School of Mathematics and Statistics

MAS8391 Project

Analysis of AB/BA Crossover Trials with Missing Data

Author:
Rosie Mellors

Supervisor:
Prof. John Matthews

1st May 2014

Abstract

Investigators of clinical crossover trials are often met with the problem of how to deal with missing data caused by patients dropping out of a study prematurely. This dissertation introduces the AB/BA crossover clinical trial and presents different methods of analysis on both the complete case (the data set where patients with incomplete data are omitted) and the full case (the data set where all patients' readings have been included, regardless of whether or not they completed the trial). The data used in this trial are from a study on analgesic pain killers for neuropathic pain by Frank *et al.* [1]. A consideration of different types of missing data also provides an insight to selection model analysis for crossover data. Conclusions show differences between treatment estimators for the complete and full case analyses, but a discussion shows that it may be pertinent to use the complete case in practical applications.

Contents

Chapter 1

Introduction to Crossover Trials

1.1 Randomised Controlled Clinical Trials

In the area of medical research and drug development, a clinical trial is an experiment performed on a sample of human subjects with a particular condition used to assess the efficacy of a new treatment or medication. There are many classifications and designs of clinical trials, including Randomised Controlled Trials (RCTs). The presence of statistical ideas in RCTs, such as randomisation, has established them as the most reputable, and sometimes only, source of evidence for the assessment of new treatments. The use of empirical data eliminates the risk implied by use of opinions or uncontrolled environments.

A number of ethical implications arise through using human subjects in an experiment, creating three points which must be considered by investigators involved in an RCT:

- A patient must never be given a treatment which is known to be inferior to another.
- A patient must be made aware of all circumstances and implications of the trial and treatment.
- A patient is able withdraw from the trial at any time.

1.1.1 Structure

The structure of an RCT is as follows:

- A group of patients which meet a particular set of eligibility criteria are recruited. This criteria is in place to find a group of patients with relevant characteristics. The consideration of eligibility criteria is important to ensure generalisability of results.
- A sample of patients who formally consent to being entered into the trial are recruited from the set of eligible patients.

- Members of this sample of recruited patients are randomly allocated to one of two (or more) groups; the treated group, which receives the new treatment, and a control group which receives the current most commonly used treatment for the condition. Randomisation of allocation ensures the two treatment groups are comparable with each other.
- Once data is collected for each participant, the outcomes are compared and the efficacy of the treatment is assessed.

1.2 Crossover Trials

Generally, in a situation where a treatment is intended to cure an ailment, patients in a trial are only allocated one of the possible treatments during the study; this is called a *parallel group design*. For conditions where treatments are not intended to cure but to ease or control symptoms of chronic diseases, such as diabetes and asthma, subjects in a study could receive two or more treatments. Such studies are called *crossover trials*.

The main feature distinguishing crossover trials from other clinical trial designs is that a measurement from one patient receiving treatment A is compared to the measurement of treatment B from the same patient. There are many varying characteristics between human subjects and so this method provides the possibility of a more precise treatment comparison. This is because the variability which exists between subjects which can be removed quite simply from the comparison due to the nature of the crossover model (see Section 1.2.3).

1.2.1 The AB/BA Design

In the comparison of two treatments, it would seem to be simplest to give all patients treatment A followed by treatment B. However, this creates ambiguity in the results. For example, consider:

- If treatment A is taken first, it may be that it appears to be better than treatment B, purely because it was taken first.
- If all patients start the trial at the same time, there may be an outside influence on the which affects all readings.

This leads to a confounding of treatment with time, which we shall call a *period effect*. Therefore, recruited patients are still randomly allocated to one of two groups but patients in one group will receive treatment A in the first treatment period and cross over to treatment B in the second while patients in the other group receive the treatments in order BA. This is an AB/BA crossover design and is the simplest of the crossover trial designs.

1.2.2 Carryover

One problem with crossover trials is the issue of *carryover effect*, i.e. the idea that the benefits of the first treatment may continue into the second period, thus affecting the results. Investigation has been done into overcoming carryover and a common approach is to use non-statistical knowledge to introduce an appropriate "wash-out period" in between treatments to eliminate the problem, or start over with a different trial design.

1.2.3 Notation

Let n_{AB} be the number of subjects in the group taking treatment A in treatment period one followed by treatment B in treatment period two and n_{BA} for the group with the treatments in the reverse order. For group AB, we use $y_{AB_{ij}}$ to denote the outcome for subject i $(i = 1, ..., n_{AB})$ in period j, $(j = 1, 2)$, with $y_{BA_{ij}}$ similarly defined for group BA. Then, we have:

$$y_{AB_{i1}} = \mu + \pi_1 + \tau_A + \xi_i + \epsilon_{i1}, \qquad y_{AB_{i2}} = \mu + \pi_2 + \tau_B + \xi_i + \epsilon_{i2}, \quad (1.1)$$

$$y_{BA_{i1}} = \mu + \pi_1 + \tau_B + \xi_i + \epsilon_{i1}, \qquad y_{BA_{i2}} = \mu + \pi_2 + \tau_A + \xi_i + \epsilon_{i2}. \quad (1.2)$$

Where,

- μ is the general mean,
- π_j is the period effect of period j,
- τ_A, τ_B are the treatment effects for treatment A and B respectively,
- ξ_i is the patient effect of subject i,
- ϵ_{ij} is the error term, with ϵ_{ij} with mean 0 and variance σ_W^2.

The subject effect, ξ_i, can be removed from the analysis easily due to the nature of crossover design. This term is eliminated by taking differences of readings from each patient, i.e.

$$\begin{aligned} d_{AB_i} &= y_{AB_{i1}} - y_{AB_{i2}} \\ &= \pi_1 - \pi_2 + \tau_A - \tau_B + \xi_i - \xi_i + \epsilon_{i1} - \epsilon_{i2} \\ &= 2\pi + 2\tau + \eta_i \qquad (1.3) \\ \text{and } d_{BA_i} &= y_{BA_{i1}} - y_{BA_{i2}} \\ &= 2\pi - 2\tau + \eta_i., \qquad (1.4) \end{aligned}$$

Where,

- $\pi = \pi_1, -\pi = \pi_2$, when π is the semi-difference of treatment periods,
- $\tau = \tau_A, -\tau = \tau_B$, when τ is the semi-difference between treatments and
- $\eta_i = \epsilon_{i1} - \epsilon_{i2}$ is another error term, this time η_i with mean 0 and variance $2\sigma_W^2 = \tilde{\sigma}^2$.

Equations (1.3) and (1.4) can be used to determine treatment differences. Analysis involves a t-test, discussed further in Chapter 2.

1.3 The Pain Trial: Introduction

The paper by Frank *et al.* [1] outlines a crossover trial which compares the efficacy and side effects of nabilone (N), a synthetic cannabinoid, versus dihydrocodeine (D), a weak opiate, for treating neuropathic pain. The study, which lasted three months, took place at three hospitals in the UK. Eligible patients had neuropathic pain according to the diagnostic criteria (e.g. stabbing pain), were aged between 18 and 90 and were already taking a steady dose of pain relief. Patients excluded from the study included those with epilepsy, liver disease, psychosis, bipolar disorder, history of substance abuse, renal failure or those which display adverse effects to dihydrocodeine or nabilone. Pain scores were given as measurements on a $0-100$mm visual analogue scale.

Figure 1.1 illustrates a flow chart describing the study. Ninety-six patients were randomised to group DN, consisting of those who take dihydrocodeine in treatment period one and nabilone in treatment period two, or group ND, consisting of those taking the treatments in the reverse order. The model used included a fixed patient effect and the assumption of normal errors. Each treatment period lasted six weeks with a washout period of two weeks in between. Data were collected from the last two weeks of each treatment period (weeks 5-6 and weeks 13-14) to enable the exclusion of carryover from the model.

For the analyses in this report, we will only be considering the 82 patients who

	Number of patients with complete data	Number of patients with missing period two data	
Group DN	32	5	37
Group ND	35	10	45
	67	15	

Table 1.1: The group allocation for patients with complete and incomplete data in the pain trial.

either have pain readings for both treatment periods or only have a pain reading for period one. Table 1.1 shows a break down of the patients randomised to each group, and whether they have complete or incomplete data (missing readings). Table 1.2 shows the means of these measurements for both treatments in each group sequence, for the 82 patients.

	Mean of treatment D (mm)	Mean of treatment N (mm)
Group DN	51.41 (*54.14*)	59.16 (*59.16*)
Group ND	58.47 (*58.47*)	63.68 (*64.87*)

Table 1.2: The average pain scores on each treatment (mm), for the complete cases only (in *italics*) and the full data set.

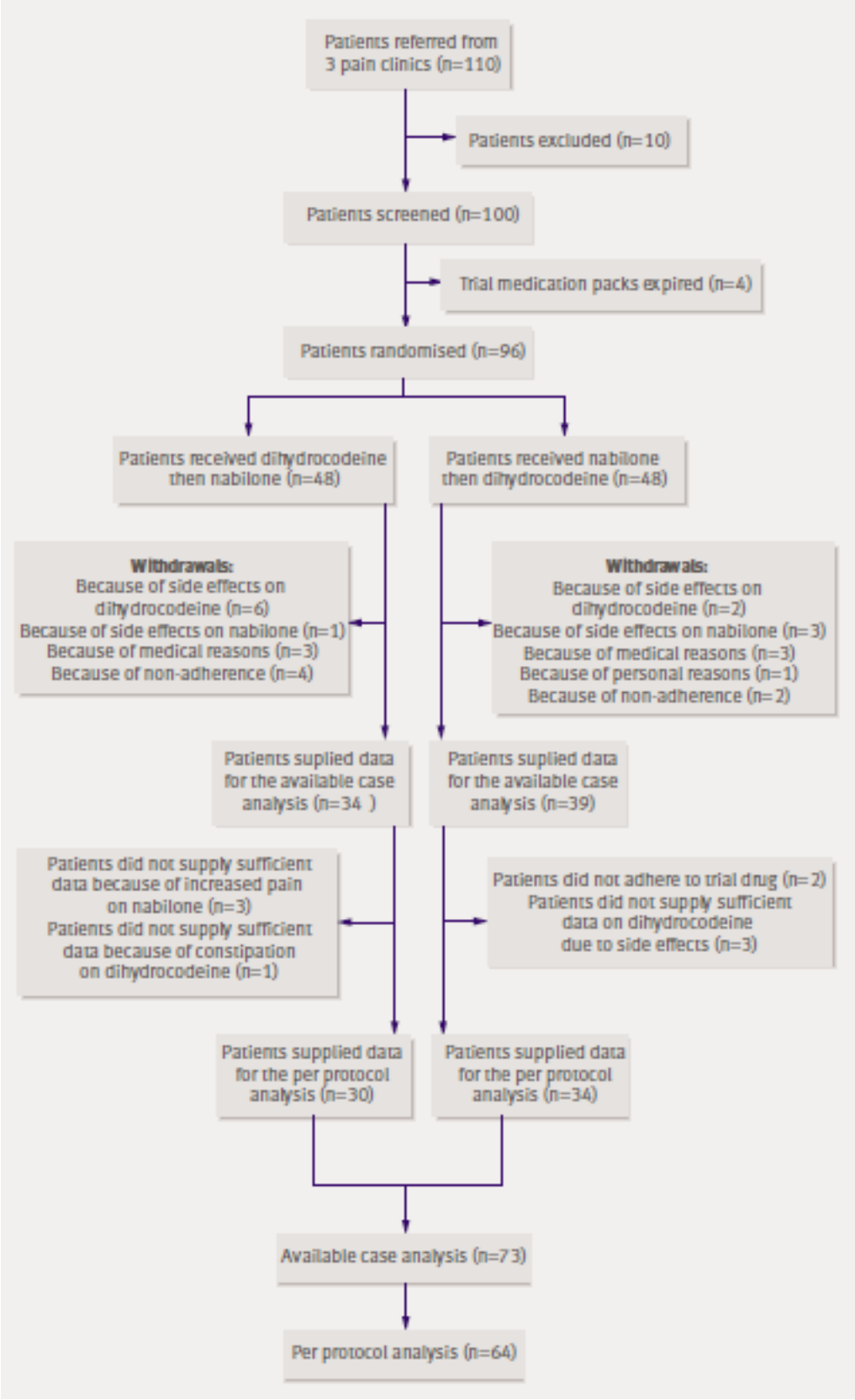

Figure 1.1: A flow chart from Frank *et al.* [1] describing the structure of the pain trial.

Chapter 2

Simple Analysis of AB/BA

2.1 Two Sample t-test

If we refer to Equations (1.3) and (1.4) in Section 1.2.3, we can see that

$$\mathrm{E}\left(d_{AB_i}\right) = 2\pi + 2\tau \quad \text{and} \quad \mathrm{E}\left(d_{BA_i}\right) = 2\pi - 2\tau.$$

If there is no treatment difference ($\tau = 0$) these two expectations are the same. To assess the difference between treatments we use a two sample t-test to test $H_0 : \tau = 0$.

2.1.1 Estimating τ

If we denote the mean sample difference in each group as

$$\bar{d}_{AB} = 2\pi + 2\tau + \bar{\eta}_{AB} \quad \text{and} \quad \bar{d}_{BA} = 2\pi - 2\tau + \bar{\eta}_{BA},$$

then $\mathrm{E}\big(\bar{d}_{AB} - \bar{d}_{BA}\big) = 4\tau$ and an estimate for the treatment difference is

$$\hat{\tau} = \frac{1}{4}\left(\bar{d}_{AB} - \bar{d}_{BA}\right). \tag{2.1}$$

If we take the expectation of this estimator, we show that it is unbiased (i.e. $\mathrm{E}(\hat{\tau}) = \tau$):

$$\begin{aligned}
\mathrm{E}(\hat{\tau}) &= \frac{1}{4}\mathrm{E}\left(\bar{d}_{AB} - \bar{d}_{BA}\right) \\
&= \frac{1}{4}\mathrm{E}\left(2\pi + 2\tau + \bar{\eta}_{AB} - 2\pi + 2\tau - \bar{\eta}_{BA}\right) \\
&= \frac{1}{4}\mathrm{E}\left(4\tau + \bar{\eta}_{AB} - \bar{\eta}_{BA}\right) \\
&= \tau + \frac{1}{4}\mathrm{E}\left(\bar{\eta}_{AB} - \bar{\eta}_{BA}\right) \\
&= \tau.
\end{aligned}$$

So, $\hat{\tau}$ is an unbiased estimator, with variance:

$$\begin{aligned} \text{Var}(\hat{\tau}) = \text{SE}^2 &= \frac{1}{16}\text{Var}\left(\bar{d}_{AB} - \bar{d}_{BA}\right) \\ &= \frac{1}{16}\text{Var}\left(\bar{\eta}_{AB} - \bar{\eta}_{BA}\right) \\ &= \frac{1}{16}\tilde{\sigma}^2\left(\frac{1}{n_{AB}} + \frac{1}{n_{BA}}\right). \end{aligned} \tag{2.2}$$

This variance is only dependent on $\tilde{\sigma}^2$, the variance of the error term, not the variance due to the patient, σ_B^2. This is the main statistical manifestation of the benefit of using a patient as his/her own control, mentioned in Section 1.2. Note also, that for n_{AB} or $n_{BA} = 0$, this variance is infinite, i.e. there must be at least one patient in each sequence for the estimator in Equation (2.1) to be valid.

2.1.2 The Pain Trial: Unpaired t-test

Using this simple analysis method, we can investigate the treatment effects for the pain trial discussed in Section 1.3. As we are using patient differences, d_{DN} and d_{ND}, we are only able to consider the patients with complete data. If we conduct an unpaired t-test on the complete cases in R, we receive the following output:

```
> t.test(dDN, dND, paired=F, var.equal=T)

Two Sample t-test

data:  dDN and dND
t = -2.3535, df = 65, p-value = 0.02163
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -21.105968  -1.728743
sample estimates:
mean of x mean of y
-5.019981  6.397374
```

Here, we have a p-value of 0.02119 (< 0.05) for the null hypothesis $H_0 : \tau = 0$, impying a difference between treatments which is not solely due to chance (at the 95% level). The output also gives values $\bar{d}_{DN} = -5.019981$ and $\bar{d}_{ND} = 6.397374$. If we recall the estimator in Equation (2.1), we can use these values to calculate a treatment difference estimate of

$$\begin{aligned} \hat{\tau} &= \frac{1}{4}\left(\bar{d}_{DN} - \bar{d}_{ND}\right) \\ &= \frac{1}{4}\left(-5.019981 - 6.397374\right) \\ &= -2.8543, \end{aligned}$$

which has a 95% confidence interval of $(-5.2679, -0.4408)$. This suggests that nabilone (N) is more beneficial than dihydrocodeine (D) for reducing pain score, as moving from treatment D to treatment N reduces pain measurement by 2.85mm.

2.2 Paired t-test

At first sight, if we did not consider a period effect, it may seem intuitive to use a paired t-test on the data. In this case our treatment difference estimator is the mean of the differences;

$$\hat{\tau}_p = \frac{\sum^{n_{AB}} d_i - \sum^{n_{BA}} d_i}{2\left(n_{AB} + n_{BA}\right)} \tag{2.3}$$

where $\sum^{n_{AB}}$ denotes the sum over all the patients in group AB and $\sum^{n_{AB}}$ denotes the sum over all the patients in group BA.

2.2.1 Expectation and Variance of $\hat{\tau}_p$

If we take the expectation of Equation (2.3), we can see that it is, in general, biased;

$$\begin{aligned}
\mathrm{E}(\hat{\tau}_p) &= \mathrm{E}\left[\frac{2n_{AB}\left(\pi+\tau\right) + \sum^{n_{AB}} \eta_i - 2n_{BA}\left(\pi-\tau\right) - \sum^{n_{BA}} \eta_i}{2(n_{AB}+n_{BA})}\right] \\
&= \mathrm{E}\left[\frac{2\pi(n_{AB}-n_{BA}) + 2\tau(n_{AB}+n_{BA}) + \sum^{n} \eta_i}{2n}\right] \\
&= \frac{\pi(n_{AB}-n_{BA})}{n} + \tau + \mathrm{E}\left(\frac{\sum^{n}\eta_i}{n}\right) \\
&= \frac{\pi(n_{AB}-n_{BA})}{n} + \tau, \qquad (2.4)
\end{aligned}$$

(where $n = n_{AB} + n_{BA}$). Note that the distribution of $\sum^{n_{AB}} \eta_i - \sum^{n_{BA}} \eta_i$ is the same as that of $\sum^{n} \eta_i$ as η is symmetric about 0.

This has a variance of

$$\begin{aligned}
\mathrm{Var}(\hat{\tau}_p) = \mathrm{SE}_p^2 &= \mathrm{Var}\left(\frac{\sum^{n}\eta_i}{2n}\right) \\
&= \frac{\tilde{\sigma}^2}{4n}.
\end{aligned}$$

We can see from Equation (2.4) that, when there is no period effect or when group sizes are the same ($n_{AB} = n_{BA}$), $\mathrm{E}\left(\hat{\tau}_p\right) = \tau$, i.e. $\hat{\tau}_p$ is an unbiased estimator. In the case when $n_{AB} = n_{BA}$, Equation (2.3) becomes $\hat{\tau}_p = \frac{1}{4}\left(\bar{d}_{AB} - \bar{d}_{BA}\right)$ and the analysis for the unpaired t-test and paired t-test are the same.

2.2.2 The Pain Trial: Paired t-test

Overleaf shows the output for a paired t-test on the complete cases from the pain trial data. d and n are vectors of readings obtained when patients were on dihydrocodeine and nabilone respectively, regardless of the treatment period. Here, we have multiplied these vectors by 0.5 as we have been working with the *semi*-treatment difference, τ.

```
> t.test(0.5*d, 0.5*n, paired=T)

Paired t-test

data:  0.5 * d and 0.5 * n
t = -2.3852, df = 66, p-value = 0.01995
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -5.2718788 -0.4676361
sample estimates:
mean of the differences
              -2.869757
```

On this occasion, the p-value is 0.01995 (< 0.05) for the null hypothesis of $H_0 : \tau = 0$, which still suggests a difference between treatments which is not solely due to chance (at the 95% level). The treatment estimate is

$$\hat{\tau}_p \;=\; -2.8698,$$

with a 95% confidence interval of $(-5.2719, -0.4676)$ and Nabilone (N) is still the better treatment for reducing pain score. If we refer to Table 2.1, we can see that this treatment estimate and its confidence interval are very similar to those found using the unpaired t-test which suggests that the period effect, π, is very small.

	$\hat{\tau}$	95% confidence interval
Unpaired t-test	-2.8543	(-5.2679, -0.4408)
Paired t-test	-2.8698	(-5.2719, -0.4676)

Table 2.1: The treatment effects and confidence intervals for the pain trial data, found using unpaired and paired t-tests.

2.3 Analytic Comparison of Methods

Despite having a biased estimate for τ, there may be some situations where it is better to use a paired t-test over an unpaired t-test. We investigate this further using the ratio of the standard errors.

2.3.1 Ratio of Standard Errors

The standard errors of $\hat{\tau}$ and $\hat{\tau}_p$ are denoted SE and SE_p respectively. The ratio is

$$\begin{aligned}\left(\frac{\text{SE}}{\text{SE}_p}\right)^2 &= \frac{1}{4}\frac{(n_{AB}+n_{BA})^2}{n_{AB}n_{BA}} \\ &= \frac{1}{4}\left(\frac{n_{AB}}{n_{BA}}+2+\frac{n_{BA}}{n_{AB}}\right).\end{aligned}$$

We know that, for any $x > 0$,

$$\left(x + \frac{1}{x}\right) \geq 2,$$

and so,

$$\left(\frac{n_{AB}}{n_{BA}} + \frac{n_{BA}}{n_{AB}}\right) \geq 2,$$
$$\frac{1}{4}\left(\frac{n_{AB}}{n_{BA}} + 2 + \frac{n_{BA}}{n_{AB}}\right) \geq 1,$$

that is,

$$\frac{\text{SE}}{\text{SE}_p} \geq 1 \quad \Rightarrow \quad \text{SE} \geq \text{SE}_p$$

So the paired method has a smaller standard error which implies that it may give results which are better representative of the population than the unpaired method. However, this standard error does not take into account the bias of τ_p. To allow for this, we use the mean square error (MSE).

2.3.2 Ratio of Mean Square Errors

The MSE of an estimator is a measure of performance of that estimator, which takes into account both accuracy and precision. It measures the average squared difference between an estimator and the parameter. It is defined by

$$\begin{aligned}
\text{MSE} &= \text{E}\left[\left(\hat{\theta} - \theta\right)^2\right] \\
&= \text{E}\left[\left(\hat{\theta} - \text{E}(\hat{\theta}) + \text{E}(\hat{\theta}) - \theta\right)^2\right] \\
&= \text{Var}\left(\hat{\theta}\right) + \text{bias}\left(\hat{\theta}\right)^2 + 2\left(E(\hat{\theta}) - \theta\right)\text{E}\left[\hat{\theta} - \text{E}(\hat{\theta})\right] \\
&= \text{Var}\left(\hat{\theta}\right) + \text{bias}\left(\hat{\theta}\right)^2.
\end{aligned}$$

So for the unpaired method,

$$\text{MSE} = \text{Var}(\hat{\tau}) = \frac{1}{16}\tilde{\sigma}^2\left(\frac{1}{n_{AB}} + \frac{1}{n_{BA}}\right),$$

whereas for the paired,

$$\begin{aligned}
\text{MSE}_p &= \text{Var}\left(\hat{\tau}_p\right) + \text{bias}\left(\hat{\tau}_p\right)^2 \\
&= \frac{\tilde{\sigma}^2}{4(n_{AB} + n_{BA})} + \left[\frac{\pi(n_{AB} - n_{BA})}{n_{AB} + n_{BA}}\right]^2 \\
&= \frac{\tilde{\sigma}^2}{4(n_{AB} + n_{BA})} + \frac{\pi^2(n_{AB} - n_{BA})^2}{(n_{AB} + n_{BA})^2}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\frac{\text{MSE}_p}{\text{MSE}} &= \left(\frac{16}{\tilde{\sigma}^2}\frac{n_{AB}n_{BA}}{n_{AB}+n_{BA}}\right)\left[\frac{\tilde{\sigma}^2}{4(n_{AB}+n_{BA})}+\frac{\pi^2(n_{AB}-n_{BA})^2}{(n_{AB}+n_{BA})^2}\right] \\
&= \frac{4n_{AB}n_{BA}}{(n_{AB}+n_{BA})^2}+\frac{16\pi^2 n_{AB}n_{BA}(n_{AB}-n_{BA})^2}{\tilde{\sigma}^2(n_{AB}+n_{BA})^3} \\
&= \frac{4n_{AB}n_{BA}}{(n_{AB}+n_{BA})^2}\left[1+\frac{4\pi^2(n_{AB}-n_{BA})^2}{\tilde{\sigma}^2(n_{AB}+n_{BA})}\right] \\
&= 4\theta(1-\theta)\left[1+\frac{4\pi^2}{\tilde{\sigma}^2}n\left(\frac{n_{AB}-n_{BA}}{n_{AB}+n_{BA}}\right)^2\right] \\
&= 4\theta(1-\theta)\left[1+k(2\theta-1)^2\right],
\end{aligned}$$

where

$$\theta = \frac{n_{AB}}{n_{AB}+n_{BA}},$$

and

$$k = \frac{4\pi^2}{\tilde{\sigma}^2}n. \tag{2.5}$$

If $\theta = \frac{1}{2}$, i.e. the group sizes are the same size, then $\text{MSE}_p/\text{MSE} = 1$ and the two methods are just as effective as each other. This in agreement with Section 2.2.1 which states that the two methods of analysis are the same if the group sizes are the same.

The unpaired method is better than the paired method when $\text{MSE}_p/\text{MSE} \geq 1$, i.e.

$$4\theta\,(1-\theta)\left[1+k(2\theta-1)^2\right] \geq 1$$

$$\begin{aligned}
\Rightarrow k &\geq \frac{1-4\theta\,(1-\theta)}{(2\theta-1)^2 4\theta\,(1-\theta)} \\
&\geq \frac{4\theta^2-4\theta+1}{(2\theta-1)^2 4\theta\,(1-\theta)} \\
&\geq \frac{1}{4\theta(1-\theta)}.
\end{aligned} \tag{2.6}$$

The result in Equation (2.6) is better illustrated in Figure 2.1, a plot of MSE_p/MSE against θ for varying k. Values below the red line correspond to situations where the paired method is more effective and above the red line corresponds to the unpaired method being more effective.

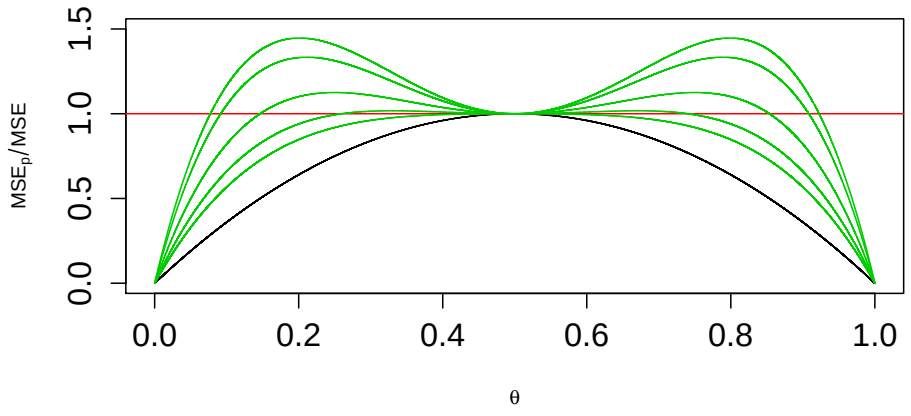

Figure 2.1: A graph of MSE_p/MSE against θ for $k = 0, 0.9, 1.3, 2, 3, 3.5$.

We can see a number of things from Figure 2.1;

- **When k=0 (the black quadratic curve), MSE_p/MSE is under the red line, which means that it is better to use the paired method to analyse crossover data.** $k = 0$ corresponds to $\pi = 0$ (See Equation (2.5)), i.e. no period effect. We have seen before that a zero period effect gives $\text{E}(\hat{\tau}_p) = \text{E}(\hat{\tau})$ and $\hat{\tau}_p$ is unbiased in Section 2.2.1.

- **Near θ=0,1, the curve is always below the line and the paired method is more favourable.** This is mentioned in Section 2.1.1. If either n_{AB} or $n_{BA} = 0$, the variance of $\hat{\tau}$ is infinite (see Equation 2.2) and so is the MSE of $\hat{\tau}$. Then, MSE_p is smaller and it is better to use the paired method. However, this is unrealistic, as the definition of a crossover trial would require patients in both groups.

- **Regardless of the size of k, when θ=$\frac{1}{2}$ the paired method and unpaired method are equally effective.** When the groups are equal sizes, the two methods coincide. Generally, we aim for group sizes to be equal in clinical trials as this minimises the variance given in Equation (2.2). However, patients in a study are able to drop out and groups are very often left unequal. Refer to Chapters 4 and 6 for more information on patient drop outs.

- **As k increases, more values reach over the red line and the unpaired method is more favourable than the paired.** Large k is likely, as $k \propto n$, and n is large.

Usually in a crossover trial, there are patients in both treatment groups and, whilst the aim would be for the groups to be equal size, they are often left unequal. Also, n would be large, which implies large k. So, an unpaired t-test is generally more advisable.

Chapter 3

Likelihood Estimation for Complete Data

In Section 2.1.1 we produced an estimate of $\hat{\tau} = \frac{1}{4}\left(\bar{d}_{AB} - \bar{d}_{BA}\right)$ for the treatment difference when using an unpaired t-test. However, this method did not consider patients with incomplete data. To assess the the role of this partial information, we can estimate τ and π using maximum likelihood estimation. As a preliminary, we consider this approach on only the complete cases.

3.1 Likelihood Function

Here, we recall Equations (1.1) and (1.2) from Section 1.2.3

$$\begin{aligned} y_{AB_{i1}} &= \mu + \pi_1 + \tau_A + \xi_i + \epsilon_{i1}, & y_{AB_{i2}} &= \mu + \pi_2 + \tau_B + \xi_i + \epsilon_{i2}, \\ y_{BA_{i1}} &= \mu + \pi_1 + \tau_B + \xi_i + \epsilon_{i1}, & y_{BA_{i2}} &= \mu + \pi_2 + \tau_A + \xi_i + \epsilon_{i2}. \end{aligned}$$

Recalling the definitions of μ, π and τ from Section 1.2.3, we can write the above's means as:

$$\begin{aligned} \mathrm{E}\left(y_{AB_i}\right) &= X_{AB}\beta \\ &= \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \mu \\ \pi \\ \tau \end{pmatrix} \\ \mathrm{E}\left(y_{BA_i}\right) &= X_{BA}\beta \\ &= \begin{pmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \pi \\ \tau \end{pmatrix}, \end{aligned}$$

where y_{AB_i} is the $[2 \times 1]$ vector of responses on subject i in group AB, similarly for y_{BA_i}. Previously we did not specify the form of ξ_i. For maximum likelihood estimation, the number of parameters cannot increase with the number of observations, so we must not take ξ_i to be a fixed effect. We therefore assume ξ_i is a normal random effect with mean 0 and variance σ_B^2.If we recall from Section 1.2.3 that

$\text{Var}(\epsilon_i) = \sigma_W^2$, then

$$\begin{aligned} \text{Var}\left(y_{AB_i}\right) &= \sigma_B^2 + \sigma_W^2 \\ &= \sigma^2. \end{aligned}$$

We have a sample of independent observations and so, the likelihood function, $\boldsymbol{L}$, is the product of the joint density function at each observation, that is,

$$\begin{aligned} \boldsymbol{L} &= \mathcal{L}\left(\beta | y_1, \ldots, y_n\right) \\ &= \prod_{i=1}^{n} f\left(y_i|\beta\right). \end{aligned}$$

So, assuming normality of the error terms, the log-likelihood is

$$\begin{aligned} l &= \log(\boldsymbol{L}) \\ &= -n\log(2\pi) - \frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n_{AB}} \left(y_{AB_i} - X_{AB}\beta\right)^\top \Sigma^{-1}\left(y_{AB_i} - X_{AB}\beta\right) \\ &\quad -\frac{1}{2}\sum_{i=1}^{n_{BA}} \left(y_{BA_i} - X_{BA}\beta\right)^\top \Sigma^{-1}\left(y_{BA_i} - X_{BA}\beta\right) \\ &= -n\log(2\pi) - \frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n_{AB}} \left({y_{AB_i}}^\top \Sigma^{-1} y_{AB_i} - 2{y_{AB_i}}^\top \Sigma^{-1} X_{AB}\beta + \beta^\top {X_{AB}}^\top \Sigma^{-1} X_{AB}\beta\right) \\ &\quad -\frac{1}{2}\sum_{i=1}^{n_{BA}} \left({y_{BA_i}}^\top \Sigma^{-1} y_{BA_i} - 2{y_{BA_i}}^\top \Sigma^{-1} X_{BA}\beta + \beta^\top {X_{BA}}^\top \Sigma^{-1} X_{BA}\beta\right). \end{aligned} \tag{3.1}$$

3.2 Maximisation and Estimation

To estimate the parameters in $\beta = (\mu, \pi, \tau)^\top$, we wish to maximise the log-likelihood function found in Equation (3.1). First, we differentiate the log-likelihood function, with respect to β;

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{n_{AB}} \left({X_{AB}}^\top \Sigma^{-1} y_{AB_i} - {X_{AB}}^\top \Sigma^{-1} X_{AB}\beta\right) + \sum_{i=1}^{n_{BA}} \left({X_{BA}}^\top \Sigma^{-1} y_{BA_i} - {X_{BA}}^\top \Sigma^{-1} X_{BA}\beta\right). \tag{3.2}$$

Then, to maximise, set $\partial l / \partial \beta = 0$ to give

$$\sum_{i=1}^{n_{AB}} {X_{AB}}^\top \Sigma^{-1} y_{AB_i} + \sum_{i=1}^{n_{BA}} {X_{BA}}^\top \Sigma^{-1} y_{BA_i} = n_{AB}{X_{AB}}^\top \Sigma^{-1} X_{AB}\hat{\beta} + n_{BA}{X_{BA}}^\top \Sigma^{-1} X_{BA}\hat{\beta}$$

If we denote the mean of the $y_{AB_{ij}}$ and $y_{BA_{ij}}$, for period $j = 1, 2$, to be $\bar{y}_{AB_j}$ and $\bar{y}_{BA_j}$ respectively, we have

$$n_{AB}{X_{AB}}^\top \Sigma^{-1} \bar{y}_{AB_j} + n_{BA}{X_{BA}}^\top \Sigma^{-1} \bar{y}_{BA_j} = n_{AB}{X_{AB}}^\top \Sigma^{-1} X_{AB}\hat{\beta} + n_{BA}{X_{BA}}^\top \Sigma^{-1} X_{BA}\hat{\beta} \tag{3.3}$$

Here,

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

which implies

$$\Sigma^{-1} = \frac{\sigma^2}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}.$$

We also set, for simplicity of notation,

$$\begin{aligned} n^+ &= n_{AB} + n_{BA} \\ n^- &= n_{AB} - n_{BA}. \end{aligned}$$

Then the RHS of Equation (3.3) becomes

$$\begin{aligned} &\left(n_{AB} {X_{AB}}^\top \Sigma^{-1} X_{AB} + n_{BA} {X_{BA}}^\top \Sigma^{-1} X_{BA}\right) \hat{\beta} \\ &= \frac{\sigma^2}{1-\rho^2} \left[n_{AB} \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \end{pmatrix} \right. \\ &\qquad \left. + n_{BA} \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix} \right] \hat{\beta} \\ &= \frac{\sigma^2}{1-\rho^2} \left[\begin{pmatrix} 2n_{AB}(1-\rho) & 0 & 0 \\ 0 & 2n_{AB}(1+\rho) & 2n_{AB}(1+\rho) \\ 0 & 2n_{AB}(1+\rho) & 2n_{AB}(1+\rho) \end{pmatrix} \right. \\ &\qquad \left. + \begin{pmatrix} 2n_{BA}(1-\rho) & 0 & 0 \\ 0 & 2n_{BA}(1+\rho) & -2n_{BA}(1+\rho) \\ 0 & -2n_{BA}(1+\rho) & 2n_{BA}(1+\rho) \end{pmatrix} \right] \hat{\beta} \\ &= \frac{\sigma^2}{1-\rho^2} \begin{pmatrix} 2(1-\rho)n^+ & 0 & 0 \\ 0 & 2(1+\rho)n^+ & 2(1+\rho)n^- \\ 0 & 2(1+\rho)n^- & 2(1+\rho)n^+ \end{pmatrix} \hat{\beta}, \end{aligned}$$

and the LHS of Equation (3.3) becomes

$$\begin{aligned} &n_{AB} {X_{AB}}^\top \Sigma^{-1} \bar{y}_{AB_j} + n_{BA} {X_{BA}}^\top \Sigma^{-1} \bar{y}_{BA_j} \\ &= \frac{\sigma^2}{1-\rho^2} \left[n_{AB} \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \bar{y}_{AB_j} + n_{BA} \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \bar{y}_{BA_j} \right] \\ &= \frac{\sigma^2}{1-\rho^2} \left[n_{AB} \begin{pmatrix} 1-\rho & 1-\rho \\ 1+\rho & -(1+\rho) \\ 1+\rho & -(1+\rho) \end{pmatrix} \begin{pmatrix} \bar{y}_{AB_1} \\ \bar{y}_{AB_2} \end{pmatrix} \right. \\ &\qquad \left. + n_{BA} \begin{pmatrix} 1-\rho & 1-\rho \\ 1+\rho & -(1+\rho) \\ -(1+\rho) & 1+\rho \end{pmatrix} \begin{pmatrix} \bar{y}_{BA_1} \\ \bar{y}_{BA_2} \end{pmatrix} \right] \\ &= \frac{\sigma^2}{1-\rho^2} \left[n_{AB} \begin{pmatrix} (1-\rho)\bar{s}_{AB} \\ (1+\rho)\bar{d}_{AB} \\ (1+\rho)\bar{d}_{AB} \end{pmatrix} + n_{BA} \begin{pmatrix} (1-\rho)\bar{s}_{BA} \\ (1+\rho)\bar{d}_{BA} \\ -(1+\rho)\bar{d}_{BA} \end{pmatrix} \right]. \end{aligned}$$

Where $\bar{d}_{AB}$ and $\bar{d}_{BA}$ are the mean sample differences in each group defined in Section 2.1.1. We also define $\bar{s}_{AB} = \bar{y}_{AB_1} + \bar{y}_{AB_2}$ and $\bar{s}_{BA} = \bar{y}_{BA_1} + \bar{y}_{BA_2}$. So, combining the LHS and RHS, we obtain

$$\begin{pmatrix} \hat{\mu} \\ \hat{\pi} \\ \hat{\tau} \end{pmatrix} = \begin{pmatrix} 2(1-\rho)n^+ & 0 & 0 \\ 0 & 2(1+\rho)n^+ & 2(1+\rho)n^- \\ 0 & 2(1+\rho)n^- & 2(1+\rho)n^+ \end{pmatrix}^{-1} \left[n_{AB} \begin{pmatrix} (1-\rho)\bar{s}_{AB} \\ (1+\rho)\bar{d}_{AB} \\ (1+\rho)\bar{d}_{AB} \end{pmatrix} + n_{BA} \begin{pmatrix} (1-\rho)\bar{s}_{BA} \\ (1+\rho)\bar{d}_{BA} \\ -(1+\rho)\bar{d}_{BA} \end{pmatrix} \right].$$

As we are only interested in $\hat{\pi},\ \hat{\tau}$, we look at:

$$\begin{aligned} \begin{pmatrix} \hat{\pi} \\ \hat{\tau} \end{pmatrix} &= \begin{pmatrix} 2(1+\rho)n^+ & 2(1+\rho)n^- \\ 2(1+\rho)n^- & 2(1+\rho)n^+ \end{pmatrix}^{-1} \left[n_{AB} \begin{pmatrix} (1+\rho)\bar{d}_{AB} \\ (1+\rho)\bar{d}_{AB} \end{pmatrix} + n_{BA} \begin{pmatrix} (1+\rho)\bar{d}_{BA} \\ -(1+\rho)\bar{d}_{BA} \end{pmatrix} \right] \\ &= \frac{1}{2\cancel{(1+\rho)}(n^{+2} - n^{-2})} \begin{pmatrix} n^+ & -n^- \\ -n^- & n^+ \end{pmatrix} \left[n_{AB} \begin{pmatrix} \cancel{(1+\rho)}\bar{d}_{AB} \\ \cancel{(1+\rho)}\bar{d}_{AB} \end{pmatrix} + n_{BA} \begin{pmatrix} \cancel{(1+\rho)}\bar{d}_{BA} \\ -\cancel{(1+\rho)}\bar{d}_{BA} \end{pmatrix} \right] \end{aligned}$$

where

$$n^{+2} - n^{-2} = (n_{AB} + n_{BA})^2 - (n_{AB} - n_{BA})^2 = 4n_{AB}n_{BA}$$

implies,

$$\begin{aligned} \begin{pmatrix} \hat{\pi} \\ \hat{\tau} \end{pmatrix} &= \frac{1}{8} \left[\frac{1}{n_{BA}} \begin{pmatrix} n^+ & -n^- \\ -n^- & n^+ \end{pmatrix} \begin{pmatrix} \bar{d}_{AB} \\ \bar{d}_{AB} \end{pmatrix} + \frac{1}{n_{AB}} \begin{pmatrix} n^+ & -n^- \\ -n^- & n^+ \end{pmatrix} \begin{pmatrix} \bar{d}_{BA} \\ -\bar{d}_{BA} \end{pmatrix} \right] \\ &= \frac{1}{8} \left[\frac{1}{n_{BA}} \begin{pmatrix} n^+\bar{d}_{AB} - n^-\bar{d}_{AB} \\ n^+\bar{d}_{AB} - n^-\bar{d}_{AB} \end{pmatrix} + \frac{1}{n_{AB}} \begin{pmatrix} n^+\bar{d}_{BA} + n^-\bar{d}_{BA} \\ -n^-\bar{d}_{BA} - n^+\bar{d}_{BA} \end{pmatrix} \right]. \end{aligned}$$

Then, an estimator for the period effect is:

$$\begin{aligned} \hat{\pi} &= \frac{1}{8} \left(\frac{1}{n_{BA}} n^+ \bar{d}_{AB} - \frac{1}{n_{BA}} n^- \bar{d}_{AB} + \frac{1}{n_{AB}} n^+ \bar{d}_{BA} + \frac{1}{n_{AB}} n^- \bar{d}_{BA} \right) \\ &= \frac{1}{8} \left[(n_{AB} + n_{BA}) \frac{1}{n_{BA}} \bar{d}_{AB} - (n_{AB} - n_{BA}) \frac{1}{n_{BA}} \bar{d}_{AB} \right. \\ &\qquad \left. + (n_{AB} + n_{BA}) \frac{1}{n_{AB}} \bar{d}_{BA} + (n_{AB} - n_{BA}) \frac{1}{n_{AB}} \bar{d}_{BA} \right] \\ &= \frac{1}{4} \left(\bar{d}_{AB} + \bar{d}_{BA} \right), \end{aligned}$$

and an estimator for the treatment effect is

$$\begin{aligned} \hat{\tau} &= \frac{1}{8} \left[(n_{AB} + n_{BA}) \frac{1}{n_{BA}} \bar{d}_{AB} - (n_{AB} - n_{BA}) \frac{1}{n_{BA}} \bar{d}_{AB} \right. \\ &\qquad \left. - (n_{AB} + n_{BA}) \frac{1}{n_{AB}} \bar{d}_{BA} - (n_{AB} - n_{BA}) \frac{1}{n_{AB}} \bar{d}_{BA} \right] \\ &= \frac{1}{4} \left(\bar{d}_{AB} - \bar{d}_{BA} \right). \end{aligned}$$

This estimator is in agreement with the estimator for the unpaired t-test found in Section 2.1.1.

Chapter 4

Likelihood Estimation for Incomplete Data

4.1 Patient Drop Outs and Missing Data

Ideally, when conducting a clinical trial, patients would turn up to every clinic to provide the appropriate readings or measurements. However, it may be the case that a patient decides to drop out of the trial, resulting in a missing reading for that patient. There are four possible situations that can arise in this trial design:

1. patients which have an observation in each treatment period,
2. patients which have no observation for the first treatment period but an observation for the second,
3. patients which have an observation for the first treatment period, but none for the second,
4. patients which have no observations for either treatment period.

Those in situation 4 are not considered in analysis. Situation 2 is very rare in practice, and so not considered in this dissertation. We will be considering data with situations 1 (denoted the *complete cases*) and 3 (denoted the *incomplete cases*).

Previous methods in this paper have been 'complete case analyses' as they have ignored patients with no data in period two. However, this may mean missing out information important the results, so we can use maximum likelihood estimation to find new estimates for τ and π based on all the available data. We shall call this a *full analysis*. Further discussion into the causes of drop outs and how this data should be dealt with in a more sophisticated way is considered in Chapter 6.

4.2 Likelihood Function

As the observations on each patient are independent and identically distributed with the assumption of normality of the errors, we take the log-likelihood function to be

the sum of the complete cases, denoted with (c), and incomplete cases, denoted with (m), i.e. (ignoring factors of 2π)

$$\begin{aligned} l\left(\beta,\sigma^2,\rho\right) &= \log L\left(\beta,\sigma^2,\rho\right) \\ &= -\frac{n_{AB(c)}}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n_{AB(c)}}\left(y_{AB_i}-X_{AB}\beta\right)^\top\Sigma^{-1}\left(y_{AB_i}-X_{AB}\beta\right) \\ &\quad -\frac{n_{BA(c)}}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n_{BA(c)}}\left(y_{BA_i}-X_{BA}\beta\right)^\top\Sigma^{-1}\left(y_{BA_i}-X_{BA}\beta\right) \\ &\quad -\frac{1}{2}n_{AB(m)}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^{n_{AB(m)}}\frac{\left(y_{AB_{i1}}-X_{AB_1}\beta\right)^2}{\sigma^2} \\ &\quad -\frac{1}{2}n_{BA(m)}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^{n_{BA(m)}}\frac{\left(y_{BA_{i1}}-X_{BA_1}\beta\right)^2}{\sigma^2}. \end{aligned} \tag{4.1}$$

Here X_{AB_1} and X_{BA_1} denote the first rows of the matrices X_{AB} and X_{BA} respectively;

$$\begin{aligned} X_{AB_1} &= \begin{pmatrix}1 & 1 & 1\end{pmatrix} \\ X_{BA_1} &= \begin{pmatrix}1 & 1 & -1\end{pmatrix}. \end{aligned}$$

We write

$$\begin{aligned} \Sigma &= \sigma^2\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix} = \sigma^2 C, \\ \Sigma^{-1} &= \frac{\sigma^{-2}}{(1-\rho^2)}\begin{pmatrix}1 & -\rho\\ -\rho & 1\end{pmatrix} = \sigma^{-2}\boldsymbol{C^{-1}}, \end{aligned}$$

and

$$|\Sigma| = \sigma^4\left(1-\rho^2\right) = \sigma^4|C|,$$

so, disregarding terms constant with respect to β, ρ and σ^2, we have

$$l\left(\beta,\sigma^2,\rho\right) = -\frac{1}{2}n^+_{(c)}\log|C| - \frac{1}{2}\left(n^+_{(m)}+2n^+_{(c)}\right)\log\sigma^2 - \frac{1}{2\sigma^2}A\left(\beta,\rho\right). \tag{4.2}$$

Where

$$\begin{aligned} A\left(\beta,\rho\right) = &\sum_{i=1}^{n_{AB(c)}}\left(y_{AB_i}-X_{AB}\beta\right)^\top\boldsymbol{C^{-1}}\left(y_{AB_i}-X_{AB}\beta\right) \\ &+\sum_{i=1}^{n_{BA(c)}}\left(y_{BA_i}-X_{BA}\beta\right)^\top\boldsymbol{C^{-1}}\left(y_{BA_i}-X_{BA}\beta\right) \\ &+\sum_{i=1}^{n_{AB(m)}}\left(y_{AB_{i1}}-X_{AB_1}\beta\right)^2 + \sum_{i=1}^{n_{BA(m)}}\left(y_{BA_{i1}}-X_{BA_1}\beta\right)^2. \end{aligned}$$

4.3 Maximisation and Estimation

We can maximise this likelihood using R function `lmer`. This is discussed further in Chapter 5. First we consider the extent to which analytical progress is possible in this case, rather than apply a general approach to this particular kind of crossover study. To proceed, we note;

$$l\left(\beta, \sigma^2, \rho\right) = function\left(\rho\right) + function\left(\sigma^2\right) + function\left(\rho, \beta, \sigma^{-2}\right).$$

For fixed ρ, we can maximise the likelihood found in Equations (4.2), with respect to β, to find $\hat{\beta}_\rho$. Given this estimate, we consider $l(\hat{\beta}_\rho, \sigma^2, \rho)$ and maximise to find $\hat{\sigma}^2_\rho$. We are then left with a profile likelihood for ρ, $l(\hat{\beta}_\rho, \hat{\sigma}^2_\rho, \rho)$, which can be maximised numerically or graphically.

4.3.1 Finding $\hat{\beta}$

To maximise Equation (4.2) with respect to β we proceed in a similar way to Section 3.2. Firstly, we note that

$$\begin{aligned}(y_1 - X_1\beta) &= (y_1 - X_1\beta)^\top (y_1 - X_1\beta) \\ &= {y_1}^2 + \beta^\top {X_1}^\top X_1\beta - 2y_1X_1\beta.\end{aligned}$$

The derivative with respect to β is $2{X_1}^\top X_1\beta - 2{X_1}^\top y_1$, then,

$$\begin{aligned}\frac{\partial l}{\partial \beta} = \sum_{i=1}^{n_{AB(c)}} &\left(X_{AB}\Sigma^{-1}y_{AB_i} - {X_{AB}}^\top\Sigma^{-1}X_{AB}\beta\right) \\ &+ \sum_{i=1}^{n_{BA(c)}} \left(X_{BA}\Sigma^{-1}y_{BA_i} - {X_{BA}}^\top\Sigma^{-1}X_{BA}\beta\right) \\ &+ \frac{1}{\sigma^2}\left[\sum_{i=1}^{n_{AB(m)}} \left({X_{AB_1}}^\top y_{AB_{i1}} - {X_{AB_1}}^\top X_{AB_1}\beta\right)\right. \\ &\left. + \sum_{i=1}^{n_{BA(m)}} \left({X_{BA_1}}^\top y_{BA_{i1}} - {X_{BA_1}}^\top X_{BA_1}\beta\right)\right].\end{aligned}$$

Once again, we take the mean over all the patients in each group for period j, $\bar{y}_{AB_j}$ and $\bar{y}_{BA_j}$, to give

$$\begin{aligned}\frac{\partial l}{\partial \beta} = \frac{1}{\sigma^2}\big[&n_{AB(c)}{X_{AB}}^\top \boldsymbol{C^{-1}}\bar{y}_{AB(c)_j} - n_{AB(c)}{X_{AB}}^\top \boldsymbol{C^{-1}}X_{AB}\beta \\ &+ n_{BA(c)}{X_{BA}}^\top \boldsymbol{C^{-1}}\bar{y}_{BA(c)_j} - n_{BA(c)}{X_{BA}}^\top \boldsymbol{C^{-1}}X_{BA}\beta + n_{AB(m)}{X_{AB_1}}^\top \bar{y}_{AB(m)_1} \\ &- n_{AB(m)}{X_{AB_1}}^\top X_{AB_1}\beta + n_{BA(m)}{X_{BA_1}}^\top \bar{y}_{BA(m)_1} - n_{BA(m)}{X_{BA_1}}^\top X_{BA_1}\beta\big].\end{aligned}$$

Again, to maximise, we set ${}^{\partial l}/_{\partial\beta} = 0$, i.e.

$$\begin{aligned}&n_{AB(c)}{X_{AB}}^\top \boldsymbol{C^{-1}}\bar{y}_{AB(c)_j} + n_{BA(c)}{X_{BA}}^\top \boldsymbol{C^{-1}}\bar{y}_{BA(c)_j} + n_{AB(m)}{X_{AB_1}}^\top \bar{y}_{AB(m)_1} + n_{BA(m)}{X_{BA_1}}^\top \bar{y}_{BA(m)_1} \\ &= n_{AB(c)}{X_{AB}}^\top \boldsymbol{C^{-1}}X_{AB}\hat{\beta} + n_{BA(c)}{X_{BA}}^\top \boldsymbol{C^{-1}}X_{BA}\hat{\beta} + n_{AB(m)}{X_{AB_1}}^\top X_{AB_1}\hat{\beta} + n_{BA(m)}{X_{BA_1}}^\top X_{BA_1}\hat{\beta}\end{aligned} \tag{4.3}$$

We know from Section 3.2 that

$$n_{AB}{X_{AB}}^{\top}\Sigma^{-1}X_{AB}+n_{BA}{X_{BA}}^{\top}\Sigma^{-1}X_{BA}=\frac{1}{1-\rho^2}\begin{pmatrix}2(1-\rho)n^{+} & 0 & 0\\ 0 & 2(1+\rho)n^{+} & 2(1+\rho)n^{-}\\ 0 & 2(1+\rho)n^{-} & 2(1+\rho)n^{+}\end{pmatrix},$$

combined with

$$\begin{aligned}n_{AB(m)}{X_{AB_1}}^{\top}X_{AB_1}+n_{BA(m)}{X_{BA_1}}^{\top}X_{BA_1}&=n_{AB(m)}\begin{pmatrix}1\\1\\1\end{pmatrix}\begin{pmatrix}1 & 1 & 1\end{pmatrix}+n_{BA(m)}\begin{pmatrix}1\\1\\-1\end{pmatrix}\begin{pmatrix}1 & 1 & -1\end{pmatrix}\\&=\begin{pmatrix}n^{+}_{(m)} & n^{+}_{(m)} & n^{-}_{(m)}\\ n^{+}_{(m)} & n^{+}_{(m)} & n^{-}_{(m)}\\ n^{-}_{(m)} & n^{-}_{(m)} & n^{+}_{(m)}\end{pmatrix}\end{aligned}$$

gives the RHS of Equation (4.3):

$$\begin{aligned}&\left(n_{AB(c)}{X_{AB}}^{\top}\boldsymbol{C^{-1}}X_{AB}+n_{BA(c)}{X_{BA}}^{\top}\boldsymbol{C^{-1}}X_{BA}+n_{AB(m)}{X_{AB_1}}^{\top}X_{AB_1}+n_{BA(m)}{X_{BA_1}}^{\top}X_{BA_1}\right)\hat{\beta}\\&\quad=\left[\frac{1}{1-\rho^2}\begin{pmatrix}2(1-\rho)n^{+}_{(c)} & 0 & 0\\ 0 & 2(1+\rho)n^{+}_{(c)} & 2(1+\rho)n^{-}_{(c)}\\ 0 & 2(1+\rho)n^{-}_{(c)} & 2(1+\rho)n^{+}_{(c)}\end{pmatrix}+\begin{pmatrix}n^{+}_{(m)} & n^{+}_{(m)} & n^{-}_{(m)}\\ n^{+}_{(m)} & n^{+}_{(m)} & n^{-}_{(m)}\\ n^{-}_{(m)} & n^{-}_{(m)} & n^{+}_{(m)}\end{pmatrix}\right]\hat{\beta}\end{aligned}$$

We also know from Section 3.2:

$$X_{AB}^{\top}\Sigma^{-1}n_{AB}\bar{y}_{AB_j}+X_{BA}^{\top}\Sigma^{-1}n_{BA}\bar{y}_{BA_j}=\frac{1}{1-\rho^2}\left[n_{AB}\begin{pmatrix}(1-\rho)\bar{s}_{AB}\\(1+\rho)\bar{d}_{AB}\\(1+\rho)\bar{d}_{AB}\end{pmatrix}+n_{BA}\begin{pmatrix}(1-\rho)\bar{s}_{BA}\\(1+\rho)\bar{d}_{BA}\\-(1+\rho)\bar{d}_{BA}\end{pmatrix}\right],$$

If we combine this with

$$\begin{aligned}n_{AB(m)}{X_{AB_1}}^{\top}\bar{y}_{AB_1}+n_{BA(m)}{X_{BA_1}}^{\top}\bar{y}_{BA_1}&=n_{AB(m)}\begin{pmatrix}1\\1\\1\end{pmatrix}\bar{y}_{AB_1}+n_{BA(m)}\begin{pmatrix}1\\1\\-1\end{pmatrix}\bar{y}_{BA_1}\\&=\begin{pmatrix}n_{AB(m)}\bar{y}_{AB_1}+n_{BA(m)}\bar{y}_{BA_1}\\ n_{AB(m)}\bar{y}_{AB_1}+n_{BA(m)}\bar{y}_{BA_1}\\ n_{AB(m)}\bar{y}_{AB_1}-n_{BA(m)}\bar{y}_{BA_1}\end{pmatrix}\end{aligned}$$

we find the LHS of Equation (4.3) to be

$$\begin{aligned}&n_{AB(c)}{X_{AB}}^{\top}\boldsymbol{C^{-1}}\bar{y}_{AB(c)_j}+n_{BA(c)}{X_{BA}}^{\top}\boldsymbol{C^{-1}}\bar{y}_{BA(c)_j}\\&\qquad+n_{AB(m)}{X_{AB_1}}^{\top}\bar{y}_{AB(m)_1}+n_{BA(m)}{X_{BA_1}}^{\top}\bar{y}_{BA(m)_1}\\&=\frac{n_{AB(c)}}{1-\rho^2}\begin{pmatrix}(1-\rho)\bar{s}_{AB(c)}\\(1+\rho)\bar{d}_{AB(c)}\\(1+\rho)\bar{d}_{AB(c)}\end{pmatrix}+\frac{n_{BA(c)}}{1-\rho^2}\begin{pmatrix}(1-\rho)\bar{s}_{BA(c)}\\(1+\rho)\bar{d}_{BA(c)}\\-(1+\rho)\bar{d}_{BA(c)}\end{pmatrix}\\&\qquad+\begin{pmatrix}n_{AB(m)}\bar{y}_{AB(m)_1}+n_{BA(m)}\bar{y}_{BA(m)_1}\\ n_{AB(m)}\bar{y}_{AB(m)_1}+n_{BA(m)}\bar{y}_{BA(m)_1}\\ n_{AB(m)}\bar{y}_{AB(m)_1}-n_{BA(m)}\bar{y}_{BA(m)_1}\end{pmatrix}.\end{aligned}$$

And so,

$$\hat{\beta} = \left[\frac{1}{1-\rho^2} \begin{pmatrix} 2(1-\rho)n^+_{(c)} & 0 & 0 \\ 0 & 2(1+\rho)n^+_{(c)} & 2(1+\rho)n^-_{(c)} \\ 0 & 2(1+\rho)n^-_{(c)} & 2(1+\rho)n^+_{(c)} \end{pmatrix} + \begin{pmatrix} n^+_{(m)} & n^+_{(m)} & n^-_{(m)} \\ n^+_{(m)} & n^+_{(m)} & n^-_{(m)} \\ n^-_{(m)} & n^-_{(m)} & n^+_{(m)} \end{pmatrix} \right]^{-1} \left[\frac{n_{AB(c)}}{1-\rho^2} \begin{pmatrix} (1-\rho)\bar{s}_{AB(c)} \\ (1+\rho)\bar{d}_{AB(c)} \\ (1+\rho)\bar{d}_{AB(c)} \end{pmatrix} + \frac{n_{BA(c)}}{1-\rho^2} \begin{pmatrix} (1-\rho)\bar{s}_{BA(c)} \\ (1+\rho)\bar{d}_{BA(c)} \\ -(1+\rho)\bar{d}_{BA(c)} \end{pmatrix} + \begin{pmatrix} n_{AB(m)}\bar{y}_{AB(m)_1} + n_{BA(m)}\bar{y}_{BA(m)_1} \\ n_{AB(m)}\bar{y}_{AB(m)_1} + n_{BA(m)}\bar{y}_{BA(m)_1} \\ n_{AB(m)}\bar{y}_{AB(m)_1} - n_{BA(m)}\bar{y}_{BA(m)_1} \end{pmatrix} \right]. \tag{4.4}$$

Explicit inversion of the matrix in Equation (4.4) is awkward and so $\hat{\beta}_\rho$ is found numerically, as discussed further in Section 4.4 for the pain trial data.

4.3.2 Finding $\hat{\sigma}^2_\rho$

In order to find the equation for $\hat{\sigma}^2_\rho$ we must maximise

$$l(\hat{\beta}_\rho, \sigma^2, \rho) = -\frac{1}{2} n^+_{(c)} \log |C| - \frac{1}{2} \left(n^+_{(m)} + 2n^+_{(c)} \right) \log \sigma^2 - \frac{1}{2\sigma^2} A\left(\hat{\beta}_\rho, \rho \right).$$

Differentiation of this likelihood with respect to σ^2, gives

$$\frac{\partial l}{\partial \sigma^2} = \frac{\left(n^+_{(m)} + 2n^+_{(c)} \right)}{\sigma^2} - \frac{1}{\sigma^4} A\left(\hat{\beta}_\rho, \rho \right),$$

and $\partial l / \partial \sigma^2 = 0$ implies

$$\hat{\sigma}^2_\rho = \frac{A\left(\hat{\beta}_\rho, \rho \right)}{n^+_{(m)} + 2n^+_{(c)}}. \tag{4.5}$$

This estimator will also be discussed further in Section 4.4.

4.3.3 Profile Likelihood

Substituting the estimate $\hat{\sigma}^2_\rho$ into Equation (4.5) gives the likelihood function

$$\begin{aligned} l\left(\hat{\beta}_\rho, \hat{\sigma}^2_\rho, \rho \right) &= -\frac{1}{2} n^+_{(c)} \log |C| - \frac{1}{2} \left(n^+_{(m)} + 2n^+_{(c)} \right) \log \hat{\sigma}^2_\rho - \frac{1}{2\hat{\sigma}^2_\rho} A\left(\hat{\beta}_\rho, \rho \right) \\ &= -\frac{1}{2} n^+_{(c)} \log (1-\rho^2) - \frac{1}{2} \left(n^+_{(m)} + 2n^+_{(c)} \right) \log \hat{\sigma}^2_\rho - \frac{1}{2} \left(n^+_{(m)} + 2n^+_{(c)} \right). \end{aligned}$$

Disregarding the terms which are constant with respect to ρ, we have the profile likelihood

$$f(\rho) = -\frac{1}{2} \left[n^+_{(c)} \log (1-\rho^2) + \left(n^+_{(m)} + 2n^+_{(c)} \right) \log \hat{\sigma}^2_\rho \right]. \tag{4.6}$$

This profile likelihood function is dependent on estimates $\hat{\beta}\rho$ and $\hat{\sigma}^2_\rho$ found in Equations (4.4) and (4.5) respectively. As $\hat{\beta}$ is difficult to solve analytically, we have written a function in R which calculates $\hat{\beta}$ at fixed $\rho = r$ (Appendix A). A similar R function for the estimate $\hat{\sigma}^2$ at fixed $\rho = r$ (Appendix B). Using these, it is possible to write a third function (Appendix C), which produces the values of Equation (4.6) at $\rho = r$.

4.4 The Pain Trial: Maximum Likelihood Estimation for the Full Data

If we recall Table 1.1 in Section 1.3, we can see that 15 of the 82 patients have data missing in period two. Here, we will follow the method outlined in Section 4.3 to calculate a new treatment estimate for the pain trial which includes these un-paired period one readings. We recall the profile likelihood function for ρ in Equation (4.6);

$$f(\rho) = -\frac{1}{2}\left[n^+_{(c)} \log\left(1-\rho^2\right) + \left(n^+_{(m)} + 2n^+_{(c)}\right) \log \hat{\sigma}^2_\rho\right],$$

and its R function in Appendix C. Upon application of this R function to our pain trial data we are able to plot the profile likelihood, found in Figure 4.1. We can

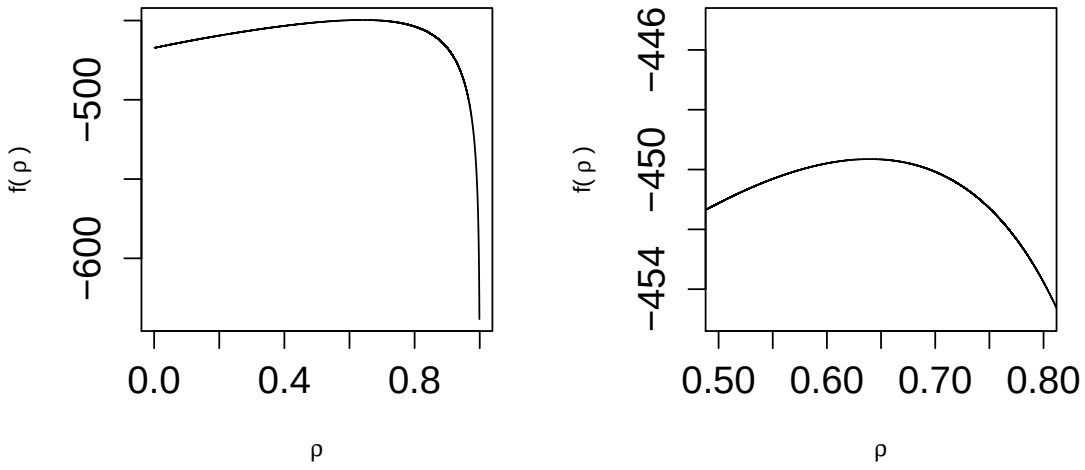

Figure 4.1: Graphical representation of the profile likelihood function for the pain trial data.

see that this likelihood is at a maximum somewhere between $\rho = 0.6$ and $\rho = 0.7$. Further numerical investigation returns $\hat{\rho} = 0.6385$. Stryhn and Christensen [2] describe finding a confidence interval for this estimator as inverting a profile

likelihood test. If we carry out an hypothesis test for ρ with $H_0 : \rho = r$ our test statistic, D, in terms of the profile likelihood, $f(\rho)$, is

$$D = 2\left\{f\left(\hat{\rho}\right) - f\left(r\right)\right\} \sim \chi^2_1.$$

So, the region in which we accept an estimate of ρ at the 0.05 significance level is

$$\left\{\rho \mid f(\hat{\rho}) - f(r) \leq \frac{1}{2}\chi^2_1\left(0.95\right)\right\},$$

i.e. $f(\rho) \geq f(\hat{\rho}) - 1.96$. For the pain trial data, this returns a 95% likelihood confidence interval of $(0.4761, 0.7566)$ illustrated in Figure 4.2. If we use this value

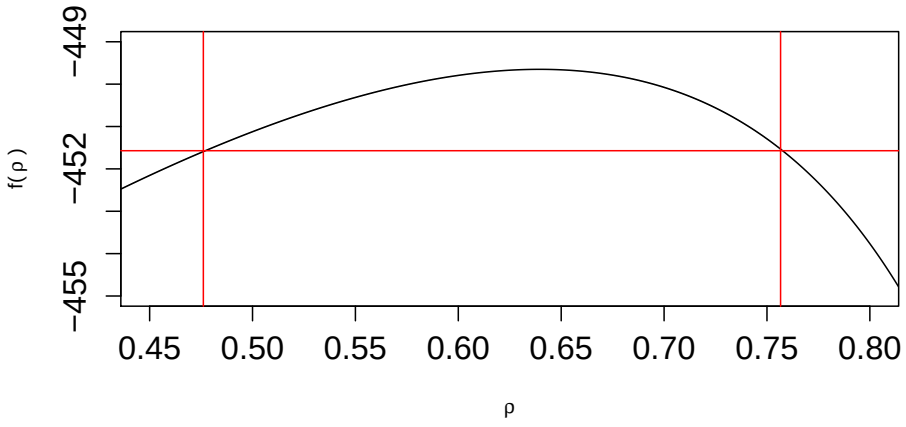

Figure 4.2: Graphical representation of the likelihood confidence interval for ρ.

$\hat{\rho} = 0.6385$ to calculate $\hat{\beta}$ and $\hat{\sigma}^2$ using the functions found in Appendices A and B we obtain

$$\begin{aligned} \hat{\beta} &= \begin{pmatrix} 57.8262 \\ 0.0142 \\ -3.1104 \end{pmatrix} \\ \hat{\sigma}^2 &= 529.0596. \end{aligned}$$

We have a new treatment difference estimate of $\hat{\tau} = -3.1104$, which is different to the estimate of $\hat{\tau} = -2.8543$, given by the complete case analyses.

Chapter 5

Linear Mixed-Effects Models

Linear mixed-effects models consist of a combination of fixed-effects and random-effects. These types of models are useful in the analysis of longitudinal data, as they account for multiple correlated readings on each subject but can also cope with unbalanced designs caused by missing readings or varying time points.

Laird and Ware [3] describe linear mixed-effects models as 'two stage random-effects models'. Stage one consists of the introduction of the fixed-effects of the model (e.g. the population parameters, individual effects and within-person variation) and stage two consists of the introduction of the random-effects (e.g. the between-person variation). So, for $j = 1, \ldots, n_i$ readings on $i = 1, \ldots, N$ individuals, the model for individual i is

$$\boldsymbol{y_i} = \boldsymbol{X_i}\boldsymbol{\beta} + \boldsymbol{Z_i}\boldsymbol{b_i} + \boldsymbol{\epsilon_i}, \tag{5.1}$$

where

- $\boldsymbol{y_i}$ is the n_i-vector of responses,
- $\boldsymbol{\beta}$ is a p-vector of unknown population parameters,
- $\boldsymbol{X_i}$ is an $[n_i \times p]$ design matrix linking $\boldsymbol{\beta}$ to $\boldsymbol{y_i}$,
- $\boldsymbol{b_i}$ is an q-vector of random effects,
- $\boldsymbol{Z_i}$ is an $[n_i \times q]$ design matrix linking $\boldsymbol{\beta}$ to $\boldsymbol{y_i}$,

and

- $\boldsymbol{\epsilon_i} \sim \mathcal{N}(0, \sigma^2 W_i)$ for $[n_i \times n_i]$ positive-definite covariance matrix W_i,
- $\boldsymbol{b_i} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ for $[q \times q]$ positive-definite covariance matrix $\boldsymbol{\Sigma}$.

5.1 The AB/BA Mixed-Effects Model

For the AB/BA crossover trial design, we are able to apply the general linear mixed-effects model in Equation (5.1) to the models provided in Equations (1.1)

and (1.2) in Section 1.2.3. The fixed-effects are the period and treatment effects and the random-effects are the patient effects so the $\boldsymbol{X_i\beta}$ term in Equation (5.1) corresponds to the

$$\begin{aligned} X_{AB}\beta &= \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \mu \\ \pi \\ \tau \end{pmatrix} \\ X_{BA}\beta &= \begin{pmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \pi \\ \tau \end{pmatrix}, \end{aligned}$$

found in Section 3.1. We also write the $\boldsymbol{Z_ib_i}$ as

$$Z_i\xi_i = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \xi_i,$$

for both groups AB and BA. Here $\xi_i \sim \mathcal{N}(0, \sigma_B^2)$. The linear mixed-effects models used in Sections 5.2 and 5.3 are

$$\begin{aligned} y_{AB_i} &= X_{AB}\beta + Z\xi_i + \epsilon_i, \\ &= \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \mu \\ \pi \\ \tau \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \xi_i + \epsilon_i \qquad (5.2) \\ y_{BA_i} &= X_{BA}\beta + Z\xi_i + \epsilon_i, \\ &= \begin{pmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \pi \\ \tau \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \xi_i + \epsilon_i. \qquad (5.3) \end{aligned}$$

5.2 The Pain Trial: Linear Mixed-Effects Regression on the Complete Cases

We have found that using maximum likelihood estimation on the complete cases derived an estimate for τ which was the same as the unpaired t-test estimator (Equation (2.1)) to use linear mixed-effects regression on the complete cases of the pain trial data we use R-function `lmer` from R package `lme4` [4]. This fits a model with fixed period (`Periodc`) and treatment (`Rxc`) effects and random patient (`Patc`) effects, like that in Equations (5.2) and (5.3). This provides the following output:

```
> complete=lmer(yc~Periodc+Rxc+(1|Patc),REML=F)
> summary(complete)
Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: yc ~ Periodc + Rxc + (1 | Patc)

      AIC       BIC    logLik  deviance
1195.5822 1210.0714 -592.7911 1185.5822
```

```
Random effects:
 Groups   Name        Variance Std.Dev.
 Patc     (Intercept) 339.4    18.42
 Residual             190.8    13.81
Number of obs: 134, groups: Patc, 67

Fixed effects:
            Estimate Std. Error t value
(Intercept)  59.2710     2.5475  23.266
Periodc       0.3443     1.1946   0.288
Rxc          -2.8543     1.1946  -2.389

Correlation of Fixed Effects:
        (Intr) Peridc
Periodc 0.000
Rxc     0.000  0.045
```

So the treatment estimate of

$$\hat{\tau} = -2.8543,$$

which is the same as that found using the unpaired t-test method and thus also the same as would be found using maximum likelihood estimation on the complete cases in the data. If we use R command `confint` on the linear mixed-effects regression we receive the output:

```
> confint(complete)
Computing profile confidence intervals ...
                 2.5 %     97.5 %
.sig01       14.629527 23.0043930
.sigma       11.768560 16.5294013
(Intercept) 54.205460 64.3364761
Periodc      -2.030938  2.7196340
Rxc          -5.229625 -0.4790531
```

and so our treatment estimate has a 95% confidence interval of $(-5.2296, -0.4791)$. This confidence interval is the same size and in roughly the same place as that found for the unpaired t-test, so these estimates are as precise as each other. We also note the period effect estimate of

$$\hat{\pi} = 0.3443,$$

with a 95% confidence interval of $(-2.0309, 2.7196)$. This is small, as was suggested in Section 2.2.2.

5.3 The Pain Trial: Linear Mixed-Effects Regression on the Full Data

As mentioned earlier in the chapter, linear mixed-effects regression copes well with unbalanced data caused by missing data, and so we can fit a linear mixed-effects

model in R to all of the pain data, complete and incomplete, with fixed period and treatment effects and random patient effect;

```
> missing=lmer(Y~Period+Rx+(1|Pat),REML=F)
> summary(missing)
Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: Y ~ Period + Rx + (1 | Pat)

      AIC       BIC    logLik  deviance
1332.1445 1347.1642 -661.0723 1322.1445

Random effects:
 Groups   Name        Variance Std.Dev.
 Pat      (Intercept) 338.7    18.40
 Residual             190.9    13.82
Number of obs: 149, groups: Pat, 82

Fixed effects:
            Estimate Std. Error t value
(Intercept) 57.82532    2.34742  24.634
Period       0.01507    1.17545   0.013
Rx          -3.10968    1.17230  -2.653

Correlation of Fixed Effects:
       (Intr) Period
Period -0.076
Rx      0.022  0.054
```

From this output, we can see that the maximum likelihood estimates for the fixed effects, $\hat{\beta}$ are

$$\hat{\beta} = \begin{pmatrix} \hat{\mu} \\ \hat{\pi} \\ \hat{\tau} \end{pmatrix} = \begin{pmatrix} 57.8253 \\ 0.0151 \\ -3.1097 \end{pmatrix}.$$

This has given us the same treatment effect as the full maximum likelihood estimation analysis, found in Section 4.4. The change to the estimate of the mean, $\hat{\mu}$, and period effect, $\hat{\pi}$, is negligible. If we use R function `confint` once more, we get:

```
> confint(missing)
Computing profile confidence intervals ...
                2.5 %     97.5 %
.sig01      14.758052 22.6355162
.sigma      11.774039 16.5176522
(Intercept) 53.156966 62.4699108
Period      -2.339166  2.3385246
Rx          -5.451721 -0.7902713
```

giving a 95% confidence interval of $(-5.4517, -0.7903)$ for $\hat{\tau}$. Recalling the definitions of σ_W^2 and σ_B^2 from Sections 1.2.3 and 3.1 respectively, the estimate for the correlation, $\hat{\rho}$, is parameterised as

$$\begin{aligned} \rho &= \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2}, \\ &= \frac{338.7}{338.7 + 190.9}, \\ &= 0.6395. \end{aligned}$$

This estimate is the same as the estimate found in Section 4.4.

Chapter 6

Selection and Pattern Mixture

6.1 Types of Missing Data

The full analysis on the pain trial data (Sections 4.4 and 5.3) obtained treatment effect estimates different to those in Sections 2.1.2, 2.2.2 and 5.2, where analysis was only on the complete cases. We have argued that, by omitting patients with missing period two data, we miss out on information important to the analysis of the study. However, Rubin [5] states that also ignoring what causes data to be missing leads to incorrect and biased results in some cases. He explains that missing data are either

- missing completely at random (MCAR), where there is no relationship between missingness and any of the missing or observed data points,
- missing at random (MAR), where the missingness of a data point is not related to the value of the missing data point, but is related to some of the observed data, or
- missing not at random (MNAR), where the missingness of the data depends on the unseen observations.

Previous analyses in this dissertation have assumed data to be MAR or MCAR. This assumption is potentially flawed, as experience of a treatment may cause a patient to drop out, which means the missing data is MNAR. If we are to avoid biased estimates in this case, we need a model which accounts for the missingness process.

Little [6] develops an analysis of MNAR data by introducing a 'missing-data indicator', $R \in \{0, 1\}$. Where Y_1 is a reading in period one and Y_2 a reading in period two, we denote the indicator to be $R = 1$ for a patient with both Y_1 and Y_2 observed and $R = 0$ for a patient with only Y_1 observed. Then the joint distributions of the data are either

$$Pr\left(Y_1, Y_2, R = 1\right) \tag{6.1}$$

or

$$Pr\left(Y_1, R = 0\right) \tag{6.2}$$

for the complete and incomplete cases respectively. Thus the log-likelihood, formed in the same way as in Section 3.1, is

$$l = \sum_{AB(c)} \log Pr\left(Y_1, Y_2, R = 1\right) + \sum_{BA(c)} \log Pr\left(Y_1, Y_2, R = 1\right) + \sum_{AB(m)} \log Pr\left(Y_1, R = 0\right) + \sum_{BA(m)} \log Pr\left(Y_1, R = 0\right).$$

$Pr\left(R = 1 \mid Y_1, Y_2\right)$ models how R depends on Y_1 and Y_2. As we consider missingness to be a result of the experience the patients have on the treatment, it is reasonable to assume that the probability should potentially depend on both Y_1 and Y_2. For simplicity, we model this situation using a linear predictor, $\theta_0 + \theta_1 Y_1 + \theta_2 Y_2$, and set

$$Pr\left(R = 1 \mid Y_1, Y_2\right) = F\left(\theta_0 + \theta_1 Y_1 + \theta_2 Y_2\right), \tag{6.3}$$

where $F\left(\cdot\right)$ denotes a function which maps $\mathbb{R} \to [0, 1]$. Note that, this model has

- $\theta_1 = \theta_2 = 0$ when the data are MCAR, and
- $\theta_2 = 0$ when the data are MAR.

Typically, logistic regression would be used for this model, however, we will use probit regression for mathematical tractability. So Equation (6.3) becomes

$$Pr\left(R = 1 \mid Y_1, Y_2\right) = \Phi\left(\theta_0 + \theta_1 Y_1 + \theta_2 Y_2\right), \tag{6.4}$$

where $\Phi\left(\cdot\right)$ denotes the standard normal distribution function.

6.2 Selection Models

There are two ways of factorising the joint distributions found in Equations (6.1) and (6.2); *selection models* and *pattern mixture models*. If we use a selection model, we have

$$Pr\left(Y_1, Y_2, R = 1\right) = Pr\left(R = 1 \mid Y_1, Y_2\right) Pr\left(Y_1, Y_2\right),$$

and

$$Pr\left(Y_1, R = 0\right) = Pr\left(R = 0 \mid Y_1\right) Pr\left(Y_1\right).$$

The likelihood function then becomes

$$l = \sum_{AB(c)} \log Pr\left(R = 1 \mid Y_1, Y_2\right) Pr\left(Y_1, Y_2\right) + \sum_{BA(c)} \log Pr\left(R = 1 \mid Y_1, Y_2\right) Pr\left(Y_1, Y_2\right) + \sum_{AB(m)} \log Pr\left(R = 0 \mid Y_1\right) Pr\left(Y_1\right) + \sum_{BA(m)} \log Pr\left(R = 0 \mid Y_1\right) Pr\left(Y_1\right). \tag{6.5}$$

We recall from Section 3.1 that Y_1 and Y_2 are bivariate normal random variables with variance $\sigma^2 C$ and means

$$\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \mu + \pi + \tau \\ \mu - \pi - \tau \end{pmatrix}$$

for group AB, and

$$\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix} = \begin{pmatrix}\mu+\pi-\tau\\ \mu-\pi+\tau\end{pmatrix}$$

for group BA. We will use μ_1 to denote the mean of the period one readings and μ_2 for period two readings, but we must keep in mind that these differ for the two treatment sequences, AB and BA. So, using this and Equation (6.4), we evaluate the terms corresponding to the complete cases in Equation (6.5) to be

$$Pr\left(R=1 \mid Y_1, Y_2\right) Pr\left(Y_1, Y_2\right) = \Phi\left(\theta_0+\theta_1 Y_1+\theta_2 Y_2\right) \phi_2\left(Y_1, Y_2\right),$$

where $\phi_2\left(Y_1, Y_2\right)$ denotes a standard bivariate normal. While this is relatively straightforward, evaluating the terms which correspond to the incomplete cases is more difficult.

6.2.1 Likelihood Contribution from the Incomplete Cases

We are unable to evaluate $Pr\left(R=0 \mid Y_1\right) Pr\left(Y_1\right)$ in the same way as the complete cases, as the *missingness* of Y_2 could depend on the *value* of Y_2. Hence, we can write

$$Pr\left(R=0 \mid Y_1\right) = \int Pr\left(R=0 \mid Y_1, Y_2\right) Pr\left(Y_2 \mid Y_1\right)\ dY_2.$$

Here,

$$\begin{aligned} Pr\left(R=0 \mid Y_1, Y_2\right) &= 1-Pr\left(R=1 \mid Y_1, Y_2\right) \\ &= \Phi\left(-\theta_0-\theta_1 Y_1-\theta_2 Y_2\right) \end{aligned}$$

and

$$Pr\left(Y_2 \mid Y_1\right) = \phi\left(Y_2 \mid Y_1\right)$$

is the conditional distribution of Y_2 given Y_1. This is normal with mean

$$\begin{aligned} E\left(Y_2 \mid Y_1\right) &= \mu_2+\rho\frac{\sigma_2}{\sigma_1}\left(Y_1-\mu_1\right) \\ &= \left(\mu_2-\rho\mu_1\right)+\rho Y_1 \\ &= \mu_{cond}, \end{aligned}$$

as $\sigma_1=\sigma_2$ in our model. Its variance is

$$\begin{aligned} Var\left(Y_2 \mid Y_1\right) &= \sigma^2\left(1-\rho^2\right) \\ &= \sigma^2_{cond}. \end{aligned}$$

Once again, we note that the conditional means differ between the two treatment sequences. Now,

$$Pr\left(R=0 \mid Y_1\right) = \int \Phi\left(-\theta_0-\theta_1 Y_1-\theta_2 Y_2\right) \phi\left(Y_2 \mid \mu_{cond}, \sigma^2_{cond}\right)\ dY_2. \tag{6.6}$$

In order to proceed, we need to introduce the *skew-normal distribution.*

6.3 Extended Skew Normal Distribution

Azzalini [7] denotes a random variable Z, which is skew-normal with parameter λ, to be $Z \sim \mathcal{SN}(\lambda)$ if Z has density

$$f(z;\lambda) = 2\phi(z)\,\Phi(\lambda z), \tag{6.7}$$

where $z \in \mathbb{R}$. Here,

- $\phi(z)$ denotes the $\mathcal{N}(0,1)$ density function,
- $\Phi(z)$ denotes the $\mathcal{N}(0,1)$ cumulative distribution function,
- λ denotes the skewness parameter, which has $\lambda \in (-\infty, \infty)$.

We note that $\lambda = 0$ corresponds to the $\mathcal{N}(0,1)$ probability density function. Often, realistic data is unimodal with some skewness and this distribution is appropriate for the analysis of such data. The skew-normal distribution can also be extended to the multivariate case. A random variable, U, is said to be distributed as an extended skew-normal if it has density

$$f(u) = \frac{\phi_p(u \mid \mu, \Omega)\,\Phi\left(\nu + \alpha^\top(u-\mu)\right)}{\Phi\left(\nu/\sqrt{1+\alpha^\top\Omega\alpha}\right)}, \tag{6.8}$$

where

- μ and α are p-dimensional vectors,
- ν is a scalar,
- $\Phi(\cdot)$ is the univariate standard normal CDF and,
- $\phi_p(\cdot \mid \mu, \Omega)$ is the PDF of a p-dimensional normal variable with mean μ and covariance Ω.

See Ho *et al.*[8] for details. As Equation (6.8) is a density, integrating it with respect to u equals one. This implies

$$\int \phi_p(u \mid \mu, \Omega)\,\Phi\left(\nu + \alpha^\top(u-\mu)\right)\,du = \Phi\left(\frac{\nu}{\sqrt{1+\alpha^\top\Omega\alpha}}\right), \tag{6.9}$$

which we can use to evaluate Equation (6.6);

$$Pr(R = 0 \mid Y_1) = \int \Phi(-\theta_0 - \theta_1 Y_1 - \theta_2 Y_2)\,\phi\left(Y_2 \mid \mu_{cond}, \sigma^2_{cond}\right)\,dY_2.$$

If we make the identifications $u = Y_2$, $\alpha = -\theta_2$, $\mu = \mu_{cond}$ and $\Omega = \sigma^2_{cond}$,

$$\begin{aligned}
\nu - \theta_2(Y_2 - \mu_{cond}) &= -\theta_0 - \theta_1 Y_1 - \theta_2 Y_2 \\
\nu + \theta_2\mu_{cond} &= -\theta_0 - \theta_1 Y_1 \\
\nu &= -\theta_0 - \theta_1 Y_1 - \theta_2\mu_{cond}.
\end{aligned}$$

So, using Equation (6.9), we get

$$\begin{aligned} Pr\left(R=0 \mid Y_1\right) &= \Phi\left(\frac{-\theta_0-\theta_1 Y_1-\theta_2\mu_{cond}}{\sqrt{1+{\theta_2}^2\sigma^2_{cond}}}\right) \\ &= \Phi\left(\frac{-\theta_0-\theta_1 Y_1-\theta_2\rho Y_1-\theta_2\left(\mu_2-\rho\mu_1\right)}{\sqrt{1+{\theta_2}^2\sigma^2\left(1-\rho^2\right)}}\right). \end{aligned}$$

6.4 Forming the Log-Likelihood

Equation (6.5) illustrated the general form of the log-likelihood for our MNAR data;

$$\begin{aligned} l = &\sum_{AB(c)} \log Pr\left(R=1 \mid Y_1,Y_2\right) Pr\left(Y_1,Y_2\right) + \sum_{BA(c)} \log Pr\left(R=1 \mid Y_1,Y_2\right) Pr\left(Y_1,Y_2\right) \\ &+ \sum_{AB(m)} \log Pr\left(R=0 \mid Y_1\right) Pr\left(Y_1\right) + \sum_{BA(m)} \log Pr\left(R=0 \mid Y_1\right) Pr\left(Y_1\right). \end{aligned}$$

This now becomes

$$\begin{aligned} l = &\sum_{AB(c)} \log\left[\Phi\left(\theta_0+\theta_1 Y_1+\theta_2 Y_2\right)\phi_2\left(Y_1,Y_2\right)\right] \\ &+ \sum_{BA(c)} \log\left[\Phi\left(\theta_0+\theta_1 Y_1+\theta_2 Y_2\right)\phi_2\left(Y_1,Y_2\right)\right] \\ &+ \sum_{AB(m)} \log\left[\Phi\left(\frac{-\theta_0-\theta_1 Y_1-\theta_2\left(\mu_2-\rho\mu_1\right)-\theta_2\rho Y_1}{\sqrt{1+{\theta_2}^2\sigma^2\left(1-\rho^2\right)}}\right)\phi\left(Y_1\right)\right] \\ &+ \sum_{BA(m)} \log\left[\Phi\left(\frac{-\theta_0-\theta_1 Y_1-\theta_2\left(\mu_2-\rho\mu_1\right)-\theta_2\rho Y_1}{\sqrt{1+{\theta_2}^2\sigma^2\left(1-\rho^2\right)}}\right)\phi\left(Y_1\right)\right]. \end{aligned}$$

If we rewrite the above as

$$\begin{aligned} l = &\sum_{AB(c)} \log\left[\phi_2\left(Y_1,Y_2\right)\right] + \sum_{BA(c)} \log\left[\phi_2\left(Y_1,Y_2\right)\right] \\ &+ \sum_{AB(m)} \log\left[\phi\left(Y_1\right)\right] + \sum_{BA(m)} \log\left[\phi\left(Y_1\right)\right] \\ &+ \sum_{AB(c)} \log\left[\Phi\left(\theta_0+\theta_1 Y_1+\theta_2 Y_2\right)\right] + \sum_{BA(c)} \log\left[\Phi\left(\theta_0+\theta_1 Y_1+\theta_2 Y_2\right)\right] \\ &+ \sum_{AB(m)} \log\left[\Phi\left(\frac{-\theta_0-\theta_1 Y_1-\theta_2\left(\mu_2-\rho\mu_1\right)-\theta_2\rho Y_1}{\sqrt{1+{\theta_2}^2\sigma^2\left(1-\rho^2\right)}}\right)\right] \\ &+ \sum_{BA(m)} \log\left[\Phi\left(\frac{-\theta_0-\theta_1 Y_1-\theta_2\left(\mu_2-\rho\mu_1\right)-\theta_2\rho Y_1}{\sqrt{1+{\theta_2}^2\sigma^2\left(1-\rho^2\right)}}\right)\right], \end{aligned} \tag{6.10}$$

we notice the first four terms are the same as the log-likelihood function for the full MAR analysis, which we will denote l_{uncorr}, found in Equation (4.1);

$$\begin{aligned} l_{uncorr} = &-\frac{n_{AB(c)}}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n_{AB(c)}}\left(y_{AB_i} - X_{AB}\beta\right)^\top \Sigma^{-1}\left(y_{AB_i} - X_{AB}\beta\right) \\ &- \frac{n_{BA(c)}}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n_{BA(c)}}\left(y_{BA_i} - X_{BA}\beta\right)^\top \Sigma^{-1}\left(y_{BA_i} - X_{BA}\beta\right) \\ &- \frac{1}{2}n_{AB(m)}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^{n_{AB(m)}}\frac{\left(y_{AB_{i1}} - X_{AB_1}\beta\right)^2}{\sigma^2} \\ &- \frac{1}{2}n_{BA(m)}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^{n_{BA(m)}}\frac{\left(y_{BA_{i1}} - X_{BA_1}\beta\right)^2}{\sigma^2}. \end{aligned}$$

The use of a selection model has given us the log-likelihood function l_{uncorr}, plus a correction term, l_{corr}, which is dependent on θ_2. Fitting this model would require maximisation of Equation (6.10) to find the relevant parameters. We notice that, for $\theta_2 = 0$ (i.e. the MAR case), the correction term becomes

$$\begin{aligned} l_{corr} = &\sum_{AB(c)}\log\left[\Phi\left(\theta_0 + \theta_1 Y_1\right)\right] + \sum_{BA(c)}\log\left[\Phi\left(\theta_0 + \theta_1 Y_1\right)\right] \\ &+ \sum_{AB(m)}\log\left[\Phi\left(-\theta_0 - \theta_1 Y_1\right)\right] + \sum_{BA(m)}\log\left[\Phi\left(-\theta_0 - \theta_1 Y_1\right)\right]. \end{aligned}$$

This term is not dependent on the model parameters and so differentiation of $l_{uncorr} + l_{corr}$ with respect to those parameters will yield the same estimators as maximum likelihood estimation on the full data, found in Chapter 4.

Chapter 7

Discussion

	$\hat{\tau}$	95% confidence interval for $\hat{\tau}$	$\hat{\pi}$
Unpaired t-test	-2.8543	$(-5.2679, -0.4408)$	
Paired t-test	-2.8698	$(-5.2719, -0.4676)$	
LMER: Complete Cases	-2.8543	$(-5.2296, -0.4791)$	0.3443

Table 7.1: The estimates and confidence intervals for the complete case analyses.

	$\hat{\tau}$	95% confidence interval for $\hat{\tau}$	$\hat{\pi}$
Likelihood Estimation	-3.1104		0.0142
LMER: Incomplete Cases	-3.1097	$(-5.4517, -0.7903)$	0.0151

Table 7.2: The estimates and confidence intervals for the full analyses.

The above tables contain the treatment and period effect estimates obtained throughout this dissertation for the pain trial data. Table 7.1 illustrates the methods of analysis where we only considered the complete cases. We can see from this table that treatment estimates differed very little between these methods of analysis. We would expect a difference between a treatment estimate obtained from the unpaired t-test compared to that of the paired t-test only if there is a significant period effect, but linear mixed-effects regression revealed the period effect to be close to zero (0.3443), which explains the lack of change between the treatment estimates from the two types of t-test. Confidence intervals for these treatment estimates are very similar in size which suggests, for this data set, the three methods provide estimates that are very similar in precision.

Table 7.2 shows the treatment estimates for the full analyses. We notice that the inclusion of this additional data has made a difference to the estimators; the treatment difference estimate has changed and could potentially be more accurate. However, the confidence interval given by the full linear mixed-effects regression analysis is similar in size to those for the three methods analysing only the complete cases, and so this estimate may be more accurate, but not more precise. The period

effect has also changed but is still close to zero.

The main difference between the methods of analysis in this dissertation is the data used. Treatment differences in the data used from the pain trial did not differ largely between the complete case analysis and full analysis, but in a trial where the proportion of drop out is high, the full analysis could be more accurate and reliable. Chapter 6 includes an analysis which uses even more information; the process causing the data to be missing. The MNAR approach to crossover trial data may seem most desirable in a statistical sense, but a thought to the practical applications of the trial may actually reveal the complete case analysis to be most pertinent. For example, Ho *et al.* [8] consider a treatment with a side-effect which causes patients to be unable to drive and so some patients refuse that treatment. There is then a subset of patients that this treatment would be administered to, and thus the treatment analysis is only relevant to those in that subset.

The concern, when faced with crossover data containing drop out, is that a patient has dropped out because they are intolerant to the treatment they were on. The modified log-likelihood in Equation (6.10) attempts to account for the missingness in the trial in a more sophisticated way. In a parallel group trial (see Section 1.1.1), a patient may actually receive a treatment different from the one they were originally assigned by randomisation, because it is considered unsuitable for them by a member of medical staff. In this case the two treatment groups are no longer comparable and analysis is no longer unbiased. Often *analysis by intention to treat* [9, p.188] is a solution to this. Here, analysis is no longer necessarily of how the patients were treated but how they were intended to be treated at the beginning of the study. Conclusions may then, for example, be in favour of surgery but knowledgeable that some patients may be inoperable and so those will be given radiotherapy. If, in our crossover trial, the concern that a patient dropping out of the study is because they cannot tolerate one of the treatments, then it becomes questionable whether those patients are relevant to the comparison under analysis by intention to treat. Then, the most pertinent analysis may only consider those who took both treatments, i.e. the complete cases.

Chapter 8

Conclusions

We have seen that the complete case analysis and full analysis have given different estimates for the pain data. From a statistical viewpoint we are able to argue that, because the full analysis provides us with more information, the estimators obtained are more reliable and representative of the total population. However, considering the basis on which the data are included, an MAR analysis is possibly naive. Even better would be the selection model analysis of the data, which would include even more information about the nature of the missingness of the data. However, we have also discussed the practical application of crossover data analysis and could argue that, while the analysis of the all the data and the consideration of the missingness process may provide an estimate more representative of the population, it may end up being less *relevant* to those who will eventually take the treatment.

Further work could be done to maximise the likelihood function for the selection model, found in Equation (6.10), and apply it to the pain trial data to investigate the impact of the MNAR analysis on the estimates. It may also be interesting to investigate the pattern mixture factorisation of the joint distributions found in Equations (6.1) and (6.2) to compare the computation and outcomes with the selection model factorisation.

Appendix A

R function: betahat

```
betahat=function(r,YAB,YBA,YABm,YBAm){

  nABc=length(YAB[,1])
  nBAc=length(YBA[,1])
  nABm=length(YABm)
  nBAm=length(YBAm)

  yABc1=mean(YAB[,1])
  yBAc1=mean(YBA[,1])
  yABc2=mean(YAB[,2])
  yBAc2=mean(YBA[,2])

  yABm1=mean(YABm)
  yBAm1=mean(YBAm)

  nplusc=nABc+nBAc
  nplusm=nABm+nBAm
  nminusc=nABc-nBAc
  nminusm=nABm-nBAm

  sABc=yABc1+yABc2
  sBAc=yBAc1+yBAc2
  dABc=yABc1-yABc2
  dBAc=yBAc1-yBAc2

  m1=matrix(data=c(2*(1-r)*nplusc,0,0,0,2*(1+r)*nplusc,2*(1+r)*nminusc,
                   0,2*(1+r)*nminusc,2*(1+r)*nplusc),nrow=3,ncol=3,byrow=T)
  m2=matrix(data=c(nplusm,nplusm,nminusm,nplusm,nplusm,nminusm,
                   nminusm,nminusm,nplusm),nrow=3,ncol=3,byrow=T)
  m3=(1/(1-r^(2)))*m1+m2
  inv=solve(m3)

  m4=matrix(data=c((1-r)*sABc,(1+r)*dABc,(1+r)*dABc),nrow=3,ncol=1,byrow=T)
  m5=matrix(data=c((1-r)*sBAc,(1+r)*dBAc,-(1+r)*dBAc),nrow=3,ncol=1,byrow=T)
  m6=matrix(data=c((nABm*yABm1)+(nBAm*yBAm1),(nABm*yABm1)+(nBAm*yBAm1),
                   (nABm*yABm1)-(nBAm*yBAm1)),nrow=3,ncol=1,byrow=T)
  LHS=(nABc/(1-r^(2)))*m4+(nBAc/(1-r^(2)))*m5+m6

  betahat=inv%*%LHS
  return(betahat)
}
```

Appendix B

R function: sigsqhat

```
sumdiag=function(r,Y){

  C=matrix(c(1,r,r,1),ncol=2,nrow=2,byrow=T)
  Cinv=solve(C)
  trans=t(Y)
  p=Y%*%Cinv%*%trans

  m=sum(diag(p))
  return(m)
}
sigsqhat=function(r,bhat,YAB,YBA,YABm,YBAm){

  nABc=length(YAB[,1])
  nBAc=length(YBA[,1])
  nABm=length(YABm)
  nBAm=length(YBAm)

  nplusc=nABc+nBAc
  nplusm=nABm+nBAm
  nminusc=nABc-nBAc
  nminusm=nABm-nBAm

  XAB=matrix(c(1,1,1,1,-1,-1),nrow=2,ncol=3,byrow=T)
  muhatAB=XAB%*%bhat
  XBA=matrix(c(1,1,-1,1,-1,1),nrow=2,ncol=3,byrow=T)
  muhatBA=XBA%*%bhat

  yABmat=matrix(nrow=length(YAB[,1]),ncol=2)
  yABmat[,1]=YAB[,1]-muhatAB[1,]
  yABmat[,2]=YAB[,2]-muhatAB[2,]

  yBAmat=matrix(nrow=length(YBA[,1]),ncol=2)
  yBAmat[,1]=YBA[,1]-muhatBA[1,]
  yBAmat[,2]=YBA[,2]-muhatBA[2,]

  XAB1=XAB[1,]
  muABm1=XAB1%*%bhat

  XBA1=XBA[1,]
  muBAm1=XBA1%*%bhat

  yBAsum=sum((YBAm-muBAm1)^2)
  yABsum=sum((YABm-muABm1)^2)

  A=sumdiag(r,yABmat)+sumdiag(r,yBAmat)+yABsum+yBAsum

  sigsqhat=A/(nplusm+2*nplusc)
  return(sigsqhat)
}
```

Appendix C

R function: prlike

```
prlike=function(r,YAB,YBA,YABm,YBAm){

  nABc=length(YAB[,1])
  nBAc=length(YBA[,1])
  nABm=length(YABm)
  nBAm=length(YBAm)

  nplusc=nABc+nBAc
  nplusm=nABm+nBAm

  bhat=betahat(r,YAB,YBA,YABm,YBAm)
  ssq=sigsqhat(r,bhat,YAB,YBA,YABm,YBAm)

  pr=-(nplusc*log(1-(r)^(2))+(nplusm+2*nplusc)*log(ssq))/2
  return(pr)
}
```

References

[1] Frank B, Serpell MG, Hughes J, Matthews JNS, Kapur D. Comparison of analgesic effect and patient tolerability of nabilone and dihydrocodeine for chronic neuropathic pain: randomised, crossover, double blind study. *British Medical Journal* 2008; **336**:199–201.

[2] Stryhn H, Christensen J. Confidence intervals by the profile likelihood method, with applications in veterinary epidemiology. *Proceedings of the 10th International Symposium on Veterinary Epidemiology and Economics*, 2003.

[3] Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38**:963–974.

[4] Bates D, Maechler M, Bolker B, Walker S. *lme4: Linear mixed-effects models using Eigen and S4* 2014. URL `http://CRAN.R-project.org/package=lme4`, r package version 1.0-6.

[5] Rubin DB. Inference and missing data. *Biometrika* 1976; **63**:581–592.

[6] Little RJA. A class of pattern-mixture models for normal incomplete data. *Biometrika* 1994; **81**(3):471–483.

[7] Azzalini A, Dalla Valle A. The multivariate skew-normal distribution. *Biometrika* 1996; **83**(4):715–726.

[8] Ho WK, Matthews JN, Henderson R, Farewell D, Rodgers LR. Dropouts in the AB/BA crossover design. *Statistics in Medicine* 2012; **31**(16):1675–1687.

[9] Matthews JNS. *Introduction to Randomized Controlled Clinical Trials, Second Edition*. CRCPress/Chapman and Hall: Boca Raton, 2006.