PhUSE 2006

Paper ST07

Analysis of incomplete two-period crossover trials with SAS PROC MIXED

Jochen Mueller-Cohrs, Accovion GmbH, Marburg, Germany

ABSTRACT

Different methods for analysing normally distributed data from a simple two-period, two-treatment crossover trial with SAS/STAT® are explored. Special emphasis is given to trials in which some data from one of the two periods are missing. The size and the power of the statistical tests are investigated by simulating data under various scenarios. To select the most suitable method for a particular trial the number and pattern of the available data and the covariance structure of the variables must be considered. The paper is rather application oriented and empirical in nature and touches on theory on few occasions only.

1. INTRODUCTION

In a simple two-period, two-treatment crossover trial experimental units are randomly assigned to two treatment sequences. For the sake of illustration I assume that the experimental units are patients in a clinical trial. The patients assigned to Sequence *AB* receive treatment *A* in Period 1 and treatment *B* in Period 2. The patients assigned to Sequence *BA* receive the treatments in reverse order. At the end of each period a target variable is measured to compare the effects of the treatments.

The analysis of crossover studies has received considerable attention in the statistical literature; see e.g. Grieve and Senn [1] and the references there. I assume that based on prior knowledge carry-over effects and treatment by period interactions are negligible (cf. Grieve and Senn [1] regarding the problem of testing for the presence of carry-over effects). In a clinical trial and also in other applications of crossover designs it is not unusual that some pairs of observations are incomplete. In this paper I will present some methods for analysing incomplete data from the simple two-period, two-treatment crossover design with the procedure MIXED of SAS/STAT® [4]. I will focus on testing the treatment effect only. Period effects are rarely of interest in itself.

Patel [3] suggested a maximum-likelihood test for models with carry-over effect. I shall consider restricted maximum likelihood (REML) tests in models without carry-over effect and examine size and power of twelve different methods of analysis.

The plan of the paper is as follows. In Section 2 I will introduce the statistical model. In Section 3 I briefly recall the standard analysis if no data is missing and compare the results with various analysis methods using PROC MIXED. Section 4 presents the results of a small simulation study for incomplete data sets and gives some hints as to what method of analysis might be most suitable in a particular trial.

2. STATISTICAL MODELS

Let $\boldsymbol{Y}_1 = [Y_{11}, Y_{12}]$ denote a pair of observations obtained from a patient who received treatment *A* in Period 1 and treatment *B* in Period 2 and let $\boldsymbol{Y}_2 = [Y_{21}, Y_{22}]$ denote a pair of observations obtained from a patient who received the treatments in reverse order. I assume that expectation and variance of the random pairs are as follows:

(1)

Treatment sequence	Random pair $\boldsymbol{Y}$	Expected value of $\boldsymbol{Y}$		Variance matrix of $\boldsymbol{Y}$
		Period 1	Period 2	
A B	$\boldsymbol{Y}_1 = [Y_{11}, Y_{12}]$	μ	$\mu + \pi + \beta$	$\boldsymbol{U} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$
B A	$\boldsymbol{Y}_2 = [Y_{21}, Y_{22}]$	$\mu + \beta$	$\mu + \pi$	$\boldsymbol{V} = \begin{bmatrix} \tau_1^2 & \tau_{12} \\ \tau_{12} & \tau_2^2 \end{bmatrix}$

The fixed effects of the model are the expected response μ to Treatment *A* in Period 1, the effect β of Treatment *B* relative to Treatment *A*, and the effect π of Period 2 relative to Period 1. I assume that treatment and period effects

are additive, particularly I do not allow for a carry-over effect. Grieve and Senn [1] provided an enlightening discussion of this issue. The delicate difference between effects of trial period and calendar time is also elucidated there.

I assume the random pairs $\boldsymbol{Y}$ to be normally distributed. Regarding the covariance structure I assume that all observed random pairs are stochastically independent, that the covariance matrix for random pairs from treatment sequence *AB* is $\boldsymbol{U}$, and that the covariance matrix for random pairs from treatment sequence *BA* is $\boldsymbol{V}$. This is a quite general covariance structure, which I will impose certain restrictions upon. A common assumption is that the covariance matrix is the same for all patients, i.e. $\boldsymbol{U}=\boldsymbol{V}$, and that the variances of the individual observations in Period 1 and Period 2 are the same. This is commonly referred to as compound symmetry covariance structure.

(2) Compound symmetry variance: $\boldsymbol{U} = \boldsymbol{V} = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$

Usually the correlation coefficient ρ for the correlation between the two observations from an individual patient is positive. This is the case, for example, if the covariance structure arises from a simple variance component model in which the patient is the only random effect besides the residual error. If the variance in the second period is different from the variance in the first period then we have what I will call a period-heterogeneous covariance structure.

(3) Period-heterogeneous variance: $\boldsymbol{U} = \boldsymbol{V} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$ $\rho = \sigma_{12} / \sigma_1 \sigma_2$.

On the other hand, there might be reasons to assume a different variability in the response to the two treatments. This would lead to the so-called treatment-heterogeneous covariance structure.

(4) Treatment-heterogeneous variance: $\boldsymbol{U} = \begin{bmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{bmatrix}$ $\boldsymbol{V} = \begin{bmatrix} \sigma_B^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_A^2 \end{bmatrix}$ $\rho = \sigma_{AB} / \sigma_A \sigma_B$.

3. ANALYSIS WITH COMPLETE DATA

The case in which no observation from any period is missing has been treated extensively in the literature. Senn [5] has provided a thorough account. In simple terms, the treatment effect can be tested by applying a two-sample *t*-test to the observed differences between Period 2 and Period 1. The expected value of the difference is $\pi+\beta$ for treatment sequence *AB* and $\pi-\beta$ for treatment sequence *BA*. The individual differences are all independent and normally distributed. They also have a common variance provided any of the above covariance structures (2), (3) or (4) holds. Thus, under these assumptions, the two-sample *t*-test is an exact test for the treatment effect β. For ease of reference I will call the two-sample *t*-test applied to the individual differences the standard analysis.

If the covariance structure is as general as in (1) then the variances of the individual differences would not necessarily be the same for the two treatment sequences and a Behrens-Fisher problem would arise.

With SAS/STAT the standard analysis can be realized with the procedure TTEST, of course. Another option that yields the same result is to do a fixed effects analysis of variance with the procedure GLM, say. To be specific I assume that the observed values are organized in a SAS® data set with the variables PAT, SEQ, PER, TRT, and VAL. The variable PAT identifies the patient and SEQ the assigned treatment sequence; PER indicates the period in which the measurement is made and TRT the corresponding treatment during that period; VAL is the value of the response variable measured at the end of that period in that patient. An appropriate analysis of variance model that produces the same test for the treatment effect as the standard analysis has the fixed effects PAT, PER, and TRT. Thus with PROC GLM the syntax may be

```
(5)   proc glm;
         class PAT PER TRT;
         model VAL = PAT PER TRT;
         lsmeans TRT;
```

A sequence effect SEQ may be included in the model (before the effect PAT) but it has no influence on the sums of squares for TRT because the sequence effect is included in the patient effect.

In case of missing values a natural approach is to employ a repeated measures analysis using the procedure MIXED of SAS. Therefore the first thing one wants to know is which mixed models yield the same result as the standard analysis in case of complete observations.

A literal translation of model (1) with compound symmetry covariance structure (2) into PROC MIXED syntax is

```
(6)   proc mixed;
         class PAT PER TRT;
         model VAL = PER TRT;
         repeated / sub=PAT type=cs;
```

This model actually produces the same test for the treatment effect as the standard analysis. The treatment effect can be estimated using the LSMEANS statement as in analysis (5) above. The REPEATED statement may be replaced by a RANDOM statement

```
random PAT;
```

reflecting the component of variance approach. The problem is, however, that if the random effects of the patients are relatively small compared to the residual error the estimate of the variance component might be zero in a particular trial. The default computation of PROC MIXED would then ignore the effect PAT and the standard error of the treatment effect would be different from the one obtained with the standard analysis. Therefore I shall not further pursue the use of the RANDOM statement here.

As in the case of a fixed effects model (5) the inclusion of a sequence effect does not alter the test of the treatment effect. Thus, the following model also produces the same result as the standard analysis.

```
proc mixed;
   class PAT SEQ PER TRT;
(7)  model VAL = SEQ PER TRT;
   repeated / sub=PAT type=cs;
```

As mentioned above the standard analysis is exact even if the covariance structure is heterogeneous as in (3) or (4). With PROC MIXED a period-heterogeneous variance (3) can be modelled with the statements

```
proc mixed;
   class PAT PER TRT;
(8)  model VAL = PER TRT;
   repeated PER / sub=PAT type=unr;
```

and a treatment-heterogeneous variance (4) can be modelled with the statements

```
proc mixed;
   class PAT PER TRT;
(9)  model VAL = PER TRT;
   repeated TRT / sub=PAT type=unr;
```

These analyses, however, do not produce the same test for the treatment effect as the standard analysis. In fact, with a small number of patients the analysis (8) would be anti-conservative for testing the treatment effect. For example, if 10 patients are included in each treatment sequence and the data have a compound symmetry covariance structure (2) with a correlation coefficient of 0.5 then the type I error probability with analysis (8) is about 6.55% as I estimated in 10 000 simulations.

In order to obtain the same result as the standard analysis a sequence effect must be added to the MODEL statement, despite the fact that the sequence effect in the statistical model is zero (1). Thus, a possible syntax for the period-heterogeneous variance (3) is

```
proc mixed;
   class PAT SEQ PER TRT;
(10) model VAL = SEQ PER TRT;
   repeated PER / sub=PAT type=unr;
```

and for the treatment-heterogeneous variance (4) it is

```
proc mixed;
   class PAT SEQ PER TRT;
(11) model VAL = SEQ PER TRT;
   repeated TRT / sub=PAT type=unr;
```

To summarize, analyses (5), (6), (7), (10), and (11) yield the same test for the treatment effect as the standard analysis in case of complete data.

A popular option of PROC MIXED is the Kenward-Roger method for calculating the degrees of freedom. It is invoked by the option DDFM=KR added to the MODEL statement. This option is aimed at improving the finite sample performance of the restricted maximum likelihood method. The Kenward-Roger method adjusts the restricted maximum likelihood estimator of the standard error according to a method of Prasad-Rao-Jeske-Kackar-Harville and then uses Satterthwaite-type degrees of freedom (see [4]).

In the analyses (6) and (7) with compound symmetry covariance structure the Kenward-Roger option has no effect if all pairs of observations are complete. This does not hold if a heterogeneous covariance structure is specified as in analyses (8) to (11). Indeed, the standard error of the treatment estimate as modified by the Kenward-Roger method may depend on the parameterization of the covariance matrix. For example, TYPE=TOEPH generally yields a slightly different standard error than TYPE=FA0(2).

Finally I should note that the comparisons made in this section are based on empirical evidence only, i.e. I analysed some more or less extreme data sets and compared the results for various PROC MIXED specifications. I have not generally verified the equivalence of any two methods in theory. This is not crucial, though, because it does not affect the results of the simulation study presented in the next section.

4. ANALYSIS WITH INCOMPLETE OBSERVATIONS

4.1 DESIGN OF THE SIMULATION STUDY

I now turn to the analysis of data from a two-period crossover trial when some of the observations are missing. I assume that the observations are missing completely at random, i.e. the probability of a value being missing is independent of both the observed and the unobserved data (cf. Little and Rubin [2]). This may not be particularly realistic in many trials but it allows judging how the methods perform under favourable circumstances. I investigated the performance of the different methods in a small simulation study. Specifically, I simulated data according to the following arbitrarily selected scenarios.

Sample size:

- 10 patients for each treatment sequence, 2 of these with missing observations in the second period.
- 10 patients for treatment sequence *AB* with 4 missing observations in the second period.
  20 patients for treatment sequence *BA* with 8 missing observations in the second period.
- 10 patients for treatment sequence *AB* with 2 missings in the first and 3 missings in the second period.
  6 patients for treatment sequence *BA* with 1 missing in the first and 2 missings in the second period.

Fixed effects:

- All zero: μ=0, π=0, β=0.

Covariance structure:

- Compound symmetry covariance structure (2) with σ = 5 and ρ = 0.2, 0.5, 0.8.
- Period-heterogeneous covariance structure (3) with $\sigma_1 = 2$, $\sigma_2 = 10$ or $\sigma_1 = 10$, $\sigma_2 = 2$;
  each combination with ρ = 0.2, 0.5, 0.8.
- Treatment-heterogeneous covariance structure (4) with $\sigma_A = 2$, $\sigma_B = 10$ or $\sigma_A = 10$, $\sigma_B = 2$;
  each combination with ρ = 0.2, 0.5, 0.8.

The three sample size configurations combined with 15 covariance structures make a total of 45 scenarios. For each scenario I generated 10 000 data sets. Thirteen different analyses were applied to all 450 000 data sets as detailed below.

If there are missing data the fixed effects model (5) is equivalent to the standard analysis applied to the complete pairs only. That is to say, if the patient effect is fixed it cannot be estimated from the data of other patients. Thus only patients who provide data for both treatments contribute information to the treatment effect.

In addition to the fixed effects model, I applied the mixed models specified in the previous section. Instead of the TYPE=CS option, however, I employed the equivalent TYPE=TOEP option for the compound symmetry covariance structure (2), and instead of the TYPE=UNR option I employed the equivalent TYPE=TOEPH option for the heterogeneous covariance structures (3) and (4).

With the PROC MIXED specifications (6) to (11) the default value for the DDFM option of the MODEL statement is DDFM=BETWITHIN or synonymously DDFM=BW. In this case the number of degrees of freedom calculated by PROC MIXED is equal to the number of complete pairs minus two. These are the same degrees of freedom as obtained with the analysis of variance model (5) and the same as used in the complete case analysis with the two-sample *t*-test. In addition to this default method I investigated the effect of the Kenward-Roger option. To summarize, I ran the following thirteen analyses for each simulated data set.

Analyses:

- Fixed effects model (5), equivalent to the standard analysis of the complete cases only.
- Compound symmetry covariance structure without sequence effect (6) and with sequence effect (7);
  both models with the DDFM=BW and the DDFM=KR option.
- Period-heterogeneous covariance structure without sequence effect (8) and with sequence effect (10);
  both models with the DDFM=BW and the DDFM=KR option.
- Treatment-heterogeneous covariance structure without sequence effect (9) and with sequence effect (11);
  both models with the DDFM=BW and the DDFM=KR option.

To compare the power of the different methods I shifted the estimated treatment effect of each analysis by some hypothesized effect δ and calculated the resulting *p*-value.

4.1 RESULTS OF THE SIMULATION STUDY

Within the simulation study I ran PROC MIXED 5.4 million times. In 136 cases the procedure did not converge, a failure rate of 25 cases per million. The maximum number of convergence failures for any scenario of 10 000 data sets was 8. Because of the small number of failures I did not try to achieve convergence in these cases but rather ignored them for that particular analysis.

Type I error probability

The results of the simulations regarding the probability of making a type I error are summarized in Tables 1 and 2. Table 1 refers to analyses with the DDFM option equal to BW (the default of PROC MIXED for the analyses considered here). Table 2 shows the observed rejection rates when the Kenward-Roger method is invoked with DDFM=KR. To facilitate the interpretation of the results I shaded the observed rejection rates below 5.5%. Observed rejection rates are shaded in bright yellow if below 4.0% and in dark yellow otherwise.

Table 1 also shows the observed rejection rates for the fixed effects model (5) although this method is known to be theoretically exact (cf. first paragraph of Section 3). Thus the somewhat high rejection rate of 5.63% in the centre column of the first row in Table 1 must be an artifact. Indeed, it shows the limitations of a simulation size of 10 000.

Table 1: Estimated type I error probability (%) for testing the treatment effect at a 5% level using the DDFM=BW option. Each estimate is based on 10 000 simulated data sets.

Analysis			Sample size [a]											
			AB: 10 (0,2)			*AB*: 10 (0,4)			*AB*: 10 (2,3)					
Sequence effect [b]	Vari-ance [c]	Data variance [d]	*BA*: 10 (0,2)			*BA*: 20 (0,8)			*BA*: 6 (1,2)				Mini-	Maxi-
			ρ=0.2	ρ=0.5	ρ=0.8	ρ=0.2	ρ=0.5	ρ=0.8	ρ=0.2	ρ=0.5	ρ=0.8	Mean	mum	mum
No or yes (5)	Indep. fixed PAT *	CS (5)	5.11	4.87	4.94	5.08	5.63	5.03	5.30	5.13	5.32	5.16	4.87	5.63
		PER (2, 10)	4.87	4.68	4.93	4.96	5.05	5.46	5.10	5.29	4.75	5.01	4.68	5.46
		PER (10, 2)	5.33	5.21	5.15	4.97	4.89	4.93	5.03	5.04	4.93	5.05	4.89	5.33
		TRT (2, 10)	5.13	5.00	4.88	5.03	4.49	5.10	4.83	5.07	5.12	4.96	4.49	5.13
		TRT (10, 2)	4.94	4.69	4.86	4.71	5.01	5.21	5.15	5.17	5.46	5.02	4.69	5.46
No (6)	CS	CS (5)	5.01	4.66	5.04	5.18	4.97	4.96	4.48	4.86	5.37	4.95	4.48	5.37
		PER (2, 10)	4.72	4.84	4.90	4.90	4.99	5.39	4.31	4.59	4.39	4.78	4.31	5.39
		PER (10, 2)	4.97	4.76	5.21	5.01	5.12	4.87	4.60	4.66	4.50	4.86	4.50	5.21
		TRT (2, 10)	4.89	4.71	4.76	3.43	3.33	4.09	3.90	4.75	4.57	4.27	3.33	4.89
		TRT (10, 2)	4.98	4.64	4.69	6.45	6.79	6.19	4.02	4.35	4.97	5.23	4.02	6.79
No (8)	PER	CS (5)	6.93	6.34	6.78	6.29	6.31	6.20	7.11	8.23	9.17	7.04	6.20	9.17
		PER (2, 10)	6.02	6.72	6.61	5.65	5.87	6.11	6.75	7.31	8.44	6.61	5.65	8.44
		PER (10, 2)	7.01	7.27	6.52	6.35	6.21	6.70	7.65	8.34	8.57	7.18	6.21	8.57
		TRT (2, 10)	8.04	7.46	8.08	13.01	12.68	14.89	10.94	12.71	13.82	11.29	7.46	14.89
		TRT (10, 2)	8.23	7.77	7.99	13.78	14.84	14.84	13.58	14.48	16.64	12.46	7.77	16.64
No (10)	TRT	CS (5)	5.02	4.87	5.22	5.37	5.22	5.26	4.72	5.45	6.29	5.27	4.72	6.29
		PER (2, 10)	4.57	4.56	4.76	7.92	7.42	8.49	5.05	5.49	6.13	6.04	4.56	8.49
		PER (10, 2)	4.86	4.69	5.11	7.58	7.71	7.75	5.45	5.66	5.77	6.06	4.69	7.75
		TRT (2, 10)	4.98	4.56	4.52	4.24	4.23	4.55	3.77	4.11	3.67	4.29	3.67	4.98
		TRT (10, 2)	4.85	4.48	4.63	4.85	5.11	4.60	3.90	3.79	3.66	4.43	3.66	5.11
Yes (7)	CS	CS (5)	4.97	4.82	5.14	5.30	5.27	4.99	4.62	4.96	5.42	5.05	4.62	5.42
		PER (2, 10)	7.41	7.06	7.44	12.11	11.32	11.90	5.63	6.06	6.00	8.33	5.63	12.11
		PER (10, 2)	3.08	3.07	3.59	1.58	1.99	1.92	3.33	3.66	3.56	2.86	1.58	3.66
		TRT (2, 10)	5.05	4.46	4.78	5.04	4.55	5.33	3.58	4.48	4.41	4.63	3.58	5.33
		TRT (10, 2)	4.90	4.73	4.66	4.63	5.02	4.45	4.43	4.83	5.52	4.80	4.43	5.52
Yes (9)	PER	CS (5)	5.13	4.94	5.38	5.59	5.49	5.39	4.82	5.57	6.18	5.39	4.82	6.18
		PER (2, 10)	5.05	4.98	5.11	5.64	5.44	5.22	4.62	4.59	4.17	4.98	4.17	5.64
		PER (10, 2)	4.83	4.49	4.67	4.49	4.50	4.33	3.92	3.86	3.33	4.27	3.33	4.83
		TRT (2, 10)	5.00	4.54	4.84	7.04	6.47	7.51	3.33	4.45	4.74	5.32	3.33	7.51
		TRT (10, 2)	4.90	4.66	4.68	3.63	4.15	4.40	4.77	5.46	5.98	4.74	3.63	5.98
Yes (11)	TRT	CS (5)	5.03	4.89	5.31	5.34	5.40	5.31	4.75	5.49	6.21	5.30	4.75	6.21
		PER (2, 10)	7.42	7.10	7.48	12.11	11.46	12.08	5.69	6.06	6.61	8.45	5.69	12.11
		PER (10, 2)	3.12	3.12	3.61	1.67	2.34	2.65	3.31	3.63	3.90	3.04	1.67	3.90
		TRT (2, 10)	5.12	4.56	4.66	5.00	4.21	4.77	3.83	4.36	3.57	4.45	3.57	5.12
		TRT (10, 2)	5.06	4.63	4.56	4.84	5.31	4.74	4.16	4.13	3.88	4.59	3.88	5.31

a) *AB: 10 (0,2) indicates that treatment sequence AB includes 10 patients, no observations are missing in period 1 and 2 observations are missing in period 2; analogous definitions apply to the other sample size specifications.*

b) *The numbers in round brackets refers to the analysis specifications in Section 3.*

c) *CS=compound symmetry, PER=period-heterogeneous, TRT= treatment-heterogeneous; * =analysis of variance with fixed effects PAT TRT PER; equivalent to complete case analysis.*

d) *CS(5) indicates a compound symmetry covariance structure in the data as in (2) with σ=5. PER(2,10) indicates a period-heterogeneous covariance structure in the data as in (3) with σ_1=2 and σ_2=10, TRT(2,10) indicates a treatment-heterogeneous covariance structure in the data as in (4) with σ_A=2 and σ_B=10, etc.*

Table 2: Estimated type I error probability (%) for testing the treatment effect at a 5% level using the Kenward-Roger option. Each estimate is based on 10 000 simulated data sets.

Analysis			Sample size [a]											
			AB: 10 (0,2)			*AB*: 10 (0,4)			*AB*: 10 (2,3)					
Sequence effect [b]	Vari-ance [c]	Data variance [d]	*BA*: 10 (0,2)			*BA*: 20 (0,8)			*BA*: 6 (1,2)					
			ρ=0.2	ρ=0.5	ρ=0.8	ρ=0.2	ρ=0.5	ρ=0.8	ρ=0.2	ρ=0.5	ρ=0.8	Mean	Mini-mum	Maxi-mum
No (6)	CS	CS (5)	5.10	4.71	5.04	5.38	5.03	4.91	5.13	5.10	5.37	5.09	4.71	5.38
		PER (2, 10)	5.05	5.08	5.22	5.47	5.45	5.86	5.32	5.34	5.11	5.32	5.05	5.86
		PER (10, 2)	4.89	4.60	5.05	4.77	4.79	4.57	4.91	5.02	4.71	4.81	4.57	5.05
		TRT (2, 10)	5.02	4.74	4.80	3.39	3.25	3.99	4.55	5.20	4.98	4.44	3.25	5.20
		TRT (10, 2)	5.11	4.60	4.71	7.00	7.04	6.39	4.76	4.86	5.29	5.53	4.60	7.04
No (8)	PER	CS (5)	5.27	4.88	5.12	5.51	5.04	4.95	5.89	5.98	6.32	5.44	4.88	6.32
		PER (2, 10)	4.87	5.32	5.09	5.38	5.22	4.94	5.82	5.71	5.84	5.36	4.87	5.84
		PER (10, 2)	5.39	5.43	4.91	5.12	4.69	5.31	5.90	6.07	5.77	5.40	4.69	6.07
		TRT (2, 10)	6.60	6.05	6.45	11.45	11.29	13.12	9.82	10.62	11.13	9.61	6.05	13.12
		TRT (10, 2)	6.76	6.28	6.29	12.97	13.79	13.24	11.91	12.30	13.56	10.79	6.28	13.79
No (10)	TRT	CS (5)	5.04	4.64	4.79	5.43	5.07	4.88	5.16	5.36	5.73	5.12	4.64	5.73
		PER (2, 10)	4.69	4.71	4.81	8.40	7.63	8.69	5.82	5.99	6.48	6.36	4.69	8.69
		PER (10, 2)	4.67	4.30	4.74	7.17	7.14	7.08	5.79	5.98	5.28	5.79	4.30	7.17
		TRT (2, 10)	5.14	4.68	4.78	4.57	4.63	5.11	4.84	5.36	5.26	4.93	4.57	5.36
		TRT (10, 2)	5.01	4.66	4.91	4.95	5.17	4.92	5.15	4.70	5.06	4.95	4.66	5.17
Yes (7)	CS	CS (5)	5.06	4.84	5.15	5.35	5.27	4.95	5.13	5.27	5.37	5.15	4.84	5.37
		PER (2, 10)	7.82	7.37	7.77	12.98	12.17	12.82	7.06	7.23	7.03	9.14	7.03	12.98
		PER (10, 2)	2.96	2.86	3.35	1.38	1.76	1.68	3.49	3.68	3.50	2.74	1.38	3.68
		TRT (2, 10)	5.02	4.46	4.82	4.88	4.42	5.15	4.15	4.84	4.76	4.72	4.15	5.15
		TRT (10, 2)	5.03	4.76	4.64	4.88	5.20	4.51	4.98	5.34	5.68	5.00	4.51	5.68
Yes (9)	PER	CS (5)	5.04	4.68	4.82	5.50	5.05	4.94	5.02	5.35	5.54	5.10	4.68	5.54
		PER (2, 10)	4.87	4.71	5.07	5.05	5.02	5.28	5.20	5.14	5.28	5.07	4.71	5.28
		PER (10, 2)	5.16	4.89	4.99	5.10	5.38	5.04	5.25	4.93	4.72	5.05	4.72	5.38
		TRT (2, 10)	4.97	4.36	4.72	7.34	6.71	7.63	3.59	4.35	4.27	5.33	3.59	7.63
		TRT (10, 2)	4.88	4.55	4.55	3.31	3.64	3.81	5.61	6.02	6.25	4.74	3.31	6.25
Yes (11)	TRT	CS (5)	4.99	4.65	4.84	5.33	5.11	4.88	4.97	5.35	5.56	5.08	4.65	5.56
		PER (2, 10)	7.73	7.33	7.69	12.82	12.05	12.64	7.07	7.14	7.27	9.08	7.07	12.82
		PER (10, 2)	2.92	2.74	3.17	1.45	1.97	2.28	3.13	3.31	3.17	2.68	1.45	3.31
		TRT (2, 10)	5.15	4.57	4.95	5.18	4.40	5.23	4.78	5.38	5.13	4.98	4.40	5.38
		TRT (10, 2)	5.15	4.65	4.73	4.94	5.47	5.02	5.23	4.82	5.17	5.02	4.65	5.47

a) AB: 10 (0,2) indicates that treatment sequence AB includes 10 patients, no observations are missing in period 1 and 2 observations are missing in period 2; analogous definitions apply to the other sample size specifications.

b) The numbers in round brackets refers to the analysis specifications in Section 3; the option DDFM=KR has to be added to the model statement of PROC MIXED.

c) CS=compound symmetry, PER=period-heterogeneous, TRT= treatment-heterogeneous.

d) CS(5) indicates a compound symmetry covariance structure in the data as in (2) with σ=5. PER(2,10) indicates a period-heterogeneous covariance structure in the data as in (3) with σ_1=2 and σ_2=10, TRT(2,10) indicates a treatment-heterogeneous covariance structure in the data as in (4) with σ_A=2 and σ_B=10, etc.

When the data have a compound symmetry covariance structure and the mixed model analysis also specifies a compound symmetry type of covariance matrix (models (6) and (7)) then the test for a treatment effect complies very well with its nominal size of 5%. For comparison I note that with 10 000 simulations the two-sided 99% Clopper-Pearson confidence interval for an observed rejection rate of 5% extends from 4.46% to 5.59% and it extends from 4.93% to 6.11% for an observed rejection rate of 5.5%.

Further, if the data have a compound symmetry covariance structure and the mixed model analysis also specifies a compound symmetry type of covariance matrix then the Kenward-Roger option (Table 2, models (6) and (7)) yielded very similar rejection rates as the default method (Table 1, models (6) and (7)). There is also no obvious difference whether a sequence effect is included, as in analysis (7), or not, as in analysis (6).

The analysis with a compound symmetry covariance matrix even copes well with a period-heterogeneous covariance structure in the data provided a sequence effect is not specified in the analysis and the default method BW is used for the degrees of freedom (Table 1, model (6) for covariance structures PER(2,10) and PER(10,2)). This does not hold if a sequence effect is specified (Table1, model (7) for the same covariance structures), in which case the test performs actually poor.

By contrast, if the data have a treatment-heterogeneous covariance structure and are being analysed by a compound symmetry covariance matrix then it would be better to specify a sequence effect (Table 1, model (7) for covariance structures TRT(2,10) and TRT(10,2)), although the omission of the sequence effect would not be too bad (Table 1, model (6)). In addition, the slightly elevated rejection rates above 6.0% for model (6) in Table 1 presumably vanish if the sample size and the number of missing observations are similar in the two sequence groups. For reasons of symmetry the rejection rates for the covariance structures TRT(2,10) and TRT(10,2) are the same if the sample size and the number of missing observations is equal in the two sequence groups.

Thus, an analysis without sequence effect and compound symmetry covariance matrix is a good choice if the true covariance structure is not too heterogeneous and if the sample sizes of treatment sequences and the number of missing values under the two treatments are similar.

On the other hand, if it is known that there is a certain degree of heterogeneity then one may want to take advantage of this fact by specifying the covariance matrix in the analysis accordingly. In this case, however, the Kenward-Roger option provides better agreement with the nominal test size than the default method. In case of period-heterogeneity a sequence effect should definitely be specified. Thus, for the extreme period-heterogeneity PER(2,10) or PER(10,2) the best choice is model (9) in Table 2 (compared to model (8) in that table or to models (9) and (8) in Table 1). For the extreme treatment-heterogeneity TRT(2,10) or TRT(10,2) the best choice is model (10) or model (11) in Table 2 (compared to the same models in Table 1).

Power

To get an idea of the difference in power between the various methods I have calculated the power for a single value δ of the treatment effect. I did not run additional simulations for the power but used the same data as above generated under the null hypothesis of no treatment effect. I shifted the estimated treatment effect by the hypothesized treatment effect δ and calculated the resulting *p*-value. In Table 3 the estimated power is shown only for those methods and scenarios that comply well with the nominal test size of 5%.

Table 3: Estimated power (%) for testing the treatment effect at a 5% level given the true treatment effect is δ. Each estimate is based on 10000 simulated data sets.

Analysis					Sample size [a]									
					AB: 10 (0,2)			*AB*: 10 (0,4)			*AB*: 10 (2,3)			
Sequence effect [b]	DDFM option	Variance [c]	Data variance [d]	Delta δ	*BA*: 10 (0,2)			*BA*: 20 (0,8)			*BA*: 6 (1,2)			
					ρ=0.2	ρ=0.5	ρ=0.8	ρ=0.2	ρ=0.5	ρ=0.8	ρ=0.2	ρ=0.5	ρ=0.8	Mean
No or	BW	Indep.	CS (5)	3	57	73	98	57	75	98	31	42	75	67
yes		fixed	PER (2, 10)	5	62	67	73	62	67	73	34	37	41	57
(5)		PAT	PER (10, 2)	5	61	66	72	62	66	74	34	38	41	57
		*	TRT (2, 10)	5	62	67	73	62	67	72	34	38	42	57
			TRT (10, 2)	5	61	66	72	62	67	73	34	37	41	57
No	BW	CS	CS (5)	3	60	76	98	65	80	98	39	49	78	71
(6)			PER (2, 10)	5	71	74	79	81	84	87	47	49	52	69
			PER (10, 2)	5	62	67	73	64	70	76	41	44	48	61
			TRT (2, 10)	5	66	71	76	71	76	80	42	47	51	64
			TRT (10, 2)	5	65	70	76	_	_	_	44	47	51	65
No	BW	TRT	TRT (2, 10)	5	67	72	79	80	84	90	49	53	63	71
(10)			TRT (10, 2)	5	66	71	79	73	80	88	49	54	63	69
Yes	KR	PER	PER (2, 10)	5	61	67	76	60	68	81	43	47	57	62
(9)			PER (10, 2)	5	70	75	81	81	85	91	49	54	63	72
Yes	KR	TRT	TRT (2, 10)	5	66	72	79	68	74	84	47	52	62	67
(11)			TRT (10, 2)	5	65	71	79	73	79	87	45	50	60	68

a) AB: 10 (0,2) indicates that treatment sequence AB includes 10 patients, no observations are missing in period 1 and 2 observations are missing in period 2; analogous definitions apply to the other sample size specifications.

b) The number in round brackets refers to the analysis specifications in Section 2´3; the option DDFM=KR has to be added in the model statement of PROC MIXED as indicated in the second column.

*c) CS=compound symmetry, PER=period-heterogeneous, TRT= treatment-heterogeneous. * =analysis of variance with fixed effects PAT TRT PER; equivalent to complete case analysis.*

d) CS(5) indicates a compound symmetry covariance structure as in (2) with σ=5. PER(2,10) indicates a period-heterogeneous covariance structure in the data as in (3) with σ_1=2 and σ_2=10, TRT(2,10) indicates a treatment-heterogeneous covariance structure in the data as in (4) with σ_A=2 and σ_B=10, etc.

In the following discussion I consider the first two sample size scenarios only where all missing observations occur in the second period. If the data have a compound symmetry covariance structure and only few observations are missing then the power of the fixed effects model (5) is only slightly lower than the power of the mixed model (6). For example with 2 of 10 observations missing in the second period and a correlation coefficient of 0.5 the power at δ=5 is 73% with the fixed effects model and 76% with the mixed model. The difference in power is similar if the covariance structure is heterogeneous and the compound symmetry model (6) is employed. More power can generally be gained in case of a heterogeneous variance if the appropriate covariance structure is specified in the analysis. There is one notable exception, though. In case of a period-heterogeneous variance where the variance in the second period is much larger than in the first period the analysis with a compound symmetry covariance matrix has higher power than the analysis with the correctly specified heterogeneous covariance matrix: compare the power of 74% for the covariance structure PER(2,10) in the second row of analysis (6) at ρ=0.5 with the power of 67% for the same data subjected to analysis (9).

Regarding the analysis of data with treatment-heterogeneous variance there is a striking contrast between analyses (10) and (11). The two methods have very similar power for the first and third sample size scenarios but for the second sample size scenario the power of analysis (10) is appreciably higher if the larger variance pertains to treatment *B*.

The few simulated examples show that power comparisons are crucially dependent on the balance of sample sizes and the pattern of missing values.

5. CONCLUSIONS

If data of a two-period crossover trial are missing completely at random then a good alternative to the complete case analysis is a restricted maximum likelihood analysis of a mixed model with fixed effects for period and treatment and a compound symmetry covariance matrix using the default degrees of freedom of PROC MIXED. This model seems to be robust against variance heterogeneity among treatments or periods, at least if the sample sizes of treatment sequences and the number of missing values under the two treatments are similar. In case of a compound symmetry covariance structure in the data the gain in power is only slight if the number of missing values is small, for example not more than two in ten.

Only if there is strong evidence for variance heterogeneity should the compound symmetry covariance matrix in the analysis replaced by a corresponding REPEATED effect with unstructured covariance matrix. The Kenward-Roger option should be applied in this case and a fixed sequence effect should be added to the model in order to keep the type I error probability at nominal level. The benefit would then be a further gain in power.

The above conclusions apply only to the scenarios investigated in this paper. The observed rejection rates and power values show a critically and sometimes unexpected dependence on sample size and missing value pattern. The ultimate recommendation is therefore to do a simulation study with the same sample sizes and missing value pattern as found in the data that are to be analysed.

REFERENCES

[1] Grieve AP, Senn SJ (1998): Estimating treatment effects in clinical crossover trials. *Journal of Biopharmaceutical Statistics,* 8: 191 - 233.

[2] Little, RJA, Rubin, DB (2002): *Statistical Analysis with Missing Data*. John Wiley & Sons Inc.

[3] Patel, HI (1985): Analysis of incomplete data in a two-period crossover design with reference to clinical trials. *Biometrika*, 72: 411-418.

[4] SAS Institute Inc. (1999): *SAS/STAT® User's Guide, Version 8*. Cary, NC, SAS Institute Inc., 1999.

[5] Senn, SJ (1993): *Cross-over Trials in Clinical Research*. John Wiley & Sons, Chichester.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Jochen Mueller-Cohrs
Accovion GmbH
Software Center 3
D-35037 Marburg, Germany
Work Phone: ++6421 9484927
Email: jochen.mueller-cohrs@accovion.com