MIXED MODELS FOR REPEATED (LONGITUDINAL) DATA

DAVID C. HOWELL   5/15/2008

When we have a design in which we have both random and fixed variables, we have what is often called a mixed model. Mixed models have begun to play an important role in statistical analysis and offer many advantages over more traditional analyses. At the same time they are more complex and the syntax for software analysis is not always easy to set up. My original plan was to put together a document that looked at many different kinds of designs and the way to use them. However I have decided that I can accomplish many of my goals by restricting myself to the analysis of repeated measures designs. This is a place where such models have important advantages. I will ignore the use of mixed models to handle nested factors (other than subjects) because that is an even more complicated system.

A large portion of this document has benefited from Chapter 15 in Maxwell & Delaney (2004) *Designing experiments and analyzing data*. They have one of the clearest discussions that I know. I am going a step beyond their example by including a between-groups factor as well as a within-subjects (repeated measures) factor. For now my purpose is to show the relationship between mixed models and the analysis of variance. The relationship is far from perfect, but it gives us a known place to start. More importantly, it allows us to see what we gain and what we lose by going to mixed models. In some ways I am going through the Maxwell & Delaney chapter backwards, because I am going to focus primarily on the use of the **repeated** command in SAS **Proc mixed.** I am doing that because it fits better with the transition from ANOVA to mixed models.

My motivation for this document came from a question asked by Rikard Wicksell at Karolinska University in Sweden. He had a randomized clinical trial with two treatment groups and measurements at pre, post, 3 months, and 6 months. His problem is that some of his data were missing. He considered a wide range of possible solutions, including "last trial carried forward," mean substitution, and listwise deletion. In some ways listwise deletion appealed most, but it would mean the loss of too much data. One of the nice things about mixed models is that we can use all of the data we have. If a score is missing, it is just missing. It has no effect on other scores from that same patient.

Another advantage of mixed models is that we don't have to be consistent about time. For example, and it does not apply in this particular example, if one subject had a follow-up test at 4 months while another had their follow-up test at 6 months, we simply enter 4 (or 6) as the time of follow-up. We don't have to worry that they couldn't be tested at the same intervals.

A third advantage of these models is that we do not have to assume sphericity or compound symmetry in the model. We can do so if we want, but we can also allow the model to select its own set of covariances or use covariance patterns that we supply. I will start by assuming sphericity because I want to show the parallels between the output from mixed models and the output from a standard repeated measures analysis of

variance. I will then delete a few scores and show what effect that has on the analysis. Finally I will use Expectation Maximization (EM) to impute missing values and then feed the newly complete data back into a repeated measures ANOVA to see how those results compare.

The Data

I have created data to have a number of characteristics. There are two groups – a Control group and a Treatment group, measured at 4 times. These times are labeled as 0 (pretest), 1 (one month posttest) 3 (3 months follow-up) and 6 (6 months follow-up). I created the treatment group to show a sharp drop at post-test and then sustain that drop (with slight regression) at 3 and 6 months. The Control group declines slowly over the 4 intervals but does not reach the low level of the Treatment group. There are noticeable individual differences in the Control group, and some subjects show a steeper slope than others. In the Treatment group there are individual differences in level but the slopes are not all that much different from one another. You might think of this as a study of depression, where the dependent variable is a depression score (e.g. Beck Depression Inventory) and the treatment is drug versus no drug. If the drug worked about as well for all subjects the slopes would be comparable and negative across time. For the control group we would expect some subjects to get better on their own and some to stay depressed, which would lead to differences in slope for that group. These facts are important because when we get to the random coefficient mixed model the individual differences will show up as variances in intercept, and any slope differences will show up as a significant variance in the slopes. For the standard ANOVA individual and for mixed models using the **repeated** command the differences in level show up as a Subject effect and we assume that the slopes are comparable across subjects.

Some of the printouts that follow were generated using SAS **Proc mixed**, but I give the SPSS commands as well. (I also give syntas for R, but I warn you that running this problem under R, even if you have Pinheiro & Bates (2000) is very difficult. I only give these commands for one analysis, but they are relatively easy to modify for related analyses.

The data follow. Notice that to set this up for ANOVA (**Proc** GLM) we read in the data one subject at a time. (You can see this is the data shown.) This will become important because we will not do that for mixed models.

Group	Subj	Time0	Time1	Time3	Time6
1	1	296	175	187	242
1	2	376	329	236	126
1	3	309	238	150	173
1	4	222	60	82	135
1	5	150	271	250	266
1	6	316	291	238	194
1	7	321	364	270	358
1	8	447	402	294	266
1	9	220	70	95	137
1	10	375	335	334	129
1	11	310	300	253	190
1	12	310	245	200	170

Group	Subj	Time0	Time1	Time3	Time6
2	13	282	186	225	134
2	14	317	31	85	120
2	15	362	104	144	114
2	16	338	132	91	77
2	17	263	94	141	142
2	18	138	38	16	95

2	19	329	62	62	6
2	20	292	139	104	184
2	21	275	94	135	137
2	22	150	48	20	85
2	23	319	68	67	12
2	24	300	138	114	174

A plot of the data follows:

Estimated Marginal Means of Dependent Variable
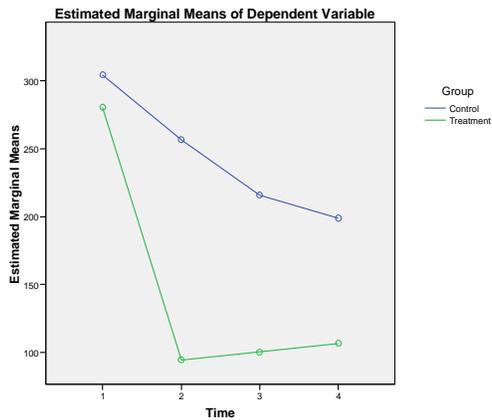
The cell means and standard deviations follow.

4. Group * Time

Measure: MEASURE_1

Group	Time	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Control	1	304.333	21.503	259.739	348.927
	2	256.667	24.061	206.768	306.566
	3	215.750	19.594	175.115	256.385
	4	197.167	18.552	158.692	235.641
Treatment	1	280.417	21.503	235.823	325.011
	2	94.500	24.061	44.601	144.399
	3	100.333	19.594	59.698	140.969
	4	106.667	18.552	68.192	145.141

Group means

Estimates

Measure: MEASURE_1

Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Control	243.896	16.405	209.874	277.918
Treatment	145.479	16.405	111.457	179.501

Grand Mean = 194.688

The results of a standard repeated measures analysis of variance with no missing data and using SAS **Proc GLM** follow. You would obtain the same results using the SPSS Univariate procedure.

```
proc GLM ;
   class group;
   model time1 time2 time3 time4 = group/ nouni;
   repeated time 4 (0, 1, 3, 6) polynomial /summary printm;
 run;
```

```
                            The GLM Procedure
                  Repeated Measures Analysis of Variance
              Tests of Hypotheses for Between Subjects Effects


    Source                DF     Type III SS      Mean Square     F Value    Pr > F
    group                  1     230496.0000      230496.0000      17.89     0.0003
    Error                 22     283514.9583       12887.0436



                            The GLM Procedure
                  Repeated Measures Analysis of Variance
             Univariate Tests of Hypotheses for Within Subject Effects


    Source          DF    Type III SS    Mean Square    F Value   Pr > F   G - G    H - F

    time             3    313917.7083    104639.2361     37.61    <.0001   <.0001   <.0001
    time*group       3     59791.7500     19930.5833      7.16    0.0003   0.0014   0.0007
    Error(time)     66    183603.5417      2781.8718


                    Greenhouse-Geisser Epsilon      0.7302
                    Huynh-Feldt Epsilon             0.8510
                            The GLM Procedure
                  Repeated Measures Analysis of Variance
              Analysis of Variance of Contrast Variables



Contrast Variable: time_1  The Linear Effect of Time (intervals = 0,1,2,3)
        Source                  DF     Type III SS    Mean Square    F Value   Pr > F
        Mean                     1     168155.6270    168155.6270     36.14    <.0001
        group                    1       1457.2401      1457.2401      0.31    0.5814
        Error                   22     102368.7996      4653.1273
```

```
Contrast Variable: time_2  The quadratic Effect of Time (intervals = 0,1,2,3)
     Source                   DF    Type III SS    Mean Square    F Value   Pr > F
     Mean                      1    96614.20130    96614.20130     48.47    <.0001
     group                     1    25235.52002    25235.52002     12.66    0.0018
     Error                    22    43851.09686     1993.23168


Contrast Variable: time_3
     Source                   DF    Type III SS    Mean Square    F Value   Pr > F
     Mean                      1    49147.88005    49147.88005     28.92    <.0001
     group                     1    33098.98990    33098.98990     19.48    0.0002
     Error                    22    37383.64520     1699.25660
```

Here we see that each of the effects in the overall analysis is significant. We don't care very much about the group effect because we expected both groups to start off equal at pre-test. What is important is the interaction, and it is significant at $p = .0003$. Clearly the drug treatment is having a differential effect on the two groups, which is what we wanted to see. The fact that the Control group seems to be dropping in the number of symptoms over time is to be expected and not exciting, although we could look at these simple effects if we wanted to. We would just run two analyses, one on each group. I would not suggest pooling the variances to calculate F, though that would be possible.

In the printout above I have included tests on linear, quadratic, and cubic trend that will be important later. However you have to read this differently than you might otherwise expect. The first test for the linear component shows an F of 36.14 for "mean" and an F of 0.31 for "group." Any other software that I have used would replace "mean" with "Time" and "group" with "Group $\times$ Time." In other words we have a significant linear trend over time, but the linear $\times$ group contrast is not significant. I don't know why they label them that way. (Well, I guess I do, but it's not the way that I would do it.) I should also note that my syntax specified the intervals for time, so that SAS is not assuming equally spaced intervals. The fact that the linear trend was not significant for the interaction means that both groups are showing about the same linear trend. But notice that there is a significant interaction for the quadratic.

Mixed Model

The use of mixed models represents a substantial difference from the traditional analysis of variance. For balanced designs the results will come out to be the same, assuming that we set the analysis up appropriately. But the actual statistical approach is quite different and ANOVA and mixed models will lead to different results whenever the data are not balanced or whenever we try to use different, and often more logical, covariance structures.

First a bit of theory. Within **Proc Mixed** the **repeated** command plays a very important role in that it allows you to specify different covariance structures, which is something

that you cannot do under **Proc GLM**. You should recall that in **Proc GLM** we assume that the covariance matrix meets our sphericity assumption and we go from there. In other words the calculations are carried out with the covariance matrix forced to sphericity. If that is not a valid assumption we are in trouble. Of course there are corrections due to Greenhouse and Geisser and Hyunh and Feldt, but they are not optimal solutions.

But what does compound symmetry, or sphericity, really represent? (The assumption is really about sphericity, but when speaking of mixed models most writers refer to compound symmetry, which is actually a bit more restrictive.) Most people know that compound symmetry means that the pattern of covariances or correlations is constant across trials. In other words, the correlation between trial 1 and trial 2 is equal to the correlation between trial 1 and trial 4 or trial 3 and trial 4, etc. But a more direct way to think about compound symmetry is to say that requires that all subjects in each group change in the same way over trials. In other words the slopes of the lines regressing the dependent variable on time are the same for all subjects. Put that way it is easy to see that compound symmetry can really be an unrealistic assumption. If some of your subjects improve but others don't, you do not have compound symmetry and make an error if you use a solution that assumes that you do. Fortunately **Proc Mixed** allows you to specify some other pattern for those covariances.

We can also get around the sphericity assumption using the MANOVA output from **Proc GLM**, but that too has its problems. Both standard univariate GLM and MANOVA GLM will insist on complete data. If a subject is missing even one piece of data, that subject is discarded. That is a problem because with a few missing observations we can lose a great deal of data and degrees of freedom.

Proc Mixed with **repeated** is different. Instead of using a least squares solution, which requires complete data, it uses a maximum likelihood solution, which does not make that assumption. (We will actually use a Restricted Maximum Likelihood (REML) solution.) When we have balanced data both least squares and REML will produce the same solution if we specify a covariance matrix with compound symmetry. But even with balanced data if we specify some other covariance matrix the solutions will differ. At first I am going to force sphericity by adding **type = cs** (which stands for compound symmetry) to the **repeated** statement. I will later relax that structure.

The first analysis below uses exactly the same data as for **Proc GLM**, though they are entered differently. Here data are entered in what is called "long form," as opposed to the "wide form" used for **Proc GLM**. This means that instead of having one line of data for each subject, we have one line of data for each observation. So with four measurement times we will have four lines of data for that subject.

Because we have a completely balanced design (equal sample sizes and no missing data) and because the time intervals are constant, the results of this analysis will come out exactly the same as those for **Proc GLM** so long as I specify **type = cs**. The data follow:

```sas
data WicksellLong;
input subj time group dv;
cards;
```

```
1   0  1.00 296.00          9   0  1.00 220.00          17   0  2.00 263.00
1   1  1.00 175.00          9   1  1.00  70.00          17   1  2.00  94.00
1   3  1.00 187.00          9   3  1.00  95.00          17   3  2.00 141.00
1   6  1.00 242.00          9   6  1.00 137.00          17   6  2.00 142.00
2   0  1.00 376.00          10  0  1.00 375.00          18   0  2.00 138.00
2   1  1.00 329.00          10  1  1.00 335.00          18   1  2.00  38.00
2   3  1.00 236.00          10  3  1.00 334.00          18   3  2.00  16.00
2   6  1.00 126.00          10  6  1.00 129.00          18   6  2.00  95.00
3   0  1.00 309.00          11  0  1.00 310.00          19   0  2.00 329.00
3   1  1.00 238.00          11  1  1.00 300.00          19   1  2.00  62.00
3   3  1.00 150.00          11  3  1.00 253.00          19   3  2.00  62.00
3   6  1.00 173.00          11  6  1.00 170.00          19   6  2.00   6.00
4   0  1.00 222.00          12  0  1.00 310.00          20   0  2.00 292.00
4   1  1.00  60.00          12  1  1.00 245.00          20   1  2.00 139.00
4   3  1.00  82.00          12  3  1.00 200.00          20   3  2.00 104.00
4   6  1.00 135.00          12  6  1.00 170.00          20   6  2.00 184.00
5   0  1.00 150.00          13  0  2.00 282.00          21   0  2.00 275.00
5   1  1.00 271.00          13  1  2.00 186.00          21   1  2.00  94.00
5   3  1.00 250.00          13  3  2.00 225.00          21   3  2.00 135.00
5   6  1.00 266.00          13  6  2.00 134.00          21   6  2.00 137.00
6   0  1.00 316.00          14  0  2.00 317.00          22   0  2.00 150.00
6   1  1.00 291.00          14  1  2.00  31.00          22   1  2.00  48.00
6   3  1.00 238.00          14  3  2.00  85.00          22   3  2.00  20.00
6   6  1.00 194.00          14  6  2.00 120.00          22   6  2.00  85.00
7   0  1.00 321.00          15  0  2.00 362.00          23   0  2.00 319.00
7   1  1.00 364.00          15  1  2.00 104.00          23   1  2.00  68.00
7   3  1.00 270.00          15  3  2.00 144.00          23   3  2.00  67.00
7   6  1.00 358.00          15  6  2.00 114.00          23   6  2.00  12.00
8   0  1.00 447.00          16  0  2.00 338.00          24   0  2.00 300.00
8   1  1.00 402.00          16  1  2.00 132.00          24   1  2.00 138.00
8   3  1.00 294.00          16  3  2.00  91.00          24   3  2.00 114.00
8   6  1.00 266.00          16  6  2.00  77.00          24   6  2.00 174.00
;
```

```sas
/* The following lines plot the data */
Symbol1 I = join v = none r = 12;
Proc gplot data = wicklong;
   Plot dv*time = subj/ nolegend;
   By group;
Run;

/* This is the main Proc Mixed procedure. */
Proc Mixed data = WicksellLong;
class group subj time;
model dv = group time group*time;
repeated time/subject = subj type = cs  rcorr;
run;
```

I have put the data in three columns to save space, but in SAS they would be entered as one long column.

The first set of commands plots the results of each individual subject broken down by groups. Earlier we saw the group means over time. Now we can see how each of the subjects stands relative the means of his or her group. In the ideal world the lines would start out at the same point on the *Y* axis (i.e. have a common intercept) and move in parallel (i.e. have a common slope). That isn't quite what happens here, but whether those are chance variations or systematic ones is something that we will look at later. We can

see in the Control group that a few subjects decline linearly over time and a few other subjects, especially those with lower scores decline at first and then increase during follow-up.

Plots (Group 1 = Control, Group 2 = Treatment)
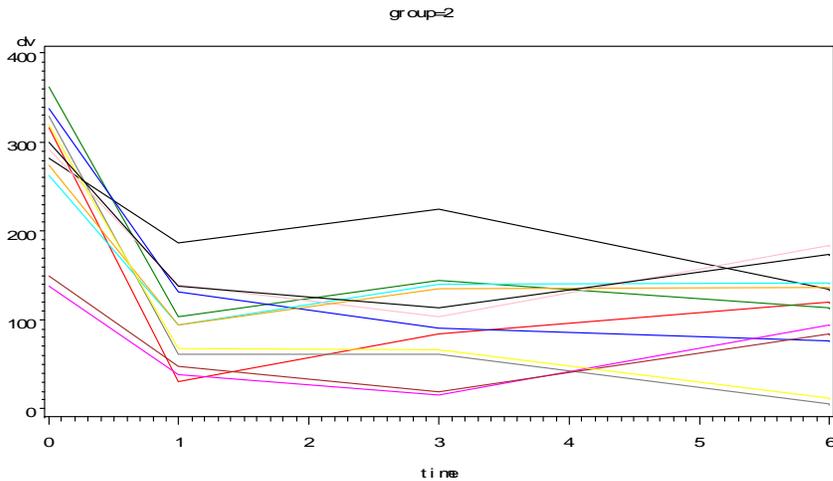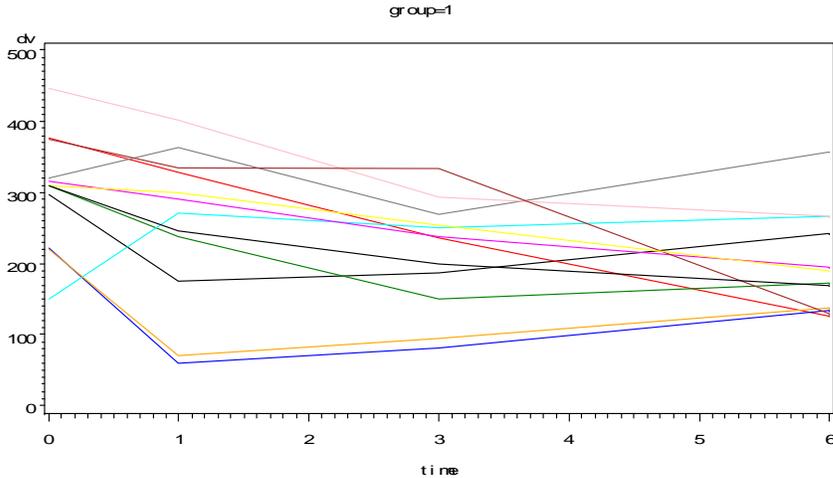
group=1

group=2

For **Proc Mixed** we need to specify that group, time, and subj are class variables. This will cause SAS to treat them as factors. The model statement tells the program that we want to treat group and time as a factorial design and generate the main effects and the interaction. (I have not appended a "/s" to the end of the model statement because I don't want to talk about the parameter estimates of treatment effects at this point, but most people would put it there.) The **repeated** command tells SAS to treat this as a repeated measures design, that the subject variable is named "subj", and that we want to treat the covariance matrix as exhibiting compound symmetry, even though in the data that I created we don't appear to come close to meeting that assumption. The specification

"rcorr" will ask for the estimated correlation matrix. (we could use "r" instead of "rcorr," but that would produce a covariance matrix, which is harder to interpret.)

The results of this analysis follow, and you can see that they very much resemble our analysis of variance approach using **Proc GLM**.

The SAS System

The Mixed Procedure

Model Information	
Data Set	WORK.WICKLONG
Dependent Variable	dv
Covariance Structure	Compound Symmetry
Subject Effect	subj
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Between-Within

Dimensions	
Covariance Parameters	2
Columns in X	15
Columns in Z	0
Subjects	24
Max Obs Per Subject	4

Number of Observations	
Number of Observations Read	96

Number of Observations	
Number of Observations Used	96
Number of Observations Not Used	0

Estimated R Correlation Matrix for subj 1				
Row	Col1	Col2	Col3	Col4
1	1.0000	0.4793	0.4793	0.4793
2	0.4793	1.0000	0.4793	0.4793
3	0.4793	0.4793	1.0000	0.4793
4	0.4793	0.4793	0.4793	1.0000

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
CS	subj	2536.07
Residual		2754.98

Fit Statistics	
-2 Res Log Likelihood	1000.6
AIC (smaller is better)	1004.6
AICC (smaller is better)	1004.8
BIC (smaller is better)	1007.0

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
1	23.47	<.0001

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
group	1	22	18.17	0.0003
time	3	66	37.58	<.0001
group*time	3	66	7.20	0.0003

On the first page of this printout there is nothing particularly exciting except that it tells us that it uses a covariance structure of compound symmetry and that the solution is via REML, the solution that we will use for most problems. On the second page we see the estimated correlations between times. These are not the actual correlations, which appear below, but the estimates that come from an assumption of compound symmetry. That assumption says that the correlations have to be equal, and what we have here are basically average correlations. The *actual* correlations, averaged over the two groups using Fisher's transformation, are:

```
            Estimated R Correlation Matrix for subj 1

     Row        Col1        Col2        Col3        Col4

      1       1.0000      0.5695      0.5351     -0.01683
      2       0.5695      1.0000      0.8612      0.4456
      3       0.5351      0.8612      1.0000      0.4202
      4      -0.01683     0.4456      0.4202      1.0000
```

Notice that they are quite different from the ones assuming compound symmetry, and that they don't look at all as if they fit that assumption. We will deal with this problem later. (I don't have a clue why the heading refers to "subj 1." It just does!)

There are also two covariance parameters. Remember that there are two sources of random effects in this design. There is our normal σ_e^2, which reflects random noise. In addition we are treating our subjects as a random sample, and there is thus random variance among subjects. Here I get to play a bit with expected mean squares. You may recall that the expected mean squares for the error term for the between-subject effect is

$E\left(MS_{w/in\_subj}\right) = \sigma_e^2 + a\sigma_\pi^2$ and our estimate of σ_e^2 is MS$_{residual}$, which is 2781.8718. The letter "*a*" stands for the number of measurement times = 4, and $MS_{w/in\ subj}$ = 12887.046. Therefore our estimate of $\sigma_\pi^2 = (12887.046\text{-}2781.8718)/4 = 2536.07$. These two estimates are our random part of the model and are given in the section headed Covariance Parameter Estimates. I don't see a situation in this example in which we would wish to make use of these values, but in other mixed designs they are useful.

You may notice one odd thing in the data. Instead of entering time as 1,2,3, & 4 I entered it as 0, 1, 3, and 6. If this were a standard ANOVA it wouldn't make any difference, and in fact it doesn't make any difference here, but when we come to looking at intercepts and slopes, it will be very important how we designated the 0 point. We could have centered time by subtracting the mean time from each entry, which would mean that the intercept is at the mean time. I have chosen to make 0 represent the pretest, which seems a logical place to find the intercept. I will say more about this later.

MISSING DATA

I have just spent considerable time discussing a balanced design where all of the data are available. Now I want to delete some of the data and redo the analysis. This is one of the areas where mixed designs have an important advantage. I am going to delete scores pretty much at random, except that I want to show a pattern of different observations over time. It is easiest to see what I have done if we look at data in the wide form, so the earlier table is presented below with "." representing missing observations. It is important to notice that data are missing completely at random, not on the basis of other observations.

Group	Subj	Time0	Time1	Time3	Time6
1	1	296	175	187	242
1	2	376	329	236	126
1	3	309	238	150	173
1	4	222	60	82	135
1	5	150	.	250	266
1	6	316	291	238	194
1	7	321	364	270	358
1	8	447	402	.	266
1	9	220	70	95	137
1	10	375	335	334	129
1	11	310	300	253	.
1	12	310	245	200	170

Group	Subj	Time0	Time1	Time3	Time6
2	13	282	186	225	134
2	14	317	31	85	120
2	15	362	104	.	.
2	16	338	132	91	77
2	17	263	94	141	142
2	18	138	38	16	95
2	19	329	.	.	6
2	20	292	139	104	.
2	21	275	94	135	137
2	22	150	48	20	85
2	23	319	68	67	.
2	24	300	138	114	174

If we treat this as a standard repeated measures analysis of variance, using Proc GLM, we have a problem. Of the 24 cases, only 17 of them have complete data. That means that our analysis will be based on only those 17 cases. Aside from a serious loss of power, there are other problems with this state of affairs. Suppose that I suspected that people

who are less depressed are less likely to return for a follow-up session and thus have missing data. To build that into the example I could deliberately deleted data from those who scored low on depression to begin with, though I kept their pretest scores. (I did not actually do this here.) Further suppose that people low in depression respond to treatment (or non-treatment) in different ways from those who are more depressed. By deleting whole cases I will have deleted low depression subjects and that will result in biased estimates of what we would have found if those original data points had not been missing. This is certainly not a desirable result.

To expand slightly on the previous paragraph, if we using **Proc GLM** , or a comparable procedure in other software, we have to assume that data are missing completely at random, normally abbreviated MCAR. (See Howell, 2008.) If the data are not missing completely at random, then the results would be biased. But if I can find a way to keep as much data as possible, and if people with low pretest scores are missing at one or more measurement times, the pretest score will essentially serve as a covariate to predict missingness. This means that I only have to assume that data are missing at random (MAR) rather than MCAR. That is a gain worth having. MCAR is quite rare in experimental research, but MAR is much more common. Using a mixed model approach requires only that data are MAR and allows me to retain considerable degrees of freedom. (That argument has been challenged by Overall & Tonidandel (2007), but in this particular example the data actually are essentially MCAR. I will come back to this issue later.)

Proc GLM results

The output from analyzing these data using **Proc GLM** follows. I give these results just for purposes of comparison, and I have omitted much of the printout.

```
                    Repeated Measures Analysis of Variance
                  Tests of Hypotheses for Between Subjects Effects

     Source                       DF     Type III SS     Mean Square    F Value    Pr > F
     group                         1     126931.1767     126931.1767       8.97    0.0091
     Error                        15     212237.4410      14149.1627


             Univariate Tests of Hypotheses for Within Subject Effects
                                                                      Adj Pr > F
Source            DF    Type III SS    Mean Square    F Value   Pr > F    G - G     H - F
time               3    201156.6493     67052.2164      27.34   <.0001   <.0001    <.0001
time*group         3     20665.5905      6888.5302       2.81   0.0502   0.0696    0.0547
Error(time)       45    110370.8507      2452.6856

                    Greenhouse-Geisser Epsilon      0.7386
                    Huynh-Feldt Epsilon             0.9300
```

Notice that we still have a group effect and a time effect, but we have lost our significant interaction, which is what I cared most about. Also notice the big drop in degrees of freedom due to the fact that we now only have 17 subjects.

Proc Mixed

Now we move to the results using **Proc mixed**. I need to modify the data file put it in its long form and to replace missing observations with a period, but that means that I just altered 9 lines out of 96 (10% of the data) instead of 7 out of 24 (29%). The syntax would look exactly the same as it did earlier. The presence of "time" on the repeated statement is not necessary if I have included missing data by using a period, but it is needed if I just remove the observation completely. (At least that is the way I read the manual.) The results follow, again with much of the printout deleted:

```
Proc Mixed data = wicklongMiss;
   class group time subj;
   model dv = group time group*time /s;
   repeated time /subject = subj type = cs  rcorr;
run;
```

Estimated R Correlation Matrix for subj 1				
Row	Col1	Col2	Col3	Col4
1	1.0000	0.4640	0.4640	0.4640
2	0.4640	1.0000	0.4640	0.4640
3	0.4640	0.4640	1.0000	0.4640
4	0.4640	0.4640	0.4640	1.0000

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
CS	subj	2558.27
Residual		2954.66

Fit Statistics	
-2 Res Log Likelihood	905.4
AIC (smaller is better)	909.4
AICC (smaller is better)	909.6
BIC (smaller is better)	911.8

Null Model Likelihood Ratio Test		
DF	**Chi-Square**	**Pr > ChiSq**
1	19.21	<.0001

Type 3 Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
group	1	22	16.53	0.0005
time	3	57	32.45	<.0001
group*time	3	57	6.09	0.0011

This is a much nicer solution, not only because we have retained our significance levels, but because it is based on considerably more data and is not reliant on an assumption that the data are missing completely at random. Again you see a fixed pattern of correlations between trials which results from my specifying compound symmetry for the analysis.

Other Covariance Structures

To this point all of our analyses have been based on an assumption of compound symmetry. (The assumption is really about sphericity, but the two are close and **Proc Mixed** refers to the solution as **type = cs**.) But if you look at the correlation matrix given earlier it is quite clear that correlations further apart in time are distinctly lower than correlations close in time, which sounds like a reasonable result. Also if you looked at Mauchly's test of sphericity (not shown) it is significant with $p = .012$. While this is not a great test, it should give us pause. We really ought to do something about sphericity.

The first thing that we could do about sphericity is to specify that the model will make no assumptions whatsoever about the form of the covariance matrix. To do this I will ask for an unstructured matrix. This is accomplished by including **type = un** in the **repeated** statement. This will force SAS to estimate all of the variances and covariances and use them in its solution. The problem with this is that there are 10 things to be estimated and therefore we will lose degrees of freedom for our tests. But I will go ahead anyway. For this analysis I will continue to use the data set with missing data, though I could have used the complete data had I wished. I will include a request that SAS use procedures due to Hotelling-Lawley-McKeon (hlm) and Hotelling-Lawley-Pillai-Samson (hlps) which do a better job of estimating the degrees of freedom for our denominators for. This is recommended for an unstructured model. The results are shown below.

Results using unstructured matrix

```
Proc Mixed data = wicklongMiss;
class group time subj;
model dv = group time group*time;
repeated time /subject = subj type = un hlm hlps rcorr;
run;
```

Estimated R Correlation Matrix for subj 1				
Row	Col1	Col2	Col3	Col4
1	1.0000	0.5858	0.5424	-0.02740
2	0.5858	1.0000	0.8581	0.3896
3	0.5424	0.8581	1.0000	0.3971
4	-0.02740	0.3896	0.3971	1.0000

Fit Statistics	
-2 Res Log Likelihood	883.7
AIC (smaller is better)	903.7
AICC (smaller is better)	906.9
BIC (smaller is better)	915.5

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
9	40.92	<.0001

\

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
group	1	22	17.95	0.0003
timefact	3	22	28.44	<.0001
group*timefact	3	22	6.80	0.0021

Type 3 Hotelling-Lawley-McKeon Statistics				
Effect	Num DF	Den DF	F Value	Pr > F
timefact	3	20	25.85	<.0001
group*timefact	3	20	6.18	0.0038

Type 3 Hotelling-Lawley-Pillai-Samson Statistics				
Effect	Num DF	Den DF	F Value	Pr > F
timefact	3	20	25.85	<.0001
group*timefact	3	20	6.18	0.0038

Notice the matrix of correlations. From posttest to the 6 month follow-up the correlation with pretest scores has dropped from .59 to -.03, and this pattern is consistent. That certainly doesn't inspire confidence in compound symmetry.

The *F*s have not changed very much from the previous model, but the degrees of freedom for within-subject terms have dropped from 57 to 22, which is a huge drop. That results from the fact that the model had to make additional estimates of covariances. Finally, the

hlm and hlps statistics further reduce the degrees of freedom to 20, but the effects are still significant. This would make me feel pretty good about the study if the data had been real data.

But we have gone from one extreme to another. We estimated two covariance parameters when we used **type = cs** and 10 covariance parameters when we used **type = un**. (Put another way, with the unstructured solution we threw up our hands and said to the program "You figure it out! We don't know what's going on." There is a middle ground (in fact there are many). We probably do know at least something about what those correlations should look like. Often we would expect correlations to decrease as the trials in question are further removed from each other. They might not decrease as fast as our data suggest, but they should probably decrease. An autoregressive model, which we will see next, assumes that correlations between any two times depend on both the correlation at the previous time and an error component. To put that differently, your score at time 3 depends on your score at time 2 and error. (This is a first order autoregression model. A second order model would have a score depend on the *two* previous times plus error.) In effect an AR(1) model assumes that if the correlation between Time 1 and Time 2 is .51, then the correlation between Time 1 and Time 3 has an expected value of $.51^2 = .26$ and between Time 1 and Time 4 has an expected value of $.51^3 = .13$. Our data look reasonably close to that. (Remember that these are expected values of r, not the actual obtained correlations.) The solution using a first order autoregressive model follows.

```
Proc Mixed data = wicklongMiss;
class group time subj;
model dv = group time group*time;
repeated time/subject = subj type = AR(1);
run;
```

Estimated R Correlation Matrix for group(subj) 1 1				
Row	Col1	Col2	Col3	Col4
1	1.0000	0.6182	0.3822	0.2363
2	0.6182	1.0000	0.6182	0.3822
3	0.3822	0.6182	1.0000	0.6182
4	0.2363	0.3822	0.6182	1.0000

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
AR(1)	group(subj)	0.6182
Residual		5350.25

Fit Statistics	
-2 Res Log Likelihood	895.1
AIC (smaller is better)	899.1
AICC (smaller is better)	899.2
BIC (smaller is better)	901.4

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
1	29.55	<.0001

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
group	1	22	17.32	0.0004
time	3	57	30.82	<.0001
group*time	3	57	7.72	0.0002

Notice the pattern of correlations. The .6182 as the correlation between adjacent trials is essentially an average of the correlations between adjacent trials in the unstructured case. The .3822 is just $.6182^2$ and $.2363 = .6182^3$. Notice that tests on within-subject effects are back up to 57 *df*, which is certainly nice, and our results are still significant. This is a far nicer solution than we had using **Proc GLM**.

Now we have three solutions, but which should we choose? One aid in choosing is to look at the "Fit Statistics" that are printout out with each solution. These statistics take into account both how well the model fits the data and how many estimates it took to get there. Put loosely, we would probably be happier with a pretty good fit based on few parameter estimates than with a slightly better fit based on many parameter estimates. If you look at the three models we have fit for the unbalanced design you will see that the AIC criterion for the **type = cs** model was 909.4, which dropped to 903.7 when we relaxed the assumption of compound symmetry. A smaller AIC value is better, so we should prefer the second model. Then when we aimed for a middle ground, by specifying the pattern or correlations but not making SAS estimate 10 separate correlations, AIC dropped again to 899.1. That model fit better, and the fact that it did so by only estimating a variance and one correlation leads us to prefer that model.

SPSS Mixed

You can accomplish the same thing using SPSS if you prefer. I will not discuss the syntax here, but the commands are given below. You can modify this syntax by replacing CS with UN or AR(1) if you wish.

```
MIXED
 dv  BY Group Time
 /CRITERIA = CIN(95) MXITER(100) MXSTEP(5) SCORING(1)
 SINGULAR(0.000000000001) HCONVERGE(0, ABSOLUTE) LCONVERGE(0,
ABSOLUTE)
 PCONVERGE(0.000001, ABSOLUTE)
 /FIXED = Group Time Group*Time  | SSTYPE(3)
 /METHOD = REML
 /PRINT = DESCRIPTIVES SOLUTION
 /REPEATED = Time | SUBJECT(Subj) COVTYPE(CS)
 /EMMEANS = TABLES(Group)
 /EMMEANS = TABLES(Time)
 /EMMEANS = TABLES(Group*Time)  .
```

Analyses Using R

The following commands will run the same analysis using the R program (or using S-Plus). The results will not be exactly the same, but they are very close. Lines beginning with # are comments.

```
# Analysis of Wicklund Data with missing values
data <- read.table(file.choose(), header = T)
attach(data)
Time = factor(Time)
Group = factor(Group)
Subj = factor(Subj)
library(nlme)
```

```
model <- lme(dv ~ Time + Group + Time*Group,  random = ~1 | Subj)
#model2 <- update(model, correlation = corCompSymm(.388,form = ~1 | Subj))
# This line above leads to weird results and I don't know why.
summary(model)
anova(model)
# This model is very close to the one produced by SAS using compound
#symmetry,
# when it comes to F values, and the log likelihood is the same. But the AIC
# and BIC are quite different. The StDev for the Random Effects are the same
# when squared. The coefficients are different because R uses the first level
# as the base, whereas SAS uses the last.
```

Mixed Models by a More Traditional Route

Because I was particularly interested in the analysis of variance, I approached the problem of mixed models first by looking at the use of the **repeated** statement in **Proc mixed**. Remember that our main problem in any repeated measures analysis is to handle the fact that when we have several observations from the same subject, our error terms are going to be correlated. This is true whether the covariances fit the compound symmetry structure or we treat them as unstructured or autoregressive. But there is another way to get at this problem. Look at the completely fictitious data shown below.

	time1	time2	time3	time4	time5
1	20.00	30.00	40.00	60.00	70.00
2	50.00	52.00	54.00	56.00	58.00
3	25.00	30.00	35.00	40.00	45.00
4	45.00	46.00	47.00	48.00	49.00
5	.	.	.	.	.

Now look at the pattern of correlations.
Correlations

	time1	time2	time3	time4	time5
time1	1	.987(*)	.902	.051	-.286
time2	.987(*)	1	.959(*)	.207	-.131
time3	.902	.959(*)	1	.472	.152
time4	.051	.207	.472	1	.942
time5	-.286	-.131	.152	.942	1

* Correlation is significant at the 0.05 level (2-tailed).

Except for the specific values, these look like the pattern we have seen before. I generated them by simply setting up data for each subject that had a different slope. For Subject 1 the scores had a very steep slope, whereas for Subject 4 the line was almost flat. In other words there was variance to the slopes. Had all of the slopes been equal (the lines parallel) the off-diagonal correlations would have been equal except for error, and

the variance of the slopes would have been 0. But when the slopes were unequal their variance was greater than 0 and the times would be differentially correlated.

As I pointed out earlier, compound symmetry is associated directly with a model in which lines representing changes for subjects over time are parallel. That means that when we assume compound symmetry, as we do in a standard repeated measures design, we are assuming that pattern for subjects. Their intercepts may differ, but not their slopes. One way to look at the analysis of mixed models is to fiddle with the expected pattern of the correlations, as we did with the **repeated** command. Another way is to look at the variances in the slopes, which we will do with the **random** command. With the appropriate selection of options that results will be the same.

We will start first with the simplest approach. We will assume that subjects differ on average (i.e. that they have different intercepts), but that they have the same slopes. This is really equivalent to our basic repeated measures ANOVA where we have a term for Subjects, reflecting subject differences, but where our assumption of compound symmetry forces us to treat the data by assuming that however subjects differ overall, they all have the same slope. I am using the missing data set here for purposes of comparison.

Here we will replace the repeated command with the random command. The "int" on the random statement tells the model to fit a different intercept for each subject, but to assume that the slopes are constant across subjects. I am requesting a covariance structure with compound symmetry.

```
Proc Mixed data = wicklongMiss;
class group time subj ;
model dv = group time group*time/s;
random int /subject = subj type = cs;
run;
```

Covariance Parameter Estimates		
Cov Parm	**Subject**	**Estimate**
Variance	**subj**	2677.70
CS	**subj**	-119.13
Residual		2954.57

Fit Statistics	
-2 Res Log Likelihood	905.4
AIC (smaller is better)	911.4

Fit Statistics	
AICC (smaller is better)	911.7
BIC (smaller is better)	914.9

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
2	19.21	<.0001

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
group	1	57	16.52	0.0001
time	3	57	32.45	<.0001
group*time	3	57	6.09	0.0011

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
time main effect	3	57	32.45	<.0001
time linear	1	57	63.90	<.0001
time quadratic	1	57	23.58	<.0001
time cubic	1	57	2.69	0.1067

These results are essentially the same as those we found using the repeated command as setting **type = cs**. By only specifying "int" as random we have not allowed the slopes to differ, and thus we have forced compound symmetry. We would have virtually the same output even if we specified that the covariance structure be "unstructured."

Now I want to go a step further. Here I am venturing into territory that I know even less well, but I think that I am correct in what follows.

Remember that when we specify compound symmetry we are specifying a pattern that results from subjects showing parallel trends over time. So when we replace our repeated statement with a random statement and specify that "int" is the only random component, we are doing essentially what the repeated statement did. We are not allowing for different slopes. But in the next analysis I am going to allow slopes to differ by entering "time" in the random statement along with "int." What I will obtain is a solution where the slope of time has a variance greater than 0. The commands for this analysis follow. Notice two differences. We suddenly have a variable called "timecont." Recall that the **class** command converts time to a factor. That is fine, but for the random variable I want time to be continuous. So I have just made a copy of the original "time," called it "timecont," and not put it in the class statement. Notice that I do not include "**type = cs**" in the following syntax because by allowing for different slopes I am allowing for a pattern of correlations that do not fit the assumption of compound symmetry.

```
Proc Mixed data = wicklongMiss ;
class group time subj ;
model dv = group time group*time;
random int timecont /subject = subj ;
run;
```

Fit Statistics	
-2 Res Log Likelihood	905.0
AIC (smaller is better)	911.0
AICC (smaller is better)	911.4
BIC (smaller is better)	914.6

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
group	1	35	15.24	0.0004
time	3	35	31.70	<.0001
group*time	3	35	6.53	0.0013

Notice that the pattern of results is similar to what we found in the earlier analyses. However we only have 35 *df* for error for each test, and our AIC fit statistic is 911.0, which is higher than for other models and represents a poorer fit. My preference would be to stay with the AR1 structure on the repeated command. That looks to me to be the best fitting model and one that makes logical sense.

There is one more approach recommended by Guerin and Stroop (2000). They suggest that when we are allowing a model that has an **AR(1)** or **UN** covariance structure, we combine the random and repeated commands in the same run. According to Littell et al., they showed that "a failure to model a separate between-subjects random effect can adversely affect inference on time and treatment $\times$ time effects."

This analysis would include both kinds of terms and is shown below:

```
proc mixed data = wicklongMiss;
class group time subj ;
model dv = group timefact group*time/solution;
random subj(group);
repeated time/ type = AR(1) subject = subj(group);
run;
```

Partial results follow:

Covariance Parameter Estimates		
Cov Parm	**Subject**	**Estimate**
subj(group)		0
AR(1)	**subj(group)**	0.6182
Residual		5349.89

Fit Statistics	
-2 Res Log Likelihood	895.1
AIC (smaller is better)	899.1
AICC (smaller is better)	899.2
BIC (smaller is better)	901.4

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
group	1	22	17.32	0.0004
time	3	57	30.82	<.0001
group*time	3	57	7.72	0.0002

You may have noticed something interesting about these results. These are exactly the same results that we obtained when we used

```
Proc Mixed data = wicklongMiss;
class group time subj;
model dv = group time group*time;
repeated time /subject = subj type = AR(1);
run;
```

which is the same commands without the random statement. Why then do we need the random statement if it is going to return the same analysis? I don't know, and I'm not alone—see below.

Solution for fixed effects

I have deliberately avoided talking about the section of the output labeled "Solution for fixed effects," and have actually left off the **solution** command in the analyses that I have run. But now is the time to at least explain what you see there.

I will use the type = AR(1) command on the **repeated** statement because that produces the best fit for our data. I will also add a command to print out least squares estimates of cell means because they will be necessary to understand what the fixed effects are. The commands and the relevant part of the printout follow. I have added the **lsmeans** command so that the program will print out the least squares means estimates for the two-way table.

```
Proc Mixed data = wicklongMiss ;
class group time subj;
model dv = group time group*time /solution;
repeated /subject = group(subj) type = AR(1) rcorr;
lsmeans group*time;
run;
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
subj(group)		0
AR(1)	subj(group)	0.6182
Residual		5349.89

Fit Statistics

-2 Res Log Likelihood	895.1
AIC (smaller is better)	899.1
AICC (smaller is better)	899.2
BIC (smaller is better)	901.4

Solution for Fixed Effects

Effect	group	time	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			106.64	23.4064	22	4.56	0.0002
group	1		95.0006	31.9183	22	2.98	0.0070
group	2		0	.	.	.	.
time		0	173.78	27.9825	57	6.21	<.0001
time		1	-8.9994	26.0818	57	-0.35	0.7313
time		3	-10.8539	21.6965	57	-0.50	0.6188
time		6	0	.	.	.	.
group*time	1	0	-71.0839	38.5890	57	-1.84	0.0707
group*time	1	1	57.7765	35.6974	57	1.62	0.1111
group*time	1	3	26.6348	29.2660	57	0.91	0.3666

Solution for Fixed Effects

Effect	group	time	Estimate	Standard Error	DF	t Value	Pr > |t|
group*time	1	6	0	.	.	.	.
group*time	2	0	0	.	.	.	.
group*time	2	1	0	.	.	.	.
group*time	2	3	0	.	.	.	.
group*time	2	6	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
group	1	22	17.32	0.0004
time	3	57	30.82	<.0001
group*time	3	57	7.72	0.0002

Least Squares Means

Effect	group	time	Estimate	Standard Error	DF	t Value	Pr > |t|
group*time	1	0	304.33	21.1146	57	14.41	<.0001
group*time	1	1	250.41	21.5400	57	11.63	<.0001
group*time	1	3	217.42	21.5410	57	10.09	<.0001
group*time	1	6	201.64	21.7007	57	9.29	<.0001
group*time	2	0	280.42	21.1146	57	13.28	<.0001
group*time	2	1	97.6361	21.6454	57	4.51	<.0001
group*time	2	3	95.7817	22.2942	57	4.30	<.0001
group*time	2	6	106.64	23.4064	57	4.56	<.0001

.

Because of the way that SAS or SPSS sets up dummy variables to represent treatment effects, the intercept will represent the cell for the last time and the last group. In other words $cell_{24}$. Notice that that cell mean is 106.64, which is also the intercept. The effect for Group 1 is the difference between the mean of the last time in the first group ($cell_{14}$ and the intercept, which equals 201.64 – 106.64 = 95.00(within rounding error). That is the treatment effect for group 1 given in the solutions section of the table. Because there is only 1 *df* for groups, we don't have a treatment effect for group 2, though we can calculate it as -95.00 because treatment effects sum to zero. For the effect of Time 0, we take the deviation of the cell for Time 0 for the last group (group 2) from the intercept, which equals 280.42 – 106.64 = 173.78. For Times 2 and 3 we would subtract 106.64 from 97.6361 and 95.7817, respectively, giving -8.9994 and -10.8539. With 3 *df* for Time, we don't have an effect for Time 4, but again we can obtain it by subtraction as 0 – (173.78-8.9994-10.8539) = -53.9267. For the interaction effects we take cell means minus row and column effects. So for $Time_{11}$ we have 304.33-(95.00 + 173.78 +106.64) = -71. Similarly for the other interaction effects.

I should probably pay a great deal more attention to these treatment effects, but I will not do so. If they were expressed in terms of deviations from the grand mean, grand mean, rather than with respect to $cell_{24}$ I could get more excited about them. (If SAS set up its design matrix differently they would come out that way. But they don't here.) I know that most statisticians will come down on my head for making such a statement, and perhaps I am being sloppy, but I think that I get more information from looking at cell means and *F* statistics.

And now the big "BUT!"

Well after this page was originally written and I thought that I had everything all figured out (well, I didn't really think that, but I hoped), I discovered that life is not as simple as we would like it to be. The classic book in the field is Littell et al. (2006). They have written about SAS in numerous books, and some of them worked on the development of **Proc Mixed.** However others who know far more statistics than I will ever learn, and who have used SAS for years, have had great difficulty in deciding on the appropriate ways of writing the syntax. An excellent paper in this regard is Overall, Ahn, Shivakumar, & Kalburgi (1999). They spent 27 pages trying to decide on the correct analysis and ended up arguing that perhaps there is a better way than using mixed models anyway. Now they did have somewhat of a special problem because they were running an analysis of covariance because missing data was dependent, in part, on baseline measures. However other forms of analyses will allow a variable to be both a dependent variable and a covariate. (If you try this with SPSS you will be allowed to enter Time1 as a covariate, but the solution is exactly the same as if you had not. I haven't yet tried this is R or S-Plus.) This points out that all of the answers are not out there. If John Overall can't figure it out, how are you and I supposed to?

That last paragraph might suggest that I should just eliminate this whole document, but that is perhaps too extreme. **Proc Mixed** is not going to go away, and we have to get used

to it. All that I suggest is a bit of caution. But if you do want to consider alternatives, look at the Overall et al. paper and read what they have to say about what they call Two-Stage models. Then look at other work that this group has done.

But I can't leave this without bringing in one more complication. Overall & Tonidandel (2007) recommend a somewhat different solution by using a continuous measure of time on the model statement. In other words, specifying time on the class variable turns time into a factor with 4 levels. If I had earlier said timecont = time in a data statement, then Overall & Tonidandel would have me specify the model statement as **model dv = group timecont group*timecont. / solution;** Littell et al. 2006 refer to this as "Comparisons using regression," but it is not clear, other than to test nonlinearity, why we would do this. It is very close to, but not exactly the same as, a test of linear and quadratic components. (For quadratic you would need to include time^2 and its interaction with group.) It yields a drastically different set of results, with 1 *df* for timecont and for timecont×group. The 1 *df* is understandable because you have one degree of freedom for each contrast. The timecont × group interaction is not close to significant, which *may* make sense if you look at the plotted data, but I'm not convinced. I am going to stick with my approach, at least for now.

The SAS printout follows based on the complete (not the missing) data.

Type 3 Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
group	1	22	9.80	0.0049
timecont	1	70	29.97	<.0001
timecont*group	1	70	0.31	0.5803

Compare this with the **Proc GLM** solution for linear trend given earlier.

```
Contrast Variable: time_1  The Linear Effect of Time (intervals = 0,1,2,3)
      Source                      DF     Type III SS     Mean Square    F Value    Pr > F
      Mean                         1     168155.6270     168155.6270      36.14    <.0001
      group                        1       1457.2401       1457.2401       0.31    0.5814
      Error                       22     102368.7996       4653.1273
```

They are not the same, though they are very close. (I don't know why they aren't the same, but I suspect that it has to do with the fact that **Proc GLM** uses a least squares solution while **Proc Mixed** uses REML.) Notice the different degrees of freedom for error, and remember that "mean" is equivalent to "timecont" and "group" is equivalent to the interaction.

What about imputation of missing values?

There are many ways of dealing with missing values (Howell, 2008), but a very common approach is known as Estimation/Maximization (EM). To describe what EM does in a very few sentences, it basically uses the means and standard deviations of the existing observations to make estimates of the missing values. Having those estimates changes the mean and standard deviation of the data, and those new means and standard deviations are used as parameter estimates to make new predictions for the (originally) missing values. Those, in turn, change the means and variances and a new set of estimated values is created. This process goes on iteratively until it stabilizes.

I used a (freely available) program called **NORM** (Shafer & Olson, 1998) to impute new data for missing values. I then took the new data, which was a complete data set, and used **Proc GLM** in SAS to run the analysis of variance on the completed data set. I repeated this several times to get an estimate of the variability in the results. The resulting Fs for three replications are shown below, along with the results of using **Proc Mixed** on the missing data with an autoregressive covariance structure and simply using the standard ANOVA with all subjects having any missing data deleted.

	Replication 1	Replication 2	Replication 3	AR1	With Missing
Group	17.011	15.674	18.709	18.03	8.97
Time	35.459	33.471	37.960	29.55	27.34
Group * Time	5.901	5.326	7.292	7.90	2.81

I will freely admit that I don't know exactly how to evaluate these results, but they are at least in line with each other except for the last column when uses casewise deletion. I find them encouraging.

I need to add a section on references. Some good ones on the web are:

http://www.uoregon.edu/~robinh/mixed_sas.html
http://www.ats.ucla.edu/stat/sas/faq/anovmix1.htm
http://www.ats.ucla.edu/STAT/SAS/library/mixedglm.pdf
http://ssc.utexas.edu/consulting/answers/sas/sas94.html
http://quiro.uab.es/jpa/pdf/Littell_mixed_JAS.pdf

Good coverage of alternative covariance structures
http://cda.morris.umn.edu/~anderson/math4601/gopher/SAS/longdata/structures.pdf

The main reference for SAS **Proc Mixed** is

Little, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., & Schabenberger, O. (2006) *SAS for mixed models*, Cary, NC SAS Institute Inc.

The Overall et al. reference that I referred to is

Overall, J. E., Ahn, C., Shivakumar, C., & Kalburgi, Y. (1999). Problematic formulations of SAS Proc.Mixed models for repeated measurements. *Journal of Biopharmaceutical Statistics, 9*, 189-216. (That probably is not a journal that you read on a monthly basis, but the article is not too technical.)

The classic reference for **R** is Penheiro, J. C. & Bates, D. M. (2000) *Mixed-effects models in S and S-Plus.* New York: Springer.

For imputation the best reference is

Shafer, J. L. & Olson, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analysts perspective. *Multivariate Behavioral Research*, 33, 545-571.