MIXED MODELS FOR REPEATED (LONGITUDINAL) DATA—PART 1

DAVID C. HOWELL 4/26/2010

FOR THE SECOND PART OF THIS DOCUMENT GO TO
www.uvm.edu/~dhowell/methods/Supplements/Mixed Models Repeated/Mixed Models for Repeated Measures2.pdf

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. I will break this paper up into multiple papers because there are a number of designs and design issues to consider. This document will deal with the use of what are called mixed models (or linear mixed models, or hierarchical linear models, or many other things) for the analysis of what we normally think of as a simple repeated measures analysis of variance. Future documents will deal with mixed models to handle single-subject design (particularly multiple baseline designs) and nested designs.

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. I will compare the standard analysis of variance model with a mixed model. 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 1 (pretest), 2 (one month posttest), 3 (3 months follow-up), and 4 (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.

The program and data used below are available at

http://www.uvm.edu/~dhowell/methods7/Supplements/Mixed Models Repeated/Wicksell.sas
http://www.uvm.edu/~dhowell/methods7/Supplements /MixedModelsRepeated/WicksellWide.dat
http://www.uvm.edu/~dhowell/methods7/Supplements/MixedModelsRepeated/WicksellLong.dat
http://www.uvm.edu/~dhowell/methods7/Supplements/MixedModelsRepeated/WicksellWideMiss.dat
http://www.uvm.edu/~dhowell/methods7/Supplements/MixedModelsRepeated/WicksellLongMiss.dat

Many of the printouts that follow were generated using SAS **Proc mixed**, but I give the SPSS commands as well. (I also give syntax 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.

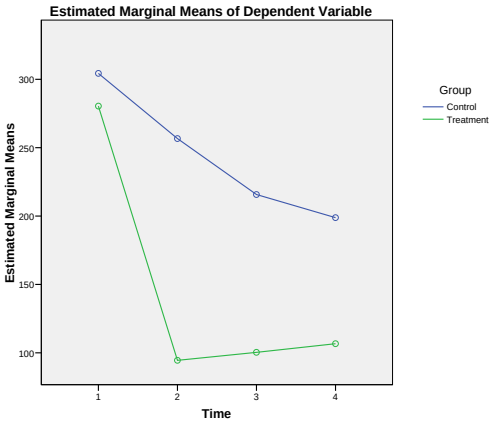
The data follow. Notice that to set this up for ANOVA (**Proc** GLM) we read in the data one subject at a time. (You can see this is the data shown.) This will become important because we will not do that for mixed models.

WicksellWide.dat

Group	Subj	Time0	Time1	Time3	Time6
1	1	296	175	187	192
1	2	376	329	236	76
1	3	309	238	150	123
1	4	222	60	82	85
1	5	150	271	250	216
1	6	316	291	238	144
1	7	321	364	270	308
1	8	447	402	294	216
1	9	220	70	95	87
1	10	375	335	334	79
1	11	310	300	253	140
1	12	310	245	200	120

Group	Subj	Time0	Time1	Time3	Time
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:



The cell means and standard errors follow.

```
---------------------------------------- group=Control ----------------------------------
                                    The MEANS Procedure

          Variable     N            Mean         Std Dev         Minimum         Maximum
          ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff
          time1       12     304.3333333      79.0642240     150.0000000     447.0000000
          time2       12     256.6666667     107.8503452      60.0000000     402.0000000
          time3       12     215.7500000      76.5044562      82.0000000     334.0000000
          time4       12     148.8333333      71.2866599      76.0000000     308.0000000
          xbar        12     231.3958333      67.9581638     112.2500000     339.7500000
          ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff


---------------------------------------- group=Treatment --------------------------------

          Variable     N            Mean         Std Dev         Minimum         Maximum
          ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff
          time1       12     280.4166667      69.6112038     138.0000000     362.0000000
          time2       12      94.5000000      47.5652662      31.0000000     186.0000000
          time3       12     100.3333333      57.9754389      16.0000000     225.0000000
          time4       12     106.6666667      55.7939934       6.0000000     184.0000000
          xbar        12     145.4791667      42.9036259      71.7500000     206.7500000
          ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff
```

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. Because I will ask for a polynomial trend analysis, I have told it to recode the levels as 0, 1, 3, 6 instead of 1, 2, 3, 4. I did not need to do this, but it seemed truer to the experimental design. It does not affect the standard summery table. (I give the entire data entry parts of the program here, but will leave it out in future code.)

```
Options nodate nonumber nocenter formdlim = '-';
libname lib
'C:\Users\Dave\Documents\Webs\StatPages\More_Stuff\MixedModelsRepeated'
;
Title 'Analysis of Wicksell complete data';

data lib.WicksellWide;
      infile
'C:\Users\Dave\Documents\Webs\StatPages\More_Stuff\MixedModelsRepeated\
WicksellWide.dat' firstobs = 2;
      input group subj time1 time2 time3 time4;
      xbar = (time1+time2+time3+time4)/4;
run;

 Proc Format;
      Value group
            1 = 'Control'
            2 = 'Treatment'
            ;
run;

Proc Means data = lib.WicksellWide;
      Format group group.;
      var time1 -- time4 xbar;
```

```
      by group;
run;

Title 'Proc GLM with Complete Data';
proc GLM ;
   class group;
   model time1 time2 time3 time4 = group/ nouni;
   repeated time 4 (0 1 3 6) polynomial /summary printm;
 run;
```

Proc GLM with Complete Data

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	177160.1667	177160.1667	13.71	0.0012
Error	22	284197.4583	12918.0663		

Proc GLM with Complete Data

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	Adj Pr > F H - F
time	3	373802.7083	124600.9028	45.14	<.0001	<.0001	<.0001
time*group	3	74654.2500	24884.7500	9.01	<.0001	0.0003	0.0001
Error(time)	66	182201.0417	2760.6218				

Greenhouse-Geisser Epsilon	0.7297
Huynh-Feldt Epsilon	0.8503

Proc GLM with Complete Data

Analysis of Variance of Contrast Variables

time_N represents the nth degree polynomial contrast for time

Contrast Variable: time_1

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Mean	1	250491.4603	250491.4603	54.27	<.0001
group	1	2730.0179	2730.0179	0.59	0.4500
Error	22	101545.1885	4615.6904		

Contrast Variable: time_2

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Mean	1	69488.21645	69488.21645	35.37	<.0001
group	1	42468.55032	42468.55032	21.62	0.0001
Error	22	43224.50595	1964.75027		

```
Contrast Variable: time_3

Source                       DF     Type III SS     Mean Square    F Value    Pr > F

Mean                          1     53823.03157     53823.03157      31.63    <.0001
group                         1     29455.68182     29455.68182      17.31    0.0004
Error
```

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 $p = .0001$. 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 54.27 for "mean" and an *F* of 0.59 for "group." Any other software that I have used would replace "mean" with "Time" and "group" with "Group × Time." In other words we have a significant linear trend over time, but the linear × 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 it 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 you 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. I have used "card" input rather than reading a file just to give an alternative approach.

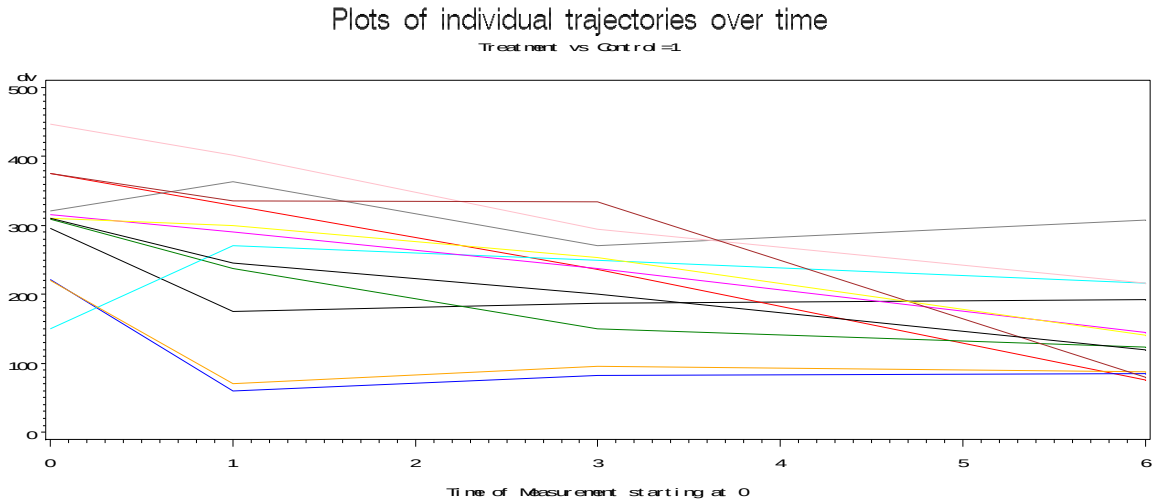
```
data WicksellLong;
input subj time group dv;
cards;
1   0  1.00  296.00
1   1  1.00  175.00
1   3  1.00  187.00
1   6  1.00  242.00
2   0  1.00  376.00
2   1  1.00  329.00
2   3  1.00  236.00
2   6  1.00  126.00
3   0  1.00  309.00
3   1  1.00  238.00
3   3  1.00  150.00
3   6  1.00  173.00
4   0  1.00  222.00
4   1  1.00   60.00
4   3  1.00   82.00
```

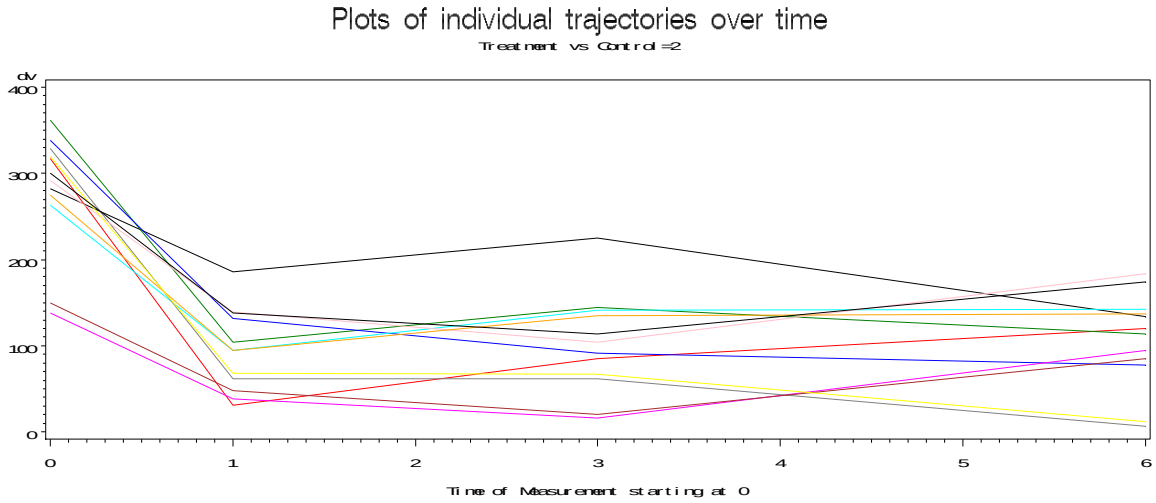
```
4   6  1.00 135.00      11  3  1.00 253.00      18  1  2.00  38.00
5   0  1.00 150.00      11  6  1.00 170.00      18  3  2.00  16.00
5   1  1.00 271.00      12  0  1.00 310.00      18  6  2.00  95.00
5   3  1.00 250.00      12  1  1.00 245.00      19  0  2.00 329.00
5   6  1.00 266.00      12  3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

/* The following lines plot the data. First I will sort to be safe. */
Proc Sort data = lib.wicklong;
        by subject group;
run;

Symbol1 I = join v = none r = 12;
Proc gplot data = wicklong;
   Plot dv*time = subject/ nolegend;
   By group;
Run;

/* This is the main Proc Mixed procedure. */
Proc Mixed data = lib.WicksellLong;
   class group subject time;
   model dv = group time group*time;
   repeated time/subject = subject 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)





For **Proc Mixed** we need to specify that group, time, and subject are class variables. (See the syntax above.) This will cause SAS to treat them as factors (nominal or ordinal variables) instead of as continuous variables. 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 "/solution" 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 "subject", 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**.

```
-------------------------------------------------------------------------------------------

Proc Mixed with complete data.

The Mixed Procedure

    Estimated R Correlation Matrix for subject 1

 Row        Col1        Col2        Col3        Col4

   1      1.0000      0.4791      0.4791      0.4791
   2      0.4791      1.0000      0.4791      0.4791
   3      0.4791      0.4791      1.0000      0.4791
   4      0.4791      0.4791      0.4791      1.0000


 Covariance Parameter Estimates

Cov Parm     Subject    Estimate

CS           subject     2539.36
Residual                 2760.62

           Fit Statistics

-2 Res Log Likelihood          1000.8
AIC (smaller is better)        1004.8
AICC (smaller is better)       1004.9
BIC (smaller is better)        1007.2


  Null Model Likelihood Ratio Test

    DF    Chi-Square      Pr > ChiSq

     1         23.45          <.0001

         Type 3 Tests of Fixed Effects
               Num     Den
Effect          DF      DF    F Value    Pr > F

group            1      22      13.71    0.0012
time             3      66      45.14    <.0001
group*time       3      66       9.01    <.0001

-------------------------------------------------------------------------------------------
```

On this printout 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 "subject 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, taken from the GLM analysis, is $MS_{residual}$, which is 2760.6218. The letter "*a*" stands for the number of measurement times = 4, and $MS_{subj\ w/in\ grps}$ = 12918.0663, again from the GLM analysis. Therefore our estimate of σ_π^2 = (12918.0663 + 2760.6218)/4 = 2539.36. 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	192
1	2	376	329	236	76
1	3	309	238	150	123
1	4	222	60	82	85
1	5	150	.	250	216
1	6	316	291	238	144
1	7	321	364	270	308
1	8	447	402	.	216
1	9	220	70	95	87

1	10	375	335	334	79
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.

```
                                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      92917.9414      92917.9414       6.57    0.0216
       Error               15     212237.4410      14149.1627

--------------------------------------------------------------------------------

                            Analysis of Wicksell missing data
                                                 13:39 Wednesday, March 31, 2010

                                The GLM Procedure
                        Repeated Measures Analysis of Variance
                  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    238578.7081     79526.2360     32.42   <.0001   <.0001   <.0001
 time*group          3     37996.4728     12665.4909      5.16   0.0037   0.0092   0.0048
 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 the *F* for our interaction has been reduced by about half, and that is what we care most about. (In a previous version I made it drop to nonsignificant, but I relented here.) 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 by putting it in its long form and to replacing 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 = lib.WicksellLongMiss;
  class group time subject;
  model dv = group time group*time /solution;
  repeated time /subject = subject type = cs  rcorr;
run;
```

```
Estimated R Correlation Matrix for subject 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	subject	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

Solution for Fixed Effects

Solution for Fixed Effects

Effect	Treatment vs Control	Time of Measurement starting at 0	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			111.87	23.7349	22	4.71	0.0001
group	1		39.6917	32.4122	22	1.22	0.2337
group	2		0	.	.	.	.
time		1	168.54	24.4208	57	6.90	<.0001
time		2	-16.0692	25.1383	57	-0.64	0.5252
time		3	-11.0431	25.5275	57	-0.43	0.6669
time		4	0	.	.	.	.
group*time	1	1	-15.7751	33.4158	57	-0.47	0.6387
group*time	1	2	118.64	34.3680	57	3.45	0.0011
group*time	1	3	75.8815	34.6537	57	2.19	0.0327
group*time	1	4	0	.	.	.	.
group*time	2	1	0	.	.	.	.
group*time	2	2	0	.	.	.	.
group*time	2	3	0	.	.	.	.
group*time	2	4	0	.	.	.	.

Analysis with random intercept and random slope.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
group	1	22	12.54	0.0018
time	3	57	38.15	<.0001
group*time	3	57	7.37	0.0003

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. This is recommended for an unstructured model. The results are shown below.

Results using unstructured matrix

```
Proc Mixed data = lib.WicksellLongMiss;
  class group time subject;
  model dv = group time group*time /solution;
  repeated time /subject = subject type = un hlm hlps  rcorr;
run;
```

```
   Estimated R Correlation Matrix for subject 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


Covariance Parameter Estimates

Cov Parm    Subject    Estimate

UN(1,1)     subject     5548.42
UN(2,1)     subject     3686.76
UN(2,2)     subject     7139.94
UN(3,1)     subject     2877.46
UN(3,2)     subject     5163.81
UN(3,3)     subject     5072.14
UN(4,1)     subject     -129.84
UN(4,2)     subject     2094.43
UN(4,3)     subject     1799.21
UN(4,4)     subject     4048.07


           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



---------------------------------------------------------------------------------------------------

Same analysis but specifying an unstructured covariance matrix.

The Mixed Procedure

  Null Model Likelihood Ratio Test

    DF    Chi-Square      Pr > ChiSq
```

```
     9         40.92          <.0001


                              Solution for Fixed Effects

                    Time of
                    Measurement
             Treatment    starting                    Standard
Effect       vs Control   at 0          Estimate       Error      DF    t Value    Pr > |t|

Intercept                                 102.94     20.5007      22       5.02      <.0001
group        1                           48.4781     27.9262      22       1.74      0.0966
group        2                                 0           .       .          .           .
time                      1               177.48     30.0714      22       5.90      <.0001
time                      2              -7.2465     26.4793      22      -0.27      0.7869
time                      3              -3.9933     24.0169      22      -0.17      0.8695
time                      4                 0           .       .          .           .
group*time   1            1              -24.5614     41.8078      22      -0.59      0.5629
group*time   1            2               111.07     36.3306      22       3.06      0.0058
group*time   1            3               71.6774     32.7021      22       2.19      0.0393
group*time   1            4                 0           .       .          .           .
group*time   2            1                 0           .       .          .           .
group*time   2            2                 0           .       .          .           .
group*time   2            3                 0           .       .          .           .
group*time   2            4                 0           .       .          .           .


         Type 3 Tests of Fixed Effects

               Num     Den
Effect          DF      DF    F Value    Pr > F

group            1      22      13.76    0.0012
time             3      22      30.47    <.0001
group*time       3      22       8.18    0.0008


   Type 3 Hotelling-Lawley-McKeon Statistics

               Num     Den
Effect          DF      DF    F Value    Pr > F

time             3      20      27.70    <.0001
group*time       3      20       7.43    0.0016

-------------------------------------------------------------------------------------

Same analysis but specifying an unstructured covariance matrix.

The Mixed Procedure

     Type 3 Hotelling-Lawley-Pillai-Samson
                 Statistics

               Num     Den
Effect          DF      DF    F Value    Pr > F

time             3      20      27.70    <.0001
group*time       3      20       7.43    0.0016
```

Notice the matrix of correlations. From pretest to the 6 month follow-up the correlation with pretest scores has dropped from .46 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.

```
Title 'Same analysis but specifying an autoregressive covariance
matrix.';
Proc Mixed data = lib.WicksellLongMiss;
      class group subject time;
      model dv = group time group*time;
      repeated time /subject = subject type = AR(1) rcorr;
run;
```

```
Same analysis but specifying an autoregressive covariance matrix.


    Estimated R Correlation Matrix for subject 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)        subject      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

Solution for Fixed Effects

Effect	Treatment vs Control	Time of Measurement starting at 0	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			106.64	23.4070	22	4.56	0.0002
group	1		45.0008	31.9192	22	1.41	0.1726
group	2		0	.	.	.	.
time		1	173.78	27.9825	57	6.21	<.0001
time		2	-8.9994	26.0814	57	-0.35	0.7313
time		3	-10.8540	21.6959	57	-0.50	0.6188

Same analysis but specifying an autoregressive covariance matrix.

The Mixed Procedure

Solution for Fixed Effects

Effect	Treatment vs Control	Time of Measurement starting at 0	Estimate	Standard Error	DF	t Value	Pr > |t|
time		4	0	.	.	.	.
group*time	1	1	-21.0841	38.5889	57	-0.55	0.5869
group*time	1	2	107.78	35.6969	57	3.02	0.0038
group*time	1	3	76.6351	29.2652	57	2.62	0.0113
group*time	1	4	0	.	.	.	.
group*time	2	1	0	.	.	.	.
group*time	2	2	0	.	.	.	.
group*time	2	3	0	.	.	.	.
group*time	2	4	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
group	1	22	13.20	0.0015
time	3	57	34.34	<.0001
group*time	3	57	9.23	<.0001

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(Subject) 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)
Subject = factor(Subject)
library(nlme)
```

```
model1 <- lme(dv ~ Time + Group + Time*Group,  random = ~1 | Subject)

summary(model1)
anova(model1)
# 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.
```

Where do we go now?

This document is sufficiently long that I am going to create a new one to handle this next question. In that document we will look at other ways of doing much the same thing. The reason why I move to alternative models, even though they do the same thing, is that the logic of those models will make it easier for you to move to what are often called single-case designs or multiple baseline designs when we have finish with what is much like a traditional analysis of variance approach to what we often think of as traditional analysis of variance designs.

If I forget to supply a link later, try the same link as this document but with a "2" after "measures."

dch
4/1/2010

References

Guerin, L., and W.W. Stroup. 2000. A simulation study to evaluate PROC MIXED analysis of repeated measures data. p. 170–203. *In* Proc. 12th Kansas State Univ. Conf. on Applied Statistics in Agriculture. Kansas State Univ., Manhattan.

Howell, D.C. (2008) The analysis of variance. In Osborne, J. I., *Best practices in Quantitative Methods*. Sage.

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

Maxwell, S. E. & Delaney, H. D. (2004) *Designing Experiments and Analyzing Data: A Model Comparison Approach, 2nd edition.* Belmont, CA. Wadsworth.

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.

Overall, J. E. & Tonindandel, S. (2002) Measuring change in controlled longitudinal studies. *British Journal of Mathematical and Statistical Psychology, 55,* 109-124.

Overall, J. E. & Tonindandel, S. (2007) Analysis of data from a controlled repeated measurements design with baseline-dependent dropouts. *Methodology, 3,* 58-66.

Pinheiro, J. C. & Bates, D. M. (2000). *Mixed-effects Models in S and S-Plus.* Springer.
Some good references on the web are:

http://www.ats.ucla.edu/stat/sas/faq/anovmix1.htm

http://www.ats.ucla.edu/stat/sas/library/mixedglm.pdf

The following is a good reference for people with questions about using SAS in general.
http://ssc.utexas.edu/consulting/answers/sas/sas94.html

For a wealth of information go to:
Downloadable Papers on Multilevel Models

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.

See also
Maxwell, S. E. & Delaney, H. D. (2004). *Designing Experiments and Analyzing Data (2nd edition).* Lawrence Erlbaum Associates.

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

Last revised: 1/1/2010
dch