, the model collapses to

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{e}_i,$$

with \mathbf{e}_{ij} independent for all i and j , for which OLS is the maximum likelihood estimator under the assumptions of normality here.

(e) Under these conditions,

$$\mathbf{W}_i = \sigma^2 \sigma^2 (\mathbf{Z}_i^T \mathbf{Z}_i)^{-1} + d \mathbf{I}.$$

Substituting in (4) and dividing through by d gives

$$\hat{\boldsymbol{\beta}} = \left\{ \sum_{i=1}^m \mathbf{A}_i^T (\sigma^2 (\mathbf{Z}_i^T \mathbf{Z}_i)^{-1} / d + \mathbf{I})^{-1} \mathbf{A}_i \right\}^{-1} \sum_{i=1}^m \mathbf{A}_i^T (\sigma^2 (\mathbf{Z}_i^T \mathbf{Z}_i)^{-1} / d + \mathbf{I})^{-1} \mathbf{A}_i^{-1} \hat{\boldsymbol{\beta}}_i.$$

Letting $d \rightarrow \infty$, we obtain

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^m \mathbf{A}_i^T \mathbf{A}_i \right)^{-1} \sum_{i=1}^m \mathbf{A}_i^T \hat{\boldsymbol{\beta}}_i.$$

This estimator has the form of OLS treating the $\hat{\boldsymbol{\beta}}_i$ as “outcome vectors.”

An intuitive interpretation of this is harder. Here is one possible interpretation. For finite d , $\hat{\boldsymbol{\beta}}$ is a weighted average of the $\hat{\boldsymbol{\beta}}_i$ that incorporates information on quality of each $\hat{\boldsymbol{\beta}}_i$ through dependence of \mathbf{W}_i on \mathbf{Z}_i (and \mathbf{R}_i and thus σ^2 here). When $d \rightarrow \infty$, the implication is that individual trajectories are completely disparate, so it is as though each individual is his/her own “population.” In this case, there is no relative quality of information on various individuals in a common population to consider. Thus, it is natural to treat the $\hat{\boldsymbol{\beta}}_i$ as the “responses” and simply combined them using OLS, which treats them all the same.

2. Loglikelihood tricks.

(a) With $\mathbf{D} = \sigma^2 \mathbf{D}_*$,

$$\mathbf{V}_i = \sigma^2 (\mathbf{I} + \mathbf{Z}_i \mathbf{D}_* \mathbf{Z}_i^T),$$

so that σ^2 becomes a multiplicative factor. In this case,

$$\partial / \partial \sigma^2 \mathbf{V}_i = (\mathbf{I} + \mathbf{Z}_i \mathbf{D}_* \mathbf{Z}_i^T),$$

so (5.38) reduces to

$$\sigma^{-2} \sum_{i=1}^m (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta})^T (\mathbf{I} + \mathbf{Z}_i \mathbf{D}_* \mathbf{Z}_i^T)^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}) - N = 0,$$

from whence it follows that

$$\sigma^2 = N^{-1} \sum_{i=1}^m (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta})^T (\mathbf{I} + \mathbf{Z}_i \mathbf{D}_* \mathbf{Z}_i^T)^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}).$$

(b) Under these conditions, (5.31) is

$$(-1/2) \sum_{i=1}^m \left\{ \log(\sigma^{2n_i} |\mathbf{I} + \mathbf{Z}_i \mathbf{D}_* \mathbf{Z}_i^T|) + \sigma^{-2} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta})^T (\mathbf{I} + \mathbf{Z}_i \mathbf{D}_* \mathbf{Z}_i^T)^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right\}.$$

Substituting the expression in (a) thus yields

$$(-1/2) \left[\sum_{i=1}^m n_i \log \{ N^{-1} \sum_{i=1}^m (\mathbf{Y}_i - \mathbf{X}_i \beta)^T (\mathbf{I} + \mathbf{Z}_i \mathbf{D}_* \mathbf{Z}_i^T)^{-1} (\mathbf{Y}_i - \mathbf{X}_i \beta) \} + \sum_{i=1}^m \log |\mathbf{I} + \mathbf{Z}_i \mathbf{D}_* \mathbf{Z}_i^T| + N \right].$$

Ignoring constants, we are left with

$$(-1/2) \left[N \log \left\{ \sum_{i=1}^m (\mathbf{Y}_i - \mathbf{X}_i \beta)^T (\mathbf{I} + \mathbf{Z}_i \mathbf{D}_* \mathbf{Z}_i^T)^{-1} (\mathbf{Y}_i - \mathbf{X}_i \beta) \right\} + \sum_{i=1}^m \log |\mathbf{I} + \mathbf{Z}_i \mathbf{D}_* \mathbf{Z}_i^T| \right]. \quad (1)$$

This depends only on β and \mathbf{D}_* .

(c) By the standard matrix inversion formula,

$$(\mathbf{I}_{n_i} + \mathbf{Z}_i \mathbf{D}_* \mathbf{Z}_i^T)^{-1} = \mathbf{I}_{n_i} - \mathbf{Z}_i (\mathbf{D}_*^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T.$$

Also, using standard results for determinants,

$$\begin{aligned} \log |\mathbf{I}_{n_i} + \mathbf{Z}_i \mathbf{D}_* \mathbf{Z}_i^T| &= \log |\mathbf{D}_* (\mathbf{D}_*^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)| \\ &= \log |\mathbf{D}_*| + \log |\mathbf{D}_*^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i| = \log (1/|\mathbf{D}_*^{-1}|). \end{aligned}$$

Substituting in the profile loglikelihood (1) gives

$$\begin{aligned} (-1/2) & \left(N \log \left[\sum_{i=1}^m (\mathbf{Y}_i - \mathbf{X}_i \beta)^T \{ \mathbf{I}_{n_i} - \mathbf{Z}_i (\mathbf{D}_*^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T \} (\mathbf{Y}_i - \mathbf{X}_i \beta) \right] \right. \\ & \left. + \sum_{i=1}^m \log |\mathbf{D}_*^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i| - N \log |\mathbf{D}_*^{-1}| \right), \end{aligned} \quad (2)$$

which depends on β and \mathbf{D}_*^{-1} .

(d) Of course, one can obtain this expression by noting that, under these conditions, (5.27) is

$$\beta = \left\{ \sum_{i=1}^m \mathbf{X}_i^T (\mathbf{I} + \mathbf{Z}_i \mathbf{D}_* \mathbf{Z}_i^T)^{-1} \mathbf{X}_i \right\}^{-1} \sum_{i=1}^m \mathbf{X}_i^T (\mathbf{I} + \mathbf{Z}_i \mathbf{D}_* \mathbf{Z}_i^T)^{-1} \mathbf{Y}_i,$$

so depends only on \mathbf{D}_* . Using the above, this can be rewritten in terms of \mathbf{D}_*^{-1} by substituting

$$(\mathbf{I}_{n_i} + \mathbf{Z}_i \mathbf{D}_* \mathbf{Z}_i^T)^{-1} = \mathbf{I}_{n_i} - \mathbf{Z}_i (\mathbf{D}_*^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T.$$

This can be substituted in the first term in (2), so that (2) is expressed in terms of \mathbf{D}_*^{-1} only.

It is of interest to note that some useful simplification occurs. By straightforward algebra, it can be shown that, substituting the expression in (5.27) and using the “stacked” notation, similar to the calculations in Section 5.4,

$$\begin{aligned} \sum_{i=1}^m (\mathbf{Y}_i - \mathbf{X}_i \beta)^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \beta) &= (\mathbf{Y} - \mathbf{X} \beta)^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X} \beta) \\ &= \mathbf{Y}^T \mathbf{V}^{-1} \mathbf{Y} - (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y})^T (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y})^T. \end{aligned} \quad (3)$$

With

$$\mathbf{V}_i^{-1} = \mathbf{I}_{n_i} - \mathbf{Z}_i (\mathbf{D}_*^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T,$$

these terms can be expressed as

$$\mathbf{Y}^T \mathbf{V}^{-1} \mathbf{Y} = \sum_{i=1}^m \mathbf{Y}_i \mathbf{V}_i^{-1} \mathbf{Y}_i = \sum_{i=1}^m \mathbf{Y}_i^T \mathbf{Y}_i - \sum_{i=1}^m (\mathbf{Z}_i^T \mathbf{Y}_i)^T (\mathbf{D}_*^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T \mathbf{Y}_i,$$

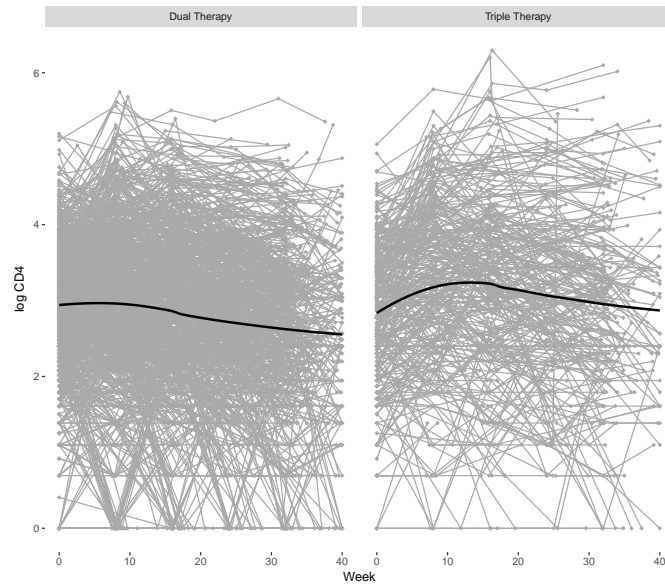
and similarly

$$\begin{aligned} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y} &= \sum_{i=1}^m \mathbf{X}_i^T \mathbf{Y}_i - \sum_{i=1}^m (\mathbf{Z}_i^T \mathbf{X}_i)^T (\mathbf{D}_*^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T \mathbf{Y}_i, \\ \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} &= \sum_{i=1}^m \mathbf{X}_i^T \mathbf{X}_i - \sum_{i=1}^m (\mathbf{Z}_i^T \mathbf{X}_i)^T (\mathbf{D}_*^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i)^{-1} \mathbf{Z}_i^T \mathbf{X}_i. \end{aligned}$$

Note that some of the quantities in these expressions need only be calculated once, e.g., $\sum_{i=1}^m \mathbf{Y}_i^T \mathbf{Y}_i$, lending computational efficiency. Thus, letting $s(\mathbf{D}_*^{-1})$ be (3) with these expressions substituted, the most computationally economical expression for the full profile loglikelihood is

$$(-1/2) \left\{ N \log s(\mathbf{D}_*^{-1}) + \sum_{i=1}^m \log |\mathbf{D}_*^{-1} + \mathbf{Z}_i^T \mathbf{Z}_i| - N \log |\mathbf{D}_*^{-1}| \right\}.$$

3. *HIV clinical trial, continued.* Spaghetti plots of the data for all 4 treatment groups are in Figure 5.3. In this problem, the investigators are regarding groups 1–3 as a single group, where the criterion for group membership is receiving a treatment regimen with two agents (“dual therapy”). It is actually quite common in research in this area to combine therapy groups, as questions are often about the relative merits of different treatment “cocktails” with fewer or greater numbers of agents, the number of agent being the key issue. So here we consider the participants who received any of regimens 1–3 as belonging to the same dual therapy group, and those in regimen 4 as the triple therapy group. Here are spaghetti plots for these two groups:



The plot for triple therapy suggests that individual trajectories might rise during the first 16 weeks but then fall, although it's hard to tell.

Based on the description and scientific questions of interest, a linear spline model at the individual subject level with a knot at 16 weeks is an appropriate framework. Letting Y_{ij} be the logarithm of CD4 count + 1 for individual i at time t_{ij} , we can posit an individual-level model as

$$Y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}(t_{ij} - 16)_+ + e_{ij},$$

where $x_+ = x$ if $x > 0$ and 0 otherwise. You probably adopted this as your basic individual-level model.

The first thing you probably did was to investigate assumptions on the within and among-individual sources of variation and correlation. As noted in class, a common approach is to adopt a basic population model that does not include relationships between the individual-specific parameters (β_{0i} , β_{1i} , and β_{2i} here) and among-individual covariates beyond an obvious categorical covariate like treatment group in a randomized study for this purpose. Under this approach, it would be reasonable to adopt the initial model

$$\begin{aligned}\beta_{0i} &= \beta_0 + b_{0i} \\ \beta_{1i} &= \beta_{10}(1 - \delta_i) + \beta_{11}\delta_i + b_{1i}, \\ \beta_{2i} &= \beta_{20}(1 - \delta_i) + \beta_{21}\delta_i + b_{2i},\end{aligned}\tag{4}$$

$$\tag{5}$$

say, where $\delta_i = 0$ if individual i is in the dual therapy and $\delta_i = 1$ if i is in the triple therapy group. Here, I have taken the mean intercept to be the same for both groups, given that this is a randomized study. We could equally well take the mean intercepts to be different in this initial model and then test for equal mean intercepts once a preferred covariance model is established. You may have adopted a fancier model involving age and gender from the start; that's fine, too.

In the attached programs, we investigate different assumptions on $\text{var}(\mathbf{e}_i|\mathbf{x}_i)$ and $\text{var}(\mathbf{b}_i|\mathbf{x}_i)$ under this basic model. Because many participants missed intended visits, we use maximum likelihood for all fits. Using SAS, we can investigate models with different \mathbf{D} matrix for each treatment group (which can't be done with `lme` in R) for $\text{var}(\mathbf{b}_i|\mathbf{x}_i)$. The time points are not equally spaced and are in fact different for different individuals, so if we want to investigate within-individual correlation, we need to use something like an exponential correlation model, ideally with a “nugget” effect, as CD4 count is likely measured with some error. Such models appear to be overkill, as it appears that the estimates of the parameters involved are driven to the boundaries of what these parameters are allowed to be.

Based on the information criteria for the models fitted in the SAS program and a preference for parsimony, I settled on the model with common \mathbf{D} and common within-individual diagonal structure. You may have picked something different.

Having decided on the basic model, I went ahead and investigated whether or not the population model (5) should be modified to include dependence of the three individual-specific parameters on age and/or gender, as suggested by (iv). Based on fitting a series of progressively more complicated models with age and gender dependence, I concluded that a reasonable final model is one that includes dependence of β_{0i} on age but no dependence of β_{1i} and β_{2i} on age or gender. Having selected this model, I then addressed (i)–(iii) in terms of this final model. You may have chosen instead to look at (i)–(iii) under a basic model excluding dependence on age and gender; this is fine, too. Basing these on the basic model

effectively averages across the distribution of age and gender so can be viewed as a “global” analysis.

From the R program, based on a likelihood ratio test as well as the fit of the model in the alternative parameterization with dual therapy as the reference treatment, there seems to be very strong evidence that the means of both β_{1i} and β_{2i} depend strongly on treatment group (dual or triple). From the fit of the final model parameterized as

$$\begin{aligned}\beta_{0i} &= \beta_{00} + \beta_{01}\text{age}_i + b_{0i} \\ \beta_{1i} &= \beta_{10}(1 - \delta_i) + \beta_{11}\delta_i + b_{1i}, \\ \beta_{2i} &= \beta_{20}(1 - \delta_i) + \beta_{21}\delta_i + b_{2i},\end{aligned}$$

the evidence seems very strong that $\beta_{10} < 0$ (dual therapy slope in the first 16 weeks) but $\beta_{11} > 0$ (triple therapy slope in the first 16 weeks). Of course, because of the missing data, Wald tests are based on potentially misleading standard errors; however, in this case, the test statistics are so large that it is likely that correct standard errors based on the observed information matrix would still lead to large test statistics. The estimates of these slopes are in the output.

We can also construct estimates of the mean slopes for each therapy in the second part of the study; see the R program, where this is done manually. There is strong evidence that these are both negative. Based on Wald inference (which is not entirely reliable given the missing data), there does not seem to be evidence that these slopes differ; given that the Wald test would be optimistic, we can probably safely say that there is not sufficient evidence of a difference – a better analysis would be to do a full vs. reduced model likelihood ratio test.

Because of the dependence of mean baseline log CD4 on age, mean at 40 weeks will also depend on age. In the R program, I show estimates of mean log CD4 at week 40 for each therapy group and then evaluate the difference. Keeping in mind that the standard errors may be optimistic, there seems to be some evidence that the mean at 40 weeks is different for the two groups.

You should have explained the process you went through to arrive at the models you selected and explained how you formulated and addressed the questions of interest in terms of your models. You should have also explained any diagnostic plots and analyses you constructed and commented on what your analyses imply about the reliability of the model assumptions (normality, etc).