The Place of Crossover Designs in Infertility Trials

Boikanyo Makubate

A Dissertation Submitted to the

University of Glasgow

for the degree of

Doctor of Philosophy in Medical Statistics

Department of Statistics

June 2009

Abstract

Many people regard infertility as an unsuitable condition in which to run cross-over designs. Because if treatments are successful, then some women will become pregnant in the first period of treatment and if they become pregnant in the first period they are very unlikely to be treated in the second period. If data from only one period are available, then it will not be possible to perform the within-patient comparison. Hence, it could be argued, such designs are inherently unsuitable since they have a built-in tendency to produce missing data. To sum up this point of view, cross-over designs in infertility are likely to produce fewer data than one would wish and should be avoided.

We see things differently, however. Suppose a parallel design is employed. If a couple is entered onto a parallel group design in infertility, they will have been allocated to one or the other treatment only. If the woman fails to achieve pregnancy, having been given that treatment, what could then be more natural than to offer the couple the chance of trying another? If another treatment is tried, then will it not be appropriate to record the outcome? Hence, in the worst case one will have all the data one would have from a parallel group trial but in practice one is likely to have more. How can more data be worse than less? Thus, to sum up our point of view, a cross-over trial is likely to produce more data than one would otherwise have had and should be encouraged.

A debate along these lines has been running for some years now (6; 24; 27; 52; 64; 74; 75; 94), with some promoting and occasionally running cross-over trials and others criticising them for doing so. In this thesis we use the logistic random effects model to illustrate that the message that the crossover design should be avoided is not the correct one. Rather, when using the crossover design one should be sure to analyze it correctly.

The study has found that treatment estimates obtained by allowing women to get pregnant twice has lower standard errors than treatment estimates obtained by conducting the realistic infertility trials. In the scenario involving no period effects the two treatment estimates are not biased. In the scenario involving period effect, the standard errors of the treatment estimate obtained in the realistic data increases rapidly, while the standard errors of treatment estimate obtained by allowing women to get pregnant twice are not dissimilar from the standard errors obtained in the scenario involving no period effect.

There is nothing wrong in conducting crossover designs in infertility provided appropriate statistical methods are employed. With the infertility crossover data set we can obtain not only conditional treatment estimates but also marginal estimates. Whereas in the parallel design we can only obtain marginal estimates. The study has found that if the treatment say $\hat{\beta}$, can be obtained using parallel design data set, then surely, $\hat{\beta}$ can be obtained using the crossover design data set, but not vice versa. Moreover the treatment estimate $\hat{\beta}$ obtained using the crossover data set will be more consistent than the treatment estimate obtained using the parallel design. We recommend that crossover designs be used in infertility trials because it will surely benefit couples as couples will be have the opportunity to try both treatments.

Acknowledgements

I am grateful to the University of Glasgow and the Association of Commonwealth Universities for providing resources and funding throughout this research and to all the staff and postgraduate students in the department of Statistics for making it a welcoming and friendly place to work.

I would like to thank Professor Stephen Senn and Professor John McColl, my supervisors for their support and guidance during the course of this research. I am heavily indebted to Professor Stephen Senn whose comments and guidance made completion of this degree possible. I could not have completed this thesis without Professor Stephen Senn especially during my struggling moments. Stephen managed to secure extra funding for myself from both our Department (Statistics Department, University Of Glasgow) and from Dr Andy Lawton (Boehringer Ingelheim) when the funding from the Association of Commonwealth Universities come to an end. I wish to thank Dr Andy Lawton (Boehringer Ingelheim) for the funds he provided during my struggling moments. Thank you very much Boehringer Ingelheim and thank you very much Stephen.

Special thanks to my wife Taboka and my three kids Ona, Tshephi and Bo for their support and encouragement. My mum, my brother (Luzar) and his family, my sister (Gladys) and her family, and my nephew (Maikano also known

as Kanda-Bongo-Man) all have given me the strength to continue.

Finally, I would like to dedicate this thesis to the memory of my sister Thatayaone Makubate who died on 2nd September 2001. Without the love and support of my caring sister, I could not have done as much in academics as I have.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Infertility

A woman is said to be infertile if she is unable to conceive while a man is infertile if he is unable to impregnate. A couple is said to be infertile if either a woman or a man is infertile. There are two types of infertility; primary and secondary infertility. Primary infertility is when the couples have never had a child while secondary infertility means that the couples have children but a medical, emotional or physical condition is currently making conception difficult or not possible. Infertility may also refer to the state of a woman who is unable to carry a pregnancy to full term. Pregnancy is the result of a complex chain of events. In order for a woman to get pregnant: Firstly, a woman must release an egg from one of her ovaries (ovulation). This egg must go through a fallopian tube and enter the uterus (womb). Secondly a man's sperm must penetrate and fertilize the egg along the way. The fertilized egg must attach to the inside of the uterus (101; 102). In order to understand how pregnancy occurs we need to have a basic understanding of the female reproductive system, which is illustrated in

Fig 1.1 below. More details on the female reproductive system can be found in (10) and (89).
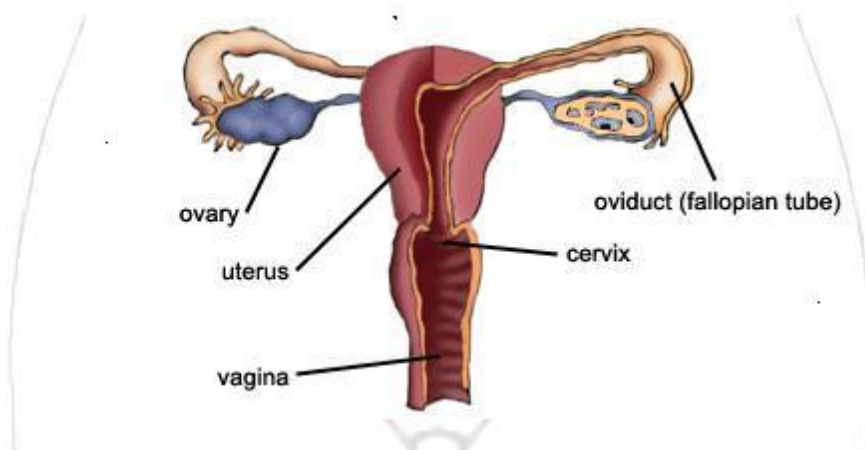
Figure 1.1. Female reproductive system

A woman does not become pregnant immediately after having sex. Fig 1.2 illustrates the process of ovulation.
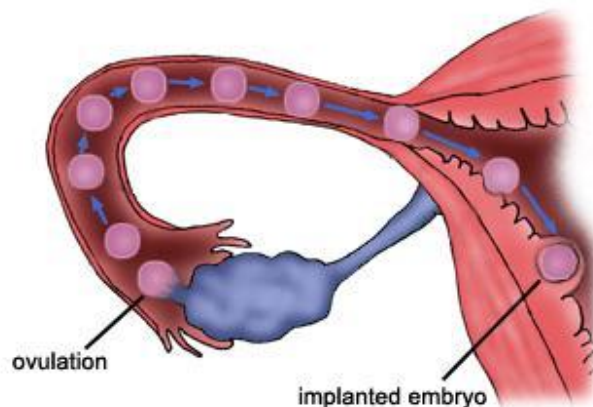
Figure 1.2. Illustrates the process of ovulation

A woman has two ovaries, and each month an egg is released from one of them (102). This process is called ovulation, and it occurs approximately 14
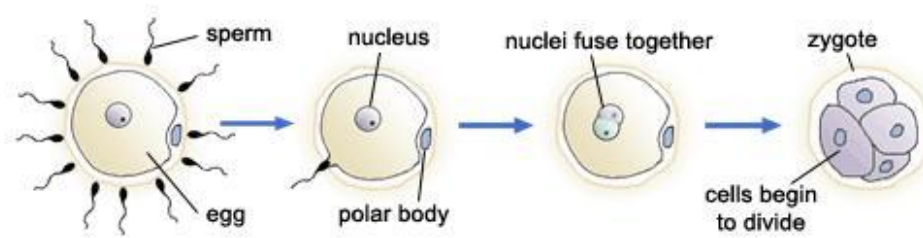
Figure 1.3. illustrates the process of fertilization

days before the onset of the next menstrual period [(12), (81)]. Evidence as to how long after ovulation the human ovum can be fertilized agree on a maximum duration of 48 hours (12). In order for pregnancy to begin, the egg must first be fertilized by sperm, and then it must successfully implant in the lining of the uterus (102). Sperm can fertilize the egg at any time as it travels from the ovary to the uterus. Sperm can also wait several days in the fallopian tubes for an egg to be released. Fig 1.3 illustrates the process of fertilization.

When a man and a woman have sex, the man ejaculates his sperm into the woman's vagina. The sperm then swim up through the uterus and into the fallopian tubes (see top diagram). If the sperm meets the egg in the fallopian tube fertilization can occur. For pregnancy to begin the fertilized egg must successfully implant in the uterus. If implantation is not successful, the woman passes the fertilized egg during menstruation. Any interference that occurs during this fertilization process may cause infertility.

1.2 Prevalence of infertility

Exact figures about the prevalence of infertility are difficult to determine. The best approximations are those which are derived from specific research studies and the numbers of childless marriages. The latter is not in itself regarded as a precise indicator as it is known that some 5% of the world's married population make a conscious decision to remain childless (76). It is apparent, however that somewhere between 1 in 10 and 1 in 6 couples experience problems with their fertility (49; 66). According to the National Center for Health Statistics of the Centers for Disease Control and Prevention (20) about 12 percent of women (7.3 million) in the United States aged $15 - 44$ had difficulty getting pregnant. Also, according to the two recent retrospective studies conducted in Aberdeen [(95; 96)], 14.1% of women had experienced difficulties in the past in becoming pregnant.

1.3 Causes of infertility

Factors such as age, stress, poor diet, athletic training, being overweight or underweight, tobacco smoking, alcohol, sexually transmitted diseases ($STDs$) and health problems that cause hormonal changes contribute considerably to fertility problems in women. [(11),(14), (15),(64)] . For a man to be fertile he must be able to produce an adequate amount of normal sperm. The sperm must be able to fertilize the woman's egg. Conditions that interfere with this process can cause infertility in men. Sometimes a man is born with the problems that affect his sperm. Other times problems start later in life due to illness or injury. For example, cystic fibrosis often causes infertility in men .

1.4   Infertility diagnosis

The cause of a couple's infertility can be diagnosed by doing a complete fertility evaluation. This process usually begins with physical exams and health and sexual histories. If there are no obvious problems, like poorly timed intercourse or absence of ovulation, tests will be needed. For a man, sperm examination and sperm count is commonly done. The number, shape, and movement of the sperm is examined microscopically in the laboratory. Sometimes doctors also suggest testing the level of a man's hormones. For a woman, the first step in testing is to find out if she is ovulating each month. A woman can track her ovulation at home by: (i) recording changes in her morning body temperature (basal body temperature) for several months, (ii) recording the texture of her cervical mucus for several months, (iii) using a home ovulation test kit (available at drug or grocery stores). Doctors can also check if a woman is ovulating by doing blood tests and an ultrasound of the ovaries.

1.5   Infertility treatment

There are various treatments for infertility depending what the problem is. Doctors recommend specific treatments for infertility based on test results, how long the couple has been trying to get pregnant, the age of both the man and woman, the overall health of the partners and preference of the partners. Intrauterine insemination (IUI) is the commonest type of treatment for infertility and is known by most people as artificial insemination. Several reviews article ((4), (13), (18), (22), (24), (39), (40), (64), (68), (70), (103)) have discussed and compared intrauterine insemination (IUI) with other treatments.

There is evidence from randomized trials to support the use of tamoxifen citrate (16), clomiphene citrate (16; 47) and anastrozole (86) in women with anovulatory infertility, though current best evidence suggests ovulation induction with gonadotrophins is an effective treatment for women with anovulatory infertility (86). Currently there is no clear evidence of using any medication to enhance male fertility (55).

1.6 Assisted reproductive technology

Assisted reproductive technology (ART) is a general term referring to methods used to achieve conception by artificial means. Assisted reproductive technology (ART) involves removing eggs from a woman's body, mixing them with sperm in the laboratory and putting the embryos back into a woman's body. Success rates vary and depend on many factors. Some things that affect the success rate of assisted reproductive technology (ART) include age of the partners, reason for infertility, clinic, type of ART, if the egg is fresh or frozen, and if the embryo is fresh or frozen. The U.S. Centers for Disease Prevention (CDC) (21) collected success rates on ART for some fertility clinics. According to (21), the average percentage of ART cycles that led to a healthy baby were as follows: 37.3% in women under the age of 35, 30.2% in women aged between 35 and 37, 20.2% in women aged between 37 and 40, and 11.0% in women aged between 41 and 42. Other methods of ART include Invitro fertilization (IVF), Zygote intrafallopian transfer ($ZIFT$) or Tubal Embryo Transfer, Gamete intrafallopian transfer (GIFT) and Intracytoplasmic sperm injection ($ICSI$). The most common complication of ART is multiple fetuses. But this is a problem that can be prevented or minimized in several different ways. The rapid development

in assisted reproduction technology (ART) has provided infertile couples with a variety of therapeutic options to achieve pregnancy (74).

1.7 Infertility designs

Two common and competing designs are available for assisted reproduction technology (ART), the crossover design and the parallel design. Crossover designs are trials in which patients are allocated to sequences of treatment with the purpose of studying differences between individual treatments (17; 84). A number of possible designs are available for the crossover studies. Under one design a patient may initially be randomized to one treatment, subsequently alternating treatment on each cycle. Alternatively, patients may initially receive one treatment for several cycles before switching to another treatment. Besides cross-over trials, there exists a parallel design. In a parallel design, a group of randomly selected subjects receives one treatment, whereas a second group receives a different treatment until the outcome of interest occurs or the trial is concluded.

1.8 Advantages of crossover designs

The crossover design to study the differences in treatments yields a more efficient comparison of treatments than a parallel design. Subjects act as their own controls, so the comparisons are usually based on within-patient variability (which is usually less than the between-subject variability). For given power and precision crossover designs require a smaller sample size Senn (84).

1.9 Disadvantages of crossover designs

Despite the appeal of having each subject serve as his/her own control, crossover studies have substantial weaknesses, as well. The potential problem in crossover design is that carryover effects may bias the direct treatment effects. Carryover effect is defined as the effect of the treatment from the previous time period on the response at the current time period. It occurs when the effect of a treatment given in the first time period persists into the second period and distorts the effect of the second treatment (84). The incorporation of washout period in the design can diminish the impact of carryover effects. A wash-out period is a period in a trial during which the effect of a treatment given previously is believed to disappear. The carry over effect should not be mistaken for the period effect. A period effect is one that occurs in a given period, irrespective of the order in which treatments are given. A carry-over effect has its origin in a preceding treatment and is thus order-dependent (84).

Another disadvantage of the crossover design is to do with censoring. Censoring can be a problem in crossover designs if a patient drops out of the trial during the first treatment period. If a Patient drops out of the trial during the first treatment period, then the other treatment will have no chance to demonstrate its effectiveness in the second period. In addition, the likelihood of becoming pregnant may not be constant from one period to the next because women who fail to conceive in the first period may have a lower probability of success in the second period. Consequently, by pooling the data obtained over the two study periods, a larger estimate of the effect of treatment is obtained than that with a parallel design trial (52). Also mentioned in the literature in regard to the crossover designs, is the problem of inconvenience to patients since patients are required

to submit to a number of treatments and the total time they spend under observation will be longer. It should also be mentioned that this particular feature can sometimes be turned to advantage in that it may be of interest for patient to have the opportunity to try out a number of treatments for himself/herself in order to gain personal experience of their effects (84).

1.10 Attitudes adopted towards carry-over and period effects?

A previously recommended method of analysing cross-over designs was to test for carryover, and if this was significant to discard the data from the second period and analyse only the data from the first period as if from a parallel group trial Grizzle (41). Senn et al. (85) do not recommend pre-testing of carryover effects. In his article, Freeman (34) showed that this method of pre-testing carry over leads to biased answers, as is generally the case when the choice between two analyses depends on the the result of a preliminary hypothesis test. In his book Senn (84) argued that the use of the crossover design is effectively built on the assumption that there is minimal carryover of the effect of a treatment into the next period. Following this philosophy, rather than testing for carryover, in this dissertation we will proceed as if there were no carryover. We will include parameters for period and estimate treatment and period effects simultaneously: that is to say, estimate treatment in the presence of period effects and vice versa.

1.11   Crossover design in Infertility trials

The commonest of crossover designs is the $AB : BA$ crossover design in which approximately half of the patients are first given treatment A and on a subsequent occasion treatment B whereas the rest of the patients are first given treatment B and on a subsequent occasion treatment A. In the standard $AB : BA$ crossover design setup, for every participant there must be two observations: observation due to treatment A and observation due to treatment B. In this standard set up, patients act as their own control. This standard crossover design represents a special situation where there is not a separate comparison group. Also, since the same subject receives both treatments, there is no possibility of covariate imbalance. Contrast this with a parallel group design where some subjects get the first treatment and different subjects get the second treatment. The standard $AB : BA$ crossover design (here and after will be called complete crossover design) has the advantage of achieving the same level of precision as the parallel design trial, but with half the number of patients because each patient contributes two observations (one for each treatment). In the complete crossover design, between-subject variation (which is inherent in a parallel design) is avoided because the comparison of the two treatments occurs in the same patient. It is the difference in outcomes within each patient that is the key observation, and these differences are pooled and analyzed at the conclusion of the design. When pregnancy is the outcome, patients who conceive in the first period are withdrawn from the study and do not receive treatment in the second period. Therefore, the essential within-subject comparison of the responses to the two treatments does not occur. Censored patients do not provide information on the treatment they were not exposed to, thereby not permitting the within patient comparison. However, in the medical research setting, a crossover design offers several practical advantages.

First, with a crossover design, all couples will be exposed to both treatments, thus couples will be motivated to participate and to complete the entire trial. Moreover, more couples will become pregnant in a crossover-design trial than in a parallel-design trial (24; 64) because all couples are offered the most effective treatment for half of the treatment period (This point is also illustrated analytically in page 21).

Since 1993, an extended and sometimes heated debate has been conducted on the place of the crossover design in infertility trials. Daya (27) opened the debate by stating that, in his opinion, the crossover design has no place in infertility trials. Khan et al. (52) published the results of a meta-analysis comparing the two designs, and concluded that within the existing literature there is an overestimation of treatment effect in crossover trials compared with parallel trials. However, the method used by Khan (52), was criticized by many statisticians including te Velde (94), who later re-analyzed the data. In re-analyzing the data, te Velde concluded that the over estimation found by Khan (52) is statistically insignificant. Ananth and Rhoads (6) also criticized the method used by Khan (52), claiming that the statistical method is inappropriate. Olive (75) agreed that the crossover design will overestimate the treatment effects but suggested that this may be due to inadequate statistical analysis. Cohlen (24), Norman (74) and McDonnell (64) conducted a series of simulations to try to answer this controversy.

Both crossover and parallel design had been extensively used in infertility trials [(24),(23), (33), (39), (40), (43), (48), (51), (65), (70), (103)]. In cases where parallel design was employed [(33), (39), (43), (65)], the statistical methods used for analysis was appropriate. On the other hand in the articles where crossover designs were employed [(24), (40), (51), (70), (103)], we feel that statistical method

used is inappropriate for a cross over design.

In assisted reproduction technology (ART), women are allowed to conceive only once. If a woman gets pregnant, the women is dropped from the trial. This dissertation aims to investigate what actually is lost by not allowing women to conceive twice. That is to say, what are we losing by moving from the complete crossover design to the incomplete crossover design? Is having some of the outcomes in the second period missing as good (may be worse) as having no outcomes in the second period? i.e., is parallel design better (worse) than the crossover design in infertility trials. We will illustrate how data set from the cross over design in infertility trials is analyzed.

1.12 Motivation

We are motivated by the fact that even though crossover designs are extensively used in infertility (24; 40; 48; 103), there are still some statisticians who feel that the use of crossover design in infertility trials is inappropriate (27; 29; 45). The critics of the use of crossover design in infertility trials, including Daya (27), are concerned that women who become pregnant after the first treatment period will leave the study. This will mean that data may not be available for within patient comparison if crossover designs is used and pregnancy is the outcome measure. We agree with them that the second outcome will be missing in the event that the woman gets pregnant in the first period. It is surprising when they put parallel design before the crossover trial with some outcomes in the second period missing. This incomplete crossover has more information than the parallel design. The question is if you were to choose between one set of data where women are exposed to either A or B (parallel), and the other set of data where

some women are exposed to both A and B, some are exposed to A only, and some are exposed to B only, which data will you choose? We tend to believe that most people will prefer the latter than the former, for the obvious reason that you are told more of what is happening in a woman, with respect to the two treatments, in the latter than in the former. Daya's comments seem strange in view that as statisticians, we will prefer to have more information than less. The crossover design in the medical setup, will surely have some missing information on those women who get pregnant on the first period, but will contain all the information that is contained in the parallel design. So, the question is why in this scenario do other statisticians like Daya opt for the data with less information.

Consider the complete crossover design data set in Table 1.1 below;

A	B	O	B	A	O
0	0	n_1	0	0	n_5
0	1	n_2	0	1	n_6
1	0	n_3	1	0	n_7
1	1	n_4	1	1	n_8

Table 1.1. Complete crossover design data set

Note that this is a theoretical situation only. In practice some of the data would be missing. It is clear that given the complete crossover data sets, one can easily predict what would have happened if an infertility crossover design was conducted. Basically, one can easily obtain the data set in Table 1.2 below;

A	B	O	B	A	O
0	0	n_1	0	0	n_5
0	1	n_2	0	1	n_6
1	*	n_3	1	*	n_7
1	*	n_4	1	*	n_8

Table 1.2. infertility crossover design data set

Similarly given the incomplete crossover data sets, we can easily predict what would have happened if the parallel design was followed. That is we can easily get the data set in Table 1.3 below;

A	B	O	B	A	O
0	*	n_1	0	*	n_5
0	*	n_2	0	*	n_6
1	*	n_3	1	*	n_7
1	*	n_4	1	*	n_8

Table 1.3. Parallel design data set

We were motivated by the way these three data sets look. Using Table 1.1, Table1.2 and Table 1.3, we expect the preference of the data sets to be in the following order: complete crossover data sets, followed by incomplete crossover data sets and finally parallel design. Of course Table 1.1 is a theoretical option only. To our surprise some statisticians prefer the data set in Table 1.3 more than the data set in Table 1.2.

Over the past decade, the crossover design has gained popularity and is being recommended for infertility research by many investigators [(24), (23),(64), (70), (74),(103)]. In this section we will look at some of the articles where crossover design has been used or recommended in assisted reproduction technology (ART). We will begin by looking at the article by Zreik et al. (103).

1.13 Review of Zreik's Paper

In this section we will look at data from Zreik (103) based on a crossover study to compare two methods of timing ovulation; Luteinizing hormone urine test (LH monitoring) and transvaginal ultrasonography. Luteinizing hormone urine test (LH) detects a rise in lutenizing hormone. Such a rise, or surge, signals the ovary to release the egg. This at-home test is often used by women to help predict ovulation. On the other hand, transvaginal ultrasonography is a a painless, harmless test that uses sound waves to produce images of the organs and structures of the body on a screen.

1.13.1 Trial Setup

The trial took place in September 1994 at the Yale Reproductive Medicine Center, New Haven, Connecticut, USA. Initially, 61 infertile couples agreed to participate, but later seven couples were excluded from the analysis for failure to return or for declining crossover during the study after randomization. Thirty-two patients with unexplained infertility or mild endometriosis, 15 patients with anovulation, and 7 patients with mild male factor infertility were enrolled. Anovulation is absence of ovulation when it would be normally expected.

The mean age of the women was 33 years with the youngest being 24 years and the oldest being 41 years. Patients underwent a complete infertility evaluation. Unexplained infertility was defined as normal findings in the infertility evaluation and included normal laparoscopic findings or minimal endometriosis. Direct laparoscopic visualization with or without biopsy of suspicious lesions was

used to confirm endometriosis.

Patients were randomized by the use of a computer generated random number table; the assignment was not known to the treating physician or the patient until consent was obtained. The patients were randomized to either group 1 or group 2. Patients randomized to group 1 began with protocol A, whereas patients randomized to group 2 began with protocol B. In protocol A, Luteinizing hormone urine test was used to time ovulation. Urinary LH levels were determined daily with the use of commercial kits (ovuQuick, Quidel, San Diego, CA), starting on day 10 of the cycle and discarding the first morning void. When urinary LH was detected, intrauterine inseminations (IUI) were performed daily for the next 2 days. In protocol B, ultrasound monitoring of folliculogenesis, starting on day 10 of the cycle, was performed until a leading follicle of $\geq 18mm$ was noted, at which human chronic gonadotropin (hCG) of $10,000$ IU was given intramuscularly and intrauterine inseminations (IUI) were performed daily for the next 2 days. If no pregnancy occurred, the couple crossed over to the alternative protocol for the next cycle, with couples having at most four cycles. For each couple , the number of LH and hCG treatments and pregnancy status were recorded. Pregnancy rate per cycle was calculated as the ratio of the number of couples pregnant to the number of cycles in a protocol. Fisher's exact test was used to compare the pregnancy rates per cycle between the two protocols in each of the three diagnostic groups (unexplained infertility, male factor infertility, and anovulatory infertility).

1.13.2 Results and Data analysis

The couples contributed 142 cycles of which one was excluded because of failure of compliance by a patient randomized to urinary LH monitoring. Six couples conceived. The data from each individual cycle in the three different groups (male factor, anovulation, and unexplained infertility) are tabulated in Table 1.4 below (103):

Cycle no	Protocol	Unexplained	Anovulation	Male factor	Total
1	hCG	$\frac{1}{18}$	$\frac{0}{8}$	$\frac{1}{2}$	$\frac{2}{28}$
	LH	$\frac{1}{13}$	$\frac{0}{7}$	$\frac{0}{5}$	$\frac{1}{25}$
2	hCG	$\frac{0}{10}$	$\frac{0}{5}$	$\frac{0}{4}$	$\frac{0}{19}$
	LH	$\frac{1}{15}$	$\frac{0}{5}$	$\frac{0}{1}$	$\frac{1}{21}$
3	hCG	$\frac{0}{8}$	$\frac{0}{3}$	$\frac{0}{0}$	$\frac{0}{11}$
	LH	$\frac{1}{9}$	$\frac{0}{3}$	$\frac{0}{3}$	$\frac{1}{15}$
4	hCG	$\frac{0}{8}$	$\frac{0}{3}$	$\frac{0}{3}$	$\frac{0}{14}$
	LH	$\frac{1}{5}$	$\frac{0}{3}$	$\frac{0}{0}$	$\frac{1}{8}$
Total		$\frac{5}{86}$	$\frac{0}{37}$	$\frac{1}{18}$	$\frac{6}{141}$

Table 1.4. Pregnancies obtained in each cycle with the two protocols

The first thing that (103) did was to stratify the data according to the infertility status and test if there is any association between the protocol and pregnancy (i.e.. they adjusted for the infertility status). The number of couples conceived (not conceived) under each infertility status was recorded. The 2×2 tables were constructed for each infertility case as shown below:

	LH	hCG
Pregnant	0	1
Not Pregnant	9	8

Table 1.5. Male Factor Infertility

	LH	hCG
Pregnant	0	0
Not Pregnant	18	19

Table 1.6. Anovulation Infertility

	LH	hCG
Pregnant	4	1
Not Pregnant	38	43

Table 1.7. Unexplained Infertility

Fisher's exact test was used to test if there is any association between the outcome and the protocol. The authors found no statistically significant differences between the two protocols. However we take issue with their method of calculation. Fisher's exact test is valid only if say N patients were randomly allocated to one treatment groups (hCG)and M patients to the other (LH). That is to say that, Fisher's Exact test was going to be the appropriate measure if say from Table 1.5, nine patients were randomly allocated to LH protocol and the other nine patients were randomly allocated to the hCG protocol. Thus Fisher's Exact test is valid in parallel design, not in cross over design.

1.14 Review of Cohlen's Paper

In their article Cohlen and his group considered a hypothetical population of subfertile couples. The population was divided into 2 equal groups and was

assigned to either treatment A or treatment B. After randomization, each group received either treatment A or treatment B continuously (parallel design) or alternated between the treatment modalities (crossover design). Each couple was offered a maximum of $n = 6$ treatment cycles. It was assumed that there is no carry over effects, no period effect, and no dropouts. Couples dropout if and only if they are pregnant. The expectations were used to estimate the per cycle probability of conception. The objective was to determine whether a crossover design results in a different estimate of treatment effect compared with a parallel design.

Table 1.8 below shows the number of couples who would have become pregnant after each treatment cycle in either parallel or crossover design trial. The total number of all couples is $2m = 2000$.

| | Parallel design | | | | | Crossover design | | | | |
| | Treatment A | | Treatment B | | | Treatment A | | Treatment B | | |
Cycle no.	No. of pregnancies/no. of couples	PR per cycle (%)	No. of pregnancies/no. of couples	PR per cycle (%)	OR	No. of pregnancies/no. of couples	PR per cycle (%)	No. of pregnancies/no. of couples	PR per cycle (%)	OR
1	100/1,000	10	200/1,000	20	2.25	100/1,000	10	200/1,000	20	2.25
2	90/900	10	160/800	20	2.25	80/800	10	180/900	20	2.25
3	81/810	10	128/640	20	2.25	72/720	10	144/720	20	2.25
4	73/729	10	102/512	20	2.25	58/576	10	130/648	20	2.25
5	66/656	10	82/410	20	2.25	52/518	10	104/518	20	2.25
6	59/590	10	66/328	20	2.25	41/414	10	93/466	20	2.25
Total:	469/4,685	10	738/3,690	20	2.25	403/4,028	10	851/4,252	20	2.25

Table 1.8. Expected number of pregnant couples, pregnancy rate per cycle, and odds ratio after each treatment cycle when a homogenous population of 2000 couples participates in a parallel or crossover design trial

Let p_A be the probability of conception under treatment A, and let $1 - p_A = q_A$ be the probability of failing to conceive under treatment A. Similarly we can define p_B and q_B. If the population is homogenous, we immediately observe that, under parallel design, we can easily deduce the following (i) the expected proportion of women pregnant at the end of the trial is

$$1 - \frac{q_A^n + q_B^n}{2} \qquad q_A = 1 - p_A, q_B = 1 - p_B$$

(ii) the expected total number of cycles involved under A, at the end of the trial is given by the expression

$$T_A = m * [\sum_{i=1}^{n} q_A^{i-1}]$$

(iii) if $\hat{p_A}$ is the ratio of the expected number of pregnancies to the expected total follow up time (no of cycles) as in (23), then $\hat{p_A} = p_A$.

Similarly, it can be shown that in the crossover designs, the expected proportion of women pregnant undergoing treatment A is

$$p_A(1 + q_B) \sum_{i=1}^{\frac{n}{2}} (q_A q_B)^{i-1}$$

and the expected proportion of women pregnant undergoing treatment B is

$$p_B(1 + q_A) \sum_{i=1}^{\frac{n}{2}} (q_A q_B)^{i-1}.$$

While the expected number of cycles under treatment A is given by

$$m * (1 + q_B) \sum_{i=0}^{\frac{n}{2}-1} (q_A q_B)^{i}$$

and the expected number of cycles under treatment B is given by

$$m * (1 + q_A) \sum_{i=0}^{\frac{n}{2}-1} (q_A q_B)^{i}$$

In addition

$$\hat{p_A} = \frac{p_A(1+q_B) \sum_{i=1}^{\frac{n}{2}} (q_A q_B)^{i-1}}{(1+q_B) \sum_{i=0}^{\frac{n}{2}-1} (q_A q_B)^{i}} = \frac{p_A \sum_{i=1}^{\frac{n}{2}} (q_A q_B)^{i-1}}{\sum_{i=1}^{\frac{n}{2}} (q_A q_B)^{i-1}} = p_A$$

and

$$\hat{p_B} = \frac{p_B(1+q_A)\sum_{i=1}^{\frac{n}{2}}(q_Aq_B)^{i-1}}{(1+q_A)\sum_{i=0}^{\frac{n}{2}-1}(q_Aq_B)^i} = \frac{p_B\sum_{i=1}^{\frac{n}{2}}(q_Aq_B)^{i-1}}{\sum_{i=1}^{\frac{n}{2}}(q_Aq_B)^{i-1}} = p_B.$$

Clearly, the estimates above are unbiased.

Under this method, it is clearly observed that, the expected number of pregnancies in the crossover trial is greater or equal to the number of pregnancies in the parallel trial since,

$$\triangle = (1 - q_A^{\frac{n}{2}}q_B^{\frac{n}{2}}) - (1 - \frac{q_A^n+q_B^n}{2})$$

$$= \frac{1}{2}(q_A^{\frac{n}{2}} - q_B^{\frac{n}{2}})^2 \geq 0.$$

The equality holds if and only if the two treatments are equivalent.

We will show that in the most effective treatment, crossover design produce more pregnancies than parallel design. While in the least effective treatment the situation is reversed. Suppose B is the most effective treatment, while A is the least effective treatment. Under crossover design, the proportion of women pregnant due to treatment B is given by

$$[\sum_{i=1}^{\frac{n}{2}}[q_B^{i-1}p_Bq_A^i + q_B^{i-1}p_Bq_A^{i-1}] \tag{1.1}$$

while under parallel design, the expected proportion of women pregnant due to treatment B is

$$1 - q_B^n \tag{1.2}$$

We will prove by induction on n (n even number). We will show that it holds for

$n = 2$. For $n = 2$, (1.1) becomes

$$p_B q_A + p_B. \tag{1.3}$$

While (1.2) becomes

$$1 - q_B^2 = (1 - q_B)(1 + q_B) = p_B + p_B q_B. \tag{1.4}$$

Clearly (1.3)>(1.4), since $q_A > q_B$.

Now we assume that (1.1)>(1.2), for any integer $n = 2, 4, 6...2x$. We will show that (1.1) > (1.2), for $n = 2x + 2$. But for $n = 2x + 2$, (1.1) becomes

$$\sum_{i=1}^{x+1} p_B q_B^{i-1} q_A^{i-1}(1 + q_A) = \sum_{i=1}^{x} p_B q_B^{i-1} q_A^{i-1}(1 + q_A) + p_B q_B^x q_A^x(1 + q_A) \tag{1.5}$$

while (1.2) becomes

$$1 - q_B^{2x+2} = (1 - q_B) \sum_{i=1}^{2x+2} q_B^{i-1} = p_B \sum_{i=1}^{2x} q_B^{i-1} + p_B q_B^{2x}(1 + q_B) \tag{1.6}$$

clearly (1.5) > (1.6), since $q_A > q_B$. Thus by induction we have proved that (1.1) is always greater than (1.2). Hence, in the most effective treatment, the expected number of pregnancies under crossover is greater than the expected number of pregnancies under parallel design (which was to be proved).

Similarly it can be shown that in the least effective treatment, the expected number of pregnancies under parallel is greater than the expected number of pregnancies under crossover design.

Hence we can immediately get Cohlen's results as follows: In their article Cohlen and his group began with $p_A = 0.1$ and $p_B = 0.2$. The sample size is

2000. Each couple was offered a maximum of six treatments. Thus the estimated per cycle probability of pregnancy for treatment A is given by

$$\hat{p_A} = \frac{1-0.9^6}{\sum_{k=1}^{6} 0.9^{i-1}} = \frac{0.468559}{4.68559} = 0.1 = p_A.$$

Similarly it could be verified that $\hat{p_B} = p_B = 0.2$.

Under a crossover trial

$$\hat{p_A} = \frac{[\sum_{n=1}^{3}[q_A^{n-1}p_A q_B^n + q_A^{n-1}p_A q_B^{n-1}]]}{[\frac{(1-q_A^3 q_B^3)(1+q_B)}{1-q_A q_B}]} = \frac{0.402912}{4.02912} = 0.1 = p_A.$$

Similarly, $\hat{p_B} = p_B = 0.2$

However crossover results in more pregnancies

$$2000 * [p_A(1 + q_B) + p_B(1 + q_A)] * \frac{\sum_{i=1}^{3}(q_A q_B)^{i-1}}{2} = 1254$$

pregnancies compared to

$$2000 * (1 - \frac{0.9^6 + 0.8^6}{2}) = 1206$$

pregnancies in the parallel design. This difference is due to the different designs and not to overestimation of the treatment effect. Crossover designs will produce more pregnancies than parallel designs simply because couples are exposed to both treatments for half of the time of trial. Crossover design results in more pregnancies with the most effective treatment $(1000*p_B(1+q_A)*\sum_{i=1}^{3}(q_A q_B)^{i-1} = 851$ versus $1000 * (1 - q_B^6) = 738)$ and a fewer pregnancies with the least effective treatment $(1000 * p_A(1 + q_B) * \sum_{i=1}^{3}(q_A q_B)^{i-1} = 403$ versus $1000(1 - q_A^6) = 469)$.

In analyzing the data, Cohlen (23) estimated the treatment (protocol) effect by the pregnancy rate per cycle, which is defined as the number of couples pregnant divided by the number of cycles in a trial. This measure pays more attention

to the couples who stayed longer in the trial. The couples who stayed longer in the trial contributes more information to this kind of measure as compared to the couples who conceived in early stages of the trial. Clearly this may not be the correct measure of treatment effect. As an example consider a scenario where 11 couples goes under an assisted reproduction technology, with couples having at most ten cycles. Suppose ten couples conceived in the first period, and the remaining one couple did not conceive at the end of the trial. Thus the contributed number of cycles is 20 and the number of pregnancies is 10, hence the pregnancy rate per cycle is $\frac{10}{20} = 0.5$. This clearly does not tell us that almost all couples conceived in the first period. For this population of 11 couples a pregnancy rate of anything in the neighborhood of 0.99 will be more plausible than 0.5. We really have an issue with this kind of measure. But the measure is not bad for descriptive statistics.

1.15 Review of Mcdonnell's Paper

McDonnell's aim was to see if the design structure and/or the presence of carryover effects lead to a bias in the statistical analysis, leading to possible under/overestimation of treatments effects. The probability of conception was assumed to be logistic and only depends on age and treatment applied.

The authors used a simulation approach to follow the progress of two theoretical cohorts entering into a trial comparing $IUI = 0$ and $IVF = 1$. The progress of the two cohorts was simulated under both the parallel and crossover designs. In the parallel arm, couples initially are assigned randomly to one treatment, staying with that treatment until leaving the trial. In the crossover arm, couples are initially assigned to one treatment, switching treatments in subsequent cycles.

Each couple was offered a maximum of six treatment cycles. The following four assumptions were considered.

(i) no couples dropped out (no censoring) and the treatment had no effect on the following cycle (no carryover effect); (ii) no censoring but there was a carryover effect; (iii) couples could be censored, but there was no carryover effect; and (iv) both censoring and carry over effects were present.

Under no-carryover and no-censoring scenarios, the per cycle probability of pregnancy is

$$p_1(age, treat) = \frac{\exp(\mu(age, treat))}{1 + \exp(\mu(age, treat))}.$$

In the carryover scenarios, the per cycle probability of pregnancy is

$$p_2(age, treat, ptreat) = \frac{\exp(\mu(age, treat) - \ln(1.5) \times ptreat}{1 + \exp(\mu(age, treat) - \ln(1.5) \times ptreat}.$$

In scenarios involving censoring, the probability of censoring was

$$p_3(treat) = \frac{\exp(\eta(treat))}{1 + \exp(\eta(treat))}.$$

where

$$\eta(treat) = -3.073850 + 1.361340 \times treat$$

$$\mu(age, treat) = -0.321865 - 6.146251 \times age + 0.330317 \times treat$$

And

$ptreat = 1$, if the previous treatment applied was IVF, otherwise $ptreat = 0$.

Also for any probability of conception p_i, $\{i = 1, 2, 3\}$ we define $q_i = 1 - p_i$.

The age of the female patient was randomly generated using the formula

$$Age = 25 \times 1.6^{U}$$

where U is a random variable uniformly distributed on [0,1]. We will like point out that the purpose of this section is to reproduce the author's results analytically (or to estimate the results analytically). The cumulative distribution of age will be

$$G(age) = P(Age < age) = P(25(1.6)^{U} < age) = P(U < \frac{\ln \frac{age}{25}}{\ln 1.6}) = \frac{\ln \frac{age}{25}}{\ln 1.6}$$

Hence the probability distribution of age is given by

$$f(age) = \frac{d(G(age))}{dage} = \frac{1}{age \ln 1.6}, \text{ for } 24 < age < 40, \text{ and zero elsewhere.}$$

Thus the median age is given

$$M = 25 * 1.6^{0.5} = 31.62278$$

Hence their population consists of women with ages between 25 and 40 only, with relatively more younger women than older women. A situation not dissimilar to that seen in practice.

Using the assumptions from (64), under no carryover effects and no censoring, we should expect the proportions of couples achieving pregnancy under parallel design and crossover design to be

$$\int_{25}^{40} 1 - \frac{\{q_1(age, 0)q_1(age, 1)\}^6}{2} f(age) dage \qquad (1.7)$$

and

$$\int_{25}^{40} 1 - \{q_1(age,0)q_1(age,1)\}^3 f(age)dage. \tag{1.8}$$

respectively.

Under a parallel design, for the carryover scenario, we should expect the proportions of couples achieving pregnancy due to IUI and IVF to be

$$\int_{25}^{40} [p_1(age,0) \sum_{n=1}^{6} q_2(age,0,0)^{n-1} f(age)]dage \tag{1.9}$$

and

$$\int_{25}^{40} [p_1(age,1) + p_2(age,1,1)q_1(age,1) \sum_{n=1}^{5} q_2(age,1,1)^{n-1} f(age)]dage \tag{1.10}$$

respectively.

Similarly, under the crossover design, the proportion of women achieving pregnancy due to IUI and IVF are

$$\int_{25}^{40} [p_1(age,0) +$$

$$q_1(age,0)q_2(age,1,0)p_2(age,0) \sum_{n=1}^{2} \{q_2(age,0,1)q_2(age,1,0)\}^{n-1}]f(age)dage$$

$$+ \int_{25}^{40} [q_1(age,1)p_2(age,0,1) \sum_{n=1}^{3} \{q_2(age,0,1)q_2(age,1,0)\}^{n-1} f(age)]dage \tag{1.11}$$

and

$$\int_{25}^{40} p_1(age,1) +$$

$$q_1(age,1)q_2(age,0,1)p_2(age,1,0) \sum_{n=1}^{2} \{q_2(age,1,0)q_2(age,0,1)\}^{n-1} f9age)dage$$

$$+ \int_{25}^{40} q_1(age,0)p_2(age,1,0) \sum_{n=1}^{3} \{q_2(age,1,0)q_2(age,0,1)\}^{n-1} f(age)dage \tag{1.12}$$

respectively.

If we consider the censoring scenario, under parallel design the proportion of women achieving pregnancy due to IUI and IVF are

$$\int\limits_{25}^{40} p_1(age,0) \sum_{n=1}^{6} \{q_1(age,0)q_3(0)\}^{n-1}) f(age) dage \tag{1.13}$$

and

$$\int\limits_{25}^{40} p_1(age,1) \sum_{n=1}^{6} \{q_1(age,1)q_3(1)\}^{n-1}) f(age) dage \tag{1.14}$$

respectively.

Similarly, under the crossover design, the proportion of women achieving pregnancy due to IUI and IVF are

$$\int_{25}^{40} p_1(age,0) \sum_{n=1}^{3} \{q_1(age,0)q_3(0)q_1(age,1)q_3(1)\}^{n-1}) f(age) dage$$

$$+ \int\limits_{25}^{40} p_1(age,0)q_1(age,1)q_3(1) \sum_{n=1}^{3} \{q_1(age,0)q_3(0)q_1(age,1)q_3(1)\}^{n-1}) f(age) dage \tag{1.15}$$

and

$$\int_{25}^{40} p_1(age,1) \sum_{n=1}^{3} \{q_1(age,0)q_3(0)q_1(age,1)q_3(1)\}^{n-1}) f(age) dage$$

$$+ \int\limits_{25}^{40} p_1(age,1)q_1(age,0)q_3(0) \sum_{n=1}^{3} \{q_1(age,0)q_3(0)q_1(age,1)q_3(1)\}^{n-1}) f(age) dage \tag{1.16}$$

respectively.

Finally if we consider the scenario involving both censoring and carryover effect, we observe that under parallel design, the proportion of women achieving pregnancy due to IUI is the same as in the scenario where we have censoring

only. Whereas the proportion of women achieving pregnancy under IVF is

$$\int_{25}^{40} [p_1(age,1) + q_1(age,1)q_3(1)p_2(age,1,1) \sum_{n=1}^{5} \{q_2(age,1,1)q_3(1)\}^{n-1} f(age)] dage$$

(1.17)

In the crossover arm, the proportion of women achieving pregnancy due to IUI and IVF are

$$\int_{25}^{40} p_1(age,0) +$$

$$q_1(age,0)q_3(0)q_2(age,1,0)q_3(1)p_2(age,0,1) \sum_{n=1}^{2} \{q_2(age,0,1)q_2(age,1,0)q_3(0)q_3(1)\}^{n-1}$$

$$+ \int_{25}^{40} q_1(age,1)q_3(1)p_2(age,0,1) \sum_{n=1}^{3} \{q_2(age,0,1)q_2(age,1,0)q_3(0)q_3(1)\}^{n-1} f(age) dage$$

(1.18)

and

$$\int_{25}^{40} p_1(age,1) +$$

$$q_1(age,1)q_3(1)q_2(age,0,1)q_3(0)p_2(age,1,0) \sum_{n=1}^{2} \{q_2(age,1,0)q_2(age,0,1)q_3(1)q_3(0)\}^{n-1}$$

$$+ \int_{25}^{40} q_1(age,0)q_3(0)p_2(age,1,0) \sum_{n=1}^{3} \{q_2(age,1,0)q_2(age,0,1)q_3(1)q_3(1)\}^{n-1} f(age) dage$$

(1.19)

Equations 1.7,1.8, 1.9,...,1.19 can be used to reproduce or estimate the results obtained by the authors.

1.15.1   Results

Table 1.9 below shows the simulation results obtained by McDonnell (64). The results obtained by evaluating expressions $1.7 - 1.117$ are shown in brackets.

censoring	carryover	state	parallel	crossover
no	no	pregnant	49.6 (49.8)	50.4 (50.1)
no	yes	pregnant	48.0 (47.9)	48.5 (48.6)
yes	no	pregnant	40.4 (40.4)	40.8 (40.8)
yes	yes	pregnant	39.3 (39.1)	39.6 (39.6)

Table 1.9. Proportion of patients (as percentage) achieving pregnancy or dropping out under each of the four baseline scenarios

The crossover design results in an increase in the pregnancy rate, as expected from expressions above. This increase remains small under other assumptions. Mcdonnell's results agree with our theoretical results. But we have a problem with the model used by these authors. The authors made a very strong assumption of independence of cycles within a couple. This strikes us as being bizarre since outcomes from the same subject are expected to be correlated. The assumption of independence of cycles within a patient, makes the likelihoods of the two designs to be the same.

1.16 Overview of later chapters

In the remaining chapters of this dissertation, Chapter 2 provides some background on the statistical methods used for modeling binary outcomes, reviews standard logistic and mixed effects logistic regression and their corresponding parameter estimation including the Hermite quadrature method. Chapter 3 describes how to estimate the parameters in the crossover design and how to compare estimates obtained from crossover design to estimate obtained from the parallel design. Chapter 4 will discuss how in practice the crossover trials should

be analyzed using the current available statistical packages. Finally we discuss

our findings in chapter 5.

Chapter 2

Mixed-Effects Regression Model
for Binary Outcomes

This chapter introduces and defines the mixed-effects regression model that is used throughout the dissertation. We begin by discussing the mixed-effects logistic regression for longitudinal (clustered) binary data. A full maximum marginal likelihood solution is illustrated for the parameter estimation. In this solution, the quadrature method is used to numerically integrate over the distribution of the random-effects.

2.1   Overview

Logistic regression is like linear regression described in Neter et al. (72) in that it is a method for modeling the effect of predictor variables on a response variable. The difference is that the response variable is binary; e.g. dead or alive, disease or non-diseased, exposed or unexposed, pregnant or not pregnant.

Usually, binary data result from a nonlinear relationship between the probability of success $p = p(\mathbf{x})$ and the covariates $\mathbf{x}$ associated with the binary outcome. A fixed change in the covariate $\mathbf{x}$ often has less impact when the probability of success p is near 0 or 1 than when p is near 0.5. In the assisted reproduction technology (ART), suppose that the probability $p(\mathbf{x})$ of enrolling in the *ART* course of treatments depends on the annual family income $\mathbf{x}$. An increase of £$36,000$ in annual income would have less effect when $\mathbf{x} = $ £$500,000$ (for which $p(\mathbf{x})$ is near 1) than when $\mathbf{x} = $ £$50,000$.

In practice, nonlinear relationships between p and $\mathbf{x}$ are often monotonic, with p increasing continuously or p decreasing continuously as $\mathbf{x}$ increases. The most important curve with this kind of behavior is the logistic curve, described more extensively in Agresti (1). It is used in many areas such as health care research and biomedical studies (9).

Logistic regression is a widely accepted technique for describing the relationship between a categorical outcome and a set of explanatory variables. The response variable is usually dichotomous or binary, but it may be polytomous, that is, have more than two response levels. These multiple-level response variables can be nominally or ordinally scaled. To provide a statistical foundation for mixed effects generalization of the logistic regression model, we now present an overview of the standard logistic regression model when the response is dichotomous.

2.2 Logistic Model

To formulate the logistic model, let p_i represent the probability of a positive outcome (i.e., $Y_i = 1$) for the *ith* individual. The probability of a negative outcome (i.e., $Y_i = 0$) is then $1 - p_i$. Denote the set of covariates as $\mathbf{x}_i = (1, x_{i1}, ..., x_{ip})$, where $\beta = (\beta_0, \beta_1, ..., \beta_p)'$ is a $(p+1) \times 1$ vector of corresponding regression coefficients. Then the logistic regression model has the form

$$\log \frac{p_i}{1 - p_i} = \mathbf{x_i'}\beta, \tag{2.1}$$

where the expression on the left-hand side is referred to as the logit or log-odds of a 1 response. The logit transformation linearizes the relationship between $p(\mathbf{x})$ and the covariates $\mathbf{x}$. It is also worth mentioning that the log-odds for a 1 response takes on values $-\infty \leq logit p(\mathbf{x}) \leq \infty$. In logistic regression, the logit is called the link function because it maps the $(0, 1)$ range of probabilities unto the $(-\infty, \infty)$ range of linear predictors. There are alternative models for dichotomous response variables (e.g. the probit regression model). One reason for the popularity of the logit model is that the coefficients have a simple interpretation in terms of the odds ratios. The odds of a 1 response is

$$\frac{p_i}{1 - p_i} = exp(\mathbf{x_i'}\beta). \tag{2.2}$$

Note that the relationship between the odds and the predictor variable $\mathbf{x}$ is non-linear.

The logit model 2.1, can be rewritten in terms of the probability of a 1 response, such that

$$p_i = \frac{exp(\mathbf{x'_i}\beta)}{1 + exp(\mathbf{x'_i}\beta)} = \frac{1}{1 + exp(-\mathbf{x'_i}\beta)} = \Psi(-\mathbf{x'_i}\beta), \quad (2.3)$$

where $\Psi(.)$ is the logistic cumulative distribution function, namely

$$\Psi(z) = \frac{1}{1 + exp(-z)}, \quad (2.4)$$

which implies that $0 \leq p(\mathbf{x}) \leq 1$; also a nonlinear function of the predictors. More details regarding the logistic regression can be found in Agresti (1) and Neter et al. (72).

2.2.1   Interpretation of parameters

The quantity β_0 in equation 2.1 is the log odds of a positive outcome for an individual with a set of covariates $\mathbf{x}_i = 0$. The parameter β_p is the increment in the log odds for a unit change in x_p holding all other covariates constant. The quantity $exp(\beta_p)$ is an odds ratio for the regressors, namely the ratio of the odds of a positive response for a unit change in x. For more understanding on the interpretation of the parameters in the logistic regression model, Allison (5) and Stokes et al. (92) are strongly recommended.

As we know from Agresti (1) that the 95% Wald confidence interval for a given regression parameter is

$$\hat{\beta} \pm z_{0.975} \times se(\hat{\beta}).$$

We would like to find a comparable confidence interval for the estimated odds

ratios. In particular suppose that we want a confidence interval for the estimated odds ratio $exp(\beta_p)$. The 95% Wald confidence interval for the odds ratio is computed as

$$exp(\beta_p \pm z_{0.975} \times se(\beta_p)).$$

That is, we compute the 95% Wald confidence interval for β_p and exponentiate the result.

2.2.2 Model Assumptions

If the dichotomous random variable Y represents the outcome from an individual subject then (i) the Y values are independent and take on values of either 0 or 1. (ii) The variable Y has a binomial distribution with $n = 1$ and $p(\mathbf{x}) = \frac{exp(\mathbf{x}'\beta)}{1+exp(\mathbf{x}'\beta)} = \frac{1}{1+exp(-\mathbf{x}'\beta)}$. Therefore $E(Y|\mathbf{x}) = p(\mathbf{x}) = \frac{exp(\mathbf{x}'\beta)}{1+exp(\mathbf{x}'\beta)} = \frac{1}{1+exp(-\mathbf{x}'\beta)}$ and $Var(Y|\mathbf{x}) = p(\mathbf{x})(1 - p(\mathbf{x}))$. That is, the conditional distribution of the response variable Y follows a binomial distribution with probability given by the conditional mean $p(\mathbf{x})$. (iii) The conditional mean is modeled as $p(\mathbf{x}) = \frac{exp(\mathbf{x}'\beta)}{1+exp(\mathbf{x}'\beta)} = \frac{1}{1+exp(-\mathbf{x}'\beta)}$. (iv) The $\mathbf{x}$ are measured without error.

In summary, the conditional mean of the regression equation must be formulated to be bounded between 0 and 1. The logistic model has this property and the parameters in the logistic model have a natural interpretation in terms of the odds ratio. Also the binomial, not the normal (as in linear regression), distribution describes the distribution of the errors and will be the statistical distribution upon which the analysis is based. The principles that guide an analysis using linear regression in Neter et al. (72) will also guide us in logistic regression.

2.2.3 Maximum Likelihood for Logistic Regression

Just as in simple linear regression, the data will be a sample of say m independent observations. In the case of p predictor variables, the data are given by

$$(\mathbf{x}_1, Y_1), (\mathbf{x}_2, Y_2), ...(\mathbf{x}_m, Y_m),$$

where Y_i represents the value of the dichotomous response variable and $\mathbf{x}_i = (x_{i1}, ..., x_{ip})$ is the value of the predictor variables for the ith subject. Furthermore, assume that the response variable has been coded as 0 or 1, representing the absence (0) or the presence (1) of the event, respectively.

In simple linear regression it was assumed in Neter et al. (72) that

$$E(Y|X) = \beta_0 + \beta_1 X,$$

and used least-squares to estimate the parameters $(\beta_0, \beta_1, \sigma^2)$ that minimized the sum of squares

$$\sum (Y_i - \beta_0 - \beta_1 X)^2.$$

For many reasons, this will not work for logistic regression. In the model 2.1 with $Y_i \sim Binomial(1, p_i(\mathbf{x}))$, we no longer have a direct connection between Y_i and $\mathbf{x}_i\beta$ (which is why we need a link function). The estimation method that is most commonly used is maximum likelihood to obtain parameter estimates. The function for the conditional probability

$$Pr(Y_i = 1|\mathbf{x}) = \frac{exp(\mathbf{x_i'}\beta)}{1 + exp(\mathbf{x_i'}\beta)} = \Psi(-\mathbf{x_i'}\beta), \tag{2.5}$$

implies that

$$Pr(Y_i = 0|\mathbf{x}) = \frac{1}{1 + exp(\mathbf{x'_i}\beta)} = 1 - \Psi(-\mathbf{x'_i}\beta). \tag{2.6}$$

Thus, for those pairs $(\mathbf{x}_i, Y_i)$ where $Y_i = 1$ the contribution to the likelihood function is $\Psi(-\mathbf{x'_i}\beta)$, and for those pairs $(\mathbf{x}_i, Y_i)$ where $Y_i = 0$ the contribution to the likelihood function is $1 - \Psi(-\mathbf{x'_i}\beta)$. A general way for describing the likelihood for a single observation is

$$Pr(Y_i) = \Psi^{Y_i}[1 - \Psi]^{1-Y_i}. \tag{2.7}$$

To see why, note that when $Y_i = 1$ the result is

$$\Psi^1[1 - \Psi]^0 = \Psi$$

and when $Y_i = 0$ the result is

$$\Psi^0[1 - \Psi]^{1-0} = 1 - \Psi.$$

Because the observations are assumed to be independent, following Hogg and Craig (46), the likelihood function for a sample of m independent observations is obtained as the product of the individual terms for each observation or

$$L = \prod_{i=1}^{m} \Psi^{Y_i}[1 - \Psi]^{1-Y_i}. \tag{2.8}$$

It is easier to manipulate the log of the likelihood function

$$\log L = \sum_{i=1}^{m} [Y_i log\Psi_i + (1 - Y_i)log(1 - \Psi_i)]. \tag{2.9}$$

If we choose the values of the parameters that maximize the log of the likelihood those same values will also maximize the likelihood. To obtain the likelihood estimates, we therefore have to take the partial derivatives of $\log L$ with respect β, set the derivatives equal to zero and solve for β. Differentiating the log likelihood function 2.9 with respect to β yields the first derivatives for the maximum likelihood (ML) solution:

$$\frac{\partial logL}{\partial \beta} = \sum_{i=1}^{m}(Y_i - \Psi_i)\mathbf{x_i} = 0. \tag{2.10}$$

This result is due to the fact that for the logistic distribution $\partial \Psi(.) = \Psi(.)(1 - \Psi(.))$. Note that

$$\frac{\partial logL}{\partial \beta} = \left(\begin{array}{ccccc} \frac{\partial logL}{\partial \beta_0} & \frac{\partial logL}{\partial \beta_1} & . & . & \frac{\partial logL}{\partial \beta_p} \end{array} \right)$$

is a vector. In the simple case of a linear effect, for one predictor variable there are two parameters to estimate, β_0 and β_1. The two resulting equations from equation 2.10 are

$$\frac{\partial logL}{\partial \beta_0} = \sum_{i=1}^{m}(Y_i - \Psi_i) = 0$$

and

$$\frac{\partial logL}{\partial \beta_1} = \sum_{i=1}^{m}(Y_i - \Psi_i)\mathbf{x_i} = 0.$$

These are not linear in the parameters. Hence iterative methods are required to solve them (1; 44). The Newton-Raphson method is an iterative method for solving nonlinear equations. To implement the Newton-Raphson method we require the second partial derivatives of log likelihood function 2.9. The second

partial derivatives are obtained as

$$\frac{\partial^2 logL}{\partial\beta\partial\beta'} = -\sum_{i=1}^{m} \Psi(1 - \Psi_i)\mathbf{x_i}\mathbf{x_i'}. \tag{2.11}$$

Note that $\frac{\partial^2 logL}{\partial\beta\partial\beta'}$ is a matrix called the Hessian matrix. In the Newton-Raphson method, provisional estimates for the vector of parameters β on iteration $\imath$ are improved by

$$\beta_{\imath+\mathbf{1}} = \beta_\imath - \frac{\partial logL}{\partial\beta_\imath}(\frac{\partial^2 logL}{\partial\beta_\imath\partial\beta_\imath'})^{-1}. \tag{2.12}$$

This iterations continue until the changes in the parameter estimates and/or likelihood value are sufficiently small. At this point the solution is said to have converged, and the large-sample variance-covariance matrix of the maximum likelihood estimator is then obtained as the negative inverse of the matrix of second derivatives. Standard errors of the parameter estimates are obtained as the square root values of the diagonal entries of this (negative inverse) matrix. The maximum likelihood estimates and their accompanying standard errors can be used to compute asymptotic $z-$statistics (*i.e.*, Wald statistics) or construct confidence intervals. More discussion regarding this maximum likelihood estimation procedure and iterative algorithm can be found in (1; 19; 28; 44) and (63).

2.2.4 Goodness of fit

Once the model has been applied, we need to assess how well it fits the data, or how close the model predicted values are to the corresponding values. Test statistics that assess fit in this manner are known as goodness-of-fit statistics. They address the difference between observed and predicted values, or their ratio, in some appropriate manner. Departures of the predicted proportions from

the observed proportions should be essentially random. The test statistics have approximate chi-square distributions when the sample size is sufficiently large. Two traditional goodness-of-fit tests are the Pearson chi-square and the likelihood ratio chi-square, also known as the deviance.

2.2.5   Statistical software

Several procedures in the statistical software can be used to perform logistic regression, including the **LOGISTIC** procedure, the **CATMOD** procedure, and the **GENMOD** procedure in SAS$^{©}$. The **LOGISTIC** procedure is designed primarily for logistic regression analysis and it provides useful information such as odds ratio estimates and model diagnosis. The **CATMOD** procedure is a general procedure designed to fit models to functions of categorical response variables. **PROC GEMOD** is a procedure for analyzing generalized linear models, of which logistic is a simple case. Other statistical software which are capable of performing logistic regression include MathCad$^{©}$, R$^{©}$, and S-plus$^{©}$. In this section, attention is focused, with the help of an example, on the use of the MathCad$^{©}$ software and on the use of the **PROC LOGISTIC** procedure in SAS$^{©}$ to perform logistic regression.

2.3   Example

The following example, from Antinori (8) is based on a study to investigate if higher pregnancy rates may be achieved by increasing the number of embryos transferred. The success of IVF depends on three major factors: embryo quality, the number of embryos transferred and uterine receptivity. Some authors have

claimed that higher pregnancy rates may be achieved by increasing the number of embryos transferred, while others have found no such linear correlation and recommend that the number of embryos transferred be reduced to two or three to decrease the possibility of multiple implantation.

To investigate these assertions, Antinori (8) compared two groups of patients who failed at least three previous IVF attempts: Group A consists of 89 patients and group B consists of 92 patients. Patients in group A received at least four embryos while patients in group B received up to three embryos. The two groups were similar in age, the duration of infertility, the number of previous IVF attempts and the indication for IVF. Embryo quality was similar for both groups. The resulting number of clinical pregnancies reached 46 in group A and 26 in group B.

We will illustrate on how to analyze this using the logistic model defined above. We employ the model

$$logit(p_{ij}) = \beta_0 + \beta_1 embryo.$$

The variable *embryo* takes the value 1 if patient receive at least four embryos and is 0 otherwise. The quantity β_0 is the log odds of pregnancy for patients who received up to three embryos. Since patients who received up to three embryos are described by the intercept, this group is known as the reference cell in this model. The parameter β_1 is the increment in log odds for patients who received at least four embryos. Table 2.1 displays the probabilities and odds predicted by this model.

We can calculate the odds ratio for patients who received at least four embryos versus patients who received at most three embryos by forming the corresponding ratio of the odds of pregnancy namely, $\frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = e^{\beta_1}$. Thus we can obtain odds

Embryo	p_{ij}	Odds of pregnancy
< 4	$\frac{e^{\beta_0}}{1+e^{\beta_0}}$	e^{β_0}
≥ 4	$\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$	$e^{\beta_0+\beta_1}$

Table 2.1. Model-Predicted Probabilities and Odds

as functions of the model in logistic regression.

To employ the maximum likelihood estimation procedure illustrated above, we will need to construct the likelihood for the data in Table 2.1. From Table 2.1, we can immediately get the contributions to the log likelihood as follows: Patients who received at most three embryos contribute to the log likelihood;

$$L_1(\beta_0, \beta_1) = 26\ln(\frac{1}{1+e^{-\beta_0}}) + 66\ln(\frac{1}{1+e^{\beta_0}}),$$

while patients who received at least four embryos with contribute

$$L_2(\beta_0, \beta_1) = 46\ln(\frac{1}{1+e^{-\beta_0-\beta_1}}) + 43\ln(\frac{1}{1+e^{\beta_0+\beta_1}}).$$

The total log-likelihood is given by

$$L(\beta_0, \beta_1) = L_1(\beta_0, \beta_1) + L_2(\beta_0, \beta_1) \tag{2.13}$$

2.3.1 Analysis using MathCad$^{©}$

The following MathCad$^{©}$ code will maximize $L(\beta_0, \beta_1)$ and produce the maximum likelihood estimates (MLE) of β_0, and β_1,.

ORIGIN:= 1

$L(\beta_0, \beta_1) := L_1(\beta_0, \beta_1) + L_2(\beta_0, \beta_1)$

$\beta_0 := 0.9$ $\qquad$ $\beta_1 := 1$

Maximize(L, β_0, β_1)

The first step **ORIGIN:=1** confirms that any counting done must begin at 1. By default, MathCad© begins counting at 0. The second step defines the function to be maximized, which is the log likelihood $L(\beta_0, \beta_1)$. The third step defines the starting values of the parameters to be estimated. MathCad© is not sensitive to the choice of the starting values. However in chapter 4 we will illustrated how in general the starting values can be estimated. The **Maximize** command will produce the maximum likelihood estimates (MLE) of the defined function. Mathcad© uses a variety of optimization techniques. If the problem is linear, the Linear method is applied. If Quadratic, the Quadratic method is used (if the Solving and Optimization Extension Pack is installed). If the problem calls for a nonlinear solver, Mathcad© uses the Conjugate Gradient solver; if that fails to converge, the Levenberg-Marquardt solver; if that too fails, the Quasi-Newton solver. These methods use different algorithms to determine the curvature and direction in which the search is to proceed.

Although Mathcad© automatically determines the kind of problem you are solving and attempts appropriate solving algorithms until one of the methods converges, you can right-click the function and choose a specific method from the menu if necessary.

The output from this analysis is as follows

$$\text{Maximize}(L, \beta_0, \beta_1) = \begin{pmatrix} -0.932 \\ 0.999 \end{pmatrix}$$

When optimizing a function of more than one parameter, Mathcad© will

return a vector of results. The first element in this vector corresponds to the first variable after the function name in the call to **MAXIMIZE**, and so on. Thus $\hat{\beta}_0 = -0.932$ and $\hat{\beta}_1 = 0.999$

We estimated the variance-covariance matrix from the observed Fisher information (46). The observed Fisher information of the data set is given by:

$$I(\beta_0, \beta_1) = \begin{pmatrix} \frac{\partial^2 L}{\partial \beta_0^2} & \frac{\partial^2 L}{\partial \beta_0 \partial \beta_1} \\ \frac{\partial^2 L}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 L}{\partial \beta_1^2} \end{pmatrix} \tag{2.14}$$

.

The approximate variances and covariances are found, respectively, in the matrix

$$\{-I(\beta_0, \beta_1)\}^{-1} \approx \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{2.15}$$

,

where $a_{11} = var(\beta_0)$, $a_{12} = a_{21} = cov(\beta_0, \beta_1)$, and $a_{22} = var(\beta_1)$.

The following MathCad$^{\copyright}$ code will yield the approximate variance-covariance matrix.

$$I(\beta_0, \beta_1) := \begin{pmatrix} \frac{\partial^2 L}{\partial \beta_0^2} & \frac{\partial^2 L}{\partial \beta_0 \partial \beta_1} \\ \frac{\partial^2 L}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 L}{\partial \beta_1^2} \end{pmatrix}$$

$$\{-I(-0.932, 0.999)\}^{-1}$$

The output for the variance-covariance matrix is as follows:

$$(-I(-0.932, 0.999))^{-1} = \begin{pmatrix} 0.054 & -0.054 \\ -0.054 & 0.099 \end{pmatrix}$$

Thus the estimated standard errors for $\hat{\beta}_0$ and $\hat{\beta}_1$ are $\sqrt{0.054} = 0.232$ and $\sqrt{0.099} = 0.315$ respectively.

2.3.2   Analysis using SAS$^©$

We will illustrate how the above example can be analyzed using **Proc Logistic** from SAS$^©$. The **LOGISTIC** procedure was designed specifically to fit logistic regression models. The response variable and the explanatory variables are specified in the **MODEL** statement, and it fits the model via maximum likelihood estimation illustrated above. **PROC LOGISTIC** produces the parameter estimates, their standard errors, and statistics to assess model fit. The following SAS$^©$ code creates the data set and invokes **PROC LOGISTIC**.

```
    data embryo;
input embryo pregnancy count @@;
0 0 66 0 1 26
1 0 43 1 1 46
;
run;
proc logistic descending;
freq count;
model pregnancy=embryo / scale=none aggregate;
run;
```

The variable pregnancy takes the value 1 if the subject is conceived and is 0 otherwise. By default, **PROC LOGISTIC** orders the response variable values alphanumerically so that, for these data, it bases its model on the probability

of the smallest value, $Pr(pregnancy = 0)$, which is the probability of no conception. The **DESCENDING** option in the **PROC LOGISTIC** statement requests that the response value ordering be reversed. For these data, this means that **PROC LOGISTIC** will model probability of conception. The **SCALE** option produces goodness-of-fit statistics; the **AGGREGATE** option requests that **PROC LOGISTIC** treat each unique combination of the explanatory variable as a distinct group in computing the goodness of fit statistics.

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.9316	0.2315	16.1864	<.0001
embryo	1	0.9990	0.3140	10.1209	0.0015

Table 2.2. Analysis of Maximum Likelihood Estimates: Antinori data

Table 2.2 lists the estimated model parameters, their standard errors, Wald chi-square tests, and $p - values$. A Wald test is a statistic that takes the form of the squared valve ratio for the estimate to its standard error; it follows an approximately chi-square distribution when the sample size is sufficiently large Phillips (79). The variable embryo is significant compared to a significance level of 0.05, with a Wald statistic of 16.1864. The model equation can be written as follows:

$$logit(p_{ij}) = -0.932 + 0.999 embryo$$

Table 2.3 lists the parameter interpretations, and Table 2.4 displays the predicted logits and odds of coronary disease.

The odds ratio is $exp(\beta_1) = exp(0.990) = 2.72$. Patients who received more

Parameter	Estimate	Standard Error	Interpretation
β_0	-0.9316	0.2315	log odds of pregnancy for patients who received at most 3 embryos
β_1	0.9990	0.3140	increment to log odds for patients who received at least 4 embryos

Table 2.3. Interpretation of parameters: Antinori data

Embryo	Logit	Odds of pregnancy
≤ 3	$\beta_0 = -0.9316$	$e^{\beta_0} = 0.394$
≥ 4	$\beta_0 + \beta_1 = 0.067$	$e^{\beta_0+\beta_1} = 1.07$

Table 2.4. Model-Predicted Logits and Odds of pregnancy: Antinori data

than three embryos, in the study have three times higher odds for pregnancy than patients who received at most three embryos, in the study. Notice that both **Proc Logistic** and MathCad$^{©}$ give similar conclusions.

2.3.3   Model Fit

Here we need to assess how close the predicted odds are to the observed values.

Embryo	observed proportion	observed odds
≤ 3	$\frac{26}{92}$	$\frac{\frac{26}{92}}{1-\frac{26}{92}} = 0.3939394$
≥ 4	$\frac{46}{89}$	$\frac{\frac{46}{89}}{1-\frac{46}{89}} = 1.069767$

Table 2.5. Observed proportions and odds

From table 2.5 we can immediately calculate the observed odds ratio as follows:

$$\frac{1.069767}{0.3939394} = 2.715564.$$

The observed odds (odds ratio) are very much similar to the predicted odds. Thus our model fits the data well.

2.4 Subject-specific models with one random effect

In longitudinal studies, repeated measurements of a response variable and a set of covariates are made on subjects across occasions. Because the within-subject measurements are likely to be positively correlated, the correlation must be accounted for by analysis appropriate to the longitudinal data. The standard logistic regression model described above, fails in its assumptions to accurately characterize the dependence in the data. Basically, the standard logistic regression model assumes that the observations are independent, which they clearly are not when they are clustered within individuals. One solution to this problem is to generalize the model to the case of a combination of fixed (e.g., treatment) and random effects. The random effects allow the correlation between the repeated measurements to be incorporated into the estimates of parameters.

2.4.1 Model fitting

To set the notation, let i denote the individuals and let j denote the treatment. Let Y_{ij} be the value of the dichotomous outcome variable, coded 0 or 1, associated with treatment j nested within individual i. The logistic regression model is written in terms of the log odds (*i.e.*, the logit) of the probability of a response, denoted p_{ij}. Considering a random-intercept model, augmenting the standard logistic regression model 2.1 with a single random effect yields:

$$\log \frac{p_{ij}}{1 - p_{ij}} = \mathbf{x}_{ij}' \beta + v_i \tag{2.16}$$

where $\mathbf{x_{ij}}$ is the $(p+1) \times 1$ covariate vector (includes a 1 for the intercept), β is the $(p+1) \times 1$ vector of unknown regression parameters, and v_i is the random subject effect. These are assumed to be distributed in the population as $N(0, \sigma^2)$.

Regression estimates are subject-specific, that is they describe the individual's response (conditional estimates, conditional on the random effect). Conditional estimates represent the effect of a regressor on the outcome controlling for or holding constant the value of the random subject effect. On the other hand, the estimates from the standard logistic regression are "marginal" or "population-averaged" estimates. Marginal estimates represent the effect of a regressor averaging over the population of subjects.

2.4.2 Estimation

To set the notation, we assume that there are $i = 1, ...m$ subjects as before, each with $j = 1, 2, ..., n_i$ repeated observations. In the case of p predictor variables, the data are given by

$$(\mathbf{x}_1, \mathbf{Y}_{1j}), (\mathbf{x}_2, \mathbf{Y}_{2,j}), ...(\mathbf{x}_m, \mathbf{Y}_{mj}),$$

where $\mathbf{Y}_{ij} = (Y_{i1}, Y_{i2}, ..., Y_{in_i})$ represents the value of the dichotomous response variables, and $x_i = (x_{i1}, ..., x_{ip})$ is the value of the predictor variables for the ith subject. The single response Y_{ij} is the jth response of the ith subject. Furthermore, $Y_{ij} = 1$ for success and $Y_{ij} = 0$ for failure. We consider estimation of a random-intercepts mixed model, that is

$$\log \frac{p_{ij}}{1 - p_{ij}} = \mathbf{x}_{ij}'\beta + v_i \tag{2.17}$$

We would like to obtain comparable functions to equations 2.5 and 2.6 respectively. A comparable function to equation 2.5 is the conditional probability of a 1 response, conditional on the random effect v_i, which is given by:

$$Pr(Y_{ij} = 1|v_i) = \Psi(z_{ij}),\tag{2.18}$$

where the standard logistic cumulative distribution is given by equation 2.2 and $z_{ij} = \mathbf{x}'_{\mathbf{ij}}\beta + v_i$. Whereas a comparable function to equation 2.6 is the probability of a 0 response which is simply

$$Pr(Y_{ij} = 0|v_i) = 1 - \Psi(z_{ij})\tag{2.19}$$

The next step is to assume that the within-subject measurements are conditionally independent given the random subject effect (i.e.,the random effects account completely for the correlation of the data within subjects). This assumption is critical and is known in Agresti (1), as the conditional independence assumption. Because the within-subject measurements are assumed to be conditionally independent, following Hogg and Craig, the conditional likelihood of n_i measurements within the ith subject is given by:

$$\ell(\mathbf{Y_i}|v) = \prod_{j=1}^{n_i}[\Psi(z_{ij})]^{Y_{ij}}[1 - \Psi(z_{ij})]^{1-Y_{ij}}\tag{2.20}$$

The conditional likelihood 2.20 and the likelihood in equation 2.8 are functionally of the same form. In the standard logistic regression case, for which equation 2.5 applies, we can multiply the probabilities from each subject together to yield the likelihood of the joint pattern of all m outcomes from the subjects. Similarly, in the random effect case, we can multiply the probabilities of each time-point

together within a subject to yield the conditional likelihood of the joint pattern of the n_i outcomes across time for that subject. Here, it is a conditional likelihood because these n_i observations are independent (and therefore can be multiplied together as in equation 2.20 only conditional on the random effect. To get to the likelihood of the n_i response patterns for all the m subjects, we need to have an expression for the likelihood of $\mathbf{Y_i}$ that does not depend on the random effects. We can arrive at such an expression by integrating over the distribution of the random effects. This yields the marginal probability for $\mathbf{Y_i}$ in the population of subjects as:

$$h(\mathbf{Y_i}) = \int_{\upsilon} \ell(\mathbf{Y_i}|\upsilon)g(\upsilon)d\upsilon \tag{2.21}$$

Where $g(\upsilon)$ represents the population distribution of the random effects υ, namely $N(0, \sigma^2)$.

The idea behind this isn't too hard to grasp. Essentially we want to consider the conditional likelihood, which depends on the random effect, for all possible values of the random effect, and thereby obtain the aggregated or marginal likelihood.

We can now form the marginal likelihood of the response patterns $\mathbf{Y_i}$ from all subjects, and thus the total sample, by multiplying each of the subject's marginal likelihoods together.

Namely,

$$L = \prod_{i=1}^{m} h(\mathbf{Y_i}) \tag{2.22}$$

or

$$\log L = \sum_{i=1}^{m} \log h(\mathbf{Y_i}). \tag{2.23}$$

It is easier to manipulate the log-likelihood in equation 2.23. If we choose the values of the parameters that maximize the log-likelihood in equation 2.23 those same values will also maximize the likelihood in equation 2.22. For this, we let the parameter vector η represent either the regressors β or the variance parameter σ, then taking derivatives

$$\frac{\partial \log L}{\partial \eta} = \sum_{i=1}^{m} h^{-1}(\mathbf{Y_i}) \frac{\partial h(\mathbf{Y_i})}{\partial \eta}. \tag{2.24}$$

We first manipulate the marginal likelihood in equation 2.21 as follows:

$$h(\mathbf{Y}_i) = \int_{\upsilon} \ell(\mathbf{Y}_i | \upsilon) g(\upsilon) d\upsilon$$

$$= \int_{\upsilon} (\prod_{j=1}^{n_i} [\Psi(z_{ij})]^{Y_{ij}} [1 - \Psi(z_{ij})]^{1-Y_{ij}}) g(\upsilon) d\upsilon$$

$$= \int_{\upsilon} [\exp(log\{\prod_{j=1}^{n_i} [\Psi(z_{ij})]^{Y_{ij}} [1 - \Psi(z_{ij})]^{1-Y_{ij}}\})] g(\upsilon) d\upsilon$$

$$= \int_{\upsilon} [\exp(\sum_{j=1}^{n_i} [Y_{ij} log\Psi(z_{ij}) + (1 - Y_{ij}) log(1 - \Psi(z_{ij}))])] g(\upsilon) d\upsilon.$$

And so, denoting $\ell(\mathbf{Y_i} | \upsilon)$ by ℓ_i, we get

$$\frac{\partial h(\mathbf{Y_i})}{\partial \eta} = \int_{\upsilon} \sum_{j=1}^{n_i} [\frac{\mathbf{Y}_{ij}}{\Psi(z_{ij})} \partial \Psi(z_{ij}) + \frac{1-\mathbf{Y}_{ij}}{1-\Psi(z_{ij})} (-\partial \Psi(z_{ij}))] \frac{\partial z_{ij}}{\partial \eta} \ell_i g(\upsilon) d\upsilon$$

$$= \int_{\upsilon} \sum_{j=1}^{n_i} \frac{\mathbf{Y}_{ij} - \Psi(z_{ij})}{\Psi(z_{ij})(1-\Psi(z_{ij}))} (\partial \Psi(z_{ij})) \frac{\partial z_{ij}}{\partial \eta} \ell_i g(\upsilon) d\upsilon,$$

yielding

$$\frac{\partial \log L}{\partial \eta} = \sum_{i=1}^{m} h^{-1}(\mathbf{Y}_i) \int_{\upsilon} \sum_{j=1}^{n_i} \frac{\mathbf{Y}_{ij} - \Psi(z_{ij})}{\Psi(z_{ij})(1 - \Psi(z_{ij}))} (\partial \Psi(z_{ij})) \frac{\partial z_{ij}}{\partial \eta} \ell_i g(\upsilon) d\upsilon, \tag{2.25}$$

where $\partial \Psi(z_{ij})$ equals the probability density function (*pdf*), which for the logistic distribution is $\Psi(z_{ij})[1 - \Psi(z_{ij})]$; and where

$$\frac{\partial z_{ij}}{\partial \beta} = \mathbf{x'_{ij}} \tag{2.26}$$

$$\frac{\partial z_{ij}}{\partial \sigma} = \frac{v_i}{\sigma}. \tag{2.27}$$

Actually this idea is also illustrated in Hedeker et.al (44) except that Hedeker have standardized the random effects. Similarly, we can obtain the second derivatives

$$\frac{\partial^2 \log L}{\partial \eta^2} = \sum_{i=1}^{m} \left\{ h^{-1}(\mathbf{Y_i}) \frac{\partial^2 h(\mathbf{Y_i})}{\partial \eta^2} - h^{-2}(\mathbf{Y_i}) \left\{ \frac{\partial h(\mathbf{Y_i})}{\partial \eta} \right\}^2 \right\}. \tag{2.28}$$

Equations 2.25 and 2.28 can now be used in the Newton-Raphson procedure outlined for the ordinary logistic regression model in the previous section to obtain the maximum likelihood estimates (MLE). At convergence, the large-sample variance covariance matrix of the parameter estimates is then obtained as the inverse of the information matrix. The square root values of the diagonal elements of this matrix can be used to obtain Wald statistics or construct asymptotic confidence intervals for the model parameters.

2.4.3 Integration over the Random Effect Distribution

In order to solve the above likelihood solutions, integration over the random effects distribution must be performed. Various approximations for evaluating the integral 2.21 have been proposed in the literature including methods based on the first or second order Taylor expansions (87). Numerical integration can also be used to perform the integration over the random-effects distribution. Specifically, if the assumed distribution is normal, Gauss Hermite quadrature can approximate the integral 2.21 to any practical degree of accuracy (93). The integration is approximated by a summation on a specified number of quadrature points Q for each dimension of the integration.

Consider the Gaussian integration formula for the Hermite polynomial in Stroud and Secrest (93). The Hermite polynomials are defined over $[-\infty, \infty]$ and the weighting function of Hermite polynomials is

$$w(x) = e^{-x^2}. \tag{2.29}$$

Therefore, Gauss Hermite quadrature naturally gives the integration for:

$$\int_{-\infty}^{\infty} f(x)e^{-x^2}dx. \tag{2.30}$$

Thus, Gauss Hermite quadrature can be naturally be associated with normal distribution as follows: Suppose we were to evaluate

$$\int_{-\infty}^{\infty} f(y)e^{-\frac{(y-u)^2}{\sqrt{2}\sigma}} dy, \tag{2.31}$$

where $f(y)$ is a function of y. Substituting

$$x = \frac{y-u}{\sqrt{2}\sigma} \tag{2.32}$$

in equation 2.31 yields

$$\int_{-\infty}^{\infty} f(y)e^{-\frac{(y-u)^2}{\sqrt{2}\sigma}} dy = \int_{-\infty}^{\infty} f(\sqrt{2}\sigma x + \mu)e^{-x^2}\sqrt{2}\sigma dx = \sqrt{2}\sigma \sum_{q=1}^{N} \psi_q f(\sqrt{2}\sigma x + \mu), \tag{2.33}$$

where following (93), $\{x_q\}_{q=1}^{N}$ are the roots of order N Hermite polynomial $P_N(x)$ and $\{\psi_q\}_{q=1}^{N}$ are:

$$\psi_q = \int_{-\infty}^{\infty} \prod_{\substack{j=1 \\ j \neq q}}^{N} \frac{x - x_j}{x_i - x_j} e^{-x^2} dx. \tag{2.34}$$

Therefore, if y is normally distributed with mean μ and variance σ^2, the expected value of $f(y)$, $(E[f(y)])$ is given by:

$$\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} f(y)e^{-\frac{(y-u)^2}{\sqrt{2}\sigma}}dy = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} f(\sqrt{2}\sigma x+\mu)e^{-x^2}\sqrt{2}\sigma dx = \frac{1}{\sqrt{\pi}}\sum_{i=1}^{N}\psi_q f(\sqrt{2}\sigma x+\mu),$$

(2.35)

where $\{x_q\}_{i=q}^{N}$ and $\{\psi_q\}_{i=q}^{N}$ are the same as above.

With the numerical Gaussian quadrature integration, the approximation to the marginal likelihood gets better as the number of quadrature points increases. However, as the dimension of the random effects increases, the number of quadrature points increases exponentially; the total number of quadrature points required for all the random effects Q^r, where r is the number of random effects. The numerical quadrature becomes computationally burdensome when there are more than 5 random effects (59).

2.4.4 Statistical software

Fortunately, we do not have to worry about how all these equations are solved; statistical software programs solve them for us. The **NLMIXED** in SAS© procedure fits nonlinear mixed models, that is, models in which both fixed and random effects are permitted to have a nonlinear relationship to the response variable. These models can take various forms, but the most common ones involve a conditional distribution for the response variable given the random effects. **PROC NLMIXED** enables us to specify such a distribution by using either a keyword for a standard form (normal, binomial, Poisson) or SAS© programming statements to specify a general distribution.

PROC NLMIXED fits the specified nonlinear mixed model by maximizing

an approximation to the likelihood integrated over the random effects. Different approximations to the integral (2.21) are available, and the two principal ones are the one we used, Gaussian quadrature and a first order Taylor series approximation. There are a variety of alternative optimization techniques; the default is the Newton Raphson described in the previous sections. Standard errors are obtained by the Delta method. For the theory and computational techniques of **PROC NLMIXED**, the book written by Pinheiro and Bates (80) is strongly recommended. Other software which are capable of handling the random effects include MathCad$^©$, R$^©$, S-Plus$^©$ and WinBugs$^©$. In this section, attention is focused, with the help of an example, on the use of the **PROC NLMIXED** procedure in SAS$^©$ and on the use of the MathCad$^©$ to perform random effects model.

2.5 Illustration

To illustrate application of the mixed-effects logistic regression model, we will present analysis from Ezzet and Whitehead (30). We will like to point out that we are using this example to illustrate how to analyze a complete crossover design. This example is not an infertility trial. As it was mentioned in the previous section, it is impossible to get a complete crossover design in infertility trials (except if no woman conceives in the first period).

Specifically, we will examine a crossover trial comparing inhalation devices for asthmatics. A crossover clinical trial has been conducted by $3M - Riker$ to compare the suitability of two new inhalation devices ($A = 1$ and $B = -1$) in patients who currently using a standard inhaler device delivering Salbutamol. Group 1 used device A for one week followed by device B for another week.

Group 2 used the devices in reverse order. No wash out period was felt necessary. Patients reported whether there were particular features which they liked about each device, and their responses were coded (1) for yes and (0) for no. The experimental design is presented in Table 2.6.

Group	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$	$(0,*)$	$(1,*)$
1	57	15	41	26	2	2
2	54	32	16	38	3	2

Table 2.6. Table of counts for the asthma inhaler study: Ezzet and Whitehead

We will analyse responses from patients who tried both devices and also we will assume that there is no evidence of carry over effects. Here, we will like to remind the reader that we emphasized in the previous sections that if there is any evidence of carry over effects, we will not recommend the crossover designs.

2.5.1 Analysis using MathCad©

Firstly, we will need to construct the likelihood of the data. Let Y_{ijk} denote the response from the ith subject in the group k during the jth period, $i = 1, ..., n_j$, $j = 1, 2$, $k = 1, 2$. Our approach utilizes a logistic fixed effect model in the form:

$$logit(p_{(ij)k}) = \beta_0 + \frac{\beta_1}{2} treatment_{(ij)k} + \frac{\beta_2}{2} period_j + \upsilon_i, \qquad (2.36)$$

where $\upsilon_i \sim N(0, \sigma^2)$. The parameters β_1 and β_2 represent treatment and period effects respectively. Denote the observed value $\mathbf{Y}_{(ij)k}$ by $y_{(ij)k}$. All the inferences are to be based on the likelihood

$$L = \prod_{k=1}^{2} \prod_{i}^{n_k} P(\mathbf{Y}_{(i1)k} = y_{(i1)k}, \mathbf{Y}_{(i2)k} = y_{(i2)k}). \qquad (2.37)$$

The tables 2.7 and 2.8 display the group, the outcomes, and the contributed likelihood for the neither period effects nor carryover effects scenario and the period effects scenario respectively.

Group	Outcome	Likelihood contribution
1	$\mathbf{Y}_1 = 0, \mathbf{Y}_2 = 0$	$L_{100} = \int_v \frac{1}{1+exp\{\beta_0+\frac{\beta_1}{2}+v\}} \frac{1}{1+exp\{\beta_0-\frac{\beta_1}{2}+v\}} g(v)dv$
1	$\mathbf{Y}_1 = 0, \mathbf{Y}_2 = 1$	$L_{101} = \int_v \frac{1}{1+exp\{\beta_0+\frac{\beta_1}{2}+v\}} \frac{exp\{\beta_0-\frac{\beta_1}{2}+v\}}{1+exp\{\beta_0-\frac{\beta_1}{2}+v\}} g(v)dv$
1	$\mathbf{Y}_1 = 1, \mathbf{Y}_2 = 0$	$L_{110} = \int_v \frac{1+exp\{\beta_0+\frac{\beta_1}{2}+v\}}{1+exp\{\beta_0+\frac{\beta_1}{2}+v\}} \frac{1}{1+exp\{\beta_0-\frac{\beta_1}{2}+v\}} g(v)dv$
1	$\mathbf{Y}_1 = 1, \mathbf{Y}_2 = 1$	$L_{111} = \int_v \frac{1+exp\{\beta_0+\frac{\beta_1}{2}+v\}}{1+exp\{\beta_0+\frac{\beta_1}{2}+v\}} \frac{1+exp\{\beta_0-\frac{\beta_1}{2}+v\}}{1+exp\{\beta_0-\frac{\beta_1}{2}+v\}} g(v)dv$
2	$\mathbf{Y}_1 = 0, \mathbf{Y}_2 = 0$	$L_{200} = \int_v \frac{1}{1+exp\{\beta_0+\frac{\beta_1}{2}+v\}} \frac{1}{1+exp\{\beta_0-\frac{\beta_1}{2}+v\}} g(v)dv$
2	$\mathbf{Y}_1 = 0, \mathbf{Y}_2 = 1$	$L_{210} = \int_v \frac{1+exp\{\beta_0+\frac{\beta_1}{2}+v\}}{1+exp\{\beta_0+\frac{\beta_1}{2}+v\}} \frac{1}{1+exp\{\beta_0-\frac{\beta_1}{2}+v\}} g(v)dv$
2	$\mathbf{Y}_1 = 1, \mathbf{Y}_2 = 0$	$L_{201} = \int_v \frac{1}{1+exp\{\beta_0+\frac{\beta_1}{2}+v\}} \frac{exp\{\beta_0-\frac{\beta_1}{2}+v\}}{1+exp\{\beta_0-\frac{\beta_1}{2}+v\}} g(v)dv$
2	$\mathbf{Y}_1 = 1, \mathbf{Y}_2 = 1$	$L_{211} = \int_v \frac{1+exp\{\beta_0+\frac{\beta_1}{2}+v\}}{1+exp\{\beta_0+\frac{\beta_1}{2}+v\}} \frac{1+exp\{\beta_0-\frac{\beta_1}{2}+v\}}{1+exp\{\beta_0-\frac{\beta_1}{2}+v\}} g(v)dv$

Table 2.7. Contributed likelihood from each outcome for the treatment only scenario

Group	Outcome	Likelihood contribution
1	$\mathbf{Y}_1 = 0, \mathbf{Y}_2 = 0$	$L_{100} = \int_v \frac{1}{1+exp\{\beta_0+\frac{\beta_1}{2}+\frac{\beta_2}{2}+v\}} \frac{1}{1+exp\{\beta_0-\frac{\beta_1}{2}-\frac{\beta_2}{2}+v\}} g(v)dv$
1	$\mathbf{Y}_1 = 0, \mathbf{Y}_2 = 1$	$L_{101} = \int_v \frac{1}{1+exp\{\beta_0+\frac{\beta_1}{2}+\frac{\beta_2}{2}+v\}} \frac{exp\{\beta_0-\frac{\beta_1}{2}-\frac{\beta_2}{2}+v\}}{1+exp\{\beta_0-\frac{\beta_1}{2}-\frac{\beta_2}{2}+v\}} g(v)dv$
1	$\mathbf{Y}_1 = 1, \mathbf{Y}_2 = 0$	$L_{110} = \int_v \frac{1+exp\{\beta_0+\frac{\beta_1}{2}+\frac{\beta_2}{2}+v\}}{1+exp\{\beta_0+\frac{\beta_1}{2}+\frac{\beta_2}{2}+v\}} \frac{1}{1+exp\{\beta_0-\frac{\beta_1}{2}-\frac{\beta_2}{2}+v\}} g(v)dv$
1	$\mathbf{Y}_1 = 1, \mathbf{Y}_2 = 1$	$L_{111} = \int_v \frac{1+exp\{\beta_0+\frac{\beta_1}{2}+\frac{\beta_2}{2}+v\}}{1+exp\{\beta_0+\frac{\beta_1}{2}+\frac{\beta_2}{2}+v\}} \frac{1+exp\{\beta_0-\frac{\beta_1}{2}-\frac{\beta_2}{2}+v\}}{1+exp\{\beta_0-\frac{\beta_1}{2}-\frac{\beta_2}{2}+v\}} g(v)dv$
2	$\mathbf{Y}_1 = 0, \mathbf{Y}_2 = 0$	$L_{200} = \int_v \frac{1}{1+exp\{\beta_0+\frac{\beta_1}{2}+\frac{\beta_2}{2}+v\}} \frac{1}{1+exp\{\beta_0-\frac{\beta_1}{2}-\frac{\beta_2}{2}+v\}} g(v)dv$
2	$\mathbf{Y}_1 = 0, \mathbf{Y}_2 = 1$	$L_{210} = \int_v \frac{1+exp\{\beta_0+\frac{\beta_1}{2}+\frac{\beta_2}{2}+v\}}{1+exp\{\beta_0+\frac{\beta_1}{2}+\frac{\beta_2}{2}+v\}} \frac{1}{1+exp\{\beta_0-\frac{\beta_1}{2}-\frac{\beta_2}{2}+v\}} g(v)dv$
2	$\mathbf{Y}_1 = 1, \mathbf{Y}_2 = 0$	$L_{201} = \int_v \frac{1}{1+exp\{\beta_0+\frac{\beta_1}{2}+\frac{\beta_2}{2}+v\}} \frac{exp\{\beta_0-\frac{\beta_1}{2}-\frac{\beta_2}{2}+v\}}{1+exp\{\beta_0-\frac{\beta_1}{2}-\frac{\beta_2}{2}+v\}} g(v)dv$
2	$\mathbf{Y}_1 = 1, \mathbf{Y}_2 = 1$	$L_{211} = \int_v \frac{1+exp\{\beta_0+\frac{\beta_1}{2}+\frac{\beta_2}{2}+v\}}{1+exp\{\beta_0+\frac{\beta_1}{2}+\frac{\beta_2}{2}+v\}} \frac{1+exp\{\beta_0-\frac{\beta_1}{2}-\frac{\beta_2}{2}+v\}}{1+exp\{\beta_0-\frac{\beta_1}{2}-\frac{\beta_2}{2}+v\}} g(v)dv$

Table 2.8. Contributed likelihood from each outcome for period effects scenario

For each scenario the total likelihood L for this model, is given by

$$\prod_{k=1}^{2} \prod_{r=0}^{1} \prod_{s=0}^{1} L_{krs}{}^{m_{krs}}, \tag{2.38}$$

where m_{krs} denotes the number of outcomes of the form (r, s) among group k. Thus evaluation of L requires the computation of eight distinct integrals of the form 2.19. These integrals are estimated using Gaussian quadrature method. We estimated the standard errors of the parameters from the observed Fisher information matrix. The results are tabulated in Table 2.15. The MathCad$^{©}$ code for this analysis is reserved for the next chapter. In this chapter, we will present the code for the practical analysis using **PROC NLMIXED** from SAS$^{©}$.

2.5.2 Analysis using SAS$^{©}$

The following SAS$^{©}$ code creates a data set **COMPLETE**. The variable **OUTCOME** is the response variable, and **PERIOD** and **TREATMENT** are the explanatory variables. The variables **OUTCOME** takes the value 1 for yes and 0 for no, **PERIOD** takes the value 1 for first period and -1 for second period, and **TREATMENT** takes the value 1 when treatment is A and -1 when treatment is B. Since **PROC NLMIXED** assumes that every time a new value of the **PATIENT** variable is encountered a new **PATIENT** is being analyzed, we need to sort our data by **PATIENT**. The **PROC SORT** arranges the data by **PATIENT**.

```
data complete;
input group patient period treatment outcome;
datalines;
1 1 1 1 0
1 2 1 1 0
1 3 1 1 0
. . . . .
. . . . .
. . . . .
2 279 -1 1 1
;
run;
proc sort data=complete;
by patient;
run;
```

The **PROC NLMIXED** statements to fit the logistic-normal model to these data sets, assuming period effects (no carryover effects) are as follows:

```
proc nlmixed data=complete;
parms beta0=-0.472 beta1=0.2141 beta2=0.3 logsigma=0.1;
sigma=exp(logsigma);
sigma2=sigma**2;
preg=beta0 + beta1*treat/2 + beta2*period/2 + u;
ppreg=exp(preg)/(1+exp(preg));
model outcome binomial(1,ppreg);
random u normal(0,sigma2) subject=patient;
predict preg out=preg;
run;
```

Where β_0, β_1 and β_2 represents constant, treatment effect and period effect respectively. The **PROC NLMIXED** statement invokes the procedure, and the **PARMS** statement defines the parameters and their starting values. In Chapter 4, we illustrate how the initial parameters are chosen. The next statements construct the variable ppreg to correspond to the probability of success, and the **MODEL** statement defines the conditional distribution of the outcomes to be binomial. The **RANDOM** statement defines **U** to be the random effect with subjects defined by the **PATIENT** variable. The **PREDICT** statement constructs predictions for each observations in the input data set. Thus, we can predict the logits, odds and probabilities for each observation. The predicted logits for this example are listed in Appendix *A*.

The output from this analysis is as follows,

The NLMIXED Procedure

Specifications

Data Set	WORK.TWO
Dependent Variable	outcome
Distribution for Dependent Variable	Binomial
Random Effects	u
Distribution for Random Effects	Normal
Subject Variable	patient
Optimization Technique	Dual Quasi-Newton
Integration Method	Adaptive Gaussian Quadrature

Table 2.9. Specification for non-linear mixed model: Ezzet and Whitehead data

The "Specifications" Table 2.9 lists some basic information about the non-linear mixed model that have been specified. Included are the input data set, dependant and patient variables, random effects, relevant distributions, and type of optimization.

Dimensions

Observations Used	558
Observations Not Used	0
Total Observations	558
Subjects	279
Max Obs Per Subject	2
Parameters	4
Quadrature Points	5

Table 2.10. Number of observations, subjects, and parameters: Ezzet and Whitehead data

The "Dimensions" Table 2.10 lists various counts related to the model, including the number of observations, subjects, and parameters. These quantities are useful for checking that data set and model are specified correctly. Also listed is the number of quadrature points that **PROC NLMIXED** has selected based on the evaluation of the log likelihood at the starting values of the parameters.

<center>Parameters</center>

beta0	beta1	beta2	logsigma	NegLogLike
-0.472	0.852	0.21	0.3	362.059054

Table 2.11. Parameters to be estimated: Ezzet and Whitehead data

The "Parameters" output in Table 2.11 lists the parameters to be estimated, their starting values, and the negative log likelihood evaluated at the these starting values.

<center>Iteration History</center>

Iter	Calls	NegLogLike	Diff	MaxGrad	Slope
1	3	362.031538	0.027516	0.646022	-7.6597
2	5	362.028531	0.003007	0.438854	-0.92625
3	7	362.027679	0.000852	0.457175	-0.13841
4	8	362.026268	0.001411	0.299867	-0.0068
5	10	362.024814	0.001454	0.071881	-0.12073
6	12	362.024644	0.00017	0.000511	-0.00034
7	14	362.024644	1.044E-8	4.794E-6	-2.08E-8

NOTE: GCONV convergence criterion satisfied.

Table 2.12. History of the minimization of the negative log likelihood: Ezzet and Whitehead data

The "Iterations" output in Table 2.12 records the history of the minimization of the negative log likelihood. For each iteration of the quasi Newton optimization, values are listed for the number of function calls, the value of the negative log likelihood, the difference from the previous iteration, the absolute value of the largest gradient, and the slope of the search direction. The note at the bottom of the table indicates that the algorithm has converged successfully according to the **GCONV** convergence criterion, a standard criterion computed using a quadratic form in the gradient and inverse Hessian.

The "Fit Statistics" output in Table 2.13 lists the final maximized value of the log likelihood as well as the information criteria of the Akaike and Schwarz.

Fit Statistics

-2 Log Likelihood	724.0
AIC (smaller is better)	732.0
AICC (smaller is better)	732.1
BIC (smaller is better)	746.6

Table 2.13. Fit statistics: Ezzet and Whitehead data

Table 2.14. Maximum likelihood estimates: Ezzet and Whitehead data

The NLMIXED Procedure

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
beta0	-0.4787	0.1386	278	-3.45	0.0006	0.05	-0.7515	-0.2059	-4.79E-6
beta1	0.8593	0.2146	278	4.00	<.0001	0.05	0.4368	1.2817	-1.35E-6
beta2	0.2106	0.2049	278	1.03	0.3048	0.05	-0.1927	0.6139	7.795E-7
logsigma	0.3473	0.1809	278	1.92	0.0559	0.05	-0.00876	0.7034	2.04E-6

The "Parameter estimates" output in Table 2.14 lists the maximum likelihood estimates and their approximate standard errors computed using the final Hessian matrix. Approximate $t - values$ and Wald-type confidence limits are also provided, with degrees of freedom equal to the number of subjects minus the number of random effects.

The above analysis is for scenario involving treatment, period effects, and the random effects. Similarly, we can get analysis for null model and also for the scenario involving treatment and random effects. Table 2.15 shows the estimates (standard errors) for all the above mentioned scenarios.

Model	Method	β_0	β_1	β_2	β_3	$log(\sigma)$	$-2logL$
Null	Ezzet	-0.444	-	-		0.228	
		(0.128)				(0.245)	
	$MathCad^{©}$	-0.445	-	-		0.226	
		(0.129)				(0.196)	
	$SAS^{©}$	-0.444	-	-		0.221	
		(0.128)				(0.194)	
Treatment	Ezzet	-0.476	0.856	-	-	0.345	
		(0.138)	(0.215)			(0.262)	
	$MathCad^{©}$	-0.478	0.856	-	-	0.348	725.049
		(0.139)	(0.214)			(0.184)	
	$SAS^{©}$	-0.476	0.855	-	-	0.341	
		(0.138)	(0.214)			(0.182)	
Treatment	Ezzet	-0.479	0.861	0.211	-	0.351	
and		(0.139)	(0.215)	(0.206)		(0.263)	
period	$MathCad^{©}$	-0.480	0.861	0.211	-	0.354	724
		(0.139)	(0.215)	(0.205)		(0.183)	
	$SAS^{©}$	-0.479	0.859	0.211	-	0.347	724
		(0.139)	(0.215)	(0.205)		(0.181)	

Table 2.15. Conditional analysis of the asthma inhaler study.

Both the Mathcad$^{©}$ and the **PROC NLMIXED** method produce essentially the same results. These two methods also agrees with the results obtained by Ezzet (30) in estimating the parameters.

Chapter 3

Crossover Designs In Infertility Trials

For some years, there has been a debate as to the place of the crossover design in assisted reproduction technology (ART). In this chapter we aim to investigate (using the models discussed in Chapter 2) whether crossover and parallel design in infertility trials result in different estimates of treatment effects. Also we will investigate what is it that the researchers are losing by conducting one design instead of the other. Specifically, we would like to find if there should be preference between parallel and crossover design when conducting infertility trials. In a crossover design, subjects randomized to different groups receive more than one treatment in a specified order (reviewed in (50) and (84)). Thus in a crossover design, repeated measurements are collected within the subject. In infertility trials, couples undergo a course of treatments. In the crossover design this course consist of repeated administration of a sequence in which the treatment changes over time (64). The outcome of interest is the binary outcome, pregnancy, coded 1 and 0 for positive (pregnant) and negative (not pregnant)

outcome respectively. This outcome will be clustered within subjects.

To understand this, we will consider the $AB : BA$ crossover design in which subjects either receive treatment A followed by treatment B or B followed by A. If a subject is randomized to say group AB, the subject will first receive treatment A, and depending on the outcome, the subject will either allowed to receive treatment B or the subject drops out of the trial. If the subject fails to conceive, the subject will go ahead to receive treatment B. But if the subject conceives, the subject drops from the trial, resulting in missing outcome for the second period. Subjects randomized to the BA group receive treatments in the reverse order. Thus, in the $AB : BA$ infertility crossover design we will have those subjects with missing outcomes in the second period, unless if all the subjects fail to conceive in the first period. That is unless if we have the standard $AB : BA$ crossover design like the one in Ezzet and Whitehead (30) discussed in Chapter 2.

For comparison purposes, we will first look at the unrealistic situation where every woman receives both treatments regardless of the outcome in the first period. That is, we will conduct the complete $AB : BA$ crossover design. We will like to emphasize that the complete $AB : BA$ crossover design is never performed in practice in infertility trials. We are merely doing it here, for the sake of comparisons.

Clearly, if we are presented with the data of the complete $AB : BA$ crossover design, we can immediately predict what could have happened if either the incomplete $AB : BA$ crossover design or the parallel design was performed. If we delete, from the complete $AB : BA$ crossover design the second outcome for every woman who conceived in the first period, then we obtain the realistic data for a cross over design. This is because by doing so, we do not allow for women to get

pregnant twice. Similarly, if we delete all the outcomes in the second period we obtain the parallel design scenario.

We will compare these three data set: the data set from the complete $AB : BA$ crossover design, the data set from the incomplete $AB : BA$ crossover design and the parallel design scenario. To this we use the maximum likelihood method discussed in Chapter 2. Specifically, we construct the likelihoods for each data set and obtain the maximum likelihood estimates (MLE) and their corresponding standard errors from each data set.

We compare the maximum likelihood estimates (and their corresponding standard errors) obtained using the complete $AB : BA$ crossover data set with the maximum likelihood estimates (and their corresponding standard errors) obtained using the incomplete $AB : BA$ crossover data set. Our aim here, being to explore what actually is it that the medical practitioners are losing (in analyzing the results) by not allowing the women to get pregnant twice. Finally, we will compare the maximum likelihood estimates (and their corresponding standard errors) obtained using the incomplete $AB : BA$ crossover data set with the maximum likelihood estimates (and their corresponding standard errors) obtained using the parallel data set. Here, we will like to explore what is it wrong with the crossover design in infertility trials. Since, if this incomplete crossover design is not employed, the only alternative design that is usually recommended is the parallel design.

The model used for the complete and incomplete crossover data set, (because of its nature to allow the correlation between the repeated outcomes to be incorporated into the estimates of parameters) is the mixed effects model described in Chapter 2. While for the parallel design the appropriate model used is the standard logistic regression model, also described in Chapter 2.

To compare two different models, we must compare analogous quantities, as is emphasized by Lee and Nelder (57). Both the maximum likelihood estimates obtained from the complete $AB : BA$ crossover data set and the maximum likelihood estimates obtained from the incomplete $AB : BA$ data set describe the individual's response (conditional estimates). Following Lee and Nelder (57) these estimates and their corresponding standard errors can be compared. The maximum likelihood estimates obtained from the parallel data set describe the marginal response to changing covariates (marginal estimates). So we will not be able to straightaway compare these estimates to the maximum likelihood estimates obtained from the crossover data set (complete or incomplete $AB : BA$).

The mixed effects model is the basic model and any mixed effects model leads to a specific marginal model. It is this specific marginal model (marginalized estimates) that allows us to compare analogous quantities. Specifically, we will marginalize the maximum likelihood estimates obtained from the mixed effects model so that they describe marginal response.

To perform all these, we suppose that we have $m = 2n$ women undergoing the complete $AB : BA$ crossover design. The experimental design can be illustrated as in Table 3.1 below, where m_{krs} denotes the number of outcomes of the form (r, s) among group k. And where $k = 1$ denotes the group AB and $k = 2$ denotes group BA.

A	B	Count	B	A	Count
0	0	m_{100}	0	0	m_{200}
0	1	m_{101}	0	1	m_{201}
1	0	m_{110}	1	0	m_{210}
1	1	m_{111}	1	1	m_{211}

Table 3.1. Experimental design for the complete $AB : BA$ infertility crossover design

If we are presented with the above data (unrealistic data set) in Table 3.1, we

can immediately predict what would have happened if the incomplete crossover design (realistic situation) was performed. Specifically, we can predict (in each group), the number of women who did not conceive, the number of women who conceived in the first period only and the total number of women who conceived in the second period only. Thus the predicted experimental design for the incomplete $AB : BA$ crossover design can be illustrated as in Table 3.2 below

A	B	Count	B	A	Count
0	0	m_{100}	0	0	m_{200}
0	1	m_{101}	0	1	m_{201}
1		$m_{110} + m_{111}$	1		$m_{210} + m_{211}$

Table 3.2. Experimental design for the predicted incomplete $AB : BA$ crossover design

Similarly, If we are presented with either the complete $AB : BA$ crossover data set, Table 3.1 or the incomplete $AB : BA$ crossover data set, Table 3.2, we can immediately predict what would have happened if the parallel design was performed. Specifically, we can predict (in each group), the total number of women who did not conceive, and the total number of women who conceived. Thus the predicted experimental design for the parallel design can be illustrated as in Table 3.3 below:

A	B	Count	B	A	Count
0		$m_{100} + m_{101}$	0		$m_{200} + m_{201}$
1		$m_{110} + m_{111}$	1		$m_{210} + m_{211}$

Table 3.3. Experimental design for the predicted parallel design

All what we are saying is that if one is presented with the complete (incomplete) $AB : BA$ data set, then one can easily analyze the data as the parallel design data set by simply neglecting the second outcomes for each subject.

We will now introduce the mentioned models for the complete and incomplete $AB : BA$ crossover design.

3.1 Mixed effects model for crossover designs

We let i, j and k denote the woman, period and group respectively. Let $p_{(ij)k}$ represent the probability that the ith woman, $i = \{1, 2, ...2n\}$ in the kth group, $k = \{1, 2\}$, conceives in period $j, j = \{1, 2\}$. Further, we let

$$\psi_{(ij)k} = \ln(\frac{p_{(ij)k}}{1 - p_{(ij)k}}) = \mathbf{x}'_{(ij)k}\beta + v_i, \tag{3.1}$$

where $\mathbf{x}_{(\mathbf{ij})\mathbf{k}}$ is the $(p+1) \times 1$ covariate vector (includes a 1 for the intercept) associated with pregnancy, β is the $(p+1) \times 1$ vector of unknown regression parameters, and v_i is the random subject effect. These are assumed to be distributed in the population as $N(0, \sigma_v^2)$.

The conditional probability (conditional on the random effect) that a woman conceives is

$$P(Y_{(ij)k} = 1 | v_i) = \frac{\exp(\psi_{(ij)k})}{1 + \exp(\psi_{(ij)k})}. \tag{3.2}$$

The conditional probability (conditional on the random effect) that a woman fails to conceive is

$$P(Y_{(ij)k} = 0 | v_i) = \frac{1}{1 + \exp(\psi_{(ij)k})}. \tag{3.3}$$

As usual, the observations within a woman are assumed independent given the random effect. As a result, suppose each woman is allowed to go under both treatments, regardless of the outcome in the first outcome, the conditional likelihoods for one woman, under this scenario (the complete crossover) is

$$\ell_C(\mathbf{Y}_{(i)k} | v) = \frac{[\exp(\psi_{(ij)k})]^{Y_{(i1)k}}}{1 + \exp(\psi_{(ij)k})} * \frac{[\exp(\psi_{(ij)k})]^{Y_{(i2)k}}}{1 + \exp(\psi_{(ij)k})}. \tag{3.4}$$

Now, suppose we consider a realistic situation, specifically, we consider a situation

where by if a woman get pregnant in the first period, that woman drops out from the trial. In this scenario (the incomplete crossover), the conditional likelihoods for one woman is

$$\ell_I(\mathbf{Y}_{(i)k}|\upsilon) = \frac{[\exp(\psi_{(ij)k})]^{Y_{(i1)k}}}{1 + \exp(\psi_{(ij)k})} * \left\{ \frac{[\exp(\psi_{(ij)k})]^{Y_{(i2)k}}}{1 + \exp(\psi_{(ij)k})} \right\}^{1-Y_{(i1)k}}. \tag{3.5}$$

Thus

$$h_C(\mathbf{Y}_{(i)k}) = \int_\upsilon \ell_C(\mathbf{Y}_{(i)k}|\upsilon)g(\upsilon)d\upsilon, \tag{3.6}$$

and

$$h_I(\mathbf{Y}_{(i)k}) = \int_\upsilon \ell_I(\mathbf{Y}_{(i)k}|\upsilon)g(\upsilon)d\upsilon, \tag{3.7}$$

are the marginal probabilities for complete and incomplete $AB : BA$ crossover respectively, where $g(\upsilon)$ represents the distribution of the random effects, namely $N(0, \sigma_\upsilon^2)$.

The marginal likelihoods for $2N - women$ for complete and incomplete cross over trials are respectively,

$$L_C = \prod_{i=1}^{2n} h_C(\mathbf{Y}_{(i)k}) \tag{3.8}$$

and

$$L_I = \prod_{i=1}^{2n} h_I(\mathbf{Y}_{(i)k}). \tag{3.9}$$

The vector $\mathbf{x}$ consists of the covariates associated with pregnancy. For the purpose of illustration, we will begin with the model involving only one covariate

(treatment). For this we let

$$\psi_{(ij)k} = \ln(\frac{p_{(ij)k}}{1 - p_{(ij)k}}) = \beta_0 + \beta_1 treatment_{(ij)k} + \exp(v_i), \qquad (3.10)$$

where β_0 is the intercept term, β_1 is the treatment term and treatment=$\{A = 1, B = -1\}$. While v_i is subject random effect and is assumed to be normally distributed with mean zero and variance σ^2. We exponentiated the random effect so that the likelihood looks more quadratic. Actually this model is taken from Ezzet and Whitehead (30). We will use the model (3.10) to obtain the likelihoods for both the complete $AB : BA$ crossover and the incomplete $AB : BA$ crossover.

3.1.1 Likelihood of the complete $AB : BA$ crossover design

In the complete $AB : BA$ crossover design, the components of the log likelihood can be viewed as the contribution by the following illustrated eight cases: Consider women randomized to the group AB. Those women will receive treatment A in the first period and treatment B in the second period. Those subjects who failed to conceive neither in the first period nor in the second period will contribute to the log-likelihood the following:

$$m_{100} \ln \int_v \left\{ \frac{1}{1 + \exp(\beta_0 + \beta_1 + v)} \frac{1}{1 + \exp(\beta_0 - \beta_1 + v)} \right\} g(v)dv. \qquad (3.11)$$

For those who failed to conceive under treatment A but conceived due to treatment B, will contribute

$$m_{101} \ln \int_v \left\{ \frac{1}{1 + \exp(\beta_0 + \beta_1 + v)} \frac{\exp(\beta_0 - \beta_1 + v)}{1 + \exp(\beta_0 - \beta_1 + v)} \right\} g(v)dv, \qquad (3.12)$$

while those who conceived due to treatment A but did not conceive under treatment B, will contribute

$$m_{110} \ln \int_v \left\{ \frac{\exp(\beta_0 + \beta_1 + v)}{1 + \exp(\beta_0 + \beta_1 + v)} \frac{1}{1 + \exp(\beta_0 - \beta_1 + v)} \right\} g(v) dv. \qquad (3.13)$$

Finally, the contribution by those subjects who conceived twice is

$$m_{111} \ln \int_v \left\{ \frac{\exp(\beta_0 + \beta_1 + v)}{1 + \exp(\beta_0 + \beta_1 + v)} \frac{\exp(\beta_0 - \beta_1 + v)}{1 + \exp(\beta_0 - \beta_1 + v)} \right\} g(v) dv. \qquad (3.14)$$

Similarly, if we consider those women who were randomized to the group BA, these women will receive treatment B in the first period and treatment A in the second period. The contribution to the log-likelihood by those women who failed to conceive neither under treatment B nor under treatment A will be:

$$m_{200} \ln \int_v \left\{ \frac{1}{1 + \exp(\beta_0 - \beta_1 + v)} \frac{1}{1 + \exp(\beta_0 + \beta_1 + v)} \right\} g(v) dv. \qquad (3.15)$$

For those who failed to conceive in the first period but conceived in the second period contribute to the likelihood:

$$m_{201} \ln \int_v \left\{ \frac{1}{1 + \exp(\beta_0 - \beta_1 + v)} \frac{\exp(\beta_0 + \beta_1 + v)}{1 + \exp(\beta_0 + \beta_1 + v)} \right\} g(v) dv. \qquad (3.16)$$

while those who conceived in the first period but did not conceive in the second period contribute:

$$m_{210} \ln \int_v \left\{ \frac{\exp(\beta_0 - \beta_1 + v)}{1 + \exp(\beta_0 - \beta_1 + v)} \frac{1}{1 + \exp(\beta_0 + \beta_1 + v)} \right\} g(v) dv, \qquad (3.17)$$

The contribution by those who conceived twice is:

$$m_{211} \ln \int_{v} \left\{ \frac{\exp(\beta_0 - \beta_1 + v)}{1 + \exp(\beta_0 - \beta_1 + v)} \frac{\exp(\beta_0 + \beta_1 + v)}{1 + \exp(\beta_0 + \beta_1 + v)} \right\} g(v) dv. \qquad (3.18)$$

As a result, the log-likelihood for the complete $AB : BA$ crossover design can be written as:

$$\Psi_C = \sum_{k=1}^{2} \sum_{r=0}^{1} \sum_{s=0}^{1} m_{krs} \ln \int_{v} \frac{\exp(\psi_{(ij)k})^r}{1 + \exp(\psi_{(ij)k})} \frac{\exp(\psi_{(ij)k})^s}{1 + \exp(\psi_{(ij)k})} g(v) dv, \qquad (3.19)$$

where m_{krs} is as defined in Table 3.1.

3.1.2 Likelihood of the infertility $AB : BA$ crossover design

The log-likelihood for the infertility crossover design, can be viewed as the addition of the following six cases: Consider the women who were randomized to the AB group. First is the contribution by those subjects who failed to conceive neither under treatment A nor under treatment B. those subjects contribute

$$m_{100} \ln \int_{v} \left\{ \frac{1}{1 + \exp(\beta_0 + \beta_1 + v)} \frac{1}{1 + \exp(\beta_0 - \beta_1 + v)} \right\} g(v) dv. \qquad (3.20)$$

For the subjects who fail to conceive in the first period but conceived in the second period will contribute

$$m_{101} \ln \int_{v} \left\{ \frac{1}{1 + \exp(\beta_0 + \beta_1 + v)} \frac{\exp(\beta_0 - \beta_1 + v)}{1 + \exp(\beta_0 - \beta_1 + v)} \right\} g(v) dv. \qquad (3.21)$$

Those who conceived in the first period, will not go under the second treatment, hence they will only contribute

$$m_{110} \ln \int_v \frac{\exp(\beta_0 + \beta_1 + v)}{1 + \exp(\beta_0 + \beta_1 + v)} g(v) dv + m_{111} \ln \int_v \frac{\exp(\beta_0 + \beta_1 + v)}{1 + \exp(\beta_0 + \beta_1 + v)} g(v) dv.$$

$$(3.22)$$

Similarly, consider those women who were randomized to the BA group. The subjects who failed to conceive neither under treatment B nor under treatment A contribute:

$$m_{200} \ln \int_v \left\{ \frac{1}{1 + \exp(\beta_0 - \beta_1 + v)} \frac{1}{1 + \exp(\beta_0 + \beta_1 + v)} \right\} g(v) dv. \qquad (3.23)$$

For the women who failed to conceive in the first period but conceived in the second will contribute to the log-likelihood the following:

$$m_{201} \ln \int_v \left\{ \frac{1}{1 + \exp(\beta_0 - \beta_1 + v)} \frac{\exp(\beta_0 + \beta_1 + v)}{1 + \exp(\beta_0 + \beta_1 + v)} \right\} g(v) dv, \qquad (3.24)$$

while those who conceived in the first period contribute to the log-likelihood the following:

$$m_{210} \ln \int_v \frac{\exp(\beta_0 - \beta_1 + v)}{1 + \exp(\beta_0 - \beta_1 + v)} g(v) dv + m_{211} \ln \int_v \frac{\exp(\beta_0 - \beta_1 + v)}{1 + \exp(\beta_0 - \beta_1 + v)} g(v) dv.$$

$$(3.25)$$

Consequently, the log-likelihood for the incomplete $AB : BA$ crossover design can be expressed as:

$$\Psi_I = \sum_{k=1}^{2} \sum_{r=0}^{1} \sum_{s=0}^{1} m_{krs} \ln \int_v \frac{\exp(\psi_{(ij)k})^r}{1 + \exp(\psi_{(ij)k})} \left\{ \frac{\exp(\psi_{(ij)k})^s}{1 + \exp(\psi_{(ij)k})} \right\}^{1-r} g(v) dv \quad (3.26)$$

where m_{krs} is as defined in Table 3.1.

We can now maximize the respective log-likelihoods (3.19) and (3.26) (by either Mathcad$^©$ or SAS$^©$) to obtain the corresponding maximum likelihood estimates. We can investigate the respective variance-covariance matrix to get the corresponding standard errors.

3.2 Comparisons of the likelihoods

If we are moving from the complete $AB : BA$ crossover design to the incomplete $AB : BA$ crossover design information is lost on women who conceived in the first period. If we are in the complete $AB : BA$ crossover design, we have the information of what happened to these women in the second period. But when we are in the incomplete $AB : BA$ crossover design we do not know what is happening in the second period with regard to those women who conceived in the first period. We will like to investigate the consequences of these loss of information by looking at the observed Fisher information and thereafter estimate the variance covariance matrix. The log-likelihoods (3.19) and (3.26) can easily be expressed as the sum of the following equations:

$$\Psi_C = \mathbf{K} + (3.13) + (3.14) + (3.17) + (3.18) \tag{3.27}$$

and

$$\Psi_I = \mathbf{K} + (3.22) + (3.25) \tag{3.28}$$

respectively. Where,

$\mathbf{K} = (3.11) + (3.12) + (3.15) + (3.16) = (3.20) + (3.21) + (3.23) + (3.24)$

3.2.1 Variance-Covariance matrix of the complete crossover

The variance -covariance matrix is estimated by the inverse of the Fisher information (46). The Fisher information of the complete $AB : BA$ crossover design is given by

$$I_C(\beta_0, \beta_1, \upsilon) = - \begin{pmatrix} \frac{\partial^2 \Psi_C}{\partial \beta_0^2} & \frac{\partial^2 \Psi_C}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 \Psi_C}{\partial \beta_0 \partial \upsilon} \\ \frac{\partial^2 \Psi_C}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 \Psi_C}{\partial \beta_1^2} & \frac{\partial^2 \Psi_C}{\partial \beta_1 \partial \upsilon} \\ \frac{\partial^2 \Psi_C}{\partial \beta_0 \partial \upsilon} & \frac{\partial^2 \Psi_C}{\partial \upsilon \partial \beta_2} & \frac{\partial^2 \Psi_C}{\partial \upsilon^2} \end{pmatrix} \tag{3.29}$$

.

The approximate variances and covariances are found, respectively, in the matrix

$$I_C^{-1}(\beta_0, \beta_1, \upsilon) \approx \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \tag{3.30}$$

,

where $a_{11} = var(\beta_0), a_{12} = a_{21} = cov(\beta_0, \beta_1), a_{13} = a_{31} = cov(\beta_0, \sigma), a_{22} = var(\beta_1)$ and $a_{33} = var(\sigma)$

3.2.2 Variance-Covariance matrix of the incomplete crossover

The Fisher information of the complete $AB : BA$ crossover design is given by

$$I_I(\beta_0, \beta_1, \upsilon) = - \begin{pmatrix} \frac{\partial^2 \Psi_I}{\partial \beta_0^2} & \frac{\partial^2 \Psi_I}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 \Psi_I}{\partial \beta_0 \partial \upsilon} \\ \frac{\partial^2 \Psi_I}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 \Psi_I}{\partial \beta_1^2} & \frac{\partial^2 \Psi_I}{\partial \beta_1 \partial \upsilon} \\ \frac{\partial^2 \Psi_I}{\partial \beta_0 \partial \upsilon} & \frac{\partial^2 \Psi_I}{\partial \upsilon \partial \beta_2} & \frac{\partial^2 \Psi_I}{\partial \upsilon^2} \end{pmatrix} \tag{3.31}$$

.

The approximate variances and covariances are found, respectively, in the matrix

$$I_I^{-1}(\beta_0, \beta_1, \upsilon) \approx \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \qquad (3.32)$$

,

where $b_{11} = var(\beta_0), b_{12} = b_{21} = cov(\beta_0, \beta_1), b_{13} = b_{31} = cov(\beta_0, \sigma), b_{22} = var(\beta_1)$ and $b_{33} = var(\sigma)$

3.3 Crossover design and Parallel design

The parallel design data set is analyzed using the standard logistic model described in the previous chapter. An obvious advantage of using the mixed effects model is that it allows conditional inferences in addition to marginal inferences (82). With the mixed effect model (3.1) we can obtain not only the conditional mean $E(Y_{ij|v_i})$ but also the marginal mean $E(Y_{ij})$, while with standard logistic model we can obtain only the marginal mean $E(Y_{ij})$. Given a conditional model, one can recover information about marginal distributions. That is, conditional model implies a marginal model, but a marginal model does not itself imply a conditional model.

We will illustrate how given a conditional model, one can recover marginal inferences.

3.3.1 Marginalized Random-Effect Model

Let β^c be the conditional estimates obtained under the random effect model described above. We can predict the marginal probability from either equation (3.6) or equation (3.7) by simply using the fact that

$$p_{ij} = \int_v \frac{\exp(\beta^c \mathbf{x}_{ij} + v)}{1 + \exp(\beta^c \mathbf{x}_{ij} + v)} g(v) dv. \tag{3.33}$$

From the p_{ij}, we can compute marginal logistic regression estimates from the mixed effects model, giving an intercept

$$\beta_0^M = \frac{logit(p_{i1}) + logit(p_{i2})}{2}, \tag{3.34}$$

and the treatment effect

$$\beta_1^M = logit(p_{i1}) - logit(p_{i2}). \tag{3.35}$$

The standard errors are easily obtained from the fact that there exist a constant **c**, such that the estimates from the fixed effects model is equal to the estimates from the random effects model multiplied by a constant **c** (more details can be seen in Agresti (1)). These marginalized estimates and their corresponding standard errors are compared with the maximum likelihood estimates obtained from the parallel data set. We expect conditional estimates (and their corresponding standard errors) to be higher than the marginal estimates (1).

3.4 Illustration

As an illustration we will look at the $3M - Riker$ example mentioned in the Chapter 2. We will like to state that we use this example for the sake of illustration only. It is because we are considering the unrealistic case; the complete cross over design in infertility, and so there is noway we can get an infertility data where women are allowed to get pregnant twice. The trial is not an infertility trial. We will first analyze the trial as the complete $AB : BA$ crossover design. We then delete the second outcome for every subject whose outcome is a 1 in the first period, thus creating an incomplete $AB : BA$ scenario, which looks more similar like the infertility trial.

3.4.1 Analysis using Mathcad$^{©}$

We will illustrate how the log-likelihoods (3.19) and (3.28) are maximized using Mathcad$^{©}$. Using the quadrature adaptive method, the log-likelihoods (3.19) and (3.28) can be estimated by

$$\Gamma_C(\beta_0, \beta_1, \sigma) = \sum_{k=1}^{2} \sum_{r=0}^{1} \sum_{s=0}^{1} \left\{ m_{krs} \ln \pi^{-\frac{1}{2}} \sum_{q=1}^{N} \psi_q \frac{\left\{ \omega_{(1)k} \right\}^r}{1 + \omega_{(1)k}} \frac{\left\{ \omega_{(2)k} \right\}^s}{1 + \omega_{(2)k}} \right\} \qquad (3.36)$$

and

$$\Gamma_I(\beta_0, \beta_1, \sigma) = \sum_{k=1}^{2} \sum_{r=0}^{1} \sum_{s=0}^{1} \left\{ m_{krs} \ln \pi^{-\frac{1}{2}} \sum_{q=1}^{N} \psi_q \frac{\left\{ \omega_{(1)k} \right\}^r}{1 + \omega_{(1)k}} \left\{ \frac{\left\{ \omega_{(2)k} \right\}^s}{1 + \omega_{(2)k}} \right\}^{1-r} \right\}$$
$$(3.37)$$

respectively, where

$$\omega_{(1)k} = \exp(\beta_0 + \tfrac{\beta_1}{2} treatment_{(1)k} + \sqrt{2} x_q \exp(\sigma))$$

and

$$\omega_{(2)k} = \exp(\beta_0 + \tfrac{\beta_1}{2} treatment_{(2)k} + \sqrt{2}x_q \exp(\sigma)).$$

The $\{x_q\}_{q=1}^N$ are the roots of order N Hermite polynomial $P_N(x)$ and $\{\psi_q\}_{q=1}^N$ are as defined in equation 2.34. Furthermore $treatment_{(1)k}$ denotes the first treatment in group k. Similarly $treatment_{(2)k}$ denotes the second treatment in group k. We maximize $\Gamma_C(\beta_0, \beta_1, \sigma)$ and $\Gamma_I(\beta_0, \beta_1, \sigma)$ using Mathcad©.

The following MathCad© code will maximize both $\Gamma_C(\beta_0, \beta_1, \sigma)$ and $\Gamma_I(\beta_0, \beta_1, \sigma)$ and produce the maximum likelihood estimates (MLE) of β_0, β_1, and σ.

ORIGIN:= 1

$$\Gamma_C(\beta_0, \beta_1, \sigma) := \sum_{k=1}^{2} \sum_{r=0}^{1} \sum_{s=0}^{1} m_{krs} \ln \pi^{-\frac{1}{2}} \sum_{q=1}^{N} \psi_q \frac{\{\omega_{(1)k}\}^r}{1+\omega_{(1)k}} \frac{\{\omega_{(2)k}\}^s}{1+\omega_{(2)k}}$$

$$\Gamma_I(\beta_0, \beta_1, \sigma) := \sum_{k=1}^{2} \sum_{r=0}^{1} \sum_{s=0}^{1} m_{krs} \ln \pi^{-\frac{1}{2}} \sum_{q=1}^{N} \psi_q \frac{\{\omega_{(1)k}\}^r}{1+\omega_{(1)k}} \left\{ \frac{\{\omega_{(2)k}\}^s}{1+\omega_{(2)k}} \right\}^{1-r}$$

$\beta_0 := 0$ $\qquad$ $\beta_1 := 1$ $\qquad$ $\sigma := 1$

Maximize$(\Gamma_C, \beta_0, \beta_1, \sigma)$

Maximize$(\Gamma_I, \beta_0, \beta_1, \sigma)$

The output of the above analysis is as follows:

$$\text{Maximize}\left(\Gamma_C, \beta_0, \beta_1, \sigma\right) = \begin{pmatrix} -0.478 \\ 0.856 \\ 0.348 \end{pmatrix} \qquad + \qquad \text{Maximize}\left(\Gamma_I, \beta_0, \beta_1, \sigma\right) = \begin{pmatrix} -0.404 \\ 0.71 \\ 0.561 \end{pmatrix}$$

The standard errors are computed from the corresponding variance-covariance matrix, as illustrated in the previous sections. Table 3.4 below shows the MLE and their standard errors for the complete and incomplete $AB : BA$ crossover for models involving treatment only and treatment with period effects.

Model	Data set	β_0	β_1	β_2	$\ln(\sigma)$
Treatment	Complete	-0.478	0.856		0.348
		(0.139)	(0.214)		(0.184)
	Incomplete	-0.404	0.710		0.561
		(0.166)	(0.272)		(0.266)
Treatment	Complete	-0.480	0.861	0.211	0.354
and		(0.139)	(0.215)	(0.205)	(0.183)
Period	Incomplete	-0.197	0.874	-0.800	1.130
		(0.637)	(0.579)	(2.456)	(1.374)

Table 3.4. conditional estimates (standard errors) from the complete crossover design and the incomplete crossover design

We now create a parallel design trial by deleting all the second outcomes. The log-likelihood for the parallel design takes the form

$$\Gamma_P(\beta_0, \beta_1, \sigma) = \sum_{k=1}^{2} \sum_{r=0}^{1} m_{kr.} \ln \frac{\left\{W_{(1)k}\right\}^r}{1 + W_{(1)k}}, \qquad (3.38)$$

where

$$W_{(1)k} = \exp(\beta_0 + \tfrac{\beta_1}{2} treatment_{(1)k}),$$

and $m_{kr.} = \sum_{s=0}^{1} m_{krs}$.

The following MathCad$^{©}$ code will maximize $\Gamma_P(\beta_0, \beta_1)$ and produce the maximum likelihood estimates (MLE) of β_0, and β_1,

ORIGIN:= 1

$$\Gamma_P(\beta_0, \beta_1) := \sum_{k=1}^{2} \sum_{r=0}^{1} m_{kr}. \ln \frac{\left\{W_{(1)k}\right\}^r}{1 + W_{(1)k}}$$

$\beta_0 := 0 \qquad \beta_1 := 1$

$\text{Maximize}(\Gamma_P, \beta_0, \beta_1)$

The output of the parallel design analysis is as follows:

$$\text{Maximize}\left(\Gamma_{.P}, \beta_{.0}, \beta_{.1}\right) = \begin{pmatrix} -0.269 \\ 0.393 \end{pmatrix} \blacksquare$$

3.4.2 Analysis using SAS$^{©}$

The following SAS$^{©}$ code creates a data set **INCOMPLETE** and **COMPLETE** respectively.

data incomplete;

```
input seq patient period treat outcome;

datalines;

1 1 1 1 0

1 2 1 1 0

1 3 1 1 0

. . . . .

. . . . .

. . . . .

2 156 -1 1 1

2 157 -1 1 1

2 158 -1 1 1

;

run;

proc sort data=incomplete;

by patient;

data complete;

input group patient period treatment outcome;

datalines;

1 1 1 1 0

1 2 1 1 0

1 3 1 1 0

. . . . .

. . . . .

. . . . .

2 277 -1 1 1

2 278 -1 1 1

2 279 -1 1 1
```

```
;
run;
proc sort data=complete;
by patient;
run;
```

The **PROC NLMIXED** statements to fit the logistic-normal model to these data sets are as follows:

```
proc nlmixed data=incomplete;
parms beta0=-0.472 beta1=0.2141 logsigma=0.1;
sigma=exp(logsigma);
sigma2=sigma**2;
preg=beta0 + beta1*treat/2 + u;
ppreg=exp(preg)/(1+exp(preg));
model outcome binomial(1,ppreg);
random u normal(0,sigma2) subject=patient;
run;
proc nlmixed data=complete;
parms beta0=-0.472 beta1=0.852 logsigma=0.3;
sigma=exp(logsigma);
sigma2=sigma**2;
preg=beta0 + beta1*treat/2 + u;
ppreg=exp(preg)/(1+exp(preg));
model outcome binomial(1,ppreg);
random u normal(0,sigma2) subject=patient;
```

run;

The "Parameter estimates" output below (for incomplete and complete cross over) lists the maximum likelihood estimates and their approximate standard errors computed using the final Hessian matrix. Approximate $t - values$ and Wald-type confidence limits are also provided, with degrees of freedom equal to the number of subjects minus the number of random effects.

The NLMIXED Procedure

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
beta0	-0.4760	0.1380	278	-3.45	0.0006	0.05	-0.7476	-0.2044	-6.76E-7
beta1	0.8549	0.2141	278	3.99	<.0001	0.05	0.4335	1.2763	-2.21E-6
logsigma	0.3413	0.1816	278	1.88	0.0613	0.05	-0.01627	0.6988	9.561E-7

Table 3.5. conditional estimates and their standard errors for the complete crossover design:model with treatment only

The NLMIXED Procedure

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
beta0	-0.4074	0.1626	278	-2.50	0.0128	0.05	-0.7275	-0.08724	-4.36E-6
beta1	0.7007	0.2684	278	2.61	0.0095	0.05	0.1724	1.2291	-0.00005
logsigma	0.5267	0.2533	278	2.08	0.0385	0.05	0.02802	1.0253	-0.00003

Table 3.6. conditional estimates and their standard errors for the incomplete crossover design:model with treatment only

The **PROC NLMIXED** statements to fit the logistic-normal model (with period effects) to these data sets are as follows:

```
    proc nlmixed data=one;
parms beta0=-0.472 beta1=0.2141 beta2=0.3 logsigma=0.1;
sigma=exp(logsigma);
sigma2=sigma**2;
preg=beta0 + beta1*treat/2 + beta2*period/2 + u;
ppreg=exp(preg)/(1+exp(preg));
model outcome binomial(1,ppreg);
random u normal(0,sigma2) subject=patient;
run;
proc nlmixed data=two;
parms beta0=-0.472 beta1=0.852 beta2=0.21 logsigma=0.3;
sigma=exp(logsigma);
sigma2=sigma**2;
preg=beta0 + beta1*treat/2 + beta2*period/2 + u;
ppreg=exp(preg)/(1+exp(preg));
model outcome binomial(1,ppreg);
random u normal(0,sigma2) subject=patient;
run;
```

Table 3.7 and Table 3.8 show the output for the incomplete and complete data analysis respectively:

The NLMIXED Procedure

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
beta0	-0.4787	0.1386	278	-3.45	0.0006	0.05	-0.7515	-0.2059	-4.79E-6
beta1	0.8593	0.2146	278	4.00	<.0001	0.05	0.4368	1.2817	-1.35E-6
beta2	0.2106	0.2049	278	1.03	0.3048	0.05	-0.1927	0.6139	7.795E-7
logsigma	0.3473	0.1809	278	1.92	0.0559	0.05	-0.00876	0.7034	2.04E-6

Table 3.7. conditional estimates and their standard errors for the complete crossover design:model with treatment and period effects

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
beta0	-0.4309	0.2449	278	-1.76	0.0796	0.05	-0.9131	0.05119	-0.00001
beta1	0.6831	0.2998	278	2.28	0.0234	0.05	0.09298	1.2733	0.000045
beta2	0.08856	0.7054	278	0.13	0.9002	0.05	-1.3000	1.4771	-8.7E-6
logsigma	0.4440	0.7543	278	0.59	0.5566	0.05	-1.0409	1.9289	-0.00001

Table 3.8. conditional estimates and their standard errors for the incomplete crossover design:model with treatment and period effects

The SAS$^{©}$ output is similar to our Mathcad$^{©}$ results. For the complete cross over design, our results are similar to the results in Ezzet and Whitehead (30). In the no period effects scenario, the maximum likelihood estimates obtained from the complete and incomplete data set are not much dissimilar. But in all the three parameters $(\beta_0, \beta_1, \sigma)$ the standard errors obtained from the incomplete data set are higher than the standard errors obtained from the complete data set. This clearly signifies loss of information. In the scenario involving period effects, the maximum likelihood estimates obtained from the incomplete data set shows evidence of bias. Also the standard errors are very much higher than the corresponding standard errors obtained under the complete data set. By moving from the no period effects scenario to the period effects scenario, we experience an increase in standard errors in the incomplete data set. This increase is so high that it makes it difficult to estimate the parameters. Unlike in the complete

data set case, there is no increase in the standard errors leading to the estimates obtained under no period effects scenario to be not dissimilar to the estimates obtained under the period effects scenario.

3.4.3 Marginalized estimates

We will use tables 3.5, 3.6, 3.7 and 3.8 to obtain the marginalized estimates that are comparable to the estimates from the parallel design. For the model involving treatment only, using equations (3.34) and (3.35), the marginalized intercept and treatment effect are estimated respectively as follows:

$$\beta_{0T}^M = \frac{\ln\frac{\lambda 1}{1-\lambda 1} + \ln\frac{\lambda 2}{1-\lambda 2}}{2} \tag{3.39}$$

and

$$\beta_{1T}^M = \ln\frac{\lambda 1}{1-\lambda 1} - \ln\frac{\lambda 2}{1-\lambda 2} \tag{3.40}$$

where

$$\lambda 1 = \pi^{\frac{-1}{2}} \sum_{q=1}^{N} \psi_q \left\{ \frac{\exp(beta0 + \frac{beta1}{2} + \sqrt{2}x_q\exp(logsigma))}{1+\exp(beta0 + \frac{beta1}{2} + \sqrt{2}x_q\exp(logsigma))} \right\}$$

and

$$\lambda 2 = \pi^{\frac{-1}{2}} \sum_{q=1}^{N} \psi_q \left\{ \frac{\exp(beta0 - \frac{beta1}{2} + \sqrt{2}x_q\exp(logsigma))}{1+\exp(beta0 - \frac{beta1}{2} + \sqrt{2}x_q\exp(logsigma))} \right\}.$$

The parameters beta0, beta1 and logsigma are the maximum likelihood estimates from either Table 3.5 or Table 3.6. Furthermore, $\{x_q\}_{q=1}^N$ are the roots of order N Hermite polynomial $P_N(x)$ and $\{\psi_q\}_{q=1}^N$ are as defined in equation 2.34.

Similarly, for the model involving treatment and period effects, the marginalized intercept and treatment effect are estimated respectively as follows:

$$\beta_{0TP}^{M} = \frac{\ln \frac{\Upsilon 1}{1 - \Upsilon 1} + \ln \frac{\Upsilon 2}{1 - \Upsilon 2}}{2} \tag{3.41}$$

and

$$\beta_{1TP}^{M} = \ln \frac{\Upsilon 1}{1 - \Upsilon 1} - \ln \frac{\Upsilon 2}{1 - \Upsilon 2} \tag{3.42}$$

where

$$\Upsilon 1 = \pi^{\frac{-1}{2}} \sum_{q=1}^{N} \psi_q \left\{ \frac{\exp(beta0 + \frac{beta1}{2} + \frac{\beta 2}{2} + \sqrt{2} x_q \exp(logsigma))}{1 + \exp(beta0 + \frac{beta1}{2} + \frac{\beta 2}{2} + \sqrt{2} x_q \exp(logsigma))} \right\}$$

and

$$\Upsilon 2 = \pi^{\frac{-1}{2}} \sum_{q=1}^{N} \psi_q \left\{ \frac{\exp(beta0 - \frac{beta1}{2} + \frac{\beta 2}{2} + \sqrt{2} x_q \exp(logsigma))}{1 + \exp(beta0 - \frac{beta1}{2} + \frac{\beta 2}{2} + \sqrt{2} x_q \exp(logsigma))} \right\}.$$

The parameters beta0, beta1, beta2 and logsigma are the maximum likelihood estimates from either Table 3.7 or Table 3.8. Table 3.9 below shows the marginalized estimates and their standard errors for the complete and incomplete $AB : BA$ crossover together with the maximum likelihood estimates obtained from the parallel data set for the two considered scenarios.

Based on our calculations, the statistical estimates of treatment effects obtained from a crossover design are not very much dissimilar from those obtained under parallel design. In both scenarios, the maximum likelihood estimates obtained from the parallel data set have the highest standard errors, followed by marginalized estimates from the incomplete data set. The marginalized estimates obtained from the complete data set has the lowest standard errors.

To further understand what is really happening we resort to simulation. Since we are satisfied that Mathcad© gives similar results to SAS© and simulation is

Model	data set	β_0^M	β_1^M
Treatment	Complete	-0.348	0.625
		(0.101)	(0.156)
	Incomplete	-0.265	0.466
		(0.109)	(0.179)
	Parallel	-0.269	0.393
		(0.121)	(0.243)
Treatment	Complete	-0.273	0.626
and		(0.079)	(0.156)
Period	Incomplete	-0.269	0.476
		(0.869)	(0.209)
	Parallel	-0.269	0.393
		(0.121)	(0.243)

Table 3.9. Marginalized estimates and the MLE from the parallel design

very much handy in Mathcad$^©$ than in SAS$^©$, we will employ Mathcad$^©$ to perform simulation.

3.5   Simulations

We simulated a two-period cross over design with the help of a Mathcad$^©$. The woman is randomized into either group AB or group BA. If a woman is randomized into group AB, she will first receive treatment A followed by treatment B. The women in the group BA receive the two treatments in the reverse order. Every woman progresses through the treatment regimen. At the first period women are randomly allocated to pregnant or not pregnant, based on their scenario-specific probability of pregnancy. Women then proceed to obtain a second treatment regardless of the outcome in the first period, again based on their scenario-specific probability of pregnancy they will be allocated to pregnant or not pregnant. For each couple the number of pregnancies were recorded. We calculated the likelihood function associated with the progress of women through this process. We then construct the realistic infertility trial by removing the

second outcomes for those women who conceived in the first period. Similarly we constructed the likelihood associated with the infertility trial. Finally, we construct a parallel design by removing the second outcome for each woman and we constructed the likelihood associated with the parallel design.

We considered two cases. One case with period effect and the other without the period effect. In each case, we simulated 2000 samples of size 300. For each sample i we record the maximum likelihood estimates (MLE), say $\hat{\beta}_i$. We estimate the true value by the average of the $\hat{\beta}_i s$. Thus

$$\beta \approx \bar{\hat{\beta}} = \frac{\sum_{i=1}^{2000} \hat{\beta}_i}{2000}.$$

The standard error of the estimator is computed as follows

$$se(\beta) = \sqrt{\frac{1}{2000} \sum_{i=1}^{2000} (\hat{\beta}_i - \bar{\hat{\beta}})^2}.$$

For the case without the period effect, the model was

$$\psi_{(ij)k} = \ln(\frac{p_{(ij)k}}{1-p_{(ij)k}}) = -0.479 + 0.861 * treatment_{(ij)k} + \exp(\mu_i),$$

whereas for the case with period effect, the model was

$$\psi_{(ij)k} = \ln(\frac{p_{(ij)k}}{1-p_{(ij)k}}) = -0.479 + 0.861 * treatment_{(ij)k} + 0.211 * period_{ij} + \exp(\mu_i),$$

where $\mu_i \sim N(0, \sigma^2)$, and $\sigma = 0.351$. Parameters β_0, β_1 and β_2 represent intercept, treatment and period effects respectively.

Finally, we again, use the above models, but with samples of larger sizes. For each case, we simulated 2000 samples of size 1000000. The following MathCad$^©$ with generate n sample of size m.

$$\underset{\sim}{\text{ORIGIN}} := 1$$

$$\beta := \begin{pmatrix} -0.479 \\ 0.861 \\ 0.211 \end{pmatrix}$$

$$\mu(\text{treatment}) := \beta_1 + \frac{\beta_2}{2} \cdot \text{treatment}$$

$$\underset{\sim}{m} := 300 \qquad\qquad n := 2000$$

$$
\text{Data} := \quad
\begin{array}{|l}
\text{for } k \in 1 \ldots n \\
\quad \begin{array}{|l}
\text{for } i \in 1 \ldots m \\
\quad \begin{array}{|l}
\text{for } j \in 1 \ldots 2 \\
\quad \text{treatment}_{i,j} \leftarrow -1 \\
\text{treatment}_{i,1} \leftarrow 1 \quad \text{if } i \leq \dfrac{m}{2} \\
\text{treatment}_{i,2} \leftarrow 1 \quad \text{otherwise} \\
a_i \leftarrow \text{rnorm}\left(1, 0, e^{0.351}\right)_1 \\
\text{age}_i \leftarrow 25 \\
\text{for } l \in 1 \ldots 2 \\
\quad \text{preg}_{i,1} \leftarrow \dfrac{\exp\left(\left(\mu\left(\text{treatment}_{i,1}\right) + a_i\right)\right)}{1 + \exp\left(\left(\mu\left(\text{treatment}_{i,1}\right) + a_i\right)\right)} \\
\text{test1} \leftarrow \text{preg}_{i,1} \geq \text{runif}(1, 0, 1)_1 \\
\text{test2} \leftarrow \text{preg}_{i,2} \geq \text{runif}(1, 0, 1)_1 \\
\text{pregnancy1}_i \leftarrow \text{test1} = 1 \\
\text{pregnancy2}_i \leftarrow \text{test2} = 1 \\
\text{count000}_i \leftarrow 1 \quad \text{if } \text{pregnancy1}_i = 0 \wedge \text{pregnancy2}_i = 0 \wedge \text{treatment}_{i,1} = -1
\end{array}
\end{array}
\end{array}
$$

$$\text{count000}_i \leftarrow 0 \quad \text{otherwise}$$

$$\text{count010}_i \leftarrow 1 \quad \text{if } \text{pregnancy1}_i = 0 \wedge \text{pregnancy2}_i = 1 \wedge \text{treatment}_{i,1} = -1$$

$$\text{count010}_i \leftarrow 0 \quad \text{otherwise}$$

$$\text{count100}_i \leftarrow 1 \quad \text{if } \text{pregnancy1}_i = 1 \wedge \text{pregnancy2}_i = 0 \wedge \text{treatment}_{i,1} = -1$$

$$\text{count100}_i \leftarrow 0 \quad \text{otherwise}$$

$$\text{count110}_i \leftarrow 1 \quad \text{if } \text{pregnancy1}_i = 1 \wedge \text{pregnancy2}_i = 1 \wedge \text{treatment}_{i,1} = -1$$

$$\text{count110}_i \leftarrow 0 \quad \text{otherwise}$$

$$\text{count001}_i \leftarrow 1 \quad \text{if } \text{pregnancy1}_i = 0 \wedge \text{pregnancy2}_i = 0 \wedge \text{treatment}_{i,1} = 1$$

$$\text{count001}_i \leftarrow 0 \quad \text{otherwise}$$

$$\text{count011}_i \leftarrow 1 \quad \text{if } \text{pregnancy1}_i = 0 \wedge \text{pregnancy2}_i = 1 \wedge \text{treatment}_{i,1} = 1$$

$$\text{count011}_i \leftarrow 0 \quad \text{otherwise}$$

$$\text{count101}_i \leftarrow 1 \quad \text{if } \text{pregnancy1}_i = 1 \wedge \text{pregnancy2}_i = 0 \wedge \text{treatment}_{i,1} = 1$$

$$\text{count101}_i \leftarrow 0 \quad \text{otherwise}$$

$$\text{count111}_i \leftarrow 1 \quad \text{if } \text{pregnancy1}_i = 1 \wedge \text{pregnancy2}_i = 1 \wedge \text{treatment}_{i,1} = 1$$

$$\text{count111}_i \leftarrow 0 \quad \text{otherwise}$$

$$S_{k,1} \leftarrow \text{mean}(\text{count001}) \cdot m$$

$$S_{k,2} \leftarrow \text{mean}(\text{count000}) \cdot m$$

$$S_{k,3} \leftarrow \text{mean}(\text{count011}) \cdot m$$

$$S_{k,4} \leftarrow \text{mean}(\text{count010}) \cdot m$$

$$S_{k,5} \leftarrow \text{mean}(\text{count101}) \cdot m$$

$$S_{k,6} \leftarrow \text{mean}(\text{count100}) \cdot m$$

$$S_{k,7} \leftarrow \text{mean}(\text{count111}) \cdot m$$

$$S_{k,8} \leftarrow \text{mean}(\text{count110}) \cdot m$$

S

In this code we can change the values of m and n to satisfy the samples number and their size.

3.6 Simulation results

3.6.1 Model with no Period Effects

For the model with no period effects, the conditional estimates are not very much dissimilar. The standard errors for the estimates obtained from the incomplete crossover design are relatively higher than their corresponding estimates obtained from the complete crossover design, as it is shown in Table 3.10.

Treat	$\bar{\hat{\beta}}_0(Std)$	$\bar{\hat{\beta}}_1$	$\bar{\hat{\sigma}}$
Complete	$-0.486(0.138)$	$0.862(0.214)$	$0.346(0.182^)$
Incomplete	$-0.483(0.15)$	$0.875(0.272)$	$0.28(0.589)$

Table 3.10. conditional estimates (standard errors) from the complete crossover design and the incomplete crossover design for the scenario involving no period effects

We can predict the marginal probability and thus compute marginal logistic regression estimates as shown in Table 3.11 below:

Treat	β_0^M	β_1^M
Full	$-0.352(0.097)$	$0.624(0.149))$
Incomplete	$-0.353(0.122)$	$0.627(0.179)$
Parallel	$-0.354(0.123)$	$0.63(0.241)$

Table 3.11. Predicted marginal regression estimates for the scenario involving no period effects

The predicted marginal regression estimates from both the complete crossover design and the incomplete design agree with the marginal estimates obtained from the parallel design. The standard errors for the treatment estimates are 0.149, 0.179 and 0.241 for the complete crossover, incomplete crossover and parallel

model respectively.

3.6.2 Model with Period Effects

The conditional regression estimates obtained from this model are very much dissimilar for the complete crossover design and incomplete crossover design as illustrated in Table 3.12. The estimates obtained under incomplete crossover design are biased with very high standard errors, while their corresponding estimates obtained from the full crossover design are unbiased with the standard errors similar to the standard errors obtained from the scenario involving no period effects.

Parameter	Full	Incomplete
$\bar{\hat{\beta}}_0$	$-0.487(0.136)$	$-0.389(0.417)$
$\bar{\hat{\beta}}_1$	$0.87(0.213)$	$1.073(0.604)$
$\bar{\hat{\beta}}_2$	$0.215(0.206)$	$-0.384(1.668)$
$\bar{\hat{\sigma}}$	$0.354(0.18)$	$-0.175(1.805)$

Table 3.12. conditional estimates from the full cross over design and the infertility cross over design for the scenario involving period effects

The predicted marginal regression estimates from both the complete crossover design and the incomplete design agree with the marginal estimates obtained from the parallel design as illustrated in Table 3.13. The standard errors for the treatment estimates are 0.146, 0.212 and 0.240 for the complete crossover, incomplete crossover and parallel model respectively.

Treat	β_0^M	β_1^M
Full	$-0.274(0.121)$	$0.626(0.146)$
Incomplete	$-0.274(0.121)$	$0.615(0.212)$
Parallel	-0.274(0.121)	0.631(0.240)

Table 3.13. Predicted marginal regression estimates for the scenario involving period effects

3.6.3 Larger samples

Using the same models, for each case, we simulated 2000 samples of size 1000000. Here we will like to investigate precision and not bias. Table 3.14 shows the maximum likelihood estimates for the complete and incomplete scenarios.

Treat	$\bar{\hat{\beta}}_0(Std \times 10^3)$	$\bar{\hat{\beta}}_1$	$\bar{\hat{\sigma}}$
Full	$-0.479(2.315)$	$0.861(3.63)$	$0.351(3.084)$
Incomplete	$-0.479(2.445)$	$0.861(4.344)$	$0.351(5.283)$

Table 3.14. conditional estimates (standard errors) from the complete crossover design and the incomplete crossover design for the scenario involving no period effects

The maximum likelihood obtained from the complete crossover design have lower standard errors than the maximum likelihood estimates obtained using the incomplete crossover design. The marginalized estimates for the scenario involving treatment only are shown in Table 3.15.

Treat	$\beta_0^M(Std \times 10^3)$	β_1^M
Full	$-0.349(1.6136)$	$0.628(2.5352)$
Incomplete	$-0.349(2.029)$	$0.628(2.9948)$
Parallel	$-0.349(2.0321)$	$0.628(4.0538)$

Table 3.15. Predicted marginal regression estimates for the scenario involving no period effects

The predicted marginal regression parameters for the incomplete crossover design, have higher standard errors than the marginalized estimates from the complete crossover design. But the predicted marginal regression parameters for the incomplete crossover design have smaller standard errors than the estimates obtained under the parallel design.

The conditional regression estimates and their corresponding predicted marginal estimates for the scenario involving period effects are shown in Table 3.16 and Table 3.17 respectively.

Parameter($Std \times 10^3$)	Full	Incomplete
$\bar{\hat{\beta}}_0$	$-0.479(2.371)$	$-0.479(5.399)$
$\bar{\hat{\beta}}_1$	$0.861(3.594)$	$0.861(6.491)$
$\bar{\hat{\beta}}_2$	$0.211(3.415)$	$0.211(18)$
$\bar{\hat{\sigma}}$	$0.351(3.114)$	$0.351(23)$

Table 3.16. conditional estimates from the full cross over design and the infertility cross over design for the scenario involving period effects

Treat	$\beta_0^M(Std \times 10^3)$	β_1^M
Full	$-0.272(2.0367)$	$0.627(2.5033)$
Incomplete	$-0.272(2.03653)$	$0.627(4.0679)$
Parallel	$-0.272(2.0365)$	$0.627(4.0865)$

Table 3.17. Predicted marginal regression estimates for the scenario involving period effects

It shows that in general,the conditional regression parameters for the incomplete crossover design, have higher standard errors than the conditional estimates obtained under complete crossover design. The predicted marginal estimates under incomplete crossover design have smaller standard errors than the marginal

estimates obtained under parallel design.

3.7 Sufficiency

Suppose that in the model involving treatment only, $Y_i = \tau_{Ci}, X_i = \tau_{Ii}$ and $Z_i = \tau_{Pi}$, $i = \{1, 2, ..., 2000\}$, are the treatment estimates obtained using the complete data set, the incomplete data set and the parallel data set respectively. That is for every sample we obtained the treatment estimate under the complete data set, the incomplete data set and the parallel data set. Here we will like to investigate the sufficiency of each of the three treatment estimates. That is to say instead of listing all the three treatment estimates, can we base our inferences only on one of them. In statistics, a statistic is said to be sufficient with respect to a parameter, when no other statistic which can be calculated from the same sample provides any additional information as to the value of the parameter. That is to say, conditional on the value of a sufficient statistic, the distribution of data is not a function of the underlying parameter the statistic is sufficient for. We will employ the latter definition to illustrate which of the three estimates (estimate from the parallel design, estimate from incomplete data set and estimate from the complete data set) is sufficient.

3.7.1 Conditional estimates

Here we will compare the treatment estimate from the incomplete data set with the treatment estimate from the complete set to see which of the two is

sufficient. To perform all these, we will consider the linear regression models:

$$E(\tau_I|\tau_C) = constant + \frac{Cov(\tau_C, \tau_I)}{Var(\tau_C)} * \tau_C, \quad (3.43)$$

and

$$E(\tau_C|\tau_I) = constant + \frac{Cov(\tau_C, \tau_I)}{Cov(\tau_I)} * \tau_I. \quad (3.44)$$

The following regression coefficient and constant were observed,

Parameter	$\frac{Cov(\tau_C,\tau_I)}{Var(\tau_C)}$	constant
$\hat{\beta}_0$	0.994	$-2.946 * 10^{-3}$
$\hat{\beta}_1$	0.994	$5.016 * 10^{-3}$
$\hat{\sigma}$	1.042	-0.015

Table 3.18. Regression coefficients for the model $E(\tau_I|\tau_C) = constant + \frac{Cov(\tau_C,\tau_I)}{Var(\tau_C)} * \tau_C$

Parameter	$\frac{Cov(\tau_C,\tau_I)}{Var(\tau_I)}$	constant
$\hat{\beta}_0$	0.891	-0.052
$\hat{\beta}_1$	0.694	0.263
$\hat{\sigma}$	0.355	0.226

Table 3.19. Regression coefficients for the model $E(\tau_C|\tau_I) = constant + \frac{Cov(\tau_C,\tau_I)}{Cov(\tau_I)} * \tau_I$

The constant term in the model $E(\tau_I|\tau_C) = constant + \frac{Cov(\tau_C,\tau_I)}{Var(\tau_C)} * \tau_C$ is $5.016 * 10^{-3} \approx 0$ and the slope term is $0.994 \approx 1$ signifying that the treatment estimate obtained using the complete data set are sufficient. That is to say that the model $E(\tau_I|\tau_C) = constant + \frac{Cov(\tau_C,\tau_I)}{Var(\tau_C)} * \tau_C$ can be simply be replaced by the model $E(\tau_I|\tau_C) = \tau_C$ without losing the meaning associated with the entire set of observations. On the other hand, the constant term in the model $E(\tau_C|\tau_I) = constant + \frac{Cov(\tau_C,\tau_I)}{Cov(\tau_I)} * \tau_I$ is $0.263 \neq 0$ and the slope term is $0.694 \neq 1$ signifying that

the treatment estimate obtained using the incomplete data set are not sufficient. We observe that here, we can not replace the model $E(\tau_C|\tau_I) = constant + \frac{Cov(\tau_C,\tau_I)}{Cov(\tau_I)} * \tau_I$ by the model $E(\tau_C|\tau_I) = \tau_I$. All what we are saying is that if you are presented with the treatment estimate obtained from the complete data set τ_C, the expectation (or best guess) of the treatment estimate obtained from the incomplete data set will be still τ_C. But not vice versa.

3.7.2 Marginalized estimates

Here we will first compare the marginalized treatment estimate from the complete data set with the treatment estimate obtained under parallel design. We consider the following linear regression models:

$$E(\tau_P|\tau_C) = constant + \frac{Cov(\tau_{MC}, \tau_P)}{Var(\tau_{MC})} * \tau_{MC}, \qquad (3.45)$$

and

$$E(\tau_{MC}|\tau_P) = constant + \frac{Cov(\tau_{MC}, \tau_P)}{Cov(\tau_P)} * \tau_P. \qquad (3.46)$$

The following slopes and constants were obtained.

Parameter	$\frac{Cov(\tau_{MC},\tau_P)}{Var(\tau_{MC})}$	constant
$\hat{\beta_0^M}$	0.991	-3.296×10^{-3}
$\hat{\beta_1^M}$	0.956	0.028

Table 3.20. Regression coefficients for the model $E(\tau_P|\tau_C) = constant + \frac{Cov(\tau_{MC},\tau_P)}{Var(\tau_{MC})} * \tau_{MC}$

The constant term in the model $E(\tau_P|\tau_{MC}) = constant + \frac{Cov(\tau_{MC},\tau_P)}{Var(\tau_{MC})} * \tau_{MC}$ is

Parameter	$\frac{Cov(\tau_{MC}, \tau_P)}{Var(\tau_P)}$	constant
$\hat{\beta}_0^M$	0.625	-0.131
$\hat{\beta}_1^M$	0.374	0.393

Table 3.21. Regression coefficients for the model $E(\tau_{MC}|\tau_P) = constant + \frac{Cov(\tau_{MC}, \tau_P)}{Cov(\tau_P)} * \tau_P$

0.028 and the slope term is $0.956 \approx 1$ signifying that the marginalized treatment estimate obtained using the complete data set are sufficient. On the other hand, the constant term in the model $E(\tau_{MC}|\tau_P) = constant + \frac{Cov(\tau_{MC}, \tau_P)}{Cov(\tau_P)} * \tau_P$ is $0.393 \neq 0$ and the slope term is $0.625 \neq 1$ signifying that the treatment estimate obtained using the parallel data set are not sufficient.

Secondly we compare the marginalized treatment estimate obtained under incomplete data set with the treatment estimate obtain under parallel design by considering the models:

$$E(\tau_P|\tau MI) = constant + \frac{Cov(\tau_{MI}, \tau_P)}{Var(\tau_{MI})} * \tau_{MI}, \qquad (3.47)$$

and

$$E(\tau_{MI}|\tau_P) = constant + \frac{Cov(\tau_{MI}, \tau_P)}{Cov(\tau_P)} * \tau_P, \qquad (3.48)$$

The following slopes and constants were obtained.

Parameter	$\frac{Cov(\tau_I, \tau_P)}{Var(\tau_I)}$	c
$\hat{\beta}_0$	0.997	-8.849×10^{-4}
$\hat{\beta}_1$	0.97	0.019

Table 3.22. Regression coefficients for the model $E(\tau_P|\tau MI) = constant + \frac{Cov(\tau_{MI}, \tau_P)}{Var(\tau_{MI})} * \tau_{MI}$

Parameter	$\frac{Cov(\tau_I, \tau_P)}{Var(\tau_P)}$	c
$\hat{\beta}_0$	0.994	-1.944×10^{-3}
$\hat{\beta}_1$	0.529	0.293

Table 3.23. Regression coefficients for the model $E(\tau_{MI}|\tau_P) = constant + \frac{Cov(\tau_{MI}, \tau_P)}{Cov(\tau_P)} * \tau_P$

The constant term in the model $E(\tau_P|\tau_{MI}) = constant + \frac{Cov(\tau_{MI}, \tau_P)}{Var(\tau_{MI})} * \tau_{MI}$ is 0.019 and the slope term is $0.97 \approx 1$ signifying that the marginalized treatment estimate obtained using the complete data set are sufficient. On the other hand, the constant term in the model $E(\tau_{MI}|\tau_P) = constant + \frac{Cov(\tau_{MI}, \tau_P)}{Cov(\tau_P)} * \tau_P$ is $0.293 \neq 0$ and the slope term is $0.529 \neq 1$ signifying that the treatment estimate obtained using the parallel data set are not sufficient. That is to say, if we are presented with the treatment estimate obtained from the incomplete data set τ_I, the expectation (or best guess) of the treatment estimate obtained from the parallel data set will be still τ_I. But not vice versa as is illustrated in Table 3.22 and Table 3.23.

Chapter 4

Analysis of crossover design in infertility trials

This section presents examples of cross over trials in infertility trials. The purpose of this chapter is to demonstrate how in practise the cross over trials in infertility should be analyzed using different statistical packages (fitting models in practice). To emphasize this, we will employ four statistical packages, namely, $SAS^©$, $R^©$ $WinBugs^©$ and GenStat$^©$. The first example that we will consider is an $AB:BA$ crossover trial from Gregoriou et al. (40) based on a crossover study to compare the pregnancy rates achieved by intrauterine insemination (IUI) and timed intercourse (TI) in gonadotrophin (hCG) stimulated cycles. Secondly, we will look at the crossover trial where couples were initially randomized to one treatment, subsequently alternating treatment on each cycle. These data are from Cohlen et al. (24) based on a crossover study to investigate whether the use of controlled ovarian hyper stimulation with human menopausal gonadotrophin in couples with male subfertility leads to a higher probability of conception when intrauterine insemination (IUI) is applied.

In $SAS^{©}$, $R^{©}$ and GenStat$^{©}$, we will carry out an analysis of deviance, described by Senn (84), while in $WinBugs^{©}$ we will make use of the deviance information criterion (DIC; Spiegelhalter et al. (90)). We adopt the attitude from Senn (84) that the investigator should determine on a *priori* grounds which form of analysis he/she favours. In both examples, we will report the analysis involving period effects and the analysis excluding period effects, and and we will not perform pre-testing of period effects.

4.1 Example I

In this section we look at the data from Gregoriou et al. (40) based on a crossover study to compare the pregnancy rates achieved by intrauterine insemination (IUI) and timed intercourse (TI) in gonadotrophin (hCG) stimulated cycles. Sixty-two couples were randomly equally divided into two groups; group A or group B. Couples randomized to group A will begin with protocol 1 before switching to protocol 2. Couples randomized to group B receive protocols in the reverse order. For all couples, controlled ovarian hyper-stimulation (COH) was performed with the help of gonadotrophin (hCG), and either timed intercourse (TI) or intrauterine insemination (IUI) was employed. In protocol 1 timed intercourse (TI) was employed, while in protocol 2 intrauterine insemination (IUI) was employed. Couples stayed in the same protocol for at most 3 cycles before they can switch to the alternative protocol, with each couple receiving in total at most 6 cycles of protocols.

Table 4.1 shows the achieved pregnancy rates in each group of attempts by the two protocols.

	Method of treatment									
	COH + TI					COH + IUI				
Attempts	Patients	Canceled cycles	Com-pleted cycles	Concep-tions	PR per cycle (%)	Patients	Canceled cycles	Com-pleted cycles	Concep-tions	PR per cycle (%)
1–3	Group A 31	15	73	4	5.5	Group B 31	14	70	8	11.4
4–6	Group B 23	14	55	1	1.4	Group A 27	13	60	7	11.7
Total	54	29	128	5	3.9	58	27	130	15	11.5

Table 4.1. Pregnancy rates achieved in each group of attempts by the two protocols

Of the 62 couples enrolled, 20 couples conceived, of which 12 conceived during the first 3 cycles and 8 conceived in the second 3 cycles. Couples left the study if and only if they conceived. We can analyze the data as the $AB : BA$ crossover design described earlier, where the first 3 cycles constitute the first period and the second 3 cycles constitute the second period. That is to say that if one couple conceives under cycle 1 and the other couple conceives under cycle 2 or 3, then those two couples are regarded as having conceived under the first period. Similarly, if one couple conceives under cycle 4 and the other couple conceives under cycle 5 or 6, then those two couples are regarded as having conceived under the second period. It is possible to have treated each cycle as a period of its own. Here, we can not do that as we are restricted by the way the data is presented. The way the data is presented does not allow us to treat each cycle as a period. We propose to analyze this data using the mixed effects model. We first

consider a scenario where the investigator believes that there is enough evidence to include period effects. Thus we assume that the probability that a couple $i, i = \{1, 2, ..., 62\}$ in group $k, k = \{A, B\}$ conceives under period $j, j = \{1, 2\}$ is

$$p_{(ij)k} = \frac{\exp(intercept + \beta_1 * protocol_{(ij)k} + \beta_2 * period_j + \upsilon_i)}{1 + \exp(intercept + \beta_1 * protocol_{(ij)k} + \beta_2 * period_j + \upsilon_i)} \qquad (4.1)$$

where as usual υ_i is the random couple effect, assumed to be distributed in the population as $N(0, \sigma^2)$. We code our protocol as 0 and 1 for protocol 1 and protocol 2 respectively. We store our data set in the F :drive under the file name **gregoriou.txt**. The file consists of five columns labelled: group, patient, period, protocol and response. The variable **RESPONSE** has two levels: 1 indicates conception and 0 indicates fail to conceive. Also, the variable **PERIOD** has two levels: 0 signifies the first period (cycles 1, 2 or 3) and 1 signifies the second period (cycles 4, 5 or 6). The factors **GROUP**, **PATIENT** and **PROTOCOL** are as defined above.

4.1.1 Analysis using $SAS^{©}$

The following statement may be used in SAS$^{©}$ to import the stored data set "**gregoriou.txt**" from the F :drive. The argument **DATAFILE=** specifies the complete path and filename of the input file. The **OUT=** argument identifies the output $SAS^{©}$ data set. If the specified $SAS^{©}$ data set does not exist, **PROC IMPORT** creates it. The line **DBMS=** specifies the type of data to import. For example, **DBMS=TAB** specifies to import a delimited file (tab-delimited values) file, which its extension is **.TXT**. We may not need to specify **DBMS=** if the filename specified with **DATAFILE=** contains a valid extension so that **PROC IMPORT** can recognize the type of data. The statement **REPLACE**

overwrites an existing $SAS^{©}$ data set. If **REPLACE** is not specified, **PROC IMPORT** does not overwrite an existing data set.

proc import datafile="F:*gregoriou.txt***"**

out=gregoriou

dbms=TAB

replace;

run;

proc sort data=gregoriou;

by patient;

run;

This code will import the data into the SAS$^{©}$ data set "**gregoriou**" and sort it by patient, which may then be used in subsequent analysis. Since **PROC NLMIXED** assumes that every time a new value of the **PATIENT** variable is encountered a new **PATIENT** is being analyzed, we need to sort our data by **PATIENT**. The **PROC SORT** arranges the data by **PATIENT**.

The following **PROC NLMIXED** fits the model involving protocol and period effects (saturated model).

proc nlmixed data=gregoriou;

parms intercept=-2 beta1=1 beta2=0.2 sigma=0.2;

sigma2=sigma**2;

preg=intercept + beta1*protocol + beta2*period + u;

ppreg=exp(preg)/(1+exp(preg));

model outcome~ binomial(1,ppreg);

random u~ normal(0,sigma2) subject=patient;

run;

The **PROC NLMIXED** statement invokes the procedure and inputs the **gregoriou** data set. The **PARMS** statement identifies the unknown parameters and their starting values. While the **PARMS** statement is not required, it is encouraged to use it to provide **PROC NLMIXED** with accurate starting values. Parameters not listed in the **PARMS** statement are assigned an initial value of 1. We illustrate a simpler way of obtaining reasonable starting values. The proportion of couples conceived under TI is $\frac{5}{54} \approx 0.093$ and the proportion of couples conceived under IUI is $\frac{15}{58} \approx 0.259$. A reasonable starting value for the intercept will be:

$$\log\left(\frac{0.093}{1-0.093}\right) \approx -2$$

While a reasonable starting value for beta1 will be:

$$\log\left(\frac{0.093}{1-0.093}\right) - \log\left(\frac{0.259}{1-0.259}\right) \approx 1$$

For the starting value of beta2, we consider the proportion of conception on each period. The proportion of conception in the first period (that is conception in cycles 1, 2 or 3) is $\frac{12}{62} \approx 0.194$, while the proportion of conception in the second period (that is conception in cycles 4, 5 or 6) is $\frac{8}{50} = 0.16$. Thus a reasonable starting value for beta2 will be:

$$\log\left(\frac{0.194}{0.806}\right) - \log\left(\frac{0.16}{0.84}\right) \approx 0.2$$

For sigma we will choose to use 0.2. Thus we opt to choose $-2, 1, 0.2$ and 0.2 as the starting values for the intercept, beta1, beta2 and sigma respectively.

The next three statements construct the variable **PPREG** to correspond to the p_{ij} and the **MODEL** statement defines the conditional distribution (given

the random effects) of the variable **OUTCOME**. Here a binomial conditional distribution is specified with $n = 1$ and $p = ppreg$; that is a binomial distribution with mean $ppreg$ and variance $ppreg(1 - ppreg)$. The **RANDOM** statement defines the random effect to be **U**, and specifies that it follows a normal distribution with mean zero and variance sigma2. The **SUBJECT=** argument defines a variable indicating when the random effect obtains new realizations: in this case, it changes according to the values of the **PATIENT** variable; that is, all observations from the same patient occur sequentially in the imported data set. Table 4.2 displays the fit statistics and the maximum likelihood estimates for the above **PROC NLMIXED** procedure (model with treatment and period effect).

Fit Statistics

-2 Log Likelihood	99.0
AIC (smaller is better)	107.0
AICC (smaller is better)	107.3
BIC (smaller is better)	115.5

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
intercept	-2.4719	0.9475	61	-2.61	0.0114	0.05	-4.3665	-0.5772	8.624E-6
beta1	1.3278	0.6392	61	2.08	0.0420	0.05	0.04963	2.6060	0.000012
beta2	-0.1793	0.6314	61	-0.28	0.7774	0.05	-1.4419	1.0833	-2.91E-6
sigma	0.9734	1.2454	61	0.78	0.4374	0.05	-1.5168	3.4637	3.699E-6

Table 4.2. Maximum likelihood estimates for the model with treatment and period effect

To carry out an analysis of deviance, we will fit a model with the period effect only and look at the change in deviance when then fitting the protocol effect as well. The following **PROC NLMIXED** fits the model with period effect only:

proc nlmixed data=gregoriou;

```
parms intercept=-2 beta2=0.2 sigma=0.2;

sigma2=sigma**2;

preg=intercept + beta2*period + u;

ppreg=exp(preg)/(1+exp(preg));

model outcome~binomial(1,ppreg);

random u~normal(0,sigma2) subject=patient;

run;
```

Table 4.3 shows the output for the model with period effect only.

Fit Statistics

-2 Log Likelihood	104.5
AIC (smaller is better)	110.5
AICC (smaller is better)	110.7
BIC (smaller is better)	116.9

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
intercept	-1.6293	0.7975	61	-2.04	0.0454	0.05	-3.2240	-0.03463	-1.48E-7
beta2	-0.09364	0.6677	61	-0.14	0.8889	0.05	-1.4287	1.2414	-9.79E-7
sigma	0.8110	1.4689	61	0.55	0.5829	0.05	-2.1262	3.7481	2.882E-7

Table 4.3. Maximum likelihood estimates for the model with period effect only

The difference in $-2logL$ for the model with period effect only and the model with protocol and period effect is $104.5 - 99 = 5.5$; since the difference in the number of parameters in these models is 1, this value should be compared to a chi-square distribution with 1 degrees of freedom. The following SAS$^{©}$ code will calculate the p-value. **PROCHI(x,n)** evaluates the probability that the chi-square of **n** degrees of freedom is smaller or equal to the valve **x**.

```
    data pvalue;
pvalue=1-probchi(5.5,1);
proc print;
run;
```

The output for the above **DATA** step is as follows:

```
Obs      pvalue
 1      0.019016
```

The p-value is 0.019016. Therefore at 5% level of significance, the protocol effect is influential. For this analysis the protocol effect estimate is 1.3278 with a 95% confidence interval of $(0.04963, 2.6060)$. The odds ratio for IUI compared to TI is the ratio of predicted odds of IUI versus TI, which is $exp(1.3278) = 3.8$. The estimated 95% confidence interval for the odds ratio is $exp(1.3278 \pm t_{(0.975,61)} * 0.6398) = exp(1.3278 \pm 1.99962 * 0.6398) = (1.050896, 13.54418)$. These odds are in favour of IUI. Couples undergoing IUI have nearly four times higher odds of conception than couples undergoing TI.

Suppose the investigator believes that there is no period effects. Here, we illustrate how the data must be analyzed if the investigator feels that there is no need to include period effect. We first fit the the model with protocol only. The following **PROC NLMIXED** fits the model with protocol only:

```
proc nlmixed data=gregoriou;
parms intercept=-2 beta1=1 sigma=0.2;
sigma2=sigma**2;
preg=intercept + beta1*protocol + u;
ppreg=exp(preg)/(1+exp(preg));
```

model outcome~binomial(1,ppreg);

random u~normal(0,sigma2) subject=patient;

run;

Table 4.4 contains the fit statistics and the maximum likelihood estimates for the

model with protocol only.

Fit Statistics

-2 Log Likelihood	99.0
AIC (smaller is better)	105.0
AICC (smaller is better)	105.3

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
intercept	-2.6333	0.8390	61	-3.14	0.0026	0.05	-4.3110	-0.9556	-0.00002
beta1	1.3451	0.6539	61	2.06	0.0440	0.05	0.03762	2.6525	-0.00001
sigma	1.1499	1.0726	61	1.07	0.2879	0.05	-0.9948	3.2946	-0.00001

Table 4.4. Maximum likelihood estimates for the model with protocol only

To carry out an analysis of deviance, we will fit the null model (model excluding

both protocol and period effect) and look at the change in deviance when then

fitting the protocol effect as well. The following **PROC NLMIXED** fits the

null model:

proc nlmixed data=gregoriou;

parms intercept=-2 sigma=0.2;

sigma2=sigma2;**

preg=intercept + u;

ppreg=exp(preg)/(1+exp(preg));

model outcome~binomial(1,ppreg);

random u~normal(0,sigma2) subject=patient;

run;

Table 4.5 contains the fit statistics and the maximum likelihood estimates for the null model.

Fit Statistics

-2 Log Likelihood	104.5
AIC (smaller is better)	108.5

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
intercept	-1.7140	0.5055	61	-3.39	0.0012	0.05	-2.7248	-0.7031	-3.19E-6
sigma	0.9280	1.0100	61	0.92	0.3618	0.05	-1.0916	2.9476	1.634E-7

Table 4.5. Maximum likelihood estimates for the null model

If we take the difference in $-2 \log L$ for the null and model involving protocol effects, $104.5 - 99 = 5.5$, we get the log likelihood ratio test for the protocol effects. The p-value is 0.019016. Therefore at 5% significance level we reject the hypothesis that the null model is as good as the model involving the protocol only. Here, the protocol effect estimate is 1.3451 with a 95% confidence interval of $(0.03762, 2.6525)$. The odds ratio for IUI compared to TI is $exp(1.3451) = 3.8$. The estimated 95% confidence interval for the odds ratio is $exp(1.3451 \pm t_{(0.975,61)} * 0.6539) = exp(1.3451 \pm 1.99962 * 0.6539) = (1.038366, 14.18736)$. These odds are in favour of IUI. Couples undergoing IUI have nearly four times higher odds of conception than couples undergoing TI.

4.1.2 Analysis using $R^{©}$

We begin by importing the stored data set **"gregoriou.txt"**. The following statement may be used in $R^{©}$ to import the data set **"gregoriou.txt"** from the

F :drive into the $R^{©}$ data set **"gregoriou"**. The **"F:/gregoriou.txt"** specifies the complete path and filename of the input file. the **HEADER=T** option specifies that the first line is a line of headings, and hence, by implication from the form of the file, that no explicit row labels are given.

> **gregoriou <- read.table("F:/gregoriou.txt", header=T)**.

The *lmer* function in the *lme4* package is designed to fit generalized mixed-effects model. We first assume that the investigator believes that there is period effect. The following **lmer** code fits the model with protocol and period effect.

> **fit <- lmer(response~protocol+period+(1|patient),family=binomial, data=gregoriou)**
> **fit**

We start by looking at the terms just after the function **lmer**. The term on the right hand side of $\sim$, **response** indicates the outcome to be modelled, which must be binary. The terms on the other side of $\sim$ are the terms that should be included in the model. The **(1|patient)** declares **Patient** to be the random effect. The **family** indicates the distribution of the responses which should be binomial. The **data** indicates the data that was read in $R^{©}$.

The output for the parameter estimates and their p-values for the above **lmer** analysis are shown in Table 4.6.

```
Generalized linear mixed model fit by the Laplace approximation
Formula: response ~ beta1 + beta2 + (1 | patient)
   Data: greg
   AIC   BIC logLik deviance
 107.0 117.8 -49.48   98.96
Random effects:
 Groups  Name          Variance Std.Dev.
 patient (Intercept) 0.94758  0.97344
Number of obs: 111, groups: patient, 62

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.4719     0.5700  -4.336 1.45e-05 ***
beta1         1.3278     0.6255   2.123   0.0338 *
beta2        -0.1793     0.5727  -0.313   0.7542
```

Table 4.6. lmer analysis for the model with protocol and period effect from the

The following **lmer** code will fit the reduced model with period effect only.

> **fit1 < − lmer(response∼period+(1|patient),family=binomial,data=gregoriou)**

> **fit1**

The output for the parameter estimates and their p-values are as shown in Table 4.7.

```
Generalized linear mixed model fit by the Laplace approximation
Formula: response ~ beta2 + (1 | patient)
   Data: greg
   AIC   BIC logLik deviance
 110.5 118.6 -52.25    104.5
Random effects:
 Groups  Name          Variance Std.Dev.
 patient (Intercept) 0.65756  0.8109
Number of obs: 111, groups: patient, 62

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.62926    0.35508  -4.588 4.47e-06 ***
beta2       -0.09371    0.53619  -0.175    0.861
```

Table 4.7. lmer analysis for the model with period effect only

The difference in $-2logL$ for the model with period effect only and the model with protocol and period effect is $104 - 98.96 \approx 5.5$ as in the $SAS^{©}$ analysis; Therefore at 5% level of significance, the protocol effect is influential. The protocol estimate is 1.3278 with 95% confidence interval of $(1.3278 \pm 1.96 * 0.6255) = (0.10182, 2.55378)$. The odds ratio for IUI compared to TI is $exp(1.3278) = 3.8$. The estimated 95% confidence interval for the odds ratio is $exp(1.3278 \pm 1.96 * 0.6255) = (1.105684, 12.89366)$. These odds are in favour of IUI. Couples undergoing IUI have nearly four times higher odds of conception than couples undergoing TI.

Next we assume that there is no period effect. The following **lmer** code will fit the model with protocol only.

> **fit1 $< -$ lmer(response~protocol+(1|patient),family=binomial, data=gregoriou)**

> **fit1**

The output for the parameter estimates and their p-values are as shown in Table 4.8.

```
Generalized linear mixed model fit by the Laplace approximation
Formula: response ~ beta1 + (1 | patient)
   Data: greg
   AIC   BIC logLik deviance
 105.0 113.2 -49.52    99.04
Random effects:
 Groups  Name          Variance Std.Dev.
 patient (Intercept) 1.3223   1.1499
Number of obs: 111, groups: patient, 62

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.6333     0.5657  -4.655 3.24e-06 ***
beta1         1.3450     0.6391   2.105  0.0353 *
```

Table 4.8. lmer analysis for the model with protocol only

Now we fit the null model excluding both period effect and the protocol. The following $R^©$ code will fit the null model.

> **fit1 < − lmer(response∼(1|patient),family=binomial,data=gregoriou)**

> **fit1**

The output for the parameter estimates and their p-values for the null model are as shown in Table 4.9.

```
Generalized linear mixed model fit by the Laplace approximation
Formula: response ~ (1 | patient)
   Data: greg
   AIC   BIC logLik deviance
 108.5 113.9 -52.26   104.5
Random effects:
 Groups  Name          Variance Std.Dev.
 patient (Intercept) 0.86086  0.92783
Number of obs: 111, groups: patient, 62

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.7139     0.2908  -5.894 3.78e-09 ***
```

Table 4.9. lmer analysis for the null model

If we take the difference in $-2 \log L$ for the null and model involving protocol effect only, $104.5 - 99.04 = 5.46$, we get the log likelihood ratio test for the protocol effects. Since the difference is 5.46 with 1 degrees of freedom, the p-value is 0.019457. This test shows that the model with protocol only provides a better fit than the null model. Thus, we reject the hypothesis that the null model is as good as the model involving the protocol only. The treatment effect estimate is 1.3450 with 95% confidence interval of $(1.3450 \pm 1.96 * 0.6391) = (0.092364, 2.597636)$. The odds ratio for IUI compared to TI is $exp(1.3450) = 3.879088$. The estimated 95% confidence interval for the odds ratio is $exp(1.3450 \pm 1.96 * 0.6391) = (1.099580, 13.68461)$. These odds are in favour of IUI. Couples undergoing IUI have nearly four times higher odds of conception than couples undergoing TI.

4.1.3 Analysis using $WinBugs^{©}$

We first assume that there is period effect. The following $WinBugs^{©}$ code will specify the random effects model involving protocol and period effects.

model

{ for(i in 1 : N1) {

$outcome1[i] \sim dbern(p1[i])$

logit(p1[i]) $<-$ intercept+beta1*protocol1[i]+ beta2*period1[i]+$v[i]$

$v[i] \sim dnorm(0, tau)$

}

```
   for( i in 1 : N2){
outcome2[i] ~ dbern(p2[i])
logit(p2[i]) < − intercept+beta1*protocol2[i]+ beta2*period2[i]+υ[i]
}


   sigma< − 1/sqrt(tau)
intercept ~ dnorm(0,1.0E-3)
beta1 ~ dnorm(0,1.0E-3)
beta2 ~ dnorm(0,1.0E-3)
tau ~ dgamma(1,1.0E-4)
log.sigma< − log(sigma)



   }
```

The $WinBugs^{©}$ language use the $\sim$ to denote stochastic (probabilistic) rela-
tionships, and the left arrow $< −$ to denote deterministic (logical) relationships.
The stochastic parameters **intercept, beta1, beta2**, and **tau** are given proper
but minimally informative prior distributions, while the logical expression for
sigma allows the standard deviation (of the random effects distribution) to be
estimated. The results are very sensitive to the prior distributions.

The following code will specify the data to be used.

```
list(outcome1 =c(0,0,0,0,0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0, 0, 0, 0,
0, 0,0, 0, 0, 0, 0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,0, 0, 0, 0, 0, 0, 0, 0,0, 0, 1,
1, 1, 1, 1, 1, 1,1,1,1,1,1) ,
```

treatment1=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,1,0,0,0,0,1,1,1,1,1,1,1,1) ,

period1 =c(0,0,0,0,0,0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0, 0, 0, 0, 0, 0,0,0,0,0,0,0),

outcome2 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0, 0, 0, 0, 0, 0,0,0,0,0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0, 0, 1, 1, 1, 1,1,1,1,1) ,

period2=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1),

treatment2=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0, 0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,0),

N1 = 62, N2=50),

The data have to be sorted in such a way that couples who failed to conceive are listed first. For example, in the first period, we listed couples who did not conceive followed by those who conceived in the second period, followed by those who conceived in the first period. In the second period, we listed couples with outcome $(0, 0)$ (those who did not conceive in both periods) first, followed by couples with outcome $(0, 1)$ (those who conceived in the second period). This will ensure that each couple keeps its random effect.

Finally the $MCMC$ sampler must be given some initial values for each stochastic node. These can be arbitrary values, although in practice, convergence can be poor if wildly inappropriate values are chosen. Different set of initial values are needed for each chain. To check for convergence we need to specify at least two chains. The following codes will initialize our model. Here we give the same

values as in $SAS^©$

list(intercept=-1,beta1=-1, beta2=1, tau=10,b=c(0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,0))

list(intercept=-4,beta1=-4,beta2=4, tau=15,b=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0,0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,1))

To run the model, we first click once on the **Model** on the tool bar. Secondly we click once on the **Specification** option. We highlight the word model at the beginning of the code. Finally we Check the model syntax by clicking once on the **check model** button in the **Specification** window. A message saying "model is syntactically correct" should appear in the bottom left of the WinBUGS$^©$ program window.

To load the data we highlight the word **list** at the beginning of the data code and we click once on the **load data** button in the **Specification** option. A message saying "data loaded" should appear in the bottom left of the WinBUGS$^©$ program window.

Now we need to select the number of chains (i.e. sets of samples to simulate). The default is 1, but we will use 2 chains for this example, since running multiple chains is one way to check the convergence of the $MCMC$ simulations. Next we compile the model by clicking once on the **compile** button in the **Specification** option. A message saying "model compiled" should appear in the bottom left of the WinBUGS$^©$ program window. To load the initial values, we highlight the

word list at the beginning of the first set of initial values and click once on the load **inits** button in the **Specification** window. A message saying "initial values loaded: model contains uninitialized nodes (try running gen inits or loading more files)" should appear in the bottom left of the WinBUGS$^{\copyright}$ program window. We repeat this process for the second initial values. A message saying "initial values loaded: model initialized" should now appear in the bottom left of the WinBUGS$^{\copyright}$ program window. We set some monitors to store the sampled values for selected parameters. For our example, we set monitors for the parameters intercept, beta1, beata2 and sigma. For model selection we use the deviance information criterion box. Now, we select the **Update** option from the **Model** menu and we indicate the number of updates (iterations of the simulation) we require in the appropriate white box (labelled **updates**), the default value is 1000. We click once on the **update** button, the program will simulate values for each parameter in the model. This may take a few seconds, the box marked iteration will indicate how many updates have currently been completed. The number of times this value is revised depends on the value set for the refresh option in the white box above the iteration box. The default is every 100 iterations. When the updates are finished, the message "updates took $***s$" will appear in the bottom left of the WinBUGS$^{\copyright}$ program window (where $***$ is the number of seconds taken to complete the simulation).

We cheek convergence by looking at the graph of the two chains. Convergence is signified by the two graphs merging together. Once we are satisfied that convergence is reached, we simulate extra samples and our estimates will be based only on the samples simulated after convergence. One way to assess the accuracy of the posterior estimates is by calculating the Monte Carlo error for each parameter. This is an estimate of the difference between the mean of

the sampled values (which we are using as our estimate of the posterior mean for each parameter) and the true posterior mean. The simulation should be run until the Monte Carlo error (MC error) for each parameter of interest is less than about 5% of the sample standard deviation. We first ran 20000 simulation and we checked convergence for the intercept, *beta*1, *beta*2 and sigma. We look for evidence of when the simulation appears to have stabilized by examining the plots of the sample values versus iterations. Fig 4.1 shows the graphs for the intercept, *beta*1, *beta*2 and sigma. The plots of the intercept, *beta*1 and *beta*2 show chains for which convergence look reasonable. The plot of sigma shows chains which have clearly not reached convergence.
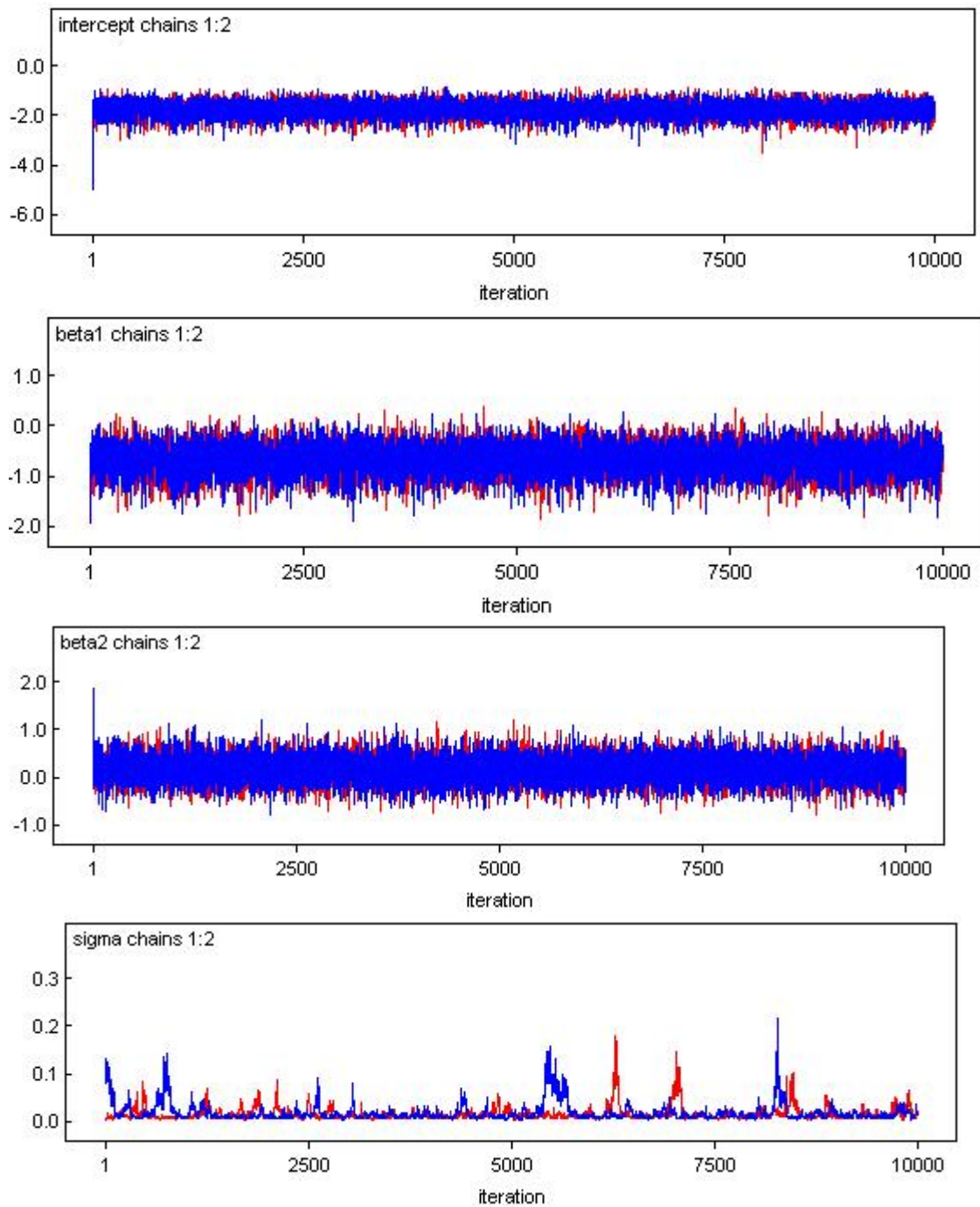
Figure 4.1. Plot of the intercept, beta1, beta2 and sigma for the model with protocol and period effects

We simulated extra 30000, and Table 4.10 below shows the summary statistics for the monitored parameters and the fit statistic (deviance information criterion)

Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes

	Dbar	Dhat	pD	DIC
outcome1	61.362	59.732	1.631	62.993
outcome2	41.032	39.631	1.401	42.433
total	102.394	99.363	3.032	105.426

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
intercept	-2.28	0.5344	0.004679	-3.42	-2.245	-1.321	20001	60000
beta1	1.32	0.5842	0.003773	0.2377	1.301	2.542	20001	60000
beta2	-0.313	0.5288	0.002317	-1.372	-0.3081	0.7094	20001	60000
sigma	0.06426	0.1967	0.005039	0.01643	0.03876	0.2426	20001	60000

Table 4.10. Parameter estimates and their standard errors

Next we fit the model with period effect only. The $WinBugs^©$ code that specify the model with period effect only, the data to be used and the initial values for the $MCMC$ sampler is in the Appendix $A.1$. We ran 20000 simulation and we checked convergence for the intercept, $beta2$ and sigma. We look for evidence of when the simulation appears to have stabilized by examining the plots of the sample values versus iterations. Fig A.1 shows the plots for the intercept, $beta2$ and sigma. The plots of the intercept and $beta2$ show chains for which convergence look reasonable. The plot of sigma shows chains which have clearly not reached convergence.

We simulated extra 30000 samples. Table 4.11 shows the summary statistics for the monitored parameters. To compare these two models: the model with protocol and period effect with the model with period effect only, here unlike in $R^©$ and $SAS^©$, we will use the deviance information criterion (DIC; (90)).

```
Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes
          Dbar      Dhat      pD        DIC
outcome1  61.937    60.924    1.013     62.950
outcome2  44.982    43.978    1.004     45.986
total     106.919   104.902   2.017     108.936
```

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
intercept	-1.462	0.326	0.00142	-2.134	-1.451	-0.8487	20001	60000
beta2	-0.245	0.512	0.002718	-1.278	-0.2433	0.7417	20001	60000
sigma	0.0532	0.05522	0.001567	0.0163	0.03759	0.1849	20001	60000

Table 4.11. Parameter estimates and their standard errors

The idea is that models with smaller deviance information criterion (DIC) should be preferred to models with larger deviance information criterion (DIC). The advantage of DIC over other criteria (likelihood ratio test), for Bayesian model selection, is that the DIC is easily calculated from the samples generated by a Markov chain Monte Carlo simulation. The likelihood ratio test require calculating the likelihood at its maximum over the unknown parameters, which is not readily available from the MCMC simulation. But to calculate DIC in $WinBugs^{©}$ we simply use the DIC tool dialog from the option **SAMPLES**. The DIC of the model with protocol and period effects is 105.426 while the DIC of the model with period effect only is 108.936. Thus our model of choice is the model with both period effects and protocol. The protocol effect is 1.32 with estimated 95% confidence interval of $1.32 \pm 1.96 * 0.5842 = (0.174968, 2.465032)$. The estimated odds ratio of IUI verses TI is $\exp(1.32) = 3.7$. The estimated 95% confidence interval for the odds ratio is $\exp(3.7 \pm 1.96 * 0.5842) = (1.19, 11.8)$. These odds are in favour of IUI. Couples undergoing IUI have nearly four times higher odds of conception than couples undergoing TI. These confidence intervals, clearly shows that protocol is statistically significant.

Suppose we assume that there is no period effect. We fit the model with protocol only. The $WinBugs^{©}$ code that specify the random effects model involving protocol only is in Appendix $A.2$. We ran 20000 simulation and we checked

convergence for the intercept, *beta*1 and sigma. We look for evidence of when the simulation appears to have stabilized by examining the plots of the sample values versus iterations. Fig A.2 shows the plots for the intercept, *beta*1 and sigma. The plots of the intercept and *beta*1 show chains for which convergence look reasonable. The plot of sigma shows chains which have clearly not reached convergence.

We simulated extra 30000 samples. Table 4.12 shows the summary statistics for the monitored parameters.

Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes

	Dbar	Dhat	pD	DIC
outcome1	61.027	59.910	1.117	62.143
outcome2	40.627	39.729	0.899	41.526
total	101.654	99.639	2.015	103.669

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
intercept	-2.38	0.491	0.003813	-3.436	-2.344	-1.503	20001	60000
beta1	1.30	0.578	0.004338	0.224	1.279	2.501	20001	60000
sigma	0.5343	0.08985	0.002816	0.01638	0.03736	0.167	20001	60000

Table 4.12. Parameter estimates and their standard errors

Next we fit the null model excluding both period effect and the protocol. The $WinBugs^{©}$ code that specify the null model is in Appendix A.3. We first ran 20000 simulation and we checked convergence for the intercept and sigma. We look for evidence of when the simulation appears to have stabilized by examining the plots of the sample values versus iterations. Fig A.3 shows the plots for the intercept and sigma. The plots of the intercept show chains for which convergence look reasonable. The plot of sigma shows chains which have clearly not reached convergence.

We simulated extra 30000. Table 4.13 shows the summary statistics for the monitored parameters.

Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes

	Dbar	Dhat	pD	DIC
outcome1	61.600	61.031	0.568	62.168
outcome2	44.497	44.039	0.458	44.955
total	106.096	105.070	1.026	107.123

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
intercept	-1.545	0.2501	0.001007	-2.048	-2.048	-1.538	20001	60000
sigma	0.01676	0.0159	6.199E-4	0.005268	0.01215	0.05921	20001	60000

Table 4.13. Parameter estimates and their standard errors

The DIC for the model with protocol only is $103,669$. While the DIC for the null model is 107.123. Thus the model with protocol only is preferred. The protocol effect is 1.30 with estimated 95% confidence interval of $1.30 \pm 1.96 * 0.578 = (0.16712, 2.43288)$. The estimated odds ratio of IUI verses TI is $\exp(1.30) = 3.7$. The estimated 95% confidence interval for the odds ratio is $\exp(1.3 \pm 1.96 * 0.578) = (1.18, 11.4)$. These odds are in favour of IUI. Couples undergoing IUI have nearly four times higher odds of conception than couples undergoing TI. These confidence intervals, clearly shows that protocol is statistically significant.

4.1.4 Analysis using $GenStat^{©}$

The following $GenStat^{©}$ will read data in parallel in five columns and declares Group, Patient, Period, and Protocol as Factors.

FACTOR Group, Patient, Period, Protocol

READ [SETNVALUES=yes; SETLEVELS=yes] Group,Patient,Period,

Protocol,Response;

FREP=2(levels),labels,*

```
1 1 1 1 1
1 2 1 1 1
1 3 1 1 1
. . . . .
. . . . .
. . . . .
2 60 -1 1 0
2 61 -1 1 0
2 62 -1 1 0:
```

The next *GenStat*© statements invokes **HGANALYSE** procedure (Lee and Nelder (56)). The terms **HGRANDOMMODEL** and **HGFIXEDMODEL** define respectively, the random model (**PATIENT**) and the fixed model (**PROTOCOL + PERIOD**) for a hierarchical generalized linear model. The **LINK** and **DISTRIBUTION** options specify their distribution and link function respectively. The variate to be analyzed is **RESPONSE** and is listed on **HGANALYSE** . The fitting process involves alternative fits of the augmented Generalized Linear Models for the mean given the current estimates of the dispersion parameters, and of the models that estimate the dispersion parameters. The convergence of the process is assessed by comparing the dispersion estimates from successive fits. The **MAXCYCLE** option can specify two scalars. The first sets a limit on the number of alternating fits (default 99), and the second controls the number of iterations in the estimation of the mean model and of the dispersion model (default 30). The **MLAPLACE** option specifies the order of Laplace approximation to use in the estimation of the mean model (0 or 1); default is 0. The **DLAPLACE** option specifies the order of Laplace approximation to use in the estimation of

the dispersion components (0, 1 or 2); default is 0. The term **NBINOMIAL** specifies the binomial totals, which is 1.

HGRANDOMMODEL [DISTRIBUTION=normal; LINK=identity] Patient

HGFIXEDMODEL [DISTRIBUTION=binomial; LINK=logit] Protocol + Period

HGANALYSE [MLAPLACE=1; DLAPLACE=1;Maxcycle=999,30] Response; NBINOMIAL=1

To present this program to a computer, we click on the **RUN** option, and click once on the **SUBMIT WINDOW** option. The parameter estimates and deviances for the saturated model described above are as follows:

Estimates from the mean model

	estimate	s.e.	t(108)
constant	-2.150	0.5076	-4.236
Treatment 1	1.220	0.5610	2.175
Period 1	-0.245	0.5205	-0.471

Estimates from the dispersion model

Estimates of parameters

Parameter	estimate	s.e.	t(*)	antilog of estimate
lambda Patient				
	-1.349	0.734	-1.84	0.2594

Likelihood statistics

-2 * h(y|v)	91.512
-2 * h	125.507
-2 * P_v(h)	99.129
-2 * P_beta,v(h)	98.760
-2 * EQD(y|v)	96.631
-2 * EQD	130.626
-2 * P_v(EQD)	104.248
-2 * P_beta,v(EQD)	103.879

Next we fit the model involving period effect only. The following *GenStat*© will fit the hierarchical model involving period effect only.

HGRANDOMMODEL [DISTRIBUTION=normal; LINK=identity] Patient

HGFIXEDMODEL [DISTRIBUTION=binomial; LINK=logit] Period

HGANALYSE [MLAPLACE=1; DLAPLACE=1;Maxcycle=999,30] Response; NBINOMIAL=1

The parameter output for the above analysis are as follows:

Estimates from the mean model

	estimate	s.e.	t(109)
constant	-1.4273	0.3218	-4.435
Period 1	-0.1960	0.4999	-0.392

Estimates from the dispersion model

Estimates of parameters

Parameter	estimate	s.e.	t(*)	antilog of estimate
lambda Patient				
	-2.88	1.49	-1.93	0.05621

Likelihood statistics

-2 * h(y|v)	102.730
-2 * h	39.103
-2 * P_v(h)	104.542
-2 * P_beta,v(h)	105.021
-2 * EQD(y|v)	107.849
-2 * EQD	44.222
-2 * P_v(EQD)	109.661
-2 * P_beta,v(EQD)	110.140

For model selection we use the information criterion based on $-2P_v(h)$ (Lee ad Nelder (56)). If we take the difference in $-2P_v(h)$ for the model with protocol and period effects and the model with period effects only, $104.542 - 99.129 = 5.413$, we get the log likelihood ratio test for the protocol (treatment) effect. Since the difference is 5.413 with 1 degrees of freedom, the p-value is 0.019987. This test shows that the difference between the two protocols is statistically significant.

Now we assume that there is no period effect and we fit the model with protocol only. The following *GenStat*© will fit the hierarchical model involving protocol only.

HGRANDOMMODEL [DISTRIBUTION=normal; LINK=identity] Patient

HGFIXEDMODEL [DISTRIBUTION=binomial; LINK=logit] Protocol

HGANALYSE [MLAPLACE=1; DLAPLACE=1;Maxcycle=999,30] Response; NBINOMIAL=1

The parameter output for the above analysis are as follows:

Estimates from the mean model

	estimate	s.e.	t(109)
constant	-2.237	0.4759	-4.701
Treatment 1	1.199	0.5588	2.146

Estimates from the dispersion model

Estimates of parameters

Parameter	estimate	s.e.	t(*)	antilog of estimate
lambda Patient				
	-1.107	0.654	-1.69	0.3306

Likelihood statistics

-2 * h(y|v)	89.754
-2 * h	139.749
-2 * P_v(h)	99.369
-2 * P_beta,v(h)	99.518
-2 * EQD(y|v)	94.873
-2 * EQD	144.868
-2 * P_v(EQD)	104.488
-2 * P_beta,v(EQD)	104.637

The null model is fit next. The following *GenStat*© will fit the null model.

HGRANDOMMODEL [DISTRIBUTION=normal; LINK=identity] Patient

HGFIXEDMODEL [DISTRIBUTION=binomial; LINK=logit]

HGANALYSE [MLAPLACE=1; DLAPLACE=1;Maxcycle=999,30] Response; NBINOMIAL=1

The parameter estimates and deviances for the null model are as follows:

Estimates from the mean model

	estimate	s.e.	t(110)
constant	-1.502	0.2517	-5.970

Estimates from the dispersion model

Estimates of parameters

Parameter lambda Patient	estimate	s.e.	t(*)	antilog of estimate
	-1.760	0.865	-2.03	0.1720

Message: s.e.s are based on dispersion parameter with value 1.

Likelihood statistics

-2 * h(y|v)	99.277
-2 * h	106.744
-2 * P_v(h)	104.695
-2 * P_beta,v(h)	105.607
-2 * EQD(y|v)	104.396
-2 * EQD	111.863
-2 * P_v(EQD)	109.814
-2 * P_beta,v(EQD)	110.726

The difference in $-2P_v(h)$ for the null model and the model with protocol only, $104.695 - 99.369 = 5.326$, the p-value is 0.021010. This test shows that the difference between the two protocols is statistically significant.

4.1.5 Summary

Table 4.14 records comparative analyses of the **gregoriou** data.

Model	Method	intercept	Beta1	Beta2	Deviance
Null	$SAS^{©}$	-1.71			104.5
		(0.506)			
	$R^{©}$	-1.71			104.5
		(0.291)			
	$WinBugs^{©}$	-1.55			-
		(0.250)			
	$GenStat^{©}$	-1.50			104.7
		(0.252)			
Protocol	$SAS^{©}$	-2.63	1.35		99.0
		(0.839)	(0.654)		
	$R^{©}$	-2.63	1.35		99.0
		(0.566)	(0.639)		
	$WinBugs^{©}$	-2.38	1.30		-
		(0.491)	(0.578)		
	$GenStat^{©}$	-2.24	1.20		99.4
		(0.476)	(0.559)		
Period	$SAS^{©}$	-1.63		-0.0936	104.5
		(0.798)		(0.668)	
	$R^{©}$	-1.63		-0.0937	104.5
		(0.355)		(0.536)	
	$WinBugs^{©}$	-1.59		-0.256	-
		(0.326)		(0.512)	
	$GenStat^{©}$	-1.43		-0.196	104.5
		(0.322)		(0.500)	
Protocol and Period	$SAS^{©}$	-2.47	1.33	-0.179	99.0
		(0.948)	(0.639)	(0.631)	
	$R^{©}$	-2.47	1.33	-0.179	99.0
		(0.570)	(0.626)	(0.573)	
	$WinBugs^{©}$	-2.28	1.32	-0.313	-
		(0.534)	(0.584)	(0.529)	
	$GenStat^{©}$	-2.15	1.22	-0.245	99.1
		(0.508)	(0.561)	(0.521)	

Table 4.14. Comparative analysis of the **Gregoriou** data

Table 4.14 gives the summary results of fitting various models for the data in

Gregoriou et al (40) using different fitting methods. The models are, in order of fitting : a null model, a model involving protocol (treatment) only, a model involving period only and a model fitting protocol and period. The most important models are the model involving protocol only and the model involving both protocol and period. The other two models are used to examine the change in deviance. Four statistical packages: $SAS^{©}$, $R^{©}$, $WinBugs^{©}$ and $GenStat^{©}$ are illustrated. $SAS^{©}$, and $R^{©}$ give similar results in all the four models since the general fitting criterion (maximum likelihood) is the same and only details of numerical implementation are different. In fact, the deviances are in good agreement between $SAS^{©}$, $R^{©}$ and $GenStat^{©}$. All the four methods ($SAS^{©}$, $R^{©}$, $WinBugs^{©}$ and $GenStat^{©}$) indicate that IUI is more effective than TI, and that couples undergoing IUI have nearly four times higher odds of conception than couples undergoing TI.

4.2 Example II

The following data are from (24) based on a randomized crossover trial to investigate whether the combination of intrauterine insemination (IUI) and ovarian stimulation improves the probability of conception in couples with male-subfertility. Seventy-four couples with subfertility possibly related to male subfertility were randomized to either group 1 or group 2. Couples randomized to group 1 were treated with intrauterine insemination (IUI) in a natural cycle, while couples randomized to group 2 were given a combination of intrauterine insemination (IUI) and ovarian stimulation. This treatment was then alternated according to a crossover design, with each couple receiving at most six treatment cycles.

Table 4.15 shows the pregnancy rate per completed IUI cycle and the number of drop-outs before starting a new treatment cycle. Altogether, nine patients dropped out before completing six cycles for reasons other than pregnancy and thirty-four couples conceived. Twenty-one couples conceived under simulated cycles, and thirteen couples conceived under natural cycle.

Cycle no.	Natural cycles		Stimulated cycles		
	No. of cycles[a]	Pregnancies (%)	No. of cycles[a]	Pregnancies (%)	No. of drop-outs
1	38 (38)[a]	4 (10.5)	36 (36)[a]	3 (8.3)	0
2	32 (33)	2 (6.3)	32 (34)	5 (15.6)	0
3	26 (29)	6 (23.1)	29 (30)	5 (17.2)	1
4	25 (25)	0	19 (21)	2 (10.5)	2
5	15 (17)	1 (6.7)	22 (23)	3 (13.6)	4
6	19 (19)	0	15 (15)	3 (20.0)	2
Total after 6 cycles	155 (161)	13 (8.4)	153 (159)	21 (13.7)	9

[a]Values in parentheses are number of started cycles

Table 4.15. Pregnancy rate per completed IUI cycle and number of drop-outs for reasons other than pregnancy before starting a new treatment cycle

To analyze these data, we first consider a scenario where the investigator believes that there is period effect. We fit the model with treatment and period effects, while regarding period as a factor. Thus the probability that couple $i, i = \{1, 2, ..., 74\}$ in group $k, k = \{1, 2\}$, conceives under period $j, j = \{1, 2, 3, 4, 5, 6\}$ is

$$p_{(ij)k} = \frac{\psi_{(ij)k}}{1 + \psi_{(ij)k}} \qquad (4.2)$$

where as usual υ_i is the random couple effect, assumed to be distributed in the population as $N(0, \sigma^2)$, and $\psi_{(ij)k} = \exp(intercept + \beta_1 treatment_{(ij)k} + \beta_2 period_2 + \beta_3 period_3 + \beta_4 period_4 + \beta_5 period_5 + \beta_6 period_6 + \upsilon_i)$.

We store the data set from Table 4.15 in the F :drive under the file name **cohlendata.txt**. The file consists of ten columns labelled group, patient, period, period2, period3 period4 period5 period6 treatment response. The variable

PERIOD indicates the period number which can be either $1, 2, 3, 4, 5$ or 6. The variable **PERIOD**$_k$, $(k = 2, 3, 4, 5, 6)$ takes the value 1 if the variable $PERIOD$ is k, otherwise it takes the value -1. The variable **TREATMENT** is -1 for IUI in natural cycle and 1 for IUI in stimulated cycle. The variables **GROUP**, **PATIENT** and **RESPONSE** are as defined previously. This stored file will be used in $SAS^©$ and $R^©$ analysis.

4.2.1   Analysis using $SAS^©$

The following SAS$^©$ code will import the data set **cohlendata** and sort it by patient.

proc import datafile="F:*cohlendata.txt***"**

out=cohlen

dbms=TAB

replace;

run;

proc sort data=cohlen;

by patient;

run;

The following **PROC NLMIXED** invocation fits the model involving period effects and treat period as a factor. The proportion of couples conceived under natural cycles is $\frac{13}{155} \approx 0.084$ and the proportion of couples conceived under stimulated cycles is $\frac{21}{153} \approx 0.137$. A reasonable starting value for the intercept will be:

$$\log(\tfrac{0.084}{1-0.084}) \approx -2$$

While a reasonable starting value for beta1 will be:

$$\log(\tfrac{0.137}{1-0.137}) - \log(\tfrac{0.084}{1-0.084}) \approx 1$$

For other parameters, we will use the default value which is 1.

proc nlmixed data=cohlen;

parms intercept=-2 beta1=0.3 beta2=1 beta3=1 beta4=1 beta5=1

beta6=1 sigma=1;

sigma2=sigma**2;

preg=intercept + beta1*treatment + beta2*period2 + beta3*period3

+ beta4*period4 + beta5*period5 + beta6*period6 + u;

ppreg=exp(preg)/(1+exp(preg));

model response~binomial(1,ppreg);

random u~normal(0,sigma2) subject=patient;

run;

The output for the parameter estimates and their 95% confidence interval are as shown in Table 4.16.

Fit Statistics

-2 Log Likelihood	207.8
AIC (smaller is better)	223.8
AICC (smaller is better)	224.3
BIC (smaller is better)	242.2

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
intercept	-3.0963	0.7865	73	-3.94	0.0002	0.05	-4.6637	-1.5289	0.000306
beta1	0.6410	0.4164	73	1.54	0.1280	0.05	-0.1889	1.4708	0.000067
beta2	0.2015	0.6260	73	0.32	0.7485	0.05	-1.0462	1.4491	-0.00003
beta3	1.1611	0.7188	73	1.62	0.1106	0.05	-0.2716	2.5937	0.000166
beta4	-0.4040	0.9893	73	-0.41	0.6842	0.05	-2.3756	1.5676	0.000073
beta5	0.5650	0.9300	73	0.61	0.5454	0.05	-1.2885	2.4185	0.000066
beta6	0.5622	1.0234	73	0.55	0.5844	0.05	-1.4773	2.6018	-0.00002
sigma	1.1625	0.6897	73	1.69	0.0962	0.05	-0.2121	2.5371	0.000017

Table 4.16. Fit statistics and parameter estimates for the model with treatment and period (treating period as a factor).

Next we fit the model with period effects only (period as a factor). The following **PROC NLMIXED** invocation fits the model involving period effect only.

proc nlmixed data=cohlen;

parms intercept=-2 beta2=1 beta3=1 beta4=1 beta5=1 beta6=1 sigma=1;

sigma2=sigma**2;

preg=intercept + beta2*period2 + beta3*period3 + beta4*period4 + beta5*period5 + beta6*period6 + u;

ppreg=exp(preg)/(1+exp(preg));

model response~binomial(1,ppreg);

random u~normal(0,sigma2) subject=patient;

run;

The output in Table 4.17 below displays the fit statistics, the parameter estimates and their 95% confidence interval.

```
                        Fit Statistics

        -2 Log Likelihood              210.3
        AIC (smaller is better)        224.3
        AICC (smaller is better)       224.7
        BIC (smaller is better)        240.5
```

```
                    Parameter Estimates
```

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
intercept	-2.6559	0.7003	73	-3.79	0.0003	0.05	-4.0516	-1.2602	0.000044
beta2	0.2321	0.6184	73	0.38	0.7085	0.05	-1.0003	1.4646	-0.00002
beta3	1.0816	0.6978	73	1.55	0.1255	0.05	-0.3091	2.4722	0.000011
beta4	-0.4697	0.9897	73	-0.47	0.6365	0.05	-2.4422	1.5028	0.000017
beta5	0.4984	0.9157	73	0.54	0.5879	0.05	-1.3266	2.3233	5.59E-6
beta6	0.4360	1.0207	73	0.43	0.6705	0.05	-1.5983	2.4704	-5.62E-7
sigma	1.0064	0.7236	73	1.39	0.1685	0.05	-0.4357	2.4485	0.000014

Table 4.17. Fit statistics and parameter estimates for the model with period effects only

If we take the difference in $-2 \log L$ for the saturated model (model with treatment and period effects, period treated as a factor) and the reduced model (model involving period effects only), $210.3 - 207.8 = 2.5$, we get the log likelihood ratio test for the treatment effects. Since the difference is 2.5 with 1 degrees of freedom, the p-value is 0.11385. This test shows that at 5% level of significance, the treatment effect is not statistically significant. Our conclusion is not different from (24); we found no statistically significant difference between the use of IUI in a natural cycle and the use of IUI in stimulated cycle.

We again consider a scenario where the investigator believes that there is period effect, but here unlike in the previous analysis, we regard period as having linear effect. The following **PROC NLMIXED** invocation fits the model involving treatment and period effects (linear effect).

proc nlmixed data=cohlen;

parms intercept=-2 beta1=0.3 beta2=1 sigma=1;

sigma2=sigma**2;

preg=intercept + beta1*treatment + beta2*period2 + u;

ppreg=exp(preg)/(1+exp(preg));

model response~binomial(1,ppreg);

random u~normal(0,sigma2) subject=patient;

run;

The output in Table 4.18 displays the parameter estimates and their 95% confidence interval for above **PROC NLMIXED** procedure.

Fit Statistics

-2 Log Likelihood	213.8
AIC (smaller is better)	221.8
AICC (smaller is better)	221.9
BIC (smaller is better)	231.0

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
intercept	-2.8033	0.7559	73	-3.71	0.0004	0.05	-4.3098	-1.2969	1.748E-6
beta1	0.5994	0.3923	73	1.53	0.1308	0.05	-0.1824	1.3813	-1.66E-7
beta2	0.06189	0.1704	73	0.36	0.7175	0.05	-0.2778	0.4016	0.000015
sigma	1.0047	0.6442	73	1.56	0.1232	0.05	-0.2791	2.2886	-1.25E-6

Table 4.18. Fit statistics and parameter estimates for the model with treatment and period effect (period having a linear effect).

Next we fit the model with period effects only (period having a linear effect). The following **PROC NLMIXED** invocation fits the model involving period effect only (period having a linear effect).

proc nlmixed data=cohlen;

parms intercept=-2 beta2=1 sigma=1;

sigma2=sigma**2;

preg=intercept + beta2*period + u;

pmreg=exp(preg)/(1+exp(preg));

pmreg=exp(preg)/(1+exp(preg));

model response~binomial(1,ppreg);

random u~normal(0,sigma2) subject=patient;

run;

The output in Table 4.19 below displays the fit statistics, the parameter estimates and their 95% confidence interval.

Fit Statistics

-2 Log Likelihood	216.2
AIC (smaller is better)	222.2
AICC (smaller is better)	222.3
BIC (smaller is better)	229.2

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
intercept	-2.4201	0.6955	73	-3.48	0.0009	0.05	-3.8062	-1.0340	0.00008
beta2	0.05080	0.1706	73	0.30	0.7668	0.05	-0.2893	0.3909	0.000357
sigma	0.9259	0.6703	73	1.38	0.1714	0.05	-0.4099	2.2617	-7.81E-7

Table 4.19. Fit statistics and parameter estimates for the model with period effects only

If we take the difference in $-2 \log L$ for the saturated model (model with treatment and period effects, period treated having linear effect) and the reduced model (model involving period effects only), $216.2 - 213.8 = 2.4$, we get the log likelihood ratio test for the treatment effects. Since the difference is 2.4 with 1 degrees of freedom, the p-value is 0.12134. This test shows that at 5% level of significance, the treatment effect is not statistically significant. Our conclusion is not different from (24); we found no statistically significant difference between the use of IUI in a natural cycle and the use of IUI in stimulated cycle.

Now we assume that there is no period effect. The following **PROC NLMIXED**

invocation fits the model involving treatment only.

    proc nlmixed data=cohlen;

parms intercept=-2 beta1=0.3 sigma=1;

sigma2=sigma**2;

preg=intercept + beta1*treatment + u;

ppreg=exp(preg)/(1+exp(preg));

model outcome ∼binomial(1,ppreg);

random u ∼normal(0,sigma2) subject=patient;

run;

The output in Table 4.20 below displays the fit statistics, the parameter estimates and their 95% confidence interval.

Fit Statistics

-2 Log Likelihood	213.9
AIC (smaller is better)	219.9
AICC (smaller is better)	220.0
BIC (smaller is better)	226.8

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
intercept	-2.5641	0.3580	73	-7.16	<.0001	0.05	-3.2775	-1.8506	3.097E-8
beta1	0.5862	0.3860	73	1.52	0.1332	0.05	-0.1831	1.3556	-7.34E-7
sigma	0.8254	0.5281	73	1.56	0.1224	0.05	-0.2271	1.8779	6.785E-9

Table 4.20. Fit statistics and parameter estimates for the reduced model

The null model that excludes both the treatment effect and the period effects is fit next. The following **PROC NLMIXED** invocation fits the reduced model.

    proc nlmixed data=cohlendata;

parms beta0=-1 sigma=1;

sigma2=sigma**2;

preg=beta0 + u;

ppreg=exp(preg)/(1+exp(preg));

model outcome ~binomial(1,ppreg);

random u~normal(0,sigma2) subject=patient;

run;

The output in Table 4.21 displays the model fit statistics for the null model. If we take the difference in $-2 \log L$ for the null and reduced models, $216.3 - 213.9 = 2.4$, we get the log likelihood ratio test for the treatment effects.

```
                        Fit Statistics

        -2 Log Likelihood               216.3
        AIC (smaller is better)         220.3
        AICC (smaller is better)        220.4
        BIC (smaller is better)         224.9
                      Parameter Estimates

                    Standard
 Parameter  Estimate   Error   DF  t Value  Pr > |t|  Alpha   Lower    Upper   Gradient

 intercept  -2.2277   0.2548   73   -8.74    <.0001    0.05  -2.7357  -1.7198  0.000112
 sigma       0.7663   0.5540   73    1.38    0.1708    0.05  -0.3378   1.8703  0.000014
```

Table 4.21. Maximum likelihood estimates for the null model

The p-value is 0.12134. At 5% significance level we fail to reject the hypothesis that the null model is as good as the model involving treatment. Therefore, at the 5% level of significance, the null model provides an adequate fit to the data. The treatment effect is not influential. Our conclusion is not different from (24); we found no statistically significant difference between the use of IUI in a natural cycle and the use of IUI in stimulated cycle.

4.2.2 Analysis using $R^{©}$

Similarly, in $R^{©}$ software, the following statement will import the data set once it is saved to, say, the $F : drive$ as it is in our case.

> **cohlen <- read.table("F:/cohlendata.txt", header=T)**

Once the data is read into the $R^{©}$ data set **cohlen**, we can use it to perform any analysis in $R^{©}$. The following code will fit in $R^{©}$, the model involving treatment and period effects (period treated as factors).

> **fit <- lmer(response~treatment+period2+period3+period4+period5 +period6+(1|patient),family=binomial,data=cohlen)**
> **fit**

The output for the parameter estimates and their p-values are as shown in Table 4.22.

```
Generalized linear mixed model fit by the Laplace approximation
Formula: response ~ beta1 + beta2 + beta3 + beta4 + beta5 + beta6 + (1 |      pat
   Data: cohlen
   AIC   BIC logLik deviance
 223.8 253.9 -103.9    207.8
Random effects:
 Groups  Name          Variance Std.Dev.
 patient (Intercept) 1.3515    1.1625
Number of obs: 320, groups: patient, 74

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.0963     0.5264  -5.882 4.04e-09 ***
beta1         0.6410     0.4264   1.503   0.1327
beta2         0.2015     0.6423   0.314   0.7538
beta3         1.1611     0.5841   1.988   0.0468 *
beta4        -0.4040     0.9469  -0.427   0.6696
beta5         0.5650     0.7516   0.752   0.4522
beta6         0.5621     0.8294   0.678   0.4979
```

Table 4.22. lmer analysis for the model with treatment and period effects, where period is regarded as a factor

Next we fit the reduced model with only period effects. The following code will fit in $R^{©}$, the model involving period effects only (period regarded as a factor).

> fit <- lmer(response~period2+period3+period4+period5+period6
+(1|patient),family=binomial,data=cohlen)
> fit

The output for the parameter estimates and their p-values are as shown in Table 4.23.

```
Generalized linear mixed model fit by the Laplace approximation
Formula: response ~ beta2 + beta3 + beta4 + beta5 + beta6 + (1 | patient)
   Data: cohlen
   AIC   BIC logLik deviance
 224.3 250.7 -105.2    210.3
Random effects:
 Groups  Name         Variance Std.Dev.
 patient (Intercept) 1.0129   1.0064
Number of obs: 320, groups: patient, 74

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.6559     0.4566  -5.816 6.01e-09 ***
beta2         0.2321     0.6228   0.373   0.7093
beta3         1.0816     0.5714   1.893   0.0584 .
beta4        -0.4697     0.9238  -0.508   0.6111
beta5         0.4984     0.7338   0.679   0.4970
beta6         0.4359     0.8066   0.540   0.5889
```

Table 4.23. lmer analysis for the model with period effects only, where period is regarded as a factor

The difference in $-2 \log L$ for the model with treatment and period effects and the reduced model involving period effects only is $210.3 - 207.8 = 2.5$. The p-value is 0.11385. This test shows that the model with period effects only provides a better fit than the model with treatment and period effects, when period is included as a factor. Therefore, at the 5% level of significance, treatment is not statistically significant. Our conclusion is not different from (24); we found no statistically significant difference between the use of IUI in a natural cycle and the use of IUI in stimulated cycle.

We again fit the model with treatment and period effects, but here, period is regarded as having a linear effect. The following code will fit in $R^{©}$, the model involving treatment and period effects (period having a linear effect).

> **fit <- lmer(response∼treatment+period+(1|patient),family=binomial, data=cohlen)**

> **fit**

The output for the parameter estimates and their p-values are as shown below in Table 4.24.

```
Generalized linear mixed model fit by the Laplace approximation
Formula: response ~ beta1 + beta2 + (1 | patient)
   Data: cohlen
   AIC   BIC logLik deviance
 221.8 236.9 -106.9    213.8
Random effects:
 Groups  Name          Variance Std.Dev.
 patient (Intercept) 1.0095    1.0047
Number of obs: 320, groups: patient, 74

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.80334    0.49501  -5.663 1.49e-08 ***
beta1        0.59945    0.40930   1.465    0.143
beta2        0.06188    0.12500   0.495    0.621
```

Table 4.24. lmer analysis for the saturated model, period having linear effect

Next we fit the reduced model with only period effects. The following code will fit in $R^{©}$, the model involving period effects only (period regarded as having a linear effect).

> **fit <- lmer(response∼period+(1|patient),family=binomial,data=cohlen)**

> **fit**

The output for the parameter estimates and their p-values are as shown in Table 4.25.

```
Generalized linear mixed model fit by the Laplace approximation
Formula: response ~ beta2 + (1 | patient)
   Data: cohlen
   AIC   BIC logLik deviance
 222.2 233.5 -108.1    216.2
Random effects:
 Groups  Name         Variance Std.Dev.
 patient (Intercept) 0.85729  0.9259
Number of obs: 320, groups: patient, 74

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.42013    0.42430  -5.704 1.17e-08 ***
beta2        0.05078    0.12281   0.413    0.679
```

Table 4.25. lmer analysis for the model with period effects only, where period is regarded as having linear effecct

The difference in $-2 \log L$ for the model with treatment and period effects and the reduced model involving period effects only is $216.2 - 213.8 = 2.4$. The p-value is 0.12134. This test shows that the model with period effects only provides a better fit than the model with treatment and period effects, when period is included as having a linear effect. Therefore, at the 5% level of significance, treatment is not statistically significant. Our conclusion is not different from (24); we found no statistically significant difference between the use of IUI in a natural cycle and the use of IUI in stimulated cycle.

We next assume that there is no period effects, and begin by fitting a model with treatment only. The following code will fit in $R^{©}$, the model involving treatment only (excludes period effects).

> **fit <- lmer(response~treatment+(1|patient),family=binomial,data=cohlen)**

> **fit**

The output for the parameter estimates and their p-values are as shown in Table 4.26 below.

```
Generalized linear mixed model fit by the Laplace approximation
Formula: response ~ beta1 + (1 | patient)
   Data: cohlen
   AIC   BIC logLik deviance
 219.9 231.2 -107.0    213.9
Random effects:
 Groups  Name          Variance Std.Dev.
 patient (Intercept) 0.68113  0.8253
Number of obs: 320, groups: patient, 74

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.5640     0.3281  -7.815 5.51e-15 ***
beta1         0.5862     0.4009   1.462    0.144
```

Table 4.26. lmer analysis for the reduced model

We now fit the null model, involving no treatment. The following $R^{©}$ will fit the null model.

> **fit <- lmer(response~ (1|patient),family=binomial,data=data.1)**

> **fit**

The output for the parameter estimates and their p-values for the null model are as shown in Table 4.27.

If we take the difference in $-2 \log L$ for the model involving treatment and the null model, $216.3 - 213.9$, we get the log likelihood ratio test for the treatment effects. Since the difference is 2.4 with 1 degrees of freedom, the p-value is 0.1213. Therefore, at the 5% level of significance, the null model provides an adequate fit to the data. The treatment effect is not influential. Our conclusion is not

```
Generalized linear mixed model fit by the Laplace approximation
Formula: response ~ (1 | patient)
   Data: cohlen
   AIC   BIC logLik deviance
 220.3 227.9 -108.2    216.3
Random effects:
 Groups  Name         Variance Std.Dev.
 patient (Intercept) 0.58709  0.76621
Number of obs: 320, groups: patient, 74

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.2278     0.2142  -10.40   <2e-16 ***
```

Table 4.27. lmer analysis for the null model

different from (24); we found no statistically significant difference between the use of IUI in a natural cycle and the use of IUI in stimulated cycle.

4.2.3 Analysis using $WinBugs^{\copyright}$

The following $WinBugs^{\copyright}$ code will specify the saturated model (model with treatment and period effects, period as a factor), the data to be used and the initial values for the $MCMC$ sampler.

model{

for(i in 1 : N1) {

$outcome1[i] \sim dbern(p1[i])$

$logit(p1[i]) < -intercept + beta1 * treatment1[i] + beta2 * period2[i] + beta3 * period3[i] + beta4 * period4[i] + beta5 * period5[i] + beta6 * period6[i] + v[i]$

b[i] ∼ dnorm(0, tau)

}

for(i in 1 : N2){

$outcome2[i] \sim dbern(p2[i])$

$logit(p2[i]) < -intercept + beta1 * treatment2[i] + beta2 * period22[i] + beta3 *$

$period3[i] + beta4 * period4[i] + beta5 * period5[i] + beta6 * period6[i] + v[i]$

}

for(i in 1 : N3){

$outcome3[i] \sim dbern(p3[i])$

$logit(p3[i]) < -intercept + beta1 * treatment3[i] + beta2 * period2[i] + beta3 *$

$period33[i] + beta4 * period4[i] + beta5 * period5[i] + beta6 * period6[i] + v[i]$

}

for(i in 1 : N4){

$outcome4[i] \sim dbern(p4[i])$

$logit(p4[i]) < -intercept + beta1 * treatment4[i] + beta2 * period2[i] + beta3 *$

$period3[i] + beta4 * period44[i] + beta5 * period5[i] + beta6 * period6[i] + v[i]$

}

or(i in 1 : N5){

$outcome5[i] \sim dbern(p5[i])$

$logit(p5[i]) < -intercept + beta1 * treatment5[i] + beta2 * period2[i] + beta3 *$

$period3[i] + beta4 * period4[i] + beta5 * period55[i] + beta6 * period6[i] + v[i]$

}

for(i in 1 : N6){

$outcome6[i] \sim dbern(p6[i])$

$logit(p6[i]) < -intercept + beta1 * treatment6[i] + beta2 * period2[i] + beta3 *$

$period3[i] + beta4 * period4[i] + beta5 * period5[i] + beta6 * period66[i] + v[i]$

}

$sigma < -1/sqrt(tau)$

$intercept \sim dnorm(0, 1.0E - 3)$

$beta1 \sim dnorm(0, 1.0E - 3)$

$beta2 \sim dnorm(0, 1.0E - 3)$

$beta3 \sim dnorm(0, 1.0E - 3)$

$beta4 \sim dnorm(0, 1.0E - 3)$

$beta5 \sim dnorm(0, 1.0E - 3)$

$beta6 \sim dnorm(0, 1.0E - 3)$

$tau \sim dgamma(1.0E - 4, 1.0E - 4)$

$log.sigma < -log(sigma)$ }

list(outcome1=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1) ,

treatment1=c(0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,1,1,1,0,1,0,0,0,0,1,1,0,0,0,0,0,0,0,1,1,1,1,1,0,0,0,0,0,0,0,0,1,1,1,0,0,0,0,0,1,1,1),

period2=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),

period3=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),

period4=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0,

0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),

period5=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),

period6=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),

outcome2 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,0) ,

treatment2=c(1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,0, 0,0,1,0,1,1,1,1,0,0,1,1,1,1,1,1,0,0,0,0,0,1,1,1,1,1,1,1,0,0,0),

period22=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1),

outcome3 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 1, 1,1,1,1,1,1,1,1,1,1,0,0) ,

treatment3=c(0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,1,1,1,0,1,

0,0,0,0,1,1,0,0,0,0,0,0,1,1,1,1,1,0,0),

period33=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1),

outcome4 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0
,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 1, 1,0,0,0,0) ,

treatment4=c(1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,0,
0,0,1,0,1,1,1,1,0,0),

period44=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1),

outcome5 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0,
0 ,0 ,0, 0, 0 ,0,0,0,0,1, 1, 1, 1,0,0) ,

treatment5=c(0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,1,
1,1,0,1),

period55=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1),

outcome6 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0,

0, 0, 0 ,0 ,0, 0, 1, 1, 1) ,

treatment6=c(1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1),

period66=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1),

N1 = 74, N2=67, N3=59,N4=46, N5=40, N6=34),

list(intercept=20,beta1=20,beta2=20,beta3=20,beta4=40,beta5=40,beta6=40, tau=3,b=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0,0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0))

list(intercept=-1,beta1=0.3,beta2=0.3,beta3=0.3,beta4=0.3,beta5=0.3,beta6=0.3, tau=1, b=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1))

As before, the data have to be sorted in such a way that for any couple i that leaves the study before couple j, then couple i is listed after we have listed couple j. For example, we have 31 couples that did not conceive but completed the study. So these couples have to be listed first, followed by those who leaves the study on the second period, followed by those who leaves the study on the third period, followed by those who leave the study on the fourth period, followed by

those who leave the study on the fifth period and followed by those who leave the study on the sixth period.

We checked convergence for the intercept, $beta1$, $beta2$, $beta3$, $beta4$, $beta5$, $beta6$ and sigma. We look for evidence of when the simulation appears to have stabilized by examining the plots of the sample values versus iterations. We ran 20000 simulations. Fig 4.2, 4.3 and 4.4 show the plots of the intercept, $beta1$, $beta2$, $beta3$, $beta4$, $beta5$, $beta6$ and sigma.
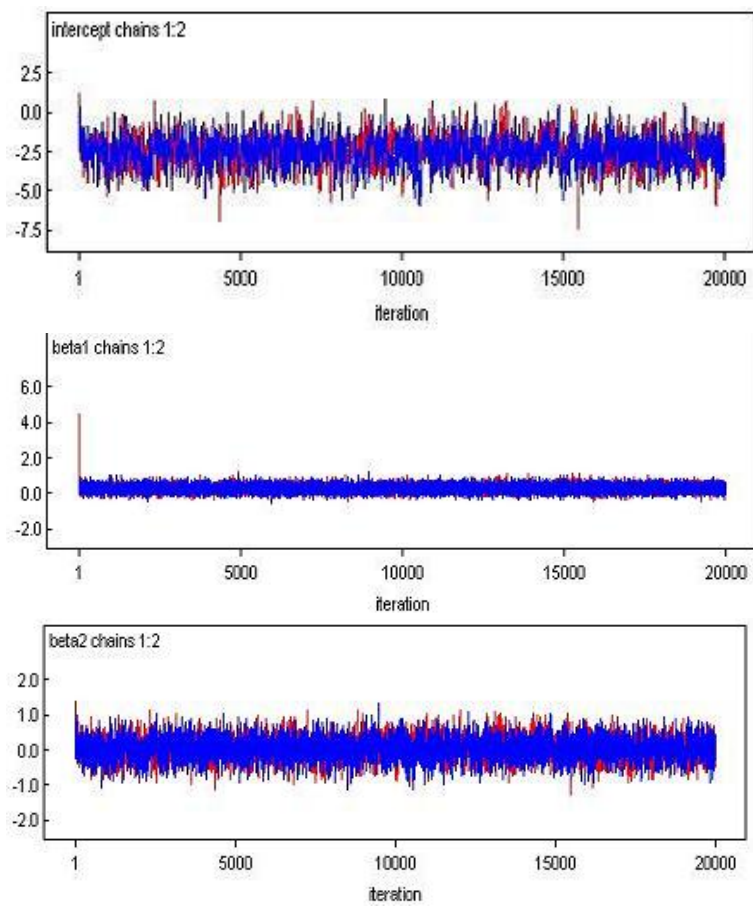
Figure 4.2. Plots of the intercept, beta1 and beta2
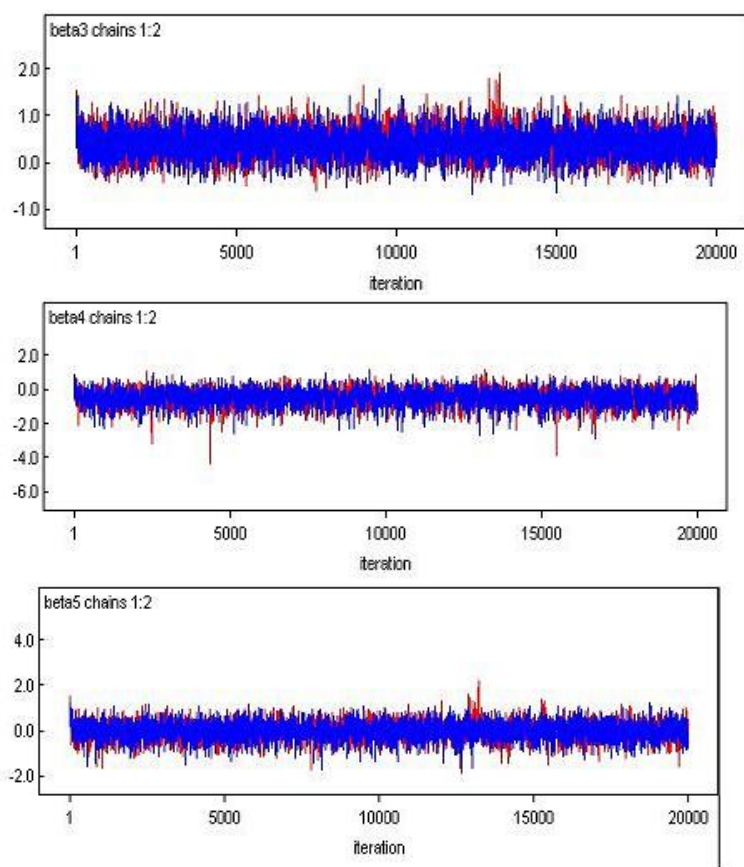
Figure 4.3. Plots of beta3, beta4 and beta5
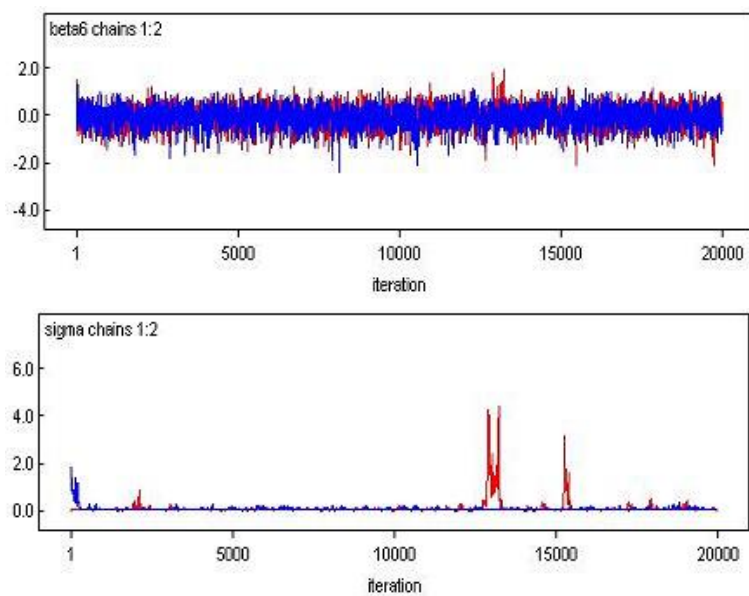
Figure 4.4. Plots of beta6 and sigma

Convergence seems reasonable in all plots except the plots of sigma. We further ran 30000 simulations and obtained our summary statistics based on those 30000 simulations. Below in table 4.28 are summary statistics for the monitored parameters based on 30000 samples.

```
Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic
nodes
              Dbar        Dhat        pD          DIC
outcome1      48.493      47.290      1.202       49.695
outcome2      44.971      43.769      1.202       46.173
outcome3      59.408      58.101      1.307       60.715
outcome4      16.536      15.527      1.009       17.545
outcome5      26.644      25.548      1.096       27.739
outcome6      19.791      18.719      1.073       20.864
total         215.842     208.954     6.888       222.730
```

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
intercept	-2.689	1.134	0.04205	-3.681	-2.624	-1.777	20001	60000
beta1	0.5623	0.4602	0.003158	-0.1841	0.5595	1.333	20001	60000
beta2	0.1437	1.165	0.04076	-1.053	0.1063	1.285	20001	60000
beta3	0.8559	1.144	0.04085	-0.2198	0.8078	1.924	20001	60000
beta4	-0.9695	1.363	0.04105	-3.013	-0.944	0.6459	20001	60000
beta5	0.01375	1.23	0.04089	-1.46	-0.007876	1.332	20001	60000
beta6	-0.1031	1.274	0.04084	-1.757	-0.1111	1.334	20001	60000
sigma	0.02282	0.1039	0.005071	0.005215	0.01216	0.05638	20001	60000

Table 4.28. Parameter estimates and their standard errors

Next we fit the reduced model with period effects only. The $WinBugs^{©}$ code that specify the model with period effects only (period as a factor), the data to be used and the initial values for the $MCMC$ sampler can be seen in appendix B.1. Convergence was reached after 20000 simulations for all the parameters except sigma. Fig B.1 shows the plots of the intercept, beta2, beta3 and beta4 and fig B.2 shows the plot of beta5, beta6 and sigma. Convergence is satisfied in all except sigma.

We simulated extra 30000 samples. Below in table4.29 are summary statistics for the monitored parameters based on 30000 samples.

```
Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes
             Dbar        Dhat        pD          DIC
 outcome1    47.032      45.174      1.858       48.891
 outcome2    45.614      44.000      1.615       47.229
 outcome3    57.441      55.663      1.779       59.220
 outcome4    17.471      16.387      1.084       18.555
 outcome5    26.913      25.698      1.216       28.129
 outcome6    21.252      20.110      1.142       22.394
 total       215.724     207.031     8.693       224.417
```

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
intercept	-2.373	0.6922	0.02372	-3.265	-2.31	-1.582	20001	60000
beta2	0.1311	0.6217	0.009971	-1.022	0.1182	1.321	20001	60000
beta3	0.853	0.6943	0.01909	-0.2049	0.8155	1.968	20001	60000
beta4	-0.9836	1.051	0.02183	-3.03	-0.9497	0.6885	20001	60000
beta5	0.06441	0.9146	0.02554	-1.395	0.04094	1.42	20001	60000
beta6	-0.1095	1.022	0.02858	-1.787	-0.1226	1.362	20001	60000
sigma	0.1067	0.5695	0.02415	0.01666	0.03892	0.2947	20001	60000

Table 4.29. Parameter estimates and their standard errors

The DIC for the saturated model with treatment and period effects is 222.730, while the DIC for the model with period effects only is 224.417. Using the DIC, the preferred model is the model with treatment and period effect. The treatment effect estimate is 0.5623 with 95% confidence interval of $0.5623 \pm 1.96 * 0.04602 = (-0.339692, 1.464292)$. The confidence interval clearly show that at 5% level of significance the treatment effect is not statistically significant.

Next we consider a scenario where period has a linear effect. The $WinBugs^{©}$ code that specify the model with treatment and period effects, period regarded as having linear effect), the data to be used and the initial values for the $MCMC$ sampler can be seen in appendix $B.2$. We checked convergence for the intercept, $beta1$, $beta2$ and sigma. We look for evidence of when the simulation appears to have stabilized by examining the plots of the sample values versus iterations. We ran 20000 simulations. Fig B.3 shows the plots for the intercept, $beta1$, $beta2$ and sigma.

Convergence seems reasonable in all plots except the plots of sigma. We

further ran 30000 simulations and obtained our summary statistics based on those 30000 simulations. Table 4.30 shows the summary statistics for the monitored parameters based on 30000 samples.

Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes

	Dbar	Dhat	pD	DIC
outcome1	48.317	47.443	0.874	49.191
outcome2	44.220	43.690	0.529	44.749
outcome3	62.171	61.784	0.387	62.558
outcome4	17.546	17.202	0.345	17.891
outcome5	25.899	25.468	0.431	26.331
outcome6	19.124	18.603	0.521	19.645
total	217.278	214.191	3.088	220.366

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
intercept	-2.346	0.4473	0.004265	-3.252	-2.334	-1.506	20001	60000
beta1	0.5648	0.379	0.002269	-0.1673	0.5607	1.325	20001	60000
beta2	-0.04505	0.1132	9.798E-4	-0.2714	-0.0438	0.1738	20001	60000
sigma	0.02064	0.06817	0.002607	0.005295	0.01268	0.06604	20001	60000

Table 4.30. Parameter estimates and their standard errors

Next we fit the reduced model with period effects only. The $WinBugs^{©}$ code that specify the random effects model with the main effect being period effects only (period having linear effect) is listed in appendix $B.3$. We checked convergence for the intercept, $beta2$ and sigma. We look for evidence of when the simulation appears to have stabilized by examining the plots of the sample values versus iterations. We ran 20000 simulations. Fig B.4 shows the plots for the intercept, $beta2$ and sigma. Convergence seems reasonable in all plots except the plots of sigma.

We further ran 30000 simulations and obtained our summary statistics based on those 30000 simulations. Table 4.31 shows the summary statistics for the monitored parameters based on 30000 samples.

The DIC for the model with period effects only is 220.797 while the DIC for

```
Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes
            Dbar        Dhat        pD          DIC
outcome1    47.098      46.325      0.773       47.871
outcome2    45.094      44.678      0.417       45.511
outcome3    60.432      60.146      0.286       60.718
outcome4    18.631      18.368      0.263       18.895
outcome5    26.273      25.919      0.354       26.628
outcome6    20.702      20.229      0.473       21.175
total       218.231     215.664     2.567       220.797
```

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
intercept	-2.032	0.4117	0.006148	-2.813	-2.018	-1.3	20001	60000
beta2	-0.0427	0.1175	0.001435	-0.2699	-0.04221	0.1786	20001	60000
sigma	0.05966	0.1595	0.005895	0.01634	0.03771	0.1924	20001	60000

Table 4.31. Parameter estimates and their standard errors

the model with treatment and period effects is 220.366 respectively. The preferred model is the model with treatment and period effects. The treatment effect estimate is 0.5648 with 95% confidence interval of $0.5648 \pm 1.96 * 0.379 = (-0.17804, 1.30764)$. The confidence interval show that, at the 5% level of significance, the treatment effect is not influential.

Next we assume that there is no period effect and fit the reduced model with treatment only. The $WinBugs^{©}$ code that specify the random effects model with the main effect being treatment only, the data to be used and the initial values for the $MCMC$ sampler is in appendix $B.4$. We first ran 20000 simulation and we checked convergence for the intercept, $beta1$ and sigma. We look for evidence of when the simulation appears to have stabilized by examining the plots of the sample values versus iterations. Fig B.5 shows the plots of the intercept, $beta1$ and sigma. The plots of the intercept and $beta1$ show chains for which convergence look reasonable. The plot of sigma shows chains which have clearly not reached convergence.

We simulated extra 30000 samples, and table 4.32 shows the summary statistics for the monitored parameters based on 30000 samples.

```
Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes
              Dbar        Dhat        pD        DIC
  outcome1    46.303      43.787      2.516     48.818
  outcome2    42.937      40.846      2.091     45.028
  outcome3    59.932      58.242      1.690     61.622
  outcome4    17.197      15.984      1.212     18.409
  outcome5    25.006      23.923      1.083     26.089
  outcome6    18.305      17.430      0.876     19.181
  total       209.679     200.212     9.468     219.147
```

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
intercept	-2.526	0.3312	0.004076	-3.242	-2.503	-1.938	20001	60000
beta1	0.5855	0.3886	0.002685	-0.1647	0.5803	1.363	20001	60000
sigma	0.3802	0.4606	0.01663	0.00994	0.172	1.609	20001	60000

Table 4.32. Parameter estimates and their standard errors

Final we fit the null model excluding both treatment and period effects. The $WinBugs^{©}$ code that specify the null model, the data to be used and the initial values for the $MCMC$ sampler is in appendix B.5. We checked convergence for the intercept and sigma. We look for evidence of when the simulation appears to have stabilized by examining the plots of the sample values versus iterations. We ran 20000 simulations. Fig B.6 shows the plots for the intercept and sigma. The plots of the intercept show chains for which convergence look reasonable. The plot of sigma shows chains which have clearly not reached convergence.

We simulated extra 30000 samples, and table 4.33 shows the summary statistics for the monitored parameters based on 30000 samples.

```
Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes
              Dbar        Dhat        pD        DIC
outcome1      45.404      43.272      2.132     47.537
outcome2      43.864      42.142      1.722     45.585
outcome3      58.478      57.062      1.416     59.894
outcome4      18.340      17.314      1.027     19.367
outcome5      25.465      24.582      0.883     26.348
outcome6      19.971      19.223      0.748     20.718
total         211.522     203.594     7.928     219.450
```

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
intercept	-2.187	0.217	0.002857	-2.658	-2.171	-1.807	20001	60000
sigma	0.3605	0.434	0.01587	0.009994	0.1653	1.513	20001	60000

Table 4.33. Parameter estimates and their standard errors

The DIC for the model with treatment only is 219.147 and the DIC for the null model is 219.450 respectively. The preferred model is the model with treatment only. The treatment effect estimate is 0.5855 with 95% confidence interval of $0.5855 \pm 1.96 * 0.3886 = (-0.176156, 1.347156)$. The confidence interval show that, at the 5% level of significance, the treatment effect is not influential.

4.2.4 Analysis using $GenStat^©$

The following $GenStat^©$ will read data in five columns and fit the hierarchical model involving treatment and period effects (period regarded as a factor);

FACTOR Group, Patient, Period, Treatment

READ [SETNVALUES=yes; SETLEVELS=yes] Group,Patient,Period,

Treatment,Response;

FREP=2(levels),labels,*

1 1 1 0 1

1 2 1 0 1

1 3 1 0 1

. . . . .

. . . . .

. . . . .

2 72 6 0 0

2 73 6 0 0

2 74 6 0 0:

HGRANDOMMODEL [DISTRIBUTION=normal; LINK=identity] Patient

HGFIXEDMODEL [DISTRIBUTION=binomial; LINK=logit] Treatment + Period

HGANALYSE [MLAPLACE=1; DLAPLACE=1;Maxcycle=999,30] Response; NBINOMIAL=1

The parameter estimates and deviances for the model involving treatment and period effects described above are as follows:

Estimates from the mean model

	estimate	s.e.	t(313)
constant	-3.397	0.5551	-6.119
Treatment 1	0.689	0.4435	1.554
Period2 1	0.287	0.6690	0.429
Period3 1	1.384	0.6101	2.268
Period4 1	-0.168	0.9837	-0.171
Period5 1	0.859	0.7852	1.093
Period6 1	0.884	0.8692	1.017

Estimates from the dispersion model

Estimates of parameters

Parameter lambda Patient	estimate	s.e.	t(*)	antilog of estimate
	0.782	0.265	2.95	2.185

Message: s.e.s are based on dispersion parameter with value 1.

Likelihood statistics

$-2 \times h(y\|v)$	141.580
$-2 \times h$	363.913
$-2 \times P_v(h)$	207.986
$-2 \times P_{\beta,v}(h)$	203.186
$-2 \times EQD(y\|v)$	130.507
$-2 \times EQD$	352.840
$-2 \times P_v(EQD)$	196.913
$-2 \times P_{\beta,v}(EQD)$	192.113

Scaled deviances

	deviance	df
Random term		
units	141.6	284.5
Patient	28.5	28.5
Total	170.1	313.0

Next we fit the hierarchical model involving period effects only (period as a factor). The following $GenStat^{©}$ will fit the model with period effects only.

HGRANDOMMODEL [DISTRIBUTION=normal; LINK=identity] Patient

HGFIXEDMODEL [DISTRIBUTION=binomial; LINK=logit] Period

HGANALYSE [MLAPLACE=1; DLAPLACE=1;Maxcycle=999,30] Response; NBINOMIAL=1

The parameter estimates and deviances for the model with period effects only are as follows:

Estimates from the mean model

	estimate	s.e.	t(314)
constant	-2.890	0.4832	-5.982
Period 2	0.318	0.6437	0.494
Period 3	1.267	0.5936	2.135
Period 4	-0.254	0.9562	-0.266
Period 5	0.750	0.7628	0.983
Period 6	0.721	0.8393	0.859

Estimates from the dispersion model

Estimates of parameters

Parameter	estimate	s.e.	t(*)	antilog of estimate
lambda Patient				
	0.505	0.285	1.77	1.658

Message: s.e.s are based on dispersion parameter with value 1.

Likelihood statistics

$-2 \times h(y\|v)$	154.508
$-2 \times h$	352.554
$-2 \times P_v(h)$	210.501
$-2 \times P_{\beta,v}(h)$	206.301
$-2 \times EQD(y\|v)$	145.119
$-2 \times EQD$	343.164
$-2 \times P_v(EQD)$	201.112
$-2 \times P_{\beta,v}(EQD)$	196.911

Scaled deviances

	deviance	df
Random term		
units	154.5	289.3
Patient	24.6	24.7
Total	179.1	314.0

The difference in $-2P_v(h)$ of the reduced and saturated models is $210.501 - 207.986 = 2.515$. The p-value is 0.11277. This test shows that the reduced model (period effects only) provides a better fit than the saturated model(model with

treatment and period effect) with period treated as a factor. At 5% level of significance, treatment is not statistically significant.

We again fit the hierarchical model involving treatment and period effect. Here we regard period as having a linear effect. The following *GenStat*© will fit the model with treatment and period effect (period having linear effect). The difference between this code and the code above(period as a factor) is that, here we do not specify the variate **PERIOD** to be a factor

FACTOR Group, Patient, Treatment

READ [SETNVALUES=yes; SETLEVELS=yes] Group,Patient,Period,

Treatment,Response;

FREP=2(levels),labels,*

1 1 1 0 1

1 2 1 0 1

1 3 1 0 1

. . . . .

. . . . .

. . . . .

2 72 6 0 0

2 73 6 0 0

2 74 6 0 0:

HGRANDOMMODEL [DISTRIBUTION=normal; LINK=identity] Patient

HGFIXEDMODEL [DISTRIBUTION=binomial; LINK=logit] Treatment + Period

HGANALYSE [MLAPLACE=1; DLAPLACE=1;Maxcycle=999,30] Response; NBI-NOMIAL=1

The parameter estimates and deviances for the model with treatment and period effects(period having a linear effect) are as follows:

Estimates from the mean model

	estimate	s.e.	t(317)
constant	-2.951	0.5087	-5.801
Treatment 1	0.615	0.4162	1.477
Period	0.092	0.1281	0.720

Estimates from the dispersion model

Estimates of parameters

Parameter lambda Patient	estimate	s.e.	t(*)	antilog of estimate
	0.296	0.301	0.98	1.344

Message: s.e.s are based on dispersion parameter with value 1.

Likelihood statistics

$-2 \times h(y|v)$	164.720
$-2 \times h$	344.659
$-2 \times P_v(h)$	213.870
$-2 \times P_{\beta,v}(h)$	217.034
$-2 \times EQD(y|v)$	156.643
$-2 \times EQD$	336.581
$-2 \times P_v(EQD)$	205.793
$-2 \times P_{\beta,v}(EQD)$	208.957

Scaled deviances

	deviance	df
Random term		
units	164.7	295.0
Patient	22.0	22.0
Total	186.8	317.0

Next we fit the hierarchical model involving period effect only (period having a linear effect)

HGRANDOMMODEL [DISTRIBUTION=normal; LINK=identity] Patient

HGFIXEDMODEL [DISTRIBUTION=binomial; LINK=logit] Period

HGANALYSE [MLAPLACE=1; DLAPLACE=1;Maxcycle=999,30] Response; NBINOMIAL=1

The parameter estimates and deviances for the reduced model are as follows:

Estimates from the mean model

	estimate	s.e.	t(318)
constant	-2.561	0.4388	-5.836
Period	0.082	0.1259	0.649

Estimates from the dispersion model

Estimates of parameters

Parameter lambda Patient	estimate	s.e.	t(*)	antilog of estimate
	0.160	0.314	0.51	1.174

Message: s.e.s are based on dispersion parameter with value 1.

Likelihood statistics

$-2 \times h(y\|v)$	171.381
$-2 \times h$	339.585
$-2 \times P_v(h)$	216.298
$-2 \times P_{\beta,v}(h)$	219.636
$-2 \times EQD(y\|v)$	163.917
$-2 \times EQD$	332.121
$-2 \times P_v(EQD)$	208.834
$-2 \times P_{\beta,v}(EQD)$	212.172

Scaled deviances

	deviance	df
Random term		
units	171.4	297.7
Patient	20.3	20.3
Total	191.7	318.0

The difference in $-2P_v(h)$ of the reduced and saturated models is $216.298 - 213.870 = 2.428$. The p-value is 0.11918. This test shows that the reduced model (period effects only) provides a better fit than the saturated model(model with treatment and period effect) with period having a linear effect. At 5% level of

significance, treatment is not statistically significant.

Now we assume that there is no period effect. We fit a model with treatment only. The following *GenStat*© will fit the hierarchical model involving treatment only (reduced model).

HGRANDOMMODEL [DISTRIBUTION=normal; LINK=identity] Patient

HGFIXEDMODEL [DISTRIBUTION=binomial; LINK=logit] Treatment

HGANALYSE [MLAPLACE=1; DLAPLACE=1;Maxcycle=999,30] Response; NBINOMIAL=1

The parameter estimates and deviances for the model with treatment only are as follows:

Estimates from the mean model

	estimate	s.e.	t(318)
constant	-2.590	0.3337	-7.762
Treatment 1	0.594	0.4046	1.467

Estimates from the dispersion model

Estimates of parameters

Parameter	estimate	s.e.	t(*)	antilog of estimate
lambda Patient				
	-0.210	0.357	-0.59	0.8109

Message: s.e.s are based on dispersion parameter with value 1.

Likelihood statistics

$-2 \times h(y\|v)$	179.944
$-2 \times h$	316.102
$-2 \times P_v(h)$	213.957
$-2 \times P_{\beta,v}(h)$	215.079
$-2 \times EQD(y\|v)$	175.958
$-2 \times EQD$	312.115
$-2 \times P_v(EQD)$	209.970
$-2 \times P_{\beta,v}(EQD)$	211.093

Scaled deviances

Random term	deviance	df
units	179.9	302.1
Patient	15.7	15.9
Total	195.6	318.0

The null model is fit next. The following *GenStat*© will fit the null model.

HGRANDOMMODEL [DISTRIBUTION=normal; LINK=identity] Patient

HGFIXEDMODEL [DISTRIBUTION=binomial; LINK=logit]

HGANALYSE [MLAPLACE=0; DLAPLACE=0;Maxcycle=999,30] Response; NBINOMIAL=1

The parameter estimates and deviances for the null model are as follows:

Estimates from the mean model

	estimate	s.e.	t(319)
constant	-2.248	0.2195	-10.24

Estimates from the dispersion model

Estimates of parameters

Parameter lambda Patient	estimate	s.e.	t(*)	antilog of estimate
	-0.348	0.375	-0.93	0.7058

Message: s.e.s are based on dispersion parameter with value 1.

Likelihood statistics

$-2 \times h(y|v)$	185.618
$-2 \times h$	310.080
$-2 \times P_v(h)$	216.344
$-2 \times P_{\beta,v}(h)$	217.539
$-2 \times EQD(y|v)$	182.449
$-2 \times EQD$	306.910
$-2 \times P_v(EQD)$	213.175
$-2 \times P_{\beta,v}(EQD)$	214.370

Scaled deviances

	deviance	df
Random term		
units	185.6	304.6
Patient	14.2	14.4
Total	199.9	319.0

The difference in $-2P_v(h)$ for the null and reduced model, $216.344 - 213.957 = 2.387$, with 1 degrees of freedom. The p-value is 0.12235. This test shows that the model with treatment only does not provide a better fit than the null model. The treatment effect is not influential. Our conclusion is not different from (24); we found no statistically significant difference between the use of IUI in a natural cycle and the use of IUI stimulated cycle.

4.3 Chapter summary

Table 4.34 records comparative analyses of the **cohlen** data.

Model	Method	intercept	treatment	period	Deviance
Null	$SAS^{©}$	-2.227			216.3
		(0.255)			
	$R^{©}$	-2.228			216.3
		(0.214)			
	$WinBugs^{©}$	-2.187			-
		(0.217)			
	$Genstat^{©}$	-2.248			216.344
		(0.2195)			
Treatment Only	$SAS^{©}$	-2.564	0.586		213.9
		(0.358)	(0.386)		
	$R^{©}$	-2.564	0.586		213.9
		(0.328)	(0.401)		
	$WinBugs^{©}$	-2.526	0.586		-
		(0.331)	(0.387)		
	$GenStat^{©}$	-2.590	0.594		213.957
		(0.3337)	(0.4046)		
Period only	$SAS^{©}$	-2.420		0.051	216.2
		(0.696)		(0.171)	
	$R^{©}$	-2.420		0.051	216.2
		(0.424)		(0.123)	
	$WinBugs^{©}$	-2.032		-0.0427	-
		(0.412)		(0.118)	
	$GenStat^{©}$	-2.561		0.082	216.298
		(0.5087)		(0.1281)	
Period and Treatment	$SAS^{©}$	-2.803	0.599	0.062	213.8
		(0.756)	(0.392)	(0.170)	
	$R^{©}$	-2.803	0.599	0.062	213.8
		(0.495)	(0.409)	(0.125)	
	$WinBugs^{©}$	-2.346	0.565	-0.0451	-
		(0.447)	(0.379)	(0.113)	
	$GenStat^{©}$	-2.951	0.615	0.092	213.870
		(0.5087)	(0.4162)	(0.1281)	

Table 4.34. Comparative analysis of the **Cohlen** data(regarding period as having linear effect)

Model	Method	intercept	treatment	period2	period3	period4	period5	period6	Deviance
Period only(factor)	$SAS^{©}$	-2.656		0.232	1.082	-0.470	0.498	0.436	210.3
		(0.700)		(0.618)	(0.698)	(0.990)	(0.916)	(1.02)	
	$R^{©}$	-2.656		0.232	1.082	-0.470	0.498	0.436	210.3
		(0.457)		(0.623)	(0.571)	(0.924)	(0.734)	(0.807)	
	$WinBugs^{©}$	-2.373		0.131	0.853	-0.984	0.064	-0.110	-
		(0.692)		(0.622)	(0.694)	(1.05)	(0.915)	(1.022)	-
	$GenStat^{©}$	-2.980		0.318	1.267	-0.254	0.750	0.721	210.501
		(0.4832)		(0.6437)	(0.5936)	(0.9562)	(0.7628)	(0.8393)	-
Period(factor)	$SAS^{©}$	-3.10	0.641	0.202	1.16	-0.404	0.565	0.562	207.8
and		(0.787)	(0.416)	(0.626)	(0.719)	(0.989)	(0.930)	(1.02)	
Treatment	$R^{©}$	-3.096	0.641	0.202	1.161	-0.404	0.565	0.562	207.8
		(0.526)	(0.426)	(0.642)	(0.584)	(0.947)	(0.752)	(0.829)	
	$WinBugs^{©}$	-2.689	0.562	0.144	0.856	-0.970	0.014	-0.103	-
		(1.134)	(0.460)	(1.165)	(1.144)	(1.363)	(1.23)	(1.274)	-
	$GenStat^{©}$	-3.397	0.689	0.287	1.384	-0.168	0.859	0.884	207.986
		(0.5551)	(0.4435)	(0.6690)	(0.6101)	(0.9837)	(0.7852)	(0.8692)	-

Table 4.35. Comparative analysis of the **Cohlen** data(regarding period as a factor)

The treatment estimates in the model involving treatment only are very much similar in all the four statistical packages. However when period is included , we observe that even though the treatment estimates are not very much dissimilar, the treatment estimate produced by $GenStat^{©}$ and the treatment estimate produced by $WinBugs^{©}$ are more different from the treatment estimate produced by $SAS^{©}$ and $R^{©}$ when period is included as a factor than when is included as having a linear effect. The deviances are similar. In fact, $SAS^{©}$, $R^{©}$ and $WinBugs^{©}$ give a change of deviance that is not significant in all the three cases.

Chapter 5

Summary

The general aim of this thesis is to investigate the place of crossover designs in infertility trials. Based on the logistic random effects model this research has investigated the importance of crossover design in infertility trials. We have shown that the use of crossover design in infertility trials is not bad as some statisticians believe. The logistic mixed effects model was employed because it allows the correlation between the repeated observation within a patient to be incorporated into the estimates of the parameters. Chapter 2 began with the overview of the standard logistic regression. We introduced and defined the mixed-effects regression model that was used throughout the dissertation. We reviewed the mixed-effects logistic regression for longitudinal (clustered) binary data. A full maximum marginal likelihood solution is outlined for the parameter estimation. In this solution, the quadrature method is used to numerically integrate over the distribution of the random-effects.

Chapter 3 mainly discusses the crossover design in infertility trials. We compared three types of data set; complete $AB : BA$ crossover design data set, incomplete $AB : BA$ crossover design data set and the parallel design data set.

The complete $AB : BA$ cross over design is the standard $AB : BA$ crossover design where every woman is exposed two both treatments regardless of the outcome in the first period. That is, women were allowed to get pregnant twice. The incomplete $AB : BA$ crossover design is the $AB : BA$ crossover design whereby whether or not the woman receives the second treatment depends on the outcome in the first period. If a positive outcome (pregnancy) is observed in the first period, the patient drops from the trial, leading to missing outcome in the second period. We first considered a complete $AB : BA$ crossover design, and from it we obtained both the incomplete $AB : BA$ crossover design data set and the parallel design data set. To obtain the incomplete $AB : BA$ crossover design data set, which is the realistic scenario, we deleted the second outcome for every woman who conceived in the first period. This is because by doing so, we do not allow women to get pregnant twice. Similarly, to get the parallel design data set, we deleted all the outcomes in the second period.

We constructed the likelihoods for each data set and obtain the maximum likelihood estimates (MLE) and their corresponding standard errors from each data set. The standard errors were obtained from the variance-covariance matrices, which were estimated from the negative inverse of the Fisher information (46). We compared the maximum likelihood estimates (and their corresponding standard errors) obtained using the complete $AB : BA$ crossover design data set with the maximum likelihood estimates (and their corresponding standard errors) obtained using the incomplete $AB : BA$ crossover design data set. These estimates were then marginalized so that they describe marginal response and compared with estimates from the parallel design data set. Specifically we looked at the data set from (30), which is not a infertility data set. The data set was merely used for illustration since we could not find a practical infertility data set

where a complete crossover design was performed. We employed MathCad$^{©}$ and **PROCNLMIXED** in SAS$^{©}$ to analyze this data set.

Next we simulated the $AB : BA$ crossover design with the help of a MathCad$^{©}$ software. We then constructed the incomplete crossover design and the parallel design like we previously did. We first considered two cases. One case with treatment only and the other with treatment and period effects. In each case, we simulated 2000 samples of size 300. Secondly we simulated 2000 samples of size 1000000. By doing so, we wanted to get a clear picture of what is happening in regard to the precision (standard errors).

Chapter 4 illustrates how in practice the cross over trials in infertility should be analyzed using different statistical packages (fitting the mixed effects logistic regression model). We used $SAS^{©}$, $R^{©}$ $WinBugs^{©}$ and GenStat$^{©}$.

5.1 Conclusion

In this thesis, we used a likelihood-based approach to the statistical analysis of pregnancy data from complete crossover design, incomplete crossover design and parallel design. This approach was based on the logistic random effects model incorporating both treatment and period effects. In both scenarios (the model with treatment only and the model with treatment and period effects) the treatment estimates obtained under complete $AB : BA$ crossover design and the treatment estimate obtained under the incomplete $AB : BA$ crossover design were not very much dissimilar. The standard errors of the treatment estimates obtained under the $AB : BA$ incomplete crossover design were higher than the standard errors of treatment estimates obtained under the complete crossover design. This did

not surprise us, since if we move from the complete $AB : BA$ crossover design to the incomplete $AB : BA$ crossover design information is lost on women who conceived in the first period. If we are in the complete $AB : BA$ crossover design, we have the information of what happened to these women in the second period. But when we are in the incomplete $AB : BA$ crossover design we do not know what is happening in the second period with regard to those women who conceived in the first period. Furthermore, this observation is expected since we are using the same model for the complete data set and the incomplete data set. The estimated marginalized treatment effects obtained using the incomplete or complete crossover design are not dissimilar to the treatment estimates obtained under parallel design. The results of this thesis do not support the conclusion of Daya that the crossover design should be avoided as inappropriate design. In both scenarios (the model with treatment only and the model with treatment and period effects), the treatment estimates obtained from the parallel design data set have the highest standard errors, followed by marginalized estimates from the incomplete crossover design data set. The marginalized estimates obtained from the complete crossover design data set have the lowest standard errors. These findings are corroborated by a set of simulations which also suggest that estimates from the parallel design will have higher standard errors than estimates obtained under crossover design.

Based on our simulations, the estimated treatment estimates obtained using the complete crossover design are sufficient relative to the treatment estimate obtained from incomplete crossover design. That is the complete crossover design should not be compared with the incomplete crossover design as it uses all available information. Suppose that one is presented with the treatment estimate obtained from the complete crossover design say τ_C, the expectation (or

best guess) of the treatment estimate τ_I, obtained using the incomplete design will be still τ_C. That is one can replace τ_I by τ_C without losing the meaning associated with the treatment estimate. But not vice versa. Likewise the treatment estimate obtained under the crossover design in infertility is sufficient relative to the treatment estimate obtained under parallel design. That is to say if we are presented with the marginalized treatment estimate obtained from the crossover design in infertility (incomplete data set) τ_{MI}, the expectation (or best guess) of the treatment estimate obtained from the parallel design data set will be still τ_{MI}. But not vice versa.

5.2 Discussion

Critics of crossover design in infertility like Daya (27) and Khan et al. (52) are mistaken by saying that the crossover design in infertility trials will lead to incomplete data set. We agree with them that crossover design in infertility trials will lead to incomplete data set only if it is compared with the standard $AB : BA$ crossover design, which is sufficient. But we believe that crossover design in infertility trials does not lead to incomplete data set if it is compared to the parallel design. The crossover design in infertility trial can be viewed as a parallel design with extra information. That is, the crossover design in infertility trials is a parallel design, but with information for those women who failed to conceive under treatment A (or treatment B) in the first period. Khan et al (52) stated that the crossover design in infertility trials may need to increase sample size, thus losing one of the major statistical advantages of the crossover design discussed in Section 1.7. However, we have shown that a sample size equivalent to a parallel design trial would give similar treatment estimate with less standard

error. Thus, there is no evidence that the crossover design presents a disadvantage in terms of sample size.

The debate in the position of crossover design in infertility trials started in more than a decade ago, when statistical softwares like **PROC NLMIXED** in $SAS^{©}$ and **lmer** procedure in $R^{©}$ were not available. Generally crossover designs are not easy to analyze especially when the outcome of interest is binary. The difficulty part is on the estimation of the integration over the random effects distribution. On the other hand the Pearson Chi-square test and student t-test have been available for the parallel design for a longer time. In the past investigators use to either analyze a crossover design as a parallel design or they do not allow the correlation between the repeated observation within a patient to be incorporated into the estimates of the parameters. McDonnell et al. (64) and Cohlen et al (24) both used models to examine the place of crossover design in infertility. Of the two models, that of McDonnell looks more like ours except that they did not account for the correlation between the repeated observation within a patient. We will like to point out that if the correlation between the repeated observation is not accounted for, the likelihood obtained under crossover design will be the same as the likelihood under parallel design as it was shown by McDonnell. The correlation between repeated observation within a patient is what makes crossover design different from the parallel design (98). Alborzi et al (3), Biacchiardi et al (13), Cohlen et al (23), Gregoriou et al (40), Kirby et al (53), Muharib et al (68), Nan et al (70), Sipe et al (86), Tiemessen et al (97) and Zreik et al (103), all used crossover design to determine if one treatment is superior to the other. However, the methods used in all these articles (3; 13; 24; 40; 53; 68; 70; 86; 97; 103) are not appropriate for the crossover design. The methods used (Chi-square and student t-test) are appropriate for a parallel design.

We are of the view that this debate was triggered by the fact that investigators did not analyze crossover design properly and not by the fact that the crossover design is inappropriate for infertility trials. The debate may be started because by that time there were no easier methods to properly handle crossover designs. Our advise to the investigators is that they should regard the information on the incomplete crossover design as extra and useful information relative to the information on the parallel design and thereby use it to derive more precise treatment estimates. We recommend the use of crossover designs in infertility trials if there is no evidence of carryover effects.

5.3 Modelling and statistical packages Issues

When data are missing, analyzing the observed data alone as if no data are missing can result in biased conditional estimates. An advantage of the methods used in Chapter 4 is that different patients can have different number of observations. The data input file has a separate line for each observation, and for longitudinal studies, computations use those times for which a patient has an observation. However, bias can arise in these methods unless the data is at least missing at random (31; 61). Little and Robin (61) called the data missing at random if missingness depends only on the observed response and not on the missing outcomes. When this is plausible, our likelihood based analysis using only observed responses is not systematically biased. We assumed that our data is missing at random and rightfully so. Often, missingness depends on the missing values. For instance, perhaps based may be on a new medical research a woman dropped out when it becomes evident that the next treatment will not be effective. Then, more complex analyses are needed that model the joint distribution

of the observed values and the missing values (60).

In this thesis the random effects were assumed to be distributed in the population as $N(0, \sigma^2)$. Neuhaus and Hauck (73) had examined the performance of the mixed effects logistic regression analysis when the random effects are misspecified. In their article, they have shown that when the random effects distribution is misspecified, estimates of model parameters including the treatment estimates are asymptotically biased. However the magnitude of the bias in the treatment estimate is small.

The **NLMIXED** procedure in $SAS^{©}$ uses exactly the procedure that is illustrated in Chapter 2. The **lmer** procedure in $R^{©}$ is only different with the **NLMIXED** on the integration over the random-effects distribution. The **lmer** procedure uses the Laplace approximations to approximate the integration. The $WinBugs^{©}$ procedure derive estimates by maximizing the same likelihood that is described in Chapter 2, but it approximates the integration over the random effects distribution by assigning proper but minimally informative prior distributions to the parameters. The likelihood in the GenStat$^{©}$ procedure is completely different from the likelihood described in Chapter 2. The GenStat$^{©}$ procedure maximizes the $h-$likelihood (56; 58) denoted by

$$h = \ell(\mathbf{Y}|\upsilon) + \ell(\upsilon), \tag{5.1}$$

where $\ell(\upsilon)$ is the logarithm of the density function for the random effect υ, and $\ell(\mathbf{Y}|\upsilon)$ is that of $\mathbf{Y}|\upsilon$. The estimates obtained from maximizing the $h-$likelihood are termed maximum h-likelihood estimates ($MHLEs$). The maximum h-likelihood estimates ($MHLEs$) are not the same as maximum likelihood estimates ($MLEs$)

obtained from either $SAS^{©}$, $R^{©}$ or $WinBugs^{©}$. The maximum likelihood estimates ($MLEs$) obtained from $SAS^{©}$, $R^{©}$ and $WinBugs^{©}$ are not exactly the same because each procedure uses a different method in approximating the integration over the random effects distribution ($SAS^{©}$ uses the Gauss-Hermite quadrature method, $R^{©}$ uses Laplace approximation or $WinBugs^{©}$ uses sampling method by assigning prior distributions to the parameters). But they are not very much dissimilar as expected.

Finally, although the negative of the matrix in equation 2.28 (Hessian matrix) must be positive definite and hence invertible to compute the variance matrix , invertible Hessian matrix do not exist for some combinations of data sets and models, and so statistical procedures sometimes fail for this reason before completion. Statistical softwares will give a warning if this is the case. When a Hessian matrix is not invertible, there is no computational trick that can make it invertible, given the model and the data chosen, since the desired inverse does not exist.

5.4 Further work

Just as two observations on the same patient might tend to be more correlated than observations on different patients, so might be observations on two patients in the same ART clinic tend to be more correlated than observations on patients from other ART clinics. So patients and ART clinics might be treated as random effects, with each referring to different levels of the model. So far we have only considered only one random effect. It will be of interest to investigate the results of multi-level modeling. It may be of interest also to consider mixed effects logistic models (one or more random effects) with unspecified random effects distribution.

This approach may help against possibly harmful misspecification effects (2).

It is also our interest to consider scenarios where missingness depends on the missing outcomes.

If we could have standardized the random effect as in (44), we observe that Monte carlo simulation can be used to estimate the parameters of interest. This will be another method that will be investigated in the future.

Appendix A

WinBugs Code for Gregoriou data

A.1 Model with Period only

model

{ for(i in 1 : N1) {

$outcome1[i] \sim dbern(p1[i])$

logit(p1[i]) $<-$ intercept+beta2*period1[i]+$v[i]$

$v[i] \sim dnorm(0, tau)$

}

for(i in 1 : N2){

outcome2[i] $\sim$ dbern(p2[i])

logit(p2[i]) $<-$ intercept+ beta2*period2[i]+$v[i]$

}

sigma$<-$ 1/sqrt(tau)

intercept $\sim$ dnorm(0,1.0E-3)

beta1 $\sim$ dnorm(0,1.0E-3)

tau $\sim$ dgamma(1,1.0E-4)

log.sigma$< -$ log(sigma)

}

list(outcome1 =c(0,0,0,0,0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0, 0, 0, 0, 0,0,

0, 0, 0, 0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,0,0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 1, 1, 1, 1,

1, 1, 1,1,1,1,1,1) ,

period1=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,

1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1),

outcome2 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0,

0, 0, 0, 0, 0,0,0,0,0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0, 0, 1, 1, 1, 1,1,1,1,1),

period2=c(-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-

1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1),

N1 = 62, N2=50),

list(intercept=-1,beta1=-1, beta2=1, tau=10,b=c(0, 0, 1, 0, 0, 0, 1, 0,

0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0,

0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,0))

list(intercept=-4,beta1=-4,beta2=4, tau=15,b=c(0, 0, 0, 0, 0, 0, 0,

0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0,

0, 0,0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,1))

Figure A.1. Plots of the intercept, beta2 and sigma for the model with period effects only

A.2 Model with Protocol only

model

{ for(i in 1 : N1) {

$outcome1[i] \sim dbern(p1[i])$

logit(p1[i]) $<-$ intercept+beta1*protocol1[i]+$v[i]$

$v[i] \sim dnorm(0, tau)$

}

for(i in 1 : N2){

outcome2[i] $\sim$ dbern(p2[i])

logit(p2[i]) $<-$ intercept+beta1*protocol2[i]+$v[i]$

}

sigma$<-$ 1/sqrt(tau)

intercept $\sim$ dnorm(0,1.0E-3)

beta1 $\sim$ dnorm(0,1.0E-3)

tau $\sim$ dgamma(1,1.0E-4)

log.sigma$<-$ log(sigma)

}

list(outcome1 =c(0,0,0,0,0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,0,0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 1, 1, 1, 1, 1, 1, 1,1,1,1,1,1) ,

protocol1=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,1,1,1,1,1,1,1,1,-1,1,1,1,1,1,-1,-1,-1,-1,-1,-1,-1,-1),

outcome2 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0, 0, 0, 0, 0, 0,0,0,0,0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0, 1, 1, 1, 1,1,1,1,1),

protocol2=c(-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,-1,-1,-1,-1,-1,-1,-1,1),

N1 = 62, N2=50),

list(intercept=-1,beta1=-1,tau=10,b=c(0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,0))

list(intercept=-4,beta1=-4, tau=15,b=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0,0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,1))

Figure A.2. Plots of the intercept, beta1 and sigma for the model with protocol only

A.3 Null model

model

{ for(i in 1 : N1) {

$outcome1[i] \sim dbern(p1[i])$

logit(p1[i]) $< -$ intercept$+v[i]$

$v[i] \sim dnorm(0, tau)$

}

for(i in 1 : N2){

outcome2[i] $\sim$ dbern(p2[i])

logit(p2[i]) $< -$ intercept$+v[i]$

}

sigma$< -$ 1/sqrt(tau)

intercept $\sim$ dnorm(0,1.0E-3)

tau $\sim$ dgamma(1,1.0E-4)

log.sigma$< -$ log(sigma)

}

list(outcome1 =c(0,0,0,0,0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 1, 1, 1, 1, 1, 1, 1,1,1,1,1,1) ,

outcome2 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0, 0, 0,

0, 0, 0,0,0,0,0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0, 0, 1, 1, 1, 1,1,1,1,1),

N1 = 62, N2=50),

list(intercept=-1,tau=10,b=c(0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0,
1, 1, 0, 1, 0, 1, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0,
0, 0, 0 ,0 ,0, 0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,0))

list(intercept=-4, tau=15,b=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0,0, 0, 0 ,0, 0, 0,
0, 0, 0 ,0 ,0, 0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,1))
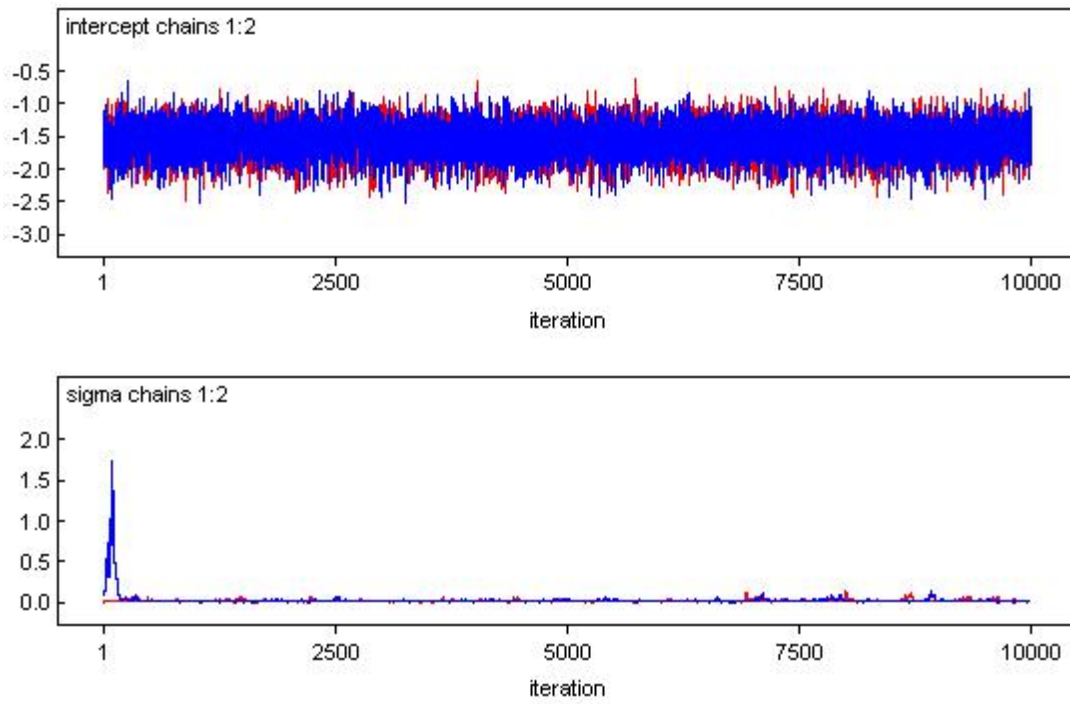
Figure A.3. Plots of the intercept and sigma for the null model

Appendix B

WinBugs Code for Cohlen data

B.1 Model with period effects only(factor)

model{

for(i in 1 : N1) {

$outcome1[i] \sim dbern(p1[i])$

$logit(p1[i]) < -intercept + beta2*period2[i] + beta3*period3[i] + beta4*period4[i] +$

$beta5 * period5[i] + beta6 * period6[i] + v[i]$

b[i] ∼ dnorm(0, tau)

}

for(i in 1 : N2){

$outcome2[i] \sim dbern(p2[i])$

$logit(p2[i]) < -intercept + beta2 * period22[i] + beta3 * period3[i] + beta4 *$

$period4[i] + beta5 * period5[i] + beta6 * period6[i] + v[i]$

}

for(i in 1 : N3){

$outcome3[i] \sim dbern(p3[i])$

$logit(p3[i]) < -intercept + beta2 * period2[i] + beta3 * period33[i] + beta4 * period4[i] + beta5 * period5[i] + beta6 * period6[i] + v[i]$

}

for(i in 1 : N4){

$outcome4[i] \sim dbern(p4[i])$

$logit(p4[i]) < -intercept + beta2*period2[i] + beta3*period3[i] + beta4*period44[i] + beta5 * period5[i] + beta6 * period6[i] + v[i]$

}

or(i in 1 : N5){

$outcome5[i] \sim dbern(p5[i])$

$logit(p5[i]) < -intercept + beta2*period2[i] + beta3*period3[i] + beta4*period4[i] + beta5 * period55[i] + beta6 * period6[i] + v[i]$

}

for(i in 1 : N6){

$outcome6[i] \sim dbern(p6[i])$

$logit(p6[i]) < -intercept] + beta2*period2[i] + beta3*period3[i] + beta4*period4[i] + beta5 * period5[i] + beta6 * period66[i] + v[i]$

}

$sigma < -1/sqrt(tau)$

$intercept \sim dnorm(0, 1.0E - 3)$

$beta2 \sim dnorm(0, 1.0E - 3)$

$beta3 \sim dnorm(0, 1.0E - 3)$

$beta4 \sim dnorm(0, 1.0E - 3)$

$beta5 \sim dnorm(0, 1.0E - 3)$

$beta6 \sim dnorm(0, 1.0E - 3)$

$tau \sim dgamma(1.0E - 4, 1.0E - 4)$

$log.sigma < -log(sigma)$ }

list(outcome1=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1) ,

period2=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),

period3=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),

period4=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),

period5=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),

period6=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0,

0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),

outcome2 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,0) ,

period22=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1),

outcome3 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 1, 1,1,1,1,1,1,1,1,1,1,0,0) ,

period33=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1),

outcome4 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 1, 1,0,0,0,0) ,

period44=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1),

outcome5 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0,0,0,0,1, 1, 1, 1,0,0) ,

period55=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1),

outcome6 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 1, 1, 1) ,

period66=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1),

N1 = 74, N2=67, N3=59,N4=46, N5=40, N6=34),

list(intercept=-4,beta2=1,beta3=1,beta4=1,beta5=1,beta6=1,tau=15,b=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0,0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0))

list(intercept=-1,beta2=0.3,beta3=0.3,beta4=0.3,beta5=0.3,beta6=0.3,tau=10, b=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1))

Figure B.1. Plots of the intercept, beta2,beta3 and beta4 for the model with period effects only(factor)
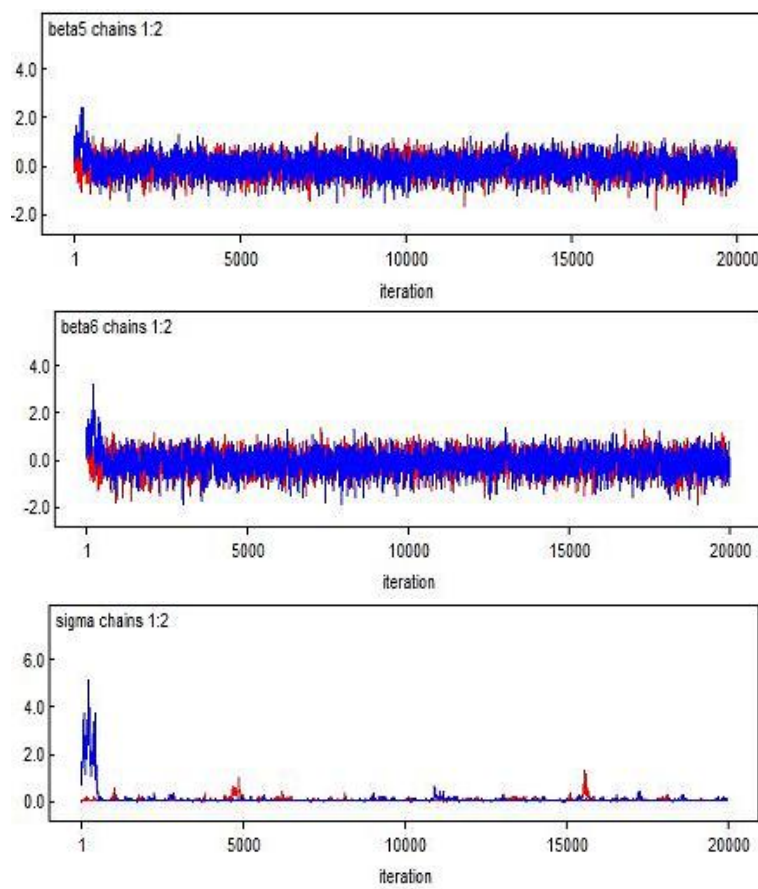
Figure B.2. Plots of the beta5,beta6 and sigma for the model with period effects only(factor)

B.2 Model with treatment and period effects(linear effect)

```
model{
for( i in 1 : N1 ) {
```

$outcome1[i] \sim dbern(p1[i])$

$logit(p1[i]) < -intercept + beta1 * treatment1[i] + beta2 * period1[i] + v[i]$

```
b[i] ~ dnorm(0, tau)
}
for( i in 1 : N2 ){
```

$outcome2[i] \sim dbern(p2[i])$

$logit(p2[i]) < -intercept + beta1 * treatment2[i] + beta2 * period2[i] + v[i]$

```
}
for( i in 1 : N3 ){
```

$outcome3[i] \sim dbern(p3[i])$

$logit(p3[i]) < -intercept + beta1 * treatment3[i] + beta2 * period3[i] + v[i]$

```
}
for( i in 1 : N4 ){
```

$outcome4[i] \sim dbern(p4[i])$

$logit(p4[i]) < -intercept + beta1 * treatment4[i] + beta2 * period4[i]] + v[i]$

```
}
or( i in 1 : N5 ){
```

$outcome5[i] \sim dbern(p5[i])$

$logit(p5[i]) < -intercept + beta1 * treatment5[i] + beta2 * period5[i] + v[i]$

```
}
for( i in 1 : N6 ){
```

$outcome6[i] \sim dbern(p6[i])$

$logit(p6[i]) < -intercept + beta1 * treatment6[i] + beta2 * period6[i] + v[i]$

}

$sigma < -1/sqrt(tau)$

$intercept \sim dnorm(0, 1.0E - 3)$

$beta1 \sim dnorm(0, 1.0E - 3)$

$beta2 \sim dnorm(0, 1.0E - 3)$

$tau \sim dgamma(1.0E - 4, 1.0E - 4)$

$log.sigma < -log(sigma)$ }

list(outcome1=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1) ,

treatment1=c(0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,1, 1,1,0,1,0,0,0,0,1,1,0,0,0,0,0,0,0,1,1,1,1,1,0,0,0,0,0,0,0,1,1,1,0,0,0,0,1,1,1),

period1=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1),

outcome2 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,0) ,

treatment2=c(1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,0,
0,0,1,0,1,1,1,1,0,0,1,1,1,1,1,1,0,0,0,0,0,1,1,1,1,1,1,1,0,0,0),

period2=c(2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,
2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2),

outcome3 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0,
0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 1, 1,1,1,1,1,1,1,1,1,1,0,0) ,

treatment3=c(0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,1,1,1,
0,1,0,0,0,0,1,1,0,0,0,0,0,0,1,1,1,1,1,0,0),

period3=c(3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,
3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3),

outcome4 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0
,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 1, 1,0,0,0,0) ,

treatment4=c(1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,0,0,0,
1,0,1,1,1,1,0,0),

period4=c(4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,
4,4,4,4,4,4,4,4),

outcome5 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0,0,0,0,1, 1, 1, 1,0,0) ,

treatment5=c(0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,1,1,1,0,1),

period5=c(5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5),

outcome6 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 1, 1, 1) ,

treatment6=c(1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1),

period6=c(6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6),

N1 = 74, N2=67, N3=59,N4=46, N5=40, N6=34),

list(intercept=-4,beta1=1,beta2=1,tau=15,b=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0,0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0))

list(intercept=-1,beta1=0.3,beta2=0.3,tau=10, b=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1))

Figure B.3. Plots for the intercept, *beta*1, *beta*2 and sigma for model with treatment and period effects (linear effect)

B.3 Model with period effects only(linear effect)

model{

for(i in 1 : N1) {

$outcome1[i] \sim dbern(p1[i])$

$logit(p1[i]) < -intercept + beta2 * period1[i] + v[i]$

b[i] $\sim$ dnorm(0, tau)

}

for(i in 1 : N2){

$outcome2[i] \sim dbern(p2[i])$

$logit(p2[i]) < -intercept + beta2 * period2[i] + v[i]$

}

for(i in 1 : N3){

$outcome3[i] \sim dbern(p3[i])$

$logit(p3[i]) < -intercept + beta2 * period3[i] + v[i]$

}

for(i in 1 : N4){

$outcome4[i] \sim dbern(p4[i])$

$logit(p4[i]) < -intercept + beta2 * period4[i]] + v[i]$

}

or(i in 1 : N5){

$outcome5[i] \sim dbern(p5[i])$

$logit(p5[i]) < -intercept + beta2 * period5[i] + v[i]$

}

for(i in 1 : N6){

$outcome6[i] \sim dbern(p6[i])$

$logit(p6[i]) < -intercept + beta2 * period6[i] + \upsilon[i]$

}

$sigma < -1/sqrt(tau)$

$intercept \sim dnorm(0, 1.0E - 3)$

$beta2 \sim dnorm(0, 1.0E - 3)$

$tau \sim dgamma(1.0E - 4, 1.0E - 4)$

$log.sigma < -log(sigma)$ }

list(outcome1=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,

0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0
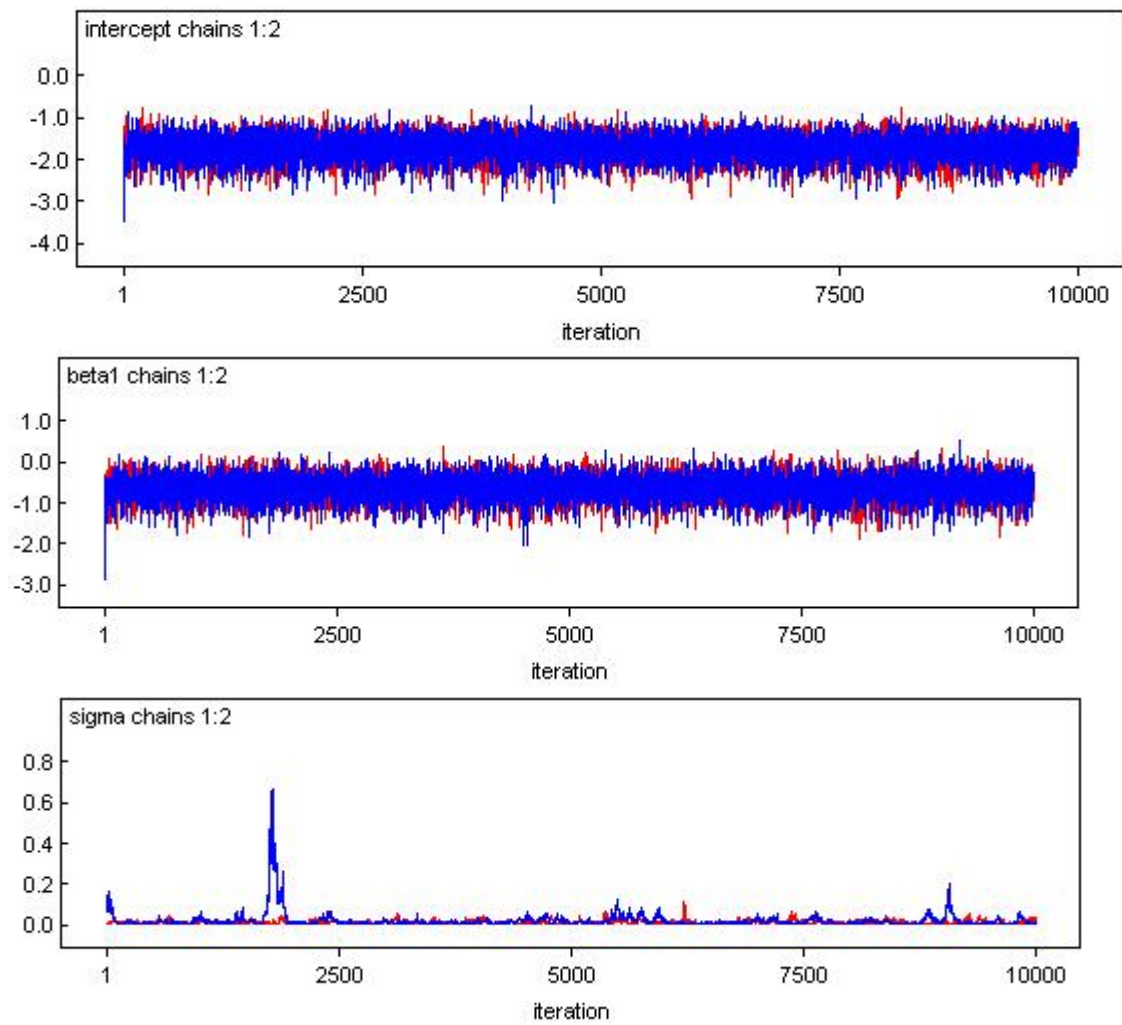
,0, 0, 0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1) ,

period1=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,

1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,

1,1),

outcome2 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,

0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0,

0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,0) ,

period2=c(2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,

2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2),

outcome3 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0

,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 1,

1,1,1,1,1,1,1,1,1,1,0,0) ,

period3=c(3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,

3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3),

outcome4 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0,

0, 0 ,0, 0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 1, 1,0,0,0,0) ,

period4=c(4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,
4,4,4,4,4,4,4,4,4,4,4,4),

outcome5 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0
,0, 0, 0, 0 ,0 ,0, 0, 0 ,0,0,0,0,1, 1, 1, 1,0,0),

period5=c(5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,
5,5,5,5,5,5,5),

outcome6 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 1, 1, 1) ,

period6=c(6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,6,
6,6),

N1 = 74, N2=67, N3=59,N4=46, N5=40, N6=34),

list(intercept=-4,beta2=1,tau=0.8,b=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0,0, 0, 0 ,0,
0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0))

list(intercept=-1,beta2=0.3,tau=0.5, b=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1))

Figure B.4. Plots of the intercept, $beta2$ and sigma for model with period effects only(linear effect)

B.4 Model with treatment only

model{

for(i in 1 : N1) {

$outcome1[i] \sim dbern(p1[i])$

$logit(p1[i]) < -intercept + beta1 * treatment1[i] + v[i]$

b[i] $\sim$ dnorm(0, tau)

}

for(i in 1 : N2){

$outcome2[i] \sim dbern(p2[i])$

$logit(p2[i]) < -intercept + beta1 * treatment2[i] + v[i]$

}

for(i in 1 : N3){

$outcome3[i] \sim dbern(p3[i])$

$logit(p3[i]) < -intercept + beta1 * treatment3[i] + v[i]$

}

for(i in 1 : N4){

$outcome4[i] \sim dbern(p4[i])$

$logit(p4[i]) < -intercept + beta1 * treatment4[i] + v[i]$

}

or(i in 1 : N5){

$outcome5[i] \sim dbern(p5[i])$

$logit(p5[i]) < -intercept + beta1 * treatment5[i] + v[i]$

}

for(i in 1 : N6){

$outcome6[i] \sim dbern(p6[i])$

$logit(p6[i]) < -intercept + beta1 * treatment6[i] + v[i]$

}

$sigma < -1/sqrt(tau)$

$intercept \sim dnorm(0, 1.0E - 3)$

$beta1 \sim dnorm(0, 1.0E - 3)$

$tau \sim dgamma(1.0E - 4, 1.0E - 4)$

$log.sigma < -log(sigma)$ }

list(outcome1=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1) ,

treatment1=c(0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,
1,1,1,0,1,0,0,0,0,1,1,0,0,0,0,0,0,0,1,1,1,1,1,0,0,0,0,0,0,0,0,1,1,1,0,0,0,0,0,1,1,1),

outcome2 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0,
0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,0) ,

treatment2=c(1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,
0,0,0,1,0,1,1,1,1,0,0,1,1,1,1,1,1,0,0,0,0,0,1,1,1,1,1,1,1,0,0,0),

outcome3 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0,
0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 1, 1,1,1,1,1,1,1,1,1,1,0,0) ,

treatment3=c(0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,1,
1,1,0,1,0,0,0,0,1,1,0,0,0,0,0,0,0,1,1,1,1,1,0,0),

outcome4 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0
,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 1, 1,0,0,0,0) ,

treatment4=c(1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,0,0,
0,1,0,1,1,1,1,0,0),

outcome5 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0,
0 ,0 ,0, 0, 0 ,0,0,0,0,1, 1, 1, 1,0,0) ,

treatment5=c(0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,
1,1,1,0,1),

outcome6 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0,
0, 0, 0 ,0 ,0, 0, 1, 1, 1) ,

treatment6=c(1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1),

N1 = 74, N2=67, N3=59,N4=46, N5=40, N6=34),

list(intercept=-4,beta1=1,tau=15,b=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0,0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0
,0, 0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0))

list(intercept=-1,beta1=0.3,tau=10, b=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1))

Figure B.5. Plots of the intercept, *beta*1 and sigma for model with treatment only

B.5 Null model

model{

for(i in 1 : N1) {

$outcome1[i] \sim dbern(p1[i])$

$logit(p1[i]) < -intercept + \upsilon[i]$

b[i] $\sim$ dnorm(0, tau)

}

for(i in 1 : N2){

$outcome2[i] \sim dbern(p2[i])$

$logit(p2[i]) < -intercept + v[i]$

}

for(i in 1 : N3){

$outcome3[i] \sim dbern(p3[i])$

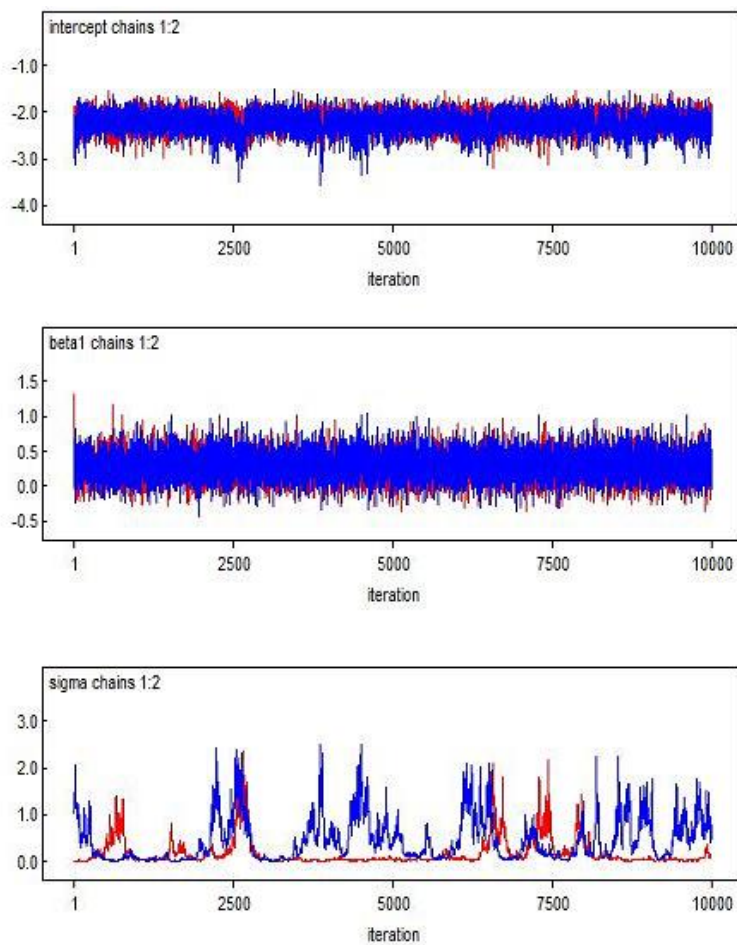$logit(p3[i]) < -intercept + v[i]$

}

for(i in 1 : N4){

$outcome4[i] \sim dbern(p4[i])$

$logit(p4[i]) < -intercept + v[i]$

}

or(i in 1 : N5){

$outcome5[i] \sim dbern(p5[i])$

$logit(p5[i]) < -intercept + v[i]$

}

for(i in 1 : N6){

$outcome6[i] \sim dbern(p6[i])$

$logit(p6[i]) < -intercept + v[i]$

}

$sigma < -1/sqrt(tau)$

$intercept \sim dnorm(0, 1.0E - 3)$

$tau \sim dgamma(1.0E - 4, 1.0E - 4)$

$log.sigma < -log(sigma)$ }

list(outcome1=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0
,0, 0, 0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1) ,

outcome2 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0,
0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,0) ,

outcome3 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 0, 0 ,0 ,0, 1,
1,1,1,1,1,1,1,1,1,1,0,0),

outcome4 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0,
0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0, 0, 0, 0 ,0, 0, 0, 0, 1, 1,0,0,0,0,0) ,

outcome5 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ,0
,0, 0, 0, 0 ,0 ,0, 0, 0 ,0,0,0,0,1, 1, 1, 1,0,0),

outcome6 =c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 1, 1, 1),

N1 = 74, N2=67, N3=59,N4=46, N5=40, N6=34),

list(intercept=-4,tau=15,b=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0 ,0 ,0, 0, 0, 0 ,0 ,0, 0, 0 ,0, 0, 0, 0, 0,0, 0, 0 ,0, 0, 0, 0,
0, 0 ,0 ,0, 0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0))

list(intercept=-1,tau=10, b=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1))

Figure B.6. Plots of the intercept and sigma for the null model

Bibliography

[1] A. Agresti. *Categorical Data Analysis*. John Wiley and Sons, New Jersey, 2nd edition edition, 1989.

[2] M. Aitkin. A general maximum likelihood likelihood analysis of variance components in generalized linear models. *Biometrics*, 55, 1999.

[3] S.A. Alborzi, S. Motazedian, M.E. Parsanezhad, and S. Jannati. Comparison of the effectiveness of a single intrauterine insemination (iui) versus double iui per cycle in infertile patients. *Fertility and Sterility*, 80, 2003.

[4] A. Allegra, A. Marino, F. Coffaro, P. Scaglione, F. Sammartano, G. Rizza, and A. Volpes. Gnrh antagonist-induced inhibition of the premature lh surge increases pregnancy rates in iui-stimulated cycles. a prospective randomized trial. *Human Reproduction*, 22:101–108, 2007.

[5] P.D. Allison. *Logistic Regression Using The SAS System, Theory and Application*. Wiley Inter-Science, 2001.

[6] C.V. Ananth and G.G. Rhoads. Future letters-'workshops on the internet' [letter]. *Fertility and Sterility*, (Sage):179, 1997.

[7] D. Anderson and M. Aitkin. Variance components models with binary response. *Journal of the Royal Statistical Society,Series B*, 47:203–210, 1985.

[8] S. Antinori, B. Caffa, H. A. Selman, C. Panci, G. Dani, and C. Versaci. Enhancement of pregnancy rates in patients with repeated ivf failures by increasing the number of embryos transferred. pages 120–121, 1996.

[9] D. Baker, G.B. Erwin and B.S. Kramer. Estimating the cumulative risk of false positive cancer screenings. *BMC Medical Research Methodology*, 3:1–6, 2003.

[10] A. H. Balen, H. S. Jacobs, and A. H. DeChernery. *Infertility in practice.* Churchill Livingstone, 1997.

[11] J. Bancroft. Impact of environment, stress, occupational, and other hazards on sexuality and sexual behavior. *Environmental Health Perspectives*, 101:101–107, 1993.

[12] J.C. Barrett and J. Marshall. The risk of conception on different days of the menstrual cycle. *Population Studies*, 23:455–461, 1969.

[13] C.P. Biacchiardi, A. Revelli, G. Gennarelli, S. Rustichelli, F. Moffa, and M. Massobrio. Fallopian tube sperm perfusion versus intrauterine insemination in unexplained infertility:a randomized, prospective, cross-over trial. *Fertility and Sterility*, 81:448–451, 2004.

[14] H. Blau, E. Freud, H. Mussaffi, M. Werner, O. Konen, and V. Rothaus. Urogenital abnormalities in male children with cystic fibrosis. *Archives of Diseases in Childhood*, 87:135–138, 2002.

[15] J. Boivin, K. Sanders, and L. Schmidt. Age and social position moderate the effect of stress on fertility. *Evolution and Human Behavior*, 27(5):345–356, 2006.

[16] R. Boostanfar, J.K. Jain, D.R. Jr. Mishell, and R.J. Paulson. A prospective randomized trial comparing clomiphene citrate with tamoxifen citrate for ovulation induction. *Fertility and Sterility*, 75(5):1024–6, 2001.

[17] M. J. Campbell. *Medical Statistics: A commonsense approach*. Wiley, third edition.

[18] N. Carroll and J.R. Palmer. A comparison of intrauterine versus intracervical insemination in fertile single women. *Fertility and Sterility*, 75(4):656–660, 2001.

[19] G. Casella and R.L. Berger. Statistical inference, second edition. *Duxbury Thomson Learning*, 2002.

[20] CDC. Cdc report on national center for health statistics. *The US Centers for Disease Control and Prevention (CDC)*, 2002.

[21] CDC. Cdc report on art. *The US Centers for Disease Control and Prevention (CDC)*, 2003.

[22] J.H. Check, E. Davies, and H. Adelson. A randomized prospective study comparing pregnancy rates following clomiphene citrate and human menopausal gonadotrophin therapy. *Human Reproduction*, 7:801–805, 1992.

[23] B.J. Cohlen, E.R. te Velde, W.W.N. Looman, R. Eijckemans, and J.D.F. Habbema. Crossover or parallel design in infertility trials? the discussion continues. *Fertility and Sterility*, 70:40–45, 1998.

[24] B.J Cohlen, E.R te Velde, R.J Van Kooij, C. W. N. Looman, and Habbema J.D.F. Controlled ovarian hyperstimulation and intrauterine insemination for treating male subfertility: a controlled study. *Human Reproduction*, 13(6):1553–1558, 1998.

[25] M.R. Conaway. Analysis of repeated categorical measurements with conditional likelihood methods. *Journal of the American Statistical Association*, 84:53–61, 1989.

[26] D.R. Cox. The analysis of binary data. *Chapman and Hall, London*, 1970.

[27] S. Daya. Is there a place for the crossover design in infertility trials? *Fertility and Sterility*, 59:6–7, 1993.

[28] A.J. Dobson. An introduction to generalized linear models. *Chapman and Hall: London*, 2nd Edition, 2002.

[29] D.R. Elbourne, D.G. Altman, J.P.T. Higgins, F. Curtin, H.V. Worthington, and Vail. A. Meta-analyses involving cross-over trials:methodological issues. *International Journal of Epidemiology*, 31:140–149, 2002.

[30] F. Ezzet and J. Whitehead. A random effects model for binary data from crossover clinical trials. *Applied Statistics*, 41:117–126, 1992.

[31] G.M. Fitzmaurice, N.M. Laird, and S.R. Lipsitz. Analysing incomplete longitudinal binary responses: A likelihood-based approach. *Biometrics*, 50(3):601–612, 1994.

[32] G.M. Fitzmaurice, N.M. Laird, and A.G. Rotnitzky. Regression models for discrete longitudinal responses. *Statistical Science*, 8:127–153, 1993.

[33] P.A. Flierman, H.V. Hogerzeil, and D.J. Hemrika. A prospective, randomized, cross-over comparison of two methods of artificial insemination by donor on the incidence of conception: intracervical insemination by straw versus cervical cap. *Human Reproduction*, 12(9):1945–1948, 1997.

[34] P. Freeman. The performance of the two-stage analysis of two-treatment, two-period crossover trials. *Statistics in Medicine*, 8:1421–1432, 1989.

[35] H. Friedl and G. Kauermann. Standard errors for em estimates in generalised linear models with random effects. *Biometrics*, 56:761–767, 2000.

[36] R.D. Gibbons and D. Hedeker. Random effects probit and logistic regression models for three-level data. *Biometrics*, 53:1527–1537, 1997.

[37] H. Goldstein. Nonlinear multilevel models, with an application to discrete response data. *Biometrika*, 78:45–51, 1991.

[38] H. Goldstein and J. Rasbash. Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society,Series A*, 159:505–513, 1996.

[39] A.J Goverde, J. McDonnell, J. P. W. Vermeiden, R. Schats, F. F. H. Rutten, and J. Schoemaker. Intrauterine insemination or in-vitro fertilisation in idiopathic subfertility and male subfertility: a randomized trial and cost effectiveness analysis. *The Lancet*, 355:13–18, 2000.

[40] O. Gregoriou, N. Vitoratos, N. Papadias, S. Konidaris, A. Gargaropoulos, and D. Rizos. Pregnancy rates in gonadotrophin stimulated cycles with timed intercourse or intrauterine insemination for the treatment of male subfertility. *European Journal of Obstetrics, Gynecology, and Reproductive Biology*, 64:213–216, 1996.

[41] J.E. Grizzle. The two-period change over design and its use in clinical trials. *Biometrics*, 21:467–80, 1965.

[42] ROCG Infertility Guideline Group. *The management of infertility in secondary care.* Royal College of Obstetricians and Gynaecologists, London, 1988.

[43] D.S. Guzick, S.A. Carson, C. Coutifaris, J.W. Overstreet, P. Factor-Litvak, M.P. Steinkampr, J.A. Hill, L. Mastroianni, J.E. Buster, S.T. Nakajima, D.L. Vogel, and R.E. Canfield. Efficacy of superovulation and intrauterine inseminiation in the treatment of infertility. *The New England Journal of Medicine*, 340(3):177–183, 1999.

[44] D. Hedeker and Gibbons R.D. *Longitudinal Data Analysis*. John Wiley and Sons, New Jersey, 2006.

[45] M. Hills and P. Armitage. The two-period cross-over clinical trial. *British journal of clinical pharmacology*, 8:7–20, 1979.

[46] R.V. Hogg and A.T. Craig. Introduction to mathematical statistics. *Prentice Hall, Upper Saddle River, New Jersey*, 5th Edition, 1995.

[47] E. Hughes, J. Collins, and P. Vandekerckhove. Clomiphene citrate versus placebo for ovulation induction in oligo-amenorrhoeic women in subfertility. *Module of Cochrane database of systematic reviews*, 1996.

[48] S.M. Hughes, Z.H. Huang, I.D. Morris, P.L. Matson, P. Buck, and B.A. Lieberman. A double-blind cross-over controlled study to evaluate the effect of human biosynthetic growth hormone on ovarian stimulation in previous poor responders to in-vitro fertilization. *Human Reproduction*, 9(1):13–18, 1994.

[49] M.G.R. Hull, C.M.A. Glazener, and N.J. Kelly. Population study of causes, treatment and outcome of infertility. *British Medical Journal*, 291:175–182, 1985.

[50] B. Jones and M.G. Kenward. Design and analysis of cross-over trials. *Chapman and Hall, London*, 1989.

[51] J.F.P. Kerin, J. Peek, G.M. Warnes, C. Kirby, R. Jeffrey, and C.D. Matthews. Improved conception rate after intrauterine insemination of washed spermatozoa from men with poor quality semen. *The Lancet*, 1:533–535, 1984.

[52] K.S. Khan, S. Daya, J.A. Collins, and Walter S.D. Empirical evidence of bias in infertility research;overestimation of the treatment effect in crossover trials using pregnancy as the outcome measure. *Fertil Steril*, 65:6–7, 1996.

[53] C.A. Kirby, S.P. Flaherty, B.M. Godfrey, G.M. Warnes, and C.D. Matthews. A prospective trial of intrauterine insemination of motile spermatozoa versus timed intercourse. *Fertil Steril*, 56, 1991.

[54] G.G. Koch, P.B. Imrey, J.M Singer, S.S. Atkinson, and Stokes M.E. Analysis of categorical data. *Montreal, Canada: Les Presses de L'Universitè de Montrel*, 1985.

[55] R. Kumar, G. Gautam, and N. Gupta. Drug therapy for idiopathic male infertility: Rationale versus evidence. *The Journal of Urology*, 176(4):1307 – 1312, 2006.

[56] Y. Lee and J.A. Nelder. Hierarchical generalized linear models. *Journal of the Royal Statistical Society,Series B*, 58:619–678, 1996.

[57] Y Lee and J.A. Nelder. Conditional and marginal models:another view. *Statistical Science*, 19:219–238, 2004.

[58] Y. Lee, J.A. Nelder, and Y. Pawitan. Generalized linear models with random effects. *Chapman and Hall*, 2006.

[59] E. Lesaffre and B. Spiessens. On the effect of the number of quadrature

points in a logistic random-effects model: An example. *Applied Statistics*, 50:325–335, 2001.

[60] R.J. Little. Missing data. *Encyclopedia of Biostatistics*, pages 2622–2635.

[61] R.J. Little and D.B. Rubin. *Statistical Analysis with Missing Data*. Wiley, New York, 1987.

[62] N.T Longford. *Random Coefficient Models*. Oxford, Oxford University Press, 1994.

[63] P. McCullagh and J.A. Nelder. Generalized linear models. *Chapman and Hall: London*, 1989.

[64] J. McDonnell, A.J. Goverde, and J.P.W. Vermeiden. The place of the crossover design in infertility trials:a maximum likelihood approach. *Hum reprod*, 19:2537–2544, 2004.

[65] G.B. Melis, A.M. Paoletti, S. Ajossa, S. Guerriero, G.F. Depau, and V. Mais. Ovulation induction with gonadotropins as sole treatment in infertilie couples with open tubes: a randomized prospective comparison between intrauterine insemination and timed vaginal intercourse. *Fertility and Sterility*, 64:1088–1093, 1995.

[66] B.E. Menning. The emotional needs of infertile couples. *Fertility and Sterility*, 34(4), 1980.

[67] B.W.J Mol and P.M.M. Bossuyt. Future letters-'workshops on the internet' [letter]. *Fertil Steril*, 67:179, 1997.

[68] N.S. Muharib, A.A. Gadir, and R.W. Shaw. Slow release intrauterine insemination versus the bolus technique in the treatment of women with cervical mucus hostility. *Human Reproduction*, 7:227–229, 1992.

[69] V. Murphy and A. Dunne. Mixed effects versus fixed effects modelling of binary data with inter-subject variability. *Journal of Pharmacokinetics and Pharmacodynamics*, 32:245–260, 2005.

[70] P.M. Nan, B.J. Cohlen, te Velde E.R., R.J. Kooij, J.M. Eimers, P. Zonneveld, and J.D. Habbema. Intra-uterine insemination or timed intercourse after ovarian stimulation for male subfertility? a controlled study. *Human Reproduction*, 9(11):2022–2026, 1994.

[71] J.A Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.

[72] J. Neter, M.H. Kutner, C.J. Nachtsheim, and W. Wasserman. Applied linear statistical models. *McGraw-Hill Companies*, 4th Edition, 1996.

[73] J.M. Neuhaus and W.W. Hauck. The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika*, 79(4):755–62, 1992.

[74] G.R. Norman and S. Daya. The alternating-sequence design (or multiple-period crossover) trial for evaluating treatment efficacy in infertility. *Fertil Steril*, 74:319–324, 2000.

[75] D.L. Olive. Future letters-'workshops on the internet' [letter]. *Fertil Steril*, 67:179, 1997.

[76] H. Page. The increasing demand for infertility treatment. *Health Trends*, 1988.

[77] Y Pawitan. In all likelihood: Statistical modelling and inference using likelihood. *Clarendon Press, Oxford*, 2001.

[78] R.J. Pepperell, B. Hudson, and C. Wood. *The Infertile Couple.* Churchill Livingstone, Endinburgh London and NewYork, 1980.

[79] P.C.B. Phillips. The exact distribution of the wald statistic. *Econometrica*, 54(4), 1986.

[80] J.C. Pinheiro and D.M. Bates. Approximations to the log-likelihood function in the nonlinear mixed effects model. *Jornal of Computational and Graphical Statistics*, 4:12–35, 1995.

[81] C. Ren, D. Sun, P.L. Speckman, C.Z. He, and S. Swan. Hierarchical models for the probabilities of conception. *Biometrical Journal*, 5:721–739, 2005.

[82] L.D. Robinson and N.P. Jewell. Some suprising results about covariate adjustment in logistic regression models. *International Statistical Review*, 58:227–240, 1991.

[83] G. Rodriguez and N. Goldman. An assessment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society,Series A*, 158:73–89, 1995.

[84] S Senn. Cross-over trials in clinical research. *Wiley, Newyork*, 2nd edn, 2002.

[85] S. Senn, G. D'Angelo, and D Potvin. Carry-over in cross-over trials in bioequivalence: theoritical concerns and empirical evidence. *Pharmaceutical statistics*, 3:133–142, 2004.

[86] C.S. Sipe, W.A. Davis, M. Maifeld, and B.J. Van Voorhis. A prospective randomized trial comparing anastrozole and clomiphene citrate in an ovulation induction protocol using gonadotropins. *Fertlity and Sterility*, 86(6):1676–1681, 2006.

[87] A. Skrondal and S. Rabe-Hesketh. Generalized latent variable modeling: Multilevel longitudinal, and structural equation modles. *Chapman and Hall/CRC, Newyork*, 2004.

[88] T. Snijders and R. Bosker. Multilevel nalysis. *Thousand Oaks*, CA(Sage), 1999.

[89] L. Speroff and M.A. Fritz. *Clinical Gynecologic Endocrinology And Infertility.* Lippincott Williams and Wilkins, 7th edition edition, 2005.

[90] D.J. Spiegelhalter, N.G. Best, B.P. Carlin, and A.V. Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical society B*, 64:583–639, 2002.

[91] R. Stiratelli, N.M. Laird, and J.H. Ware. Random-effects models for serial observations with binary response. *Biometrics*, 40:961–971, 1984.

[92] E. Stokes, S.D. Charles, and G.G. Koch. Categorical data analysis using the sas system. *Cary, NC:SAS Institute INc.,2000*, 2nd Edition, 1996.

[93] A.H. Stroud and D. Secrest. Gaussian quadrature formulas. *Prentice Hall, Englewood Cliffs*, 1966.

[94] E.R. te Velde, B.J. Cohlen, C.W.N. Looman, and Habbema J.D.F. Crossover design versus parallel studies in infertility research [letter]. *Fertility and Sterility*, 69:357–358, 1998.

[95] A. Templeton, C. Fraser, and B. Thompson. The epidemiology of infertility in aberdeen. *British Medical Journal*, 301:148–152, 1990.

[96] A. Templeton, C. Fraser, and B. Thompson. Infertility- epidemiology and referral practice. *Human Reproduction*, 6:1391–1394, 1991.

[97] C.H.J. Tiemessen, R.S.G.M. Bots, M.F. Peeters, and J.L.H. Evers. Direct intraperitoneal insemination compared to intrauterine insemination in superovulated cycles:a randomized cross-over study. *Gynecologic and Obstetric Investigation*, 44:149–152, 1997.

[98] E.R. Williams and J.A. John. Construction of crossover designs with correlated errors. *Australian and New Zealand Journal of Statistics*, 49(1):61–68, 2007.

[99] G.Y. Wong and W.M. Mason. The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association*, 80:513–524, 1985.

[100] J.R. Woods, J.G. William, and M. Tavel. The two period crossover design in medical research. *Annals of internal medicine*, 110(7):560–6, 1989.

[101] H. Zhou and C.R. Weinberg. Potential for bias in estimating human fecundability parameters:a comparison of statistical models. *Statistics in Medicine*, 18:411–422, 1999.

[102] H. Zhou, C.R. Weinberg, A.J. Wilcox, and D.D. Baird. A random-effects model for cycle viability in fertility studies. *BMC Medical Research Methodology*, 3:1–6, 2003.

[103] T.G. Zreik, J.A. Garcia-Velasco, M.S. Habboosh, D.L. Olive, and A. Arici. Prospective, randomized, crossover study to evaluate the benefit of human chorionic gonadotropin-timed versus urinary luteinizing hormone-timed intrauterine inseminations in clomiphene citrate-simulated treatment cycles. *Fertily and Sterility*, 71:1070–1074, 1999.