

Analysis of Longitudinal Data: Comparison between PROC GLM and PROC MIXED.

Maribeth Johnson, Medical College of Georgia, Augusta, GA

ABSTRACT

Longitudinal data refers to datasets with multiple measurements of a response variable on the same experimental unit made over a period of time. These types of data require special attention because they involve correlated data. The relationships between repeated measurements are important in assessing reliability and tracking of those measurements. The proper variance-covariance structure in the analysis model is essential to the understanding and interpretation of those relationships. The assumption of compound symmetry necessary for correctly using the intraclass correlation as a measure of tracking can be tested against other variance structures using PROC MIXED. This paper compares the variance, covariance and correlation estimates obtained from the GLM and MIXED procedures of SAS/STAT® on two sets of data, one of which has missing data.

INTRODUCTION

The correlations between repeated measurements on individuals enable us to quantify the inter-relationships of the measurements. If the variances and covariances are constant across all repeated measurements then the relationship can be thought of as intraclass correlation (ICC) reliability. Different covariance structures provide an opportunity to investigate and possibly quantify the tracking of measurements as well as provide the basis for other relationships.

DATA

The data used in the paper are from The Family Health Project that was conducted at the Texas site in the Studies of Child Activity and Nutrition (SCAN) program, a National Heart, Lung and Blood Institute funded multi-center longitudinal study of the role of nutrition and physical activity on the development of cardio-vascular disease risk factors and associated behaviors in families of young children (Baranowski, et al. 1993). During the initial year of study the subjects included children aged 3 or 4 years.

Table 1. Data structure

Number of records		Number of children	
		WT	HR
1 of 4		102	114
2 of 4		34	43
3 of 4		36	58
4 of 4		98	60
		270	275

Year	WT			HR		
	N	Mean	SD	N	Mean	SD
1986	243	16.1	2.4	249	93.7	10.2
1987	148	18.5	3.0	144	90.4	10.0
1988	144	21.3	4.0	111	87.5	8.3
1989	135	24.2	4.7	110	82.2	9.7

Families participated in four annual summer clinics, held at

the University of Texas, Galveston, at which time a variety of physiological and self-report measures were collected. The first annual summer clinic was in 1986. Many children were seen at the first clinic and then dropped out of the study for various reasons.

The variables used in these analyses are weight (WT) and resting heart rate (HR). WT (nearest 0.1 kg) was measured using a Detecto balance-beam scale. HR (beats/min) was obtained using an automatic Dinamap Adult/Pediatric Vital Signs Monitor (Model 845 XT/XT-IEC) following standardized protocols (in the early morning, without prior exercise, having fasted overnight, lying at rest for fifteen minutes, and using the right arm). Because of normal variability in HR, five heart rate readings were taken, one minute apart, and the last four were recorded. The mean of the four values was used in the analyses. The pattern of missing data and the within year means and standard deviations (SD) are seen in Table 1.

USING GLM WITH THE RANDOM STATEMENT

In the past, a mixed model univariate analysis of variance has been used to estimate the correlation between repeated measurements through the calculation of the ICC. For these data this was accomplished in PROC GLM using the statements:

```
proc glm; class child year;  
model wt hr = child year;  
random child;
```

The random statement outputs the expected mean square (MS) for the between child variation of the form

$$\text{Var(Error)} + k \text{ Var(CHILD)}$$

where the value of k is the average number of observations per child which is equal to the number of years for the balanced data. CHILD is treated as a fixed effect as far as the model fit is concerned. Since it is necessary to adjust for differences between years the Error MS is used as the VAR(Error), i.e. the within child variation, in the calculation. Therefore the between child variation and ICC are calculated as,

$$\text{Var(CHILD)} = \frac{\text{MS(CHILD)} - \text{MS(Error)}}{k}$$

$$\text{ICC} = \frac{\text{Var(CHILD)}}{\text{Var(CHILD)} + \text{MS(Error)}}$$

This ICC can be thought of as the reliability of a single year of measurement. While this estimator of reliability is biased, the individual components of variation are maximum-likelihood estimators when the data are balanced. See Winer (1971) for an explanation of how to obtain an unbiased estimator of the ICC from the mean square estimates.

This mixed model assumes that the error has a constant

variance and that the additive random variable for child is independent from year to year. These assumptions imply that the correlation between measurements is constant between any two years and that the total variance is the same for each year. In the notation of multivariate analysis, the variance-covariance matrix of a child's vector of responses over time is compound symmetric. Thus, if the variability of a measure changes from year to year, as happens with certain growth parameters in young children, this model is inappropriate. The standardization of variables within measurement periods can alleviate the problem of unequal variances but the assumption of equal correlation between any two years also may not be valid. Importantly, when any of the assumptions of compound symmetry do not hold, the simple expression of ICC, and therefore reliability, do not correctly assess the relationships between repeated measurements. Models are needed which provide for more general variance-covariance structures.

USING GLM WITH THE REPEATED STATEMENT

Using this type of model allows for the multivariate test of the assumption of compound symmetry (CS). The drawbacks to this analysis are that it can not handle any missing data and if the CS assumption is rejected it cannot help in determining the correct underlying covariance structure.

The REPEATED statement in PROC GLM works only when repeated measures are written as multivariate responses in the MODEL statement. This means that the repeated measurements also must appear in a multivariate mode in the dataset, i.e. with the multiple observations listed on one line for each subject. This data setup is different from both the univariate GLM and the MIXED analyses. Observations with missing data for any of the repeated measures variables are not used in this analysis.

The statements used are:

```
proc glm;
model depvar1-depvar4 = /nouni;
repeated year / print;
```

where *depvar* was the four yearly measurements for either WT or HR. The NOUNI option on the MODEL statement suppresses the univariate analyses of each year.

The PRINTE option outputs the "Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|". These are partial correlations computed from residuals after fitting the between-subjects model.

The next important piece of output from this analysis is the "Test for Sphericity". This is a test of whether the condition holds that is necessary for univariate analysis of variance. Specifically, it is a test of whether a set of orthonormal contrasts of the repeated measures variables are independent and have equal variances, i.e. are the data compound symmetric. Statistical significance of the test for sphericity tells you that this condition is not met and that the estimate of the correlation between measures from the univariate GLM analysis is not valid.

The correlations from this analysis are identical to the correlations computed from the RCORR option in the

REPEATED statement in PROC MIXED when TYPE=UN is specified as the covariance structure. The tests of fixed effects from both analyses are close. However, the REPEATED statements of each perform different functions and bear little resemblance to each other.

USING PROC MIXED

The ability to model irregular changes over time using different covariance structures available in PROC MIXED may increase our understanding of the inter-relationships of repeated measurements. The basic model that was used is as follows:

```
proc mixed; class child year;
model depvar = year;
repeated year / subject=child r rcorr
type=cov-structure ;
```

where *depvar* was either WT or HR. The REPEATED statement models the covariance structures in **R**, the variance-covariance matrix of the vector of errors. The SUBJECT=CHILD option is the mechanism for block diagonalizing **R** since subjects are considered independent. The R option of the REPEATED statement requests that the first block of the **R** matrix be printed, the RCORR options prints the correlation matrix corresponding to **R**. Four different *cov-structures* were considered, these are:

- 1) Compound symmetric (CS): This structure is the assumption of the ANOVA estimates of the variance components; therefore the univariate GLM results for the balanced data should be identical to the estimates from this analysis.
- 2) Heterogeneous compound symmetric (CSH): This structure assumes a common correlation between years but allows for different variances along the diagonal.
- 3) Heterogeneous first-order autoregressive (ARH(1)): This structure also allows for different variance parameters and produces an estimate of the autoregressive parameter so that the correlations between years separated by the same amount of time are the same, and the correlations of different amounts of time are also structured in the usual way, i.e. the correlation is ρ^m where m is the number of years between measures.
- 4) Unstructured (UN): This structure produces estimates of all four variances and six covariances in each subject block of **R**. Therefore, all of the correlations between years may be different. These estimates are identical to those found using the PRINTE option on the REPEATED statement in GLM.

Since the same fixed effect of year was included in all models a likelihood ratio test (LRT) was used to compare models for which one is a special case of the other (Wolfinger,1992).

An LRT for the significance of a more general model can be constructed if one covariance model is a submodel of another by computing -2 times the difference between their log likelihoods. Then this statistic is compared to the chi-

square distribution with degrees of freedom equal to the difference in the number of parameters for the two models.

Model comparisons can also be made using Akaike's Information Criterion (AIC) or Schwarz's Bayesian Criterion (SBC). In both cases, the model that has the largest value is the preferred model. SBC penalizes models with more covariance parameters more than AIC so the two criteria may not agree when assessing model fit.

All analyses were performed on a balanced data set of children that were measured at all four clinics, and the unbalanced data set containing observations on all children that ever came to the clinics.

RESULTS AND DISCUSSION

Weight

Both the means and variances for WT in these children are increasing over time (Table 1). In the past, the data might have been standardized within years before an estimate of the correlation between the repeated measurements was obtained.

The variances, covariances and correlations from the R and RCORR options of the REPEATED statement from the MIXED analysis for WT using the balanced data set are shown in Table 2.

All parameters estimated from the CS structure are the same as those calculated using the univariate GLM mean square estimates. The between child variance is equal to 8.86, which is the covariance between years in the MIXED setup. The within child variance is 1.90, which when added to the between child variance is the same as the year variance of 10.76 from MIXED. Therefore, the ICC is 0.823 from both analyses that assume equal year variances and common correlations between years.

When the constraint of equality of variances between years is relaxed using the CSH structure, the common correlation between years is estimated to be 0.92. The GLM analysis of WT standardized within years produced an ICC estimate identical to this common correlation estimate.

Table 2. Covariances and Correlations for WT using Different Covariance Structures, Balanced Data. (n=98)

Type = CS:				
<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	
10.76	8.86	8.86	8.86	
1.0	.823	.823	.823	
(Variance and covariances in top line, correlations below)				
Type = CSH:				
<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	
4.71	5.04	6.61	9.21	
.920	6.37	7.69	10.71	
.920	.920	10.96	14.05	
.920	.920	.920	21.27	

(Variances on diagonal, covariances above, correlations below)

Type = ARH(1):

<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>
4.49	5.12	6.44	8.44
.951	6.45	8.12	10.63
.905	.951	11.30	14.78
.861	.905	.951	21.37

(Variances on diagonal, covariances above, correlations below)

Type = UN:

<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>
4.61	5.28	6.45	8.39
.961	6.55	8.12	10.61
.900	.951	11.13	14.32
.858	.910	.942	20.77

(Variances on diagonal, covariances above, correlations below)

When the common correlation constraint is structured using ARH(1), the correlation between measurements separated by one year is 0.951, by two years is 0.905 and by three years is 0.861. The UN structure produced very similar results (that were identical to the correlations from the multivariate GLM analysis). The LRT between these two models did not show a significant difference in fit, as seen in Table 3.

Table 3. REML Likelihood Ratio Tests between Covariance Structures for WT.

Balanced Data:

Type	Parameters	-2LL	CM±	df	χ ²
CS	2	1658	-	-	-
CSH	5	1370	CS	3	288**
ARH(1)	5	1292	CS	3	366**
UN	10	1284	CSH	5	86**
UN	10	1284	ARH(1)	5	8

Unbalanced Data:

Type	Parameters	-2LL	CM±	df	χ ²
CS	2	3012	-	-	-
CSH	5	2601	CS	3	411**
ARH(1)	5	2505	CS	3	507**
UN	10	2496	CSH	5	105**
UN	10	2496	ARH(1)	5	9

±CM is the comparison model for the likelihood ratio test

**p<.0005

The simpler ARH(1) covariance structure appears to be the best fit for these data. The assumption of a common correlation between years does not hold even when the variances are allowed to differ. A significant test of sphericity from the multivariate GLM analysis indicated that the univariate estimate was not valid. What has been done in the past does not adequately describe the time relationships of this measurement. The AR(1) parameter is in fact higher than any of the other correlation coefficients from models that do not fit as well. By using improper models we are underestimating the correlation between adjacent measurements.

The results of the WT analyses from the unbalanced data set are shown in Table 4.

With unbalanced data the ANOVA estimates from GLM are

no longer maximum-likelihood estimators so the parameter estimates are not identical to the CS structure from PROC MIXED, although they are similar. The between child variance is estimated to be 9.95 from GLM and 9.17 from MIXED. The within child (error) variances are 2.03 and 2.00 from GLM and MIXED, respectively. These result in slightly different ICC estimates. The remaining results are similar to the balanced case.

Table 4. Covariances and Correlations for WT using different Covariance Structures, Unbalanced Data. (N=270)

Type = CS:				
<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	
11.17	9.17	9.17	9.17	
1.0	.820	.820	.820	
(Variance and covariances in top line, correlations below)				
Type = CSH:				
<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	
6.11	6.80	9.37	12.16	
.937	8.60	11.12	14.42	
.937	.937	16.37	19.82	
.937	.937	.937	27.53	
(Variances on diagonal, covariances above, correlations below)				
Type = ARH(1):				
<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	
5.97	7.02	9.24	11.79	
.962	8.94	11.77	15.01	
.925	.962	16.75	21.36	
.889	.925	.962	29.45	
(Variances on diagonal, covariances above, correlations below)				
Type = UN:				
<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	
5.97	7.06	8.99	11.52	
.966	8.94	11.61	14.98	
.904	.955	16.54	21.21	
.871	.925	.964	29.30	
(Variances on diagonal, covariances above, correlations below)				

By relaxing the constraint of variance equality using the CSH structure the model is seen to have a better fit over CS (Table 3). The same holds true when both the variance and correlation equality constraints are relaxed using the ARH(1) structure. As was the case using balanced data, the fit from the UN model was not significantly better than that from ARH(1). Therefore, the simpler ARH(1) model is also preferred for the unbalanced data.

In all cases the parameter estimates are very similar to those estimated from the balanced data even though there are anywhere from 37 to 145 more children measured in any one year.

Heart rate

Mean HR decreases as these children increase in age from 3-4 to 6-7 years while the within year variances remain fairly constant (Table 1).

For the balanced data (Table 5) the ANOVA estimates and those from the CS structure are again identical. The ICC estimate from both analyses is 0.553. The common correlation estimate is very similar (.556) when the variances

are allowed to be heterogeneous under the CSH structure. When a first-order autocorrelation structure with heterogeneous variances is fit using ARH(1) the correlation between measurements separated by one year is 0.616, by two years is 0.379 and by three years is 0.234. In these data, the correlation between successive years in the UN analysis is lower when children are young than when they are older. This correlation structure is different from the assumption of a common correlation and from that of autoregression.

Table 5. Covariances and Correlations for HR using Different Covariance Structures, Balanced Data (n=60).

Type = CS:				
<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	
81.1	44.9	44.9	44.9	
1.0	.553	.553	.553	
(Variance and covariances in top line, correlations below)				
Type = CSH:				
<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	
84.5	46.2	40.6	49.7	
.556	81.7	40.0	48.8	
.556	.556	63.2	43.0	
.556	.556	.556	94.4	
(Variances on diagonal, covariances above, correlations below)				
Type = ARH(1):				
<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	
80.4	52.1	26.7	20.4	
.616	88.9	45.6	34.7	
.379	.616	61.7	47.0	
.234	.379	.616	94.3	
(Variances on diagonal, covariances above, correlations below)				
Type = UN:				
<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	
78.6	49.1	25.2	41.9	
.598	85.6	42.0	59.2	
.360	.576	62.1	51.9	
.477	.646	.665	98.1	
(Variances on diagonal, covariances above, correlations below)				

The fit is significantly better for UN over the models where fewer parameters are estimated as seen in the LRT shown in Table 6. The use of PROC MIXED and the general variance-covariance structures it provides helps in the assessment of the properties of HR as a measure in children of this age.

Table 6. REML Likelihood Ratio Tests between covariance Structures for HR.

Balanced Data:					
<u>Type</u>	<u>Parameters</u>	<u>-2LL</u>	<u>CM±</u>	<u>df</u>	<u>χ²</u>
CS	2	1639	-	-	-
CSH	5	1635	CS	3	4
ARH(1)	5	1637	CS	3	2
UN	10	1622	CS	8	15*
UN	10	1622	CSH	5	13*
UN	10	1622	ARH(1)	5	15*

Unbalanced Data:

Type	Parameters	-2LL	CM‡	df	χ^2
CS	2	4389	-	-	-
CSH	5	4379	CS	3	10*
ARH(1)	5	4383	CS	3	6
UN	10	4364	CS	8	25*
UN	10	4364	CSH	5	15*
UN	10	4364	ARH(1)	5	19*

‡CM is the comparison model for the likelihood ratio test

* $p < .05$

The variance and covariance estimates of HR using the unbalanced data (Table 7) closely match those from the balanced data analyses. For the unbalanced data, the model containing the UN structure also had a significantly better fit ($p < .05$) than those containing the simpler CS, CSH and ARH(1) structures (Table 6).

The relationships between the HR measurements in children of this age are more clearly elucidated by examining the components of the unstructured variance-covariance matrix. In both the balanced and the unbalanced data, the variance was smaller in the third year, and the correlation was higher between the third and fourth years than between earlier adjacent years. Since this phenomenon holds for the balanced data set, it is not likely due to a selective bias in losing subjects. As children age, we may find a higher correlation between adjacent years. Note that the correlation between the first and third year is smaller than the correlation between the second and fourth year, also showing poorer tracking in the earlier years.

Table 7. Covariances and Correlations for HR using Different Covariance Structures, Unbalanced Data (n=275).

Type = CS:

<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>
95.9	52.3	52.3	52.3
1.0	.545	.545	.545

(Variance and covariances in top line, correlations below)

Type = CSH:

<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>
108.7	55.0	46.8	55.5
.537	96.3	44.0	52.2
.537	.537	69.8	41.9
.537	.537	.537	87.2

(Variances on diagonal, covariances above, correlations below)

Type = ARH(1):

<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>
108.5	63.4	31.4	21.3
.604	101.6	50.3	34.1
.365	.604	68.4	46.3
.220	.365	.604	86.1

(Variances on diagonal, covariances above, correlations below)

Type = UN:

<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>
105.7	56.8	30.4	44.0
.557	98.0	43.7	54.5
.358	.535	68.0	53.2
.451	.580	.679	90.1

(Variances on diagonal, covariances above, correlations below)

CONCLUSION

The more general multivariate models, which have a broader class of variance-covariance structures, yield results which seem to make more sense in the context of the problem. We would expect the variability of weight in children to increase from age 3-4 to age 6-7, and for the correlation of weight in adjacent years to be higher than the correlation of weight among years that are more separated. The ARH(1) structure appears to be useful in explaining how these measures track during the periods of growth exhibited by these children. Correlations between adjacent years are higher than previously thought and decrease as a power of the number of years of separation.

The unstructured nature of the relationships between HR measurements in children of this age aids in the assessment of the quality of this measure in very young children. During this period of growth a compound symmetric relationship provides a poor fit to the data.

When variances and covariances change over time, ICC (whether calculated from GLM or estimated from MIXED) is not useful as an estimate of the correlation between repeated measurements. In these settings, the question of subsampling to increase reliability of measurement is not valid. PROC MIXED gives us an opportunity to understand and quantify the inter-relationships between these measurements and to identify alternative correlational models to data sets.

REFERENCES

Baranowski T, Stone ET, Klesges RK, et al. (1993) "Studies of child activity and nutrition (SCAN): Longitudinal research on CVD risk factors and CVH behaviors in young children." *Cardiovascular Risk Factors*, 2, 4-16.

Littel, RC, Milliken, GA, Stroup, WW and Wolfinger, RD (1996) *SAS System for Mixed Models*, Cary, NC: SAS Institute, Inc.

SAS Institute Inc. (1989), *SAS/STAT User's Guide: Version 6, Fourth Edition, Volume 2*, Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1992), *SAS Technical Report P-229, SAS/STAT Software: Changes and Enhancements, Release 6.07*, Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1994), *SAS/STAT Software: Changes and Enhancements, Release 6.10*, Cary, NC: SAS Institute Inc.

Winer, BJ (1971), *Statistical Principles in Experimental Design*, Second Edition, New York: McGraw-Hill, Inc.

Wolfinger, RD (1992), "A tutorial on mixed models", Cary,

NC: SAS Institute, Inc. SAS and SAS/STAT are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Author Contact

Maribeth Johnson
Office of Biostatistics, AE-3035
Medical College of Georgia
Augusta, Georgia 30912-4900
Phone: (706) 721-3785
E-mail: maribeth@stat.mcg.edu