Four-way cross-over trial fitting covariance patterns -continued - dropping out the carryover term

In the previous section, we saw that the estimates of the treatment and carryover differences were dependent on whether the period or treatment structure covariance pattern was fit. Even though the carryover effect was statistically significant with the treatment structure, it was not with the period structure. The authors of the textbook felt that the inclusion of washout periods made the carryover effect unlikely and thought the analyses should be rerun without the carryover effect. The following tables summarize the results of reanalyzing the data without carryover in the model.

Covariance Pattern	-2 log likelihood	param.	AIC	AICC	BIC
Compound Symmetry	544.6	2	548.6	548.8	549.9
Unstructured - Period	528.2	10	548.2	553.9	554.5
Heterogeneous Toeplitz - Period	529.1	7	543.1	545.8	547.5
Unstructured - Treatment	527.4	10	547.4	553.2	553.8

The heterogeneous Toeplitz and two unstructured covariance patterns all have similar fits. Because the heterogeneous Toeplitz covariance pattern has fewer parameters that need to be estimated, it is the best choice for modeling the covariance structure. Following are the variances and correlations for the four models.

	Variance	Correlations			
Period	**Compound symmetry**				
1	3580.29	1			
2		0.4902	1		
3		0.4902	0.4902	1	
4		0.4902	0.4902	0.4902	1
Period	**Unstructured**				
1	7390.17	1			
2	2182.93	0.7505	1		
3	3720.95	0.4449	0.6121	1	
4	1275.33	0.4336	0.4316	0.5796	1
Period	**Heterogeneous Toeplitz**				
1	6756.49	1			
2	2091.87	0.6515	1		
3	4038.55	0.4477	0.6515	1	
4	1369.32	0.4390	0.4477	0.6515	1
Treatment	**Unstructured**				
A	2060.79	1			
B	6128.90	0.2017	1		
C	3441.08	0.3882	0.8188	1	
D	2780.88	0.1347	0.5561	0.7613	1

We see the same patterns that we saw when the crossover effect was included in the model, where in the period structures the variances in the first period are much greater than in the subsequent periods. For the treatment structure model, Treatment B, which is the placebo, again has greater variation than any of the three active treatments.

Finally, we can compare the estimated treatment differences from the three models.

Unstructured (Period structure):

Differences of Least Squares Means

Effect	treata	period	_treata	_period	Estimate	Standard Error	DF	t Value	Pr > |t|
treata	A		B		45.9105	11.8378	25.8	3.88	0.0006
treata	A		C		47.9029	11.7787	26.2	4.07	0.0004
treata	A		D		70.5208	11.5906	25.6	6.08	<.0001
treata	B		C		1.9924	11.7581	26.6	0.17	0.8667
treata	B		D		24.6103	11.6102	25.6	2.12	0.0439
treata	C		D		22.6179	11.2355	25.3	2.01	0.0549

Heterogeneous Toeplitz (Period Structure):

Differences of Least Squares Means

Effect	treata	period	_treata	_period	Estimate	Standard Error	DF	t Value	Pr > |t|
treata	A		B		43.9077	11.9391	26.4	3.68	0.0571
treata	A		C		46.5811	12.0831	29.7	3.86	0.0006
treata	A		D		69.2389	11.6293	26.8	5.95	<.0001
treata	B		C		2.6734	11.9812	28.2	0.22	0.8250
treata	B		D		25.3311	11.7215	27.8	2.16	0.0395
treata	C		D		22.6578	11.4086	27.5	1.99	0.0571

Unstructured (Treatment Structure):

Differences of Least Squares Means

Effect	treata	period	_treata	_period	Estimate	Standard Error	DF	t Value	Pr > |t|
treata	A		B		43.1429	21.9679	12.6	1.96	0.0719
treata	A		C		24.4173	15.7255	12.5	1.55	0.1455
treata	A		D		68.9882	17.3586	10.7	3.97	0.0023
treata	B		C		-18.7255	12.1806	11.5	-1.54	0.1512
treata	B		D		25.8454	17.6056	11.3	1.47	0.1693
treata	C		D		44.5709	10.5189	11.1	4.24	0.0014

First, comparing the differences in the treatments among the two period structures and the treatment structure, we notice differences in the levels of significance between the treatment comparisons for the two period structures and the treatment structure. Note that the estimates for the unstructured and heterogeneous Toeplitz are similar. Although it appears that the heterogeneous Toeplitz would be the best choice for fitting the covariance pattern, the others of the textbook did not explore this variance structure. They felt that because the unstructured for treatment structure had a smaller –2 log likelihood than the unstructured for period structure and the relatively low correlations involving treatment A, that the unstructured for treatment structure was the most plausible model.

Analysis of Categorical Data from Crossover designs

The last area that we will briefly cover in crossover designs is when the data are not normally distributed. Many of the same issues that we saw with continuous data are encountered with categorical data. The authors of the text note that, except for the case of binary data, purely categorical response variables that don't have an underlying scale are extremely rare and the standard statistical packages (i.e. SAS) currently don't have procedures available to handle them. We will go through the example from the text of a two-treatment, two-period, two-sequence binary crossover trial.

The data are originally from Jones and Kenward (1989), with some observations missing in the second period. The data are from a trial on cerebrovascular insufficiency, where the response variable was 1=normal ECG and 0=abnormal ECG. Note that the results that are presented here do not agree with the results presented in the textbook. The dataset that they provided for the problem is not the same one that they used for the example in the book. The following SAS code reads in the data and creates the sequence and outcome variables.

```
DATA a; INFILE 'i:\kathy mixed model\ex78.dat';
INPUT patient period treat $ outcome;
one=1;
RUN;
data table1;
set a;
proc sort;
by patient period;

data table1;
attrib per1 length=$1 per2 length=$1 outcomes length=$2;
retain seq per1 outcome2;
set table1;
by patient period;
if first.patient then do;
   if treat='A' then seq='AB';
   else seq='BA';
   per1=outcome;
end;
else do;
   per2=outcome;
   outcomes=per1||per2;
   output;
end;
run;
```

Note that for summarizing the data for the table, we need to create an "outcomes" variable that is the outcomes from Period 1 and Period 2 concatenated together. With the missing observations in the second period, this leads to six possible outcomes. The observations for the first five patients in each sequence are printed

Obs	patient	seq	outcomes
1	1	BA	00
2	2	BA	00
3	3	BA	00

```
 4        4       BA        00
 5        5       BA        00
51       51       AB        00
52       52       AB        00
53       53       AB        00
54       54       AB        00
55       55       AB        0.
```

The following SAS code summarizes the data

```
proc freq;
table seq*outcomes;
run;
```

which are presented in the following table.

	Outcomes						
Sequence	**(0,0)**	**(0,1)**	**(0,.)**	**(1,0)**	**(1,1)**	**(1,.)**	**Total**
AB	9	5	5	5	26	0	50
BA	12	1	1	6	26	4	50
Total	21	6	6	11	52	4	100

Prescott's Test

When the patient is treated as a fixed effect, then only those patients that have completed both periods are included in the analysis. The most powerful test for testing the null hypothesis of no treatment difference is the Prescott's Test (Prescott, 1981). The data need to be recoded to a change in score. The following SAS code performs the necessary recoding and summarization.

```
data prescott;
retain seq per1;
set prescott;
by patient period;
if first.patient then do;
   if treat='A' then seq='AB';
   else seq='BA';
   per1=outcome;
end;
else do;
   changes=per1-outcome;
   output;
end;
run;
proc sort;
by seq;
proc freq;
table seq*changes;
run;
```

The following table summarizes the change in score for the 90 patients who completed both periods of the study.

	Change Score			
Sequence	**-1**	**0**	**1**	**Total**
AB	5	35	5	45
BA	1	38	6	45
Total	6	73	11	90

Values of the change scores are compared between treatment groups by using the exact option in proc freq.

```
proc freq data=prescott;
table seq*changes/exact;
run;
```

The following output is produced:

```
                 Statistics for Table of seq by changes

 Statistic                     DF       Value      Prob
ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
Chi-Square                     2       2.8809    0.2368
Likelihood Ratio Chi-Square    2       3.1254    0.2096
Mantel-Haenszel Chi-Square     1       1.4784    0.2240
Phi Coefficient                        0.1789
Contingency Coefficient                0.1761
Cramer's V                             0.1789
  WARNING: 33% of the cells have expected counts less
          than 5. Chi-Square may not be a valid test.

                   Fisher's Exact Test
          ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
          Table Probability (P)       0.0221
          Pr <= P                     0.3113
               Effective Sample Size = 90
                Frequency Missing = 10
```

The p-value is .31. The Prescott test tells us about whether or not there is a significant difference between the treatment effects, but does not give us any estimate on the size of the treatment effect. We can use a mixed model to both recover the between-patient information and get estimates of the treatment effect. However, we have at most, only two observations per patient and a number of these patients have the same outcome for both observations, which can lead us to problems that we saw earlier when we had uniform random effect categories. If we use a repeated measures analysis, with a compound symmetry structure, we will avoid these problems. The linear model for this analysis is

$$\log(\mu_{ijk} / 1 - \mu_{ijk}) = a + t_i + p_j + s_k$$

a = intercept effect
t_i = treatment effect i

p_j = period effect j
s_k = random effect of patient ~N(0,R) where initial R is compound symmetry covariance pattern.

The following SAS code performs the analysis.

```
PROC GENMOD;
CLASS treat period patient;
MODEL outcome/one=period treat/ ERROR=B;
REPEATED SUBJECT=patient/ WITHIN=period TYPE=CS MODELSE CORRW;
RUN;
```

Note that we are dealing with a Bernoulli distribution, so we model the outcome variable over one.

We get the following output

```
                        The GENMOD Procedure
                         Model Information

                Data Set                        WORK.A
                Distribution                  Binomial
                Link Function                    Logit
                Response Variable (Events)     outcome
                Response Variable (Trials)         one
                Number of Observations Read        200
                Number of Observations Used        190
                Number of Events                   125
                Number of Trials                   190
                Missing Values                      10

                      Class Level Information
Class      Levels  Values
treat           2  A B
period          2  1 2
patient       100  1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
                   21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37
                   ...

                     Parameter Information
             Parameter       Effect       treat    period
             Prm1            Intercept
             Prm2            period               1
             Prm3            period               2
             Prm4            treat        A
             Prm5            treat        B
```

The model information lets us know that we are using the logit link and the number of observations, events and trials are what we expect. Note that there are five parameters that we are estimating; the intercept, two periods and two treatments.

```
               Criteria For Assessing Goodness Of Fit

     Criterion                 DF           Value        Value/DF

     Deviance                 187        242.0829          1.2946
     Scaled Deviance          187        242.0829          1.2946
     Pearson Chi-Square       187        189.9881          1.0160
     Scaled Pearson X2        187        189.9881          1.0160
     Log Likelihood                     -121.0414
```

```
Algorithm converged.

                        GEE Model Information

          Correlation Structure                    Exchangeable
          Within-Subject Effect                period (2 levels)
          Subject Effect                     patient (100 levels)
          Number of Clusters                                  100
          Clusters With Missing Values                         10
          Correlation Matrix Dimension                          2
          Maximum Cluster Size                                  2
          Minimum Cluster Size                                  1

Algorithm converged.
                      Working Correlation Matrix

                                   Col1         Col2
                      Row1       1.0000       0.5851
                      Row2       0.5851       1.0000

                         Exchangeable Working
                              Correlation

                       Correlation    0.5850903245
```

This large correlation indicates that the observations within a patient are highly correlated.

```
                  Analysis Of GEE Parameter Estimates
                   Empirical Standard Error Estimates

                         Standard   95% Confidence
   Parameter    Estimate    Error       Limits            Z Pr > |Z|

   Intercept     0.6842   0.2484   0.1973   1.1712    2.75   0.0059
   period    1   0.1901   0.1976  -0.1972   0.5774    0.96   0.3360
   period    2   0.0000   0.0000   0.0000   0.0000     .      .
   treat     A  -0.3213   0.1979  -0.7091   0.0665   -1.62   0.1044
   treat     B   0.0000   0.0000   0.0000   0.0000     .      .
```

```
               Analysis Of GEE Parameter Estimates
               Model-Based Standard Error Estimates

                        Standard   95% Confidence
   Parameter   Estimate    Error       Limits            Z Pr > |Z|

   Intercept     0.6842   0.2418   0.2103   1.1581    2.83   0.0047
   period    1   0.1901   0.2001  -0.2020   0.5822    0.95   0.3419
   period    2   0.0000   0.0000   0.0000   0.0000     .      .
   treat     A  -0.3213   0.2004  -0.7140   0.0715   -1.60   0.1089
   treat     B   0.0000   0.0000   0.0000   0.0000     .      .
   Scale         1.0013    .        .        .         .      .
NOTE: The scale parameter for GEE estimation was computed as the
      square root of the normalized Pearson's chi-square.
```

Our empirical and model-based standard error estimates are in fairly close agreement, indicating that we are modeling the appropriate distribution. There is not a significant period effect in this study. However, because of the design of the study, it should be kept in the analysis. The treatment effect is approaching significance. When we compare the p values from this analyses with the p value of .31 we obtained with the Prescott's Test, we can see how the mixed model analysis allows the recovery of information from patients with incomplete data.

We can build a 95% confidence interval for the odds ratio using the above estimate and 95% confidence limits

```
data b;
est=exp(-0.3213);
up=exp(0.0715);
low=exp(-0.7140);

run;
proc print;
run;
```

```
               Obs      ans         up        low

                 1    0.72521    1.07412    0.48968
```

So we have an estimated odds ratio of 0.73 of the proportion normal to the proportion abnormal in the two treatment groups, with 95% confidence limit of 0.49 to 1.07. Remember, that when one is included in the interval, then the odds ratio is not statistically significant.

General Points regarding Cross-Over Trials

The authors of the textbook close their chapter on cross-over trials reviewing some general points.

1) "Random effects models can have advantages over fixed effects models in the context of cross-over trials", especially if the data are very unbalanced or the patient variance component is small relative to the residual.
2) Random effects models are not always the best method, especially when dealing with non-normal data.

3) Although crossover trials are used extensively in earlier stage pharmaceutical trials, where the sample size is small, it is harder to get precise estimates of the variance components when mixed models are used. Therefore, it may be better in some cases to use fixed effect, rather than mixed effect models.
4) In models where carryover effects are being estimated, random effects models are better to use. However, the authors feel that Senn (1993) had expressed sound arguments against including simple carry-over terms in the model and list Senn's case at the end of Chapter 7 in the text.