Accounting for Baseline Observations in Randomized Clinical Trials

Scott S. Emerson, M.D., Ph.D.

Department of Biostatistics, University of Washington, Seattle, WA 98195, USA

August 25, 2012

Abstract

In clinical trials investigating the treatment of chronic, progressive diseases, a common design is to make some physiologic measurement of the disease state at the start of the trial, randomize patients to either the experimental treatment or some control treatment (placebo), follow the patients over some period of time, and then measure once again the physiologic measurement of the disease state. The question then arises: How should the baseline measurements made prior to randomization be used? In this manuscript, I consider several common strategies that might be used in this setting, including analyzing only the follow-up measurement, analyzing the change in measurements, and the analysis of covariance (ANCOVA) approach of adjusting for the baseline measurement in a linear regression model. I illustrate several different ways to motivate the use of ANCOVA, as well as examining the settings in which not measuring baseline might be preferred. I conclude with a discussion of how these results might change when using a randomized crossover design.

1 Introduction

In clinical trials investigating the treatment of chronic, progressive diseases, a common design is to make some physiologic measurement of the disease state at the start of the trial, randomize patients to either the experimental treatment or some control treatment (placebo), follow the patients over some period of time, and then measure once again the physiologic measurement of the disease state. Take for instance a double-blind, placebo controlled, randomized clinical trial (RCT) of a new drug in the setting of hypertension. Upon recruiting a patient into the RCT, we would typically measure the patient's systolic blood pressure (SBP) for the purposes of determining eligibility for the trial. We might further decide to stratify randomization of patients on SBP in order to ensure balance across treatment arms with respect to severity of hypertension. Then after treating the patients with either the experimental treatment or placebo for the period of, say, a year, we again measure the SBP of all patients, and analyze the data to decide whether the experimental treatment is associated with a scientifically important and statistically significant lower SBP when compared to placebo. The question is: How do we use the baseline (pre-randomization) measurement of SBP in the data analysis? Owing to the randomized nature of the study, there are several choices that will each estimate the causal effect of the treatment in an unbiased fashion:

1. *Ignore baseline and analyze the final measurements.* By virtue of randomization, the distribution of SBP at the start of the study is equal in each treatment arm. Hence, any differences in the distribution of SBP at the end of the trial is directly attributable to the effect of the experimental treatment.

2. *Analyze the change in measurements over the course of the study.* For each patient, we compute the difference between the final SBP and their baseline SBP, and then compare the treatment arms with respect to the change in measurements.

3. *Adjust for the baseline measurement as a covariate in a linear regression model of final measurement for each patient by treatment arm.*

4. *Adjust for the baseline measurement as a covariate in a linear regression model of the change in measures for each patient by treatment arm.*

In my experience, the second of these methods is the one most commonly advocated by clinical investigators. It seems only natural to them: we are interested in an intervention that would either change the pathophysiologic process for the better or (at least) lessen the change in the pathophysiologic process for the worse. In this manuscript, I explore the settings in which one of the approaches might be preferable to the others. In this exploration, I illustrate the seemingly paradoxical result that suggests that in RCT that second method (analyzing the change in measurements) is never an optimal choice, while there are situations that any of the other three is optimal.

It should be noted that the results presented herein are specific to RCT, because we make extensive use of the fact that the distribution of baseline measurements is known to be the same in both treatment groups.

2 Notation in the Homoscedastic, Independent Sample Setting

Let Y_{ktj} be the SBP measurement in the kth treatment group ($k = 1$ for experimental treatment, $k = 0$ for placebo) at time t ($t = 0$ for the baseline (pre-randomization) measurement, $t = 1$ for the end of treatment measurement) in the jth patient ($j = 1, \ldots, n_i$). We assume all subjects are independent and that the subjects in the experimental treatment group are different that those in the placebo group.

Suppose $Y_{ktj} \sim (\mu_{kt}, \sigma^2)$, meaning that for a population receiving treatment k, the average measurement pre-randomization would be μ_{k0} with a standard deviation of σ, and the average measurement post-treatment would be μ_{k1} with a standard deviation of σ. We presume that individuals are independent, and that repeat measurements made on the same subject in the kth group have correlation ρ. Thus $corr(Y_{k1j}, Y_{k0j}) = \rho$, $corr(Y_{ktj}, Y_{kt'j'}) = 0$ for $j \neq j'$, and $corr(Y_{ktj}, Y_{k't'j'}) = 0$ for $k \neq k'$. (This "homoscedastic" model presumes not only that the treatment does not affect the variability of the measurements, but also that the treatment does not affect the correlation of the baseline and follow-up measurements.)

We presume a randomized study, so we have $\mu_{10} = \mu_{00}$. We are ultimately interested in estimating $\theta = \mu_{11} - \mu_{01}$.

In order to derive general results, we consider the distribution of linear combinations of the form $\Delta_{kj} = Y_{k1j} - a_k Y_{k0j}$. Using simple properties of expectation and covariances, we find

$$E\left[\Delta_{kj}\right] = E\left[Y_{k1j} - a_i Y_{k0j}\right] = \mu_{k1} - a_k \mu_{k0}$$
$$Var\left(\Delta_{kj}\right) = Var\left(Y_{k1j} - a_k Y_{k0j}\right) = \sigma^2 + a_k^2\sigma^2 - 2a_k\rho\sigma^2 = \left(1 - 2a_k\rho + a_k^2\right)\sigma^2$$

so letting

$$\overline{\Delta}_{k\cdot} = \frac{1}{n_k}\sum_{j=1}^{n_k} \Delta_{kj}$$

we can derive moments

$$E\left[\overline{\Delta}_{k\cdot}\right] = \mu_{k1} - a_k\mu_{k0}$$
$$Var\left[\overline{\Delta}_{k\cdot}\right] = \frac{\left(1 - 2a_k\rho + a_k^2\right)\sigma^2}{n_k}$$

Now, for a specified value of a, we define $\hat{\theta}^{(a)} = \overline{\Delta}_{1\cdot} - \overline{\Delta}_{0\cdot}$, and find distribution moments

$$\begin{aligned} E\left[\hat{\theta}^{(a)}\right] &= (\mu_{11} - a_1\mu_{10}) - (\mu_{01} - a_0\mu_{00}) = \theta - (a_1\mu_{10} - a_0\mu_{00}) \\ &= \theta - (a_1 - a_0)\mu_{00} \end{aligned}$$

and $\hat{\theta}$ is unbiased for θ for arbitrary μ_{00} if and only if $a_1 = a_0 = a$.

Now we can find distribution variance

$$Var\left[\hat{\theta}^{(a)}\right] = \left(1 - 2a\rho + a^2\right)\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_0}\right)$$

Special cases of interest are

1. *Ignore baseline and analyze the final measurements.* This corresponds to $a = 0$, in which case
$$Var(\hat{\theta}^{(0)}) = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_0}\right).$$

2. *Analyze the change in measurements over the course of the study.* This corresponds to $a = 1$, in which case
$$Var(\hat{\theta}^{(1)}) = 2\sigma^2\left(1 - \rho\right)\left(\frac{1}{n_1} + \frac{1}{n_0}\right).$$

By differentiating the above expression with respect to a, we can find the form of $\hat{\theta}$ with minimal variability as having $a = \rho$, in which case
$$Var(\hat{\theta}^{(\rho)}) = \sigma^2\left(1 - \rho^2\right)\left(\frac{1}{n_1} + \frac{1}{n_0}\right).$$

Note that

1. For $\rho < 0.5$, $Var(\hat{\theta}^{(0)}) < Var(\hat{\theta}^{(1)})$, and throwing away the baseline value is more efficient than comparing the change in measurements across treatment arms.

2. For $\rho = 0.5$, $Var(\hat{\theta}^{(0)}) = Var(\hat{\theta}^{(1)})$, and throwing away the baseline value is just as efficient as comparing the change in measurements across treatment arms. (This is the only point of equality between these two estimators.)

3. For $\rho > 0.5$, $Var(\hat{\theta}^{(0)}) > Var(\hat{\theta}^{(1)})$, and throwing away the baseline value is less efficient than comparing the change in measurements across treatment arms.

4. For all $-1 \leq \rho \leq 1$, $Var(\hat{\theta}^{(\rho)}) \leq Var(\hat{\theta}^{(0)})$ and $Var(\hat{\theta}^{(\rho)}) \leq Var(\hat{\theta}^{(1)})$.

5. For $\rho = 0$, $Var(\hat{\theta}^{(0)}) = Var(\hat{\theta}^{(\rho)})$. (This is the only point of equality between these two estimators.)

6. For $\rho = 1$, $Var(\hat{\theta}^{(1)}) = Var(\hat{\theta}^{(\rho)})$. (This is the only point of equality between these two estimators.)

Note that the third observation above suggests that when ρ is known, the use of $\hat{\theta}^{(\rho)}$ is the best linear estimator, as it is always at least as good as either of the other two approaches based on $a = 0$ or $a = 1$.

3 An Alternative Derivation When Measurements are Normally Distributed

Consider the same homoscedastic setting as in section 2 with σ^2 and ρ known constants, but now suppose that it is known that vector (Y_{k0j}, Y_{k1j}) are bivariate normal. We can then write the density for our data as a function of $\vec{\theta} = (\mu_{00}, \mu_{01}, \mu_{11})$ as

$$\begin{aligned} f(\vec{Y};\vec{\theta}) &= \prod_{k=0}^{1}\prod_{j=1}^{n_k} \frac{1}{2\pi\sigma^2\sqrt{(1-\rho^2)}} \exp\left[-\frac{(Y_{k0j}-\mu_{00})^2 - 2\rho(Y_{k0j}-\mu_{00})(Y_{k1j}-\mu_{k1}) + (Y_{k1j}-\mu_{k1})^2}{2\sigma^2(1-\rho^2)}\right] \\ &= g(\theta)h(\vec{Y})\exp\left[-\sum_{\ell=1}^{3} c_\ell(\theta)T_\ell(\vec{Y})\right], \end{aligned}$$

where the familiar form of an exponential family density is found by expanding the product and grouping terms, with

$$\begin{aligned} g(\vec{\theta}) &= \left[\frac{1}{2\pi\sigma^2\sqrt{(1-\rho^2)}}\right]^{n_0+n_1} \exp\left[-\frac{(n_0+n_1)\mu_{00}^2 - 2\rho(n_0\mu_{01}+n_1\mu_{11})\mu_{00} + n_1\mu_{11}^2 + n_0\mu_{01}^2}{2\sigma^2(1-\rho^2)}\right] \\ h(\vec{Y}) &= \exp\left[-\frac{\sum_{k=0}^{1}\sum_{j=1}^{n_k}\left(Y_{k0j}^2 - 2\rho Y_{k0j}Y_{k1j} + Y_{k1j}^2\right)}{2\sigma^2(1-\rho^2)}\right] \\ c_1(\vec{\theta}) &= \frac{\mu_{00}}{\sigma^2(1-\rho^2)} \\ c_2(\vec{\theta}) &= \frac{\mu_{01}}{\sigma^2(1-\rho^2)} \\ c_3(\vec{\theta}) &= \frac{\mu_{11}}{\sigma^2(1-\rho^2)} \\ T_1(\vec{Y}) &= \sum_{k=0}^{1}\sum_{j=1}^{n_k}(Y_{k0j} - \rho Y_{k1j}) \\ T_2(\vec{Y}) &= \sum_{j=1}^{n_0}(Y_{01j} - \rho Y_{00j}) \\ T_3(\vec{Y}) &= \sum_{j=1}^{n_1}(Y_{11j} - \rho Y_{10j}). \end{aligned}$$

Note that in that exponential family density, $\vec{\theta}$ is a three dimensional vector, and the range of $\vec{\theta}$ contains an open rectangle in 3 dimensions. Hence, we know that $\vec{T}$ is a complete sufficient statistic.

Furthermore, because $\hat{\theta}^{(\rho)} = T_3(\vec{Y})/n_1 - T_2(\vec{Y})/n_0$ is unbiased for θ and a function of the complete sufficient statistic, we know that $\hat{\theta}^{(\rho)}$ is the uniform minimum variance unbiased estimator of θ.

4 Settings Where Not Measuring Baseline is Optimal

Even when ρ is known, there is a setting in which a better approach would be to use only the final measurement.

Suppose the measurement of Y_{ktj} is extremely expensive. Then the limiting factor in our experimental design might be the number of measurements we have to make.

In order to estimate $\hat{\theta}^{(\rho)}$, we must make $2(n_0 + n_1)$ measurements– one for each subject at baseline and one for each subject at the end of treatment. As noted above, the variance of $\hat{\theta}^{(\rho)}$ is given by

$$Var(\hat{\theta}^{(\rho)}) = \sigma^2 \left(1 - \rho^2\right) \left(\frac{1}{n_1} + \frac{1}{n_0}\right).$$

But suppose that we doubled the number of subjects to $n_0^* = 2n_0$ and $n_1^* = 2n_1$ and we did not measure baseline values at all. In this case, we would use $\hat{\theta}^{(0)}$ which has variance

$$Var(\hat{\theta}^{(0)}) = \sigma^2 \left(\frac{1}{n_1^*} + \frac{1}{n_0^*}\right) = .\frac{\sigma^2}{2} \left(\frac{1}{n_1} + \frac{1}{n_0}\right).$$

We can then consider the setting in which $Var(\hat{\theta}^{(0)}) < Var(\hat{\theta}^{(\rho)})$ as

$$\begin{aligned} Var(\hat{\theta}^{(0)}) &< Var(\hat{\theta}^{(\rho)}) \\ \frac{\sigma^2}{2} \left(\frac{1}{n_1} + \frac{1}{n_0}\right) &< \sigma^2 \left(1 - \rho^2\right) \left(\frac{1}{n_1} + \frac{1}{n_0}\right) \\ \frac{1}{2} &< \left(1 - \rho^2\right) \\ \rho &< \sqrt{\frac{1}{2}} = 0.7071 \end{aligned}$$

Hence, if there is no other need to measure baseline values, and if it is relatively easier and cheaper to accrue patients than to make measurements on them, then the most efficient design is to measure and use only the follow-up values at the end of treatment, unless the correlation within subjects is extremely high ($\rho > 0.7071$).

Of course, it is not unusual for patient eligibility to be based on the pre-randomization value of the measurement, in which case, the advantage of ignoring the baseline measurement is gone.

5 Settings Where ρ is Unknown: ANCOVA

The above derivations assumed that both ρ and σ^2 were known in a homoscedastic setting. It is most often the case, however, that neither ρ nor σ^2 are known, and we must use estimates. One approach to such estimation is to use a linear model in which we adjust for the baseline value as a covariate.

To further explore this approach, it is useful to modify our notation. Define outcome vector $\vec{Z}$, treatment vector $\vec{X}$, and baseline vector $\vec{W}$ for $i = 1, \ldots, n = n_0 + n_1$ by

$$Z_i = \begin{cases} Y_{01i} & i \leq n_0 \\ Y_{11j} & i = n_0 + j \end{cases} \qquad X_i = \begin{cases} 0 & i \leq n_0 \\ 1 & i > n_0 \end{cases} \qquad W_i = \begin{cases} Y_{00i} & i \leq n_0 \\ Y_{10j} & i = n_0 + j \end{cases}$$

We then fit ordinary least squares (OLS) model

$$E[Z_i \mid X_i, W_i] = \beta_0 + X_i\beta_1 + W_i\beta_2$$

obtaining OLS estimates

$$\begin{aligned}
\hat{\beta}_1 &= \frac{1}{1-r^2_{XW}}\left[\frac{V_{XZ}}{V_{XX}} - \frac{V_{XW}V_{WZ}}{V_{XX}V_{WW}}\right] \\
\hat{\beta}_2 &= \frac{1}{1-r^2_{XW}}\left[\frac{V_{WZ}}{V_{WW}} - \frac{V_{XW}V_{XZ}}{V_{XX}V_{WW}}\right],
\end{aligned}$$

where for sample means $\overline{Z}.$, $\overline{X}.$, and $\overline{W}$.

$$\begin{aligned}
r_{WX} &= \frac{S_{XW}}{\sqrt{S_{XX}S_{WW}}} \\
S_{XX} &= \sum_{i=1}^{n}(X_i - \overline{X}.)^2, \\
V_{XX} &= S_{XX}/n, \\
S_{XW} &= \sum_{i=1}^{n}(X_i - \overline{X}.)(W_i - \overline{W}.), \\
V_{XW} &= S_{XW}/n,
\end{aligned}$$

with similar definitions for V_{WW}, V_{XZ}, and V_{WZ}. Now by randomization, we know that W_k and X_k are independent. Thus under the assumption that the randomization ratio approaches some constant (i.e., $n_1/n \rightarrow \lambda \in (0,1)$ as $n \rightarrow \infty$), the consistency of sample moments for population moments and the properties of convergence in probability then provide that

$$\begin{aligned}
V_{XW} &\rightarrow_p 0 \\
r_{XW} &\rightarrow_p 0 \\
V_{XX} &\rightarrow_p \lambda(1-\lambda) \\
V_{WW} &\rightarrow_p \sigma^2 \\
V_{ZZ} &\rightarrow_p \sigma^2 + \lambda(1-\lambda)(\mu_{11} - \mu_{01})^2 \\
V_{WZ} &\rightarrow_p \rho\sigma^2 \\
V_{XZ} &\rightarrow_p \lambda(1-\lambda)(\mu_{11} - \mu_{01}) = \lambda(1-\lambda)\theta \\
\hat{\beta}_1 &\rightarrow_p \theta \\
\hat{\beta}_2 &\rightarrow_p \rho.
\end{aligned}$$

Thus, this "analysis of covariance (ANCOVA)" model is essentially using a consistent estimate for ρ as the coefficient for the baseline value.

Further statements can be made if we know that our statistical model is an accurate representation of the data generation mechanism. Because the treatment assignment is binary, the only issue is whether some linear relationship exists between the baseline and follow-up measurement within each dose group. But if this does hold, then in the homoscedastic setting (irrespective of the distribution of the errors–including irrespective of whether the errors are identically distributed) we know by the Gauss-Markov Theorem that $\hat{\beta}_1$ is the best linear unbiased estimator (BLUE) for θ.

Sometimes an investigator is extremely insistent on analyzing the change in measurements. They then fit a model in which they define

$$Z_i = \begin{cases} Y_{01i} - Y_{00i} & i \leq n_0 \\ Y_{11j} - Y_{10j} & i = n_0 + j \end{cases} \qquad X_i = \begin{cases} 0 & i \leq n_0 \\ 1 & i > n_0 \end{cases} \qquad W_i = \begin{cases} Y_{00i} & i \leq n_0 \\ Y_{10j} & i = n_0 + j \end{cases}$$

and fit ordinary least squares (OLS) model

$$E[Z_k \mid X_i, W_i] = \alpha_0 + X_i\alpha_1 + W_i\alpha_2.$$

This approach provides the exact same inference about the treatment effect, because $\alpha_0 = \beta_0$, $\alpha_1 = \beta_1$, and $\alpha_2 = \beta_2 - 1$ and OLS estimates $\hat{\alpha}_0 = \hat{\beta}_0$, $\hat{\alpha}_1 = \hat{\beta}_1$, and $\hat{\alpha}_2 = \hat{\beta}_2 - 1$.

6 ANCOVA in the General Heteroscedastic Setting

6.1 Notation

Let Y_{ktj} be the SBP measurement in the kth treatment group ($k = 1$ for experimental treatment, $k = 0$ for placebo) at time t ($t = 0$ for the baseline (pre-randomization) measurement, $t = 1$ for the end of treatment measurement) in the jth patient ($j = 1, \ldots, n_i$).

Suppose $Y_{ktj} \sim (\mu_{kt}, \sigma^2_{kt})$, meaning that for a population receiving treatment k, the average measurement pre-randomization would be μ_{k0} with a standard deviation of σ_{k0}, and the average measurement post-treatment would be μ_{k1} with a standard deviation of σ_{k1}. We presume that individuals are independent, and that repeat measurements made on the same subject in the kth group have correlation ρ_k. Thus $corr(Y_{k1j}, Y_{k0j}) = \rho_k$, $corr(Y_{ktj}, Y_{kt'j'}) = 0$ for $j \neq j'$, and $corr(Y_{ktj}, Y_{k't'j'}) = 0$ for $k \neq k'$. (In this heteroscedastic setting, we consider that the treatment can affect both the variability of the follow-up measurements, as well as the correlation between the baseline and follow-up measurements.)

6.2 Optimal Linear Combination of Measurements

We presume a randomized study, so we have $\mu_{10} = \mu_{00}$ and $\sigma_{10} = \sigma_{00}$. We are ultimately interested in estimating $\theta = \mu_{11} - \mu_{01}$.

In order to derive general results, we consider the distribution of linear combinations of the form $\Delta_{kj} = Y_{k1j} - a_k Y_{k0j}$. Using simple properties of expectation and covariances, we find

$$E\left[\Delta_{kj}\right] = E\left[Y_{k1j} - a_k Y_{k0j}\right] = \mu_{k1} - a_k\mu_{i0}$$
$$Var\left(\Delta_{kj}\right) = Var\left(Y_{k1j} - a_k Y_{k0j}\right) = \sigma^2_{k1} + a^2_k\sigma^2_{k0} - 2a_k\rho_k\sigma_{k1}\sigma_{k0}$$

so letting

$$\overline{\Delta}_{k\cdot} = \frac{1}{n_k}\sum_{j=1}^{n_k}\Delta_{kj}$$

we can derive moments

$$E\left[\overline{\Delta}_{k\cdot}\right] = \mu_{k1} - a_k\mu_{k0}$$
$$Var\left[\overline{\Delta}_{k\cdot}\right] = \frac{\sigma^2_{k1} + a^2_k\sigma^2_{k0} - 2a_k\rho_k\sigma_{k1}\sigma_{k0}}{n_k}$$

Now, we define $\hat{\theta} = \overline{\Delta}_{1\cdot} - \overline{\Delta}_{0\cdot}$, and find distribution moments

$$\begin{aligned} E\left[\hat{\theta}\right] &= (\mu_{11} - a_1\mu_{10}) - (\mu_{01} - a_0\mu_{00}) = \theta - (a_1\mu_{10} - a_0\mu_{00}) \\ &= \theta - (a_1 - a_0)\mu_{00} \end{aligned}$$

and $\hat{\theta}$ is unbiased for θ for arbitrary μ_{00} if and only if $a_1 = a_0 = a$.

Now we can find distribution variance

$$\begin{aligned} Var\left[\hat{\theta}\right] &= \frac{\sigma_{11}^2 + a^2\sigma_{10}^2 - 2a\rho_1\sigma_{11}\sigma_{10}}{n_1} + \frac{\sigma_{01}^2 + a^2\sigma_{00}^2 - 2a\rho_0\sigma_{01}\sigma_{00}}{n_0} \\ &= \frac{\sigma_{11}^2}{n_1} + \frac{\sigma_{01}^2}{n_0} + a^2\sigma_{00}^2\left(\frac{1}{n_1} + \frac{1}{n_0}\right) - 2a\sigma_{00}\left(\frac{\rho_1\sigma_{11}}{n_1} + \frac{\rho_0\sigma_{01}}{n_0}\right) \end{aligned}$$

By differentiating the above expression with respect to a, we can find the form of $\hat{\theta}$ with minimal variability as having

$$a = \left(\frac{n_0}{n_0 + n_1}\right)\rho_1\frac{\sigma_{11}}{\sigma_{10}} + \left(\frac{n_1}{n_0 + n_1}\right)\rho_0\frac{\sigma_{01}}{\sigma_{00}} = (1-\lambda)\,\rho_1\frac{\sigma_{11}}{\sigma_{00}} + \lambda\rho_0\frac{\sigma_{01}}{\sigma_{00}},$$

where $\lambda = n_1/(n_0 + n_1)$ reflects the randomization ratio.

Note that this is a weighted average of the slope parameter from a regression of Y_{11j} on Y_{10j} and the slope parameter from a regression of Y_{01j} on Y_{00j}.

6.3 An Alternative Derivation When Measurements are Normally Distributed

Consider the same heteroscedastic setting as in section 6.2 with σ_{kt}^2 and ρ_k known constants, but now suppose that it is known that vector (Y_{k0j}, Y_{k1j}) are bivariate normal. We can then write the density for our data as a function of $\vec{\mu} = (\mu_{00}, \mu_{01}, \mu_{11})$ as

$$\begin{aligned} f(\vec{Y};\vec{\mu}) &= \prod_{k=0}^{1}\prod_{j=1}^{n_k}\frac{1}{2\pi\sigma_{00}\sigma_{k1}\sqrt{(1-\rho_k^2)}}\exp\left[-\frac{(Y_{k0j}-\mu_{00})^2}{2\sigma_{00}^2(1-\rho_k^2)} + \frac{\rho_k(Y_{k0j}-\mu_{00})(Y_{k1j}-\mu_{k1})}{\sigma_{00}\sigma_{k1}(1-\rho_k^2)} - \frac{(Y_{k1j}-\mu_{k1})^2}{2\sigma_{k1}^2(1-\rho_k^2)}\right] \\ &= g(\mu)h(\vec{Y})\exp\left[-\sum_{\ell=1}^{3}c_\ell(\vec{\mu})T_\ell(\vec{Y})\right], \end{aligned}$$

where the familiar form of an exponential family density is found by expanding the product and grouping terms, with

$$
\begin{aligned}
g(\vec{\mu}) &= \prod_{k=0}^{1}\left[\left(\frac{1}{2\pi\sigma_{00}\sigma_{k1}\sqrt{(1-\rho_k^2)}}\right)^{n_k}\exp\left\{-\frac{n_k\mu_{00}^2}{2\sigma_{00}^2(1-\rho_k^2)}+\frac{n_k\rho_k\mu_{k1}\mu_{00}}{\sigma_{00}\sigma_{k1}(1-\rho_k^2)}-\frac{n_k\mu_{k1}^2}{2\sigma_{k1}^2(1-\rho_k^2)}\right\}\right]\\
h(\vec{Y}) &= \exp\left\{-\sum_{k=0}^{1}\sum_{j=1}^{n_k}\left[\frac{Y_{k0j}^2}{2\sigma_{00}^2(1-\rho_k^2)}-\frac{\rho_k Y_{k0j}Y_{k1j}}{\sigma_{00}\sigma_{k1}(1-\rho_k^2)}+\frac{Y_{k1j}^2}{2\sigma_{k1}^2(1-\rho_k^2)}\right]\right\}\\
c_1(\vec{\theta}) &= \frac{\mu_{00}}{\sigma_{00}^2}\\
c_2(\vec{\theta}) &= \frac{\mu_{01}}{\sigma_{01}^2(1-\rho_0^2)}\\
c_3(\vec{\theta}) &= \frac{\mu_{11}}{\sigma_{11}^2(1-\rho_1^2)}\\
T_1(\vec{Y}) &= \sum_{k=0}^{1}\left[\frac{1}{1-\rho_k^2}\sum_{j=1}^{n_k}\left(Y_{k0j}-\rho_k\frac{\sigma_{00}}{\sigma_{k1}}Y_{k1j}\right)\right]\\
&= \frac{n_0}{(1-\rho_0^2)}\overline{Y}_{00\cdot}-\frac{n_0\rho_0}{(1-\rho_0^2)}\frac{\sigma_{00}}{\sigma_{01}}\overline{Y}_{01\cdot}+\frac{n_1}{(1-\rho_1^2)}\overline{Y}_{10\cdot}-\frac{n_1\rho_1}{(1-\rho_1^2)}\frac{\sigma_{00}}{\sigma_{11}}\overline{Y}_{11\cdot}\\
T_2(\vec{Y}) &= \sum_{j=1}^{n_0}\left(Y_{01j}-\rho_0\frac{\sigma_{01}}{\sigma_{00}}Y_{00j}\right)=n_0\overline{Y}_{01\cdot}-n_0\rho_0\frac{\sigma_{01}}{\sigma_{00}}\overline{Y}_{00\cdot}\\
T_3(\vec{Y}) &= \sum_{j=1}^{n_1}\left(Y_{11j}-\rho_1\frac{\sigma_{11}}{\sigma_{00}}Y_{10j}\right)=n_1\overline{Y}_{11\cdot}-n_1\rho_1\frac{\sigma_{11}}{\sigma_{00}}\overline{Y}_{10\cdot}.
\end{aligned}
$$

Note that in that exponential family density, $\vec{\mu}$ is a three dimensional vector, and the range of $\vec{\mu}$ contains an open rectangle in 3 dimensions. Hence, we know that $\vec{T}$ is a complete sufficient statistic, and any unbiased estimator of a function of $\vec{\mu}$ that is only a function of the complete sufficient statistic is the uniform minimum variance unbiased estimator.

Thus for unbiased estimator $\widehat{\vec{\mu}}=(\hat{\mu}_{00}\ \hat{\mu}_{01}\ \hat{\mu}_{11})^T$ with

$$
\begin{aligned}
\hat{\mu}_{00} &= \frac{1}{n_0+n_1}\left[T_1(\vec{Y})+\frac{\rho_0}{1-\rho_0^2}\frac{\sigma_{00}}{\sigma_{01}}T_2(\vec{Y})+\frac{\rho_1}{1-\rho_1^2}\frac{\sigma_{00}}{\sigma_{11}}T_3(\vec{Y})\right]\\
&= \frac{n_0}{n_0+n_1}\overline{Y}_{00\cdot}+\frac{n_1}{n_0+n_1}\overline{Y}_{10\cdot}\\
\hat{\mu}_{01} &= \frac{1}{n_0}T_2(\vec{Y})+\rho_0\frac{\sigma_{01}}{\sigma_{00}}\hat{\mu}_{00}\\
&= \overline{Y}_{01\cdot}-\rho_0\frac{n_1}{n_0+n_1}\frac{\sigma_{01}}{\sigma_{00}}\overline{Y}_{00\cdot}+\rho_0\frac{n_1}{n_0+n_1}\frac{\sigma_{01}}{\sigma_{00}}\overline{Y}_{10\cdot}\\
\hat{\mu}_{11} &= \frac{1}{n_1}T_3(\vec{Y})+\rho_1\frac{\sigma_{11}}{\sigma_{10}}\hat{\mu}_{00}\\
&= \overline{Y}_{11\cdot}+\rho_1\frac{n_0}{n_0+n_1}\frac{\sigma_{11}}{\sigma_{00}}\overline{Y}_{00\cdot}-\rho_1\frac{n_0}{n_0+n_1}\frac{\sigma_{11}}{\sigma_{00}}\overline{Y}_{10\cdot}
\end{aligned}
$$

it is easily seen that this estimator is UMVUE for $\vec{\mu}$. It similarly follows that $\hat{\theta}=\hat{\mu}_{11}-\hat{\mu}_{01}$ is UMVUE for

$\theta = \mu_{11} - \mu_{01}$, with

$$\hat{\theta} = \left(\overline{Y}_{11\cdot} - \left(\frac{n_1}{n_0+n_1} \rho_0 \frac{\sigma_{01}}{\sigma_{00}} + \frac{n_0}{n_0+n_1} \rho_1 \frac{\sigma_{11}}{\sigma_{00}} \right) \overline{Y}_{10\cdot} \right) - \left(\overline{Y}_{01\cdot} - \left(\frac{n_1}{n_0+n_1} \rho_0 \frac{\sigma_{01}}{\sigma_{00}} + \frac{n_0}{n_0+n_1} \rho_1 \frac{\sigma_{11}}{\sigma_{00}} \right) \overline{Y}_{00\cdot} \right).$$

We can thus find that the UMVUE is exactly equal to the optimal choice of a when reducing data on each subject to a single measurement.

6.4 Use of the OLS ANCOVA model

In the realistic setting in which we do not know ρ_k or σ^2_{kt}, we could again consider the ANCOVA model.

$$E[Z_i \mid X_i, W_i] = \beta_0 + X_i\beta_1 + W_i\beta_2$$

with OLS estimates

$$\begin{aligned} \hat{\beta}_1 &= \frac{1}{1 - r^2_{XW}} \left[\frac{V_{XZ}}{V_{XX}} - \frac{V_{XW}V_{WZ}}{V_{XX}V_{WW}} \right] \\ \hat{\beta}_2 &= \frac{1}{1 - r^2_{XW}} \left[\frac{V_{WZ}}{V_{WW}} - \frac{V_{XW}V_{XZ}}{V_{XX}V_{WW}} \right]. \end{aligned}$$

In the heteroscedastic setting in which the randomization ratio is kept constant (so again assuming $n_1/n \rightarrow \lambda \in (0,1)$ as $n \rightarrow \infty$), we have

$$\begin{aligned} V_{XW} &\rightarrow_p 0, \\ r_{XW} &\rightarrow_p 0, \\ V_{XX} &\rightarrow_p \lambda(1-\lambda), \\ V_{WW} &\rightarrow_p \sigma^2_{00}, \\ V_{ZZ} &\rightarrow_p (1-\lambda)\sigma^2_{01} + \lambda\sigma^2_{11} + \lambda(1-\lambda)(\mu_{11} - \mu_{01})^2, \\ V_{WZ} &\rightarrow (1-\lambda)\rho_0\sigma_{00}\sigma_{01} + \lambda\rho_1\sigma_{00}\sigma_{11}, \\ V_{XZ} &\rightarrow_p \lambda(1-\lambda)(\mu_{11} - \mu_{01}) = \lambda(1-\lambda)\theta, \\ \hat{\beta}_1 &\rightarrow_p \theta, \\ \hat{\beta}_2 &\rightarrow_p \lambda\rho_1\frac{\sigma_{11}}{\sigma_{10}} + (1-\lambda)\,\rho_0\frac{\sigma_{01}}{\sigma_{00}}. \end{aligned}$$

Note that in this case, the OLS coefficient for the baseline measurement is consistent for the optimal value of a, when the randomization ratio is 1:1 (i.e., $\lambda = 0.5$– a very common setting) or $\rho_1\sigma_{11} = \rho_0\sigma_{01}$ (something that is rather hard to ascertain). Otherwise, the OLS coefficient is not consistent for the optimal value of a.

If we know that our statistical model is an accurate representation of the data generation mechanism, we know by the Gauss-Markov Theorem that the OLS estimator $\hat{\beta}_1$ is not the best linear unbiased estimator (BLUE) for θ unless $Var(Y_{k1j} \mid Y_{k0j})$ is a constant independent of k and j. Otherwise, the BLUE would be the weighted least squares estimate with each observation weighted by the inverse variance of the follow-up measurement conditional on its treatment group and its baseline measurement.

As noted above, the optimal linear combination of the baseline and follow-up observations based on a is of a form that would suggest performing linear regressions of the follow-up observations on the baseline observations within each group, and then using the estimated coefficients for the baseline values to adjust the follow-up measurements. That is, we could fit linear regression models

$$E[\,Y_{k1j}\,|\,Y_{k0j}\,] = \gamma_{k0} + Y_{k0j}\gamma_{k1},$$

and then define variables

$$Z_i^* = Z_i - \left(\left(\frac{n_0}{n_0+n_1}\right)\hat{\gamma}_{11} + \left(\frac{n_1}{n_0+n_1}\right)\hat{\gamma}_{01}\right)W_i,$$

and then regress Z_i^* on X_i. It should be noted, however, that the Z_i^* are now (very slightly) correlated within treatment groups, and thus the inference obtained from classical OLS on that simple linear regression model would not be entirely correct.