Paper 187-31

Analyzing Multivariate Longitudinal Data Using SAS®

Feng Gao, Paul Thompson, Chengjie Xiong, J. Philip Miller
Division of Biostatistics, Washington University School of Medicine, St. Louis, MO

ABSTRACT

Multivariate longitudinal data provides a unique opportunity in studying the joint evolution of multiple response variables over time. Comparing to the traditional univariate longitudinal data, the analysis of multivariate longitudinal data can be challenging because a) the variances of errors are likely to be different for different markers, b) the errors are likely to be correlated for the same marker measured at different occasions, and c) the errors are also likely to be correlated among markers measured at the same time. In this paper, with application to a real-world study to evaluate the joint evolution of the biomarkers for renal structure and function, we illustrate and compare 3 different approaches provided by SAS to analyze multivariate longitudinal data: the multivariate repeated measurement model with a Kronecker product covariance (PROC MIXED), the random coefficient mixed model (PROC MIXED) and the structural equation modeling approach (PROC CALIS).

INTRODUCTION

In many epidemiological studies and clinical trials, subjects are measured on several occasions with regarding to a collection of response variables. Multivariate longitudinal data of this kind would be useful to study the joint evolution of these response variables over time. Consider, as an example, the Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) (Chapman *et al* 2003). CRISP is an observational study using imaging techniques to track the progress of patients with polycystic kidney disease where the biomarkers for kidney structure and function are simultaneously measured in each subject over the time course. Similar examples include the joint modeling of CD4 and CD8 lymphocyte counts in the process of HIV infection (Thiebaut et al 2002) as well as the couple-level growth curve analysis in social sciences (Newsom 2002). However, the analysis of such a multivariate longitudinal data can be challenging because a) the variances of errors are likely to be different for different markers, b) the errors are likely to be correlated for the same marker measured at different occasions, and c) the errors are also likely to be correlated among markers measured at the same time.

In the SAS software, 3 different approaches have been provided to analyze multivariate longitudinal data: multivariate repeated measurement models with a Kronecker product covariance structure (Galecki 1994), random coefficient mixed models (Littell et al 1996) and structural equation modeling (Hatcher 1998). In this paper, we first present the CRISP data that motivates our study. Then we illustrate and compare these 3 approaches in studying the joint evolution of the renal structure and function for patients at an early stage of autosomal dominant polycystic kidney disease (ADPKD).

Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP)

Autosomal dominant polycystic kidney disease (ADPKD) is the most common genetically-caused renal disease that is characterized by gradual renal enlargement and cyst growth prior to loss of renal function. The CRISP study is a multi-center prospective observational study on subjects in early course of ADPKD. It is designed to develop innovative imaging techniques using magnetic resonance imaging (MRI) to accurately monitor the progression of cyst and renal volume. A total of 241 individuals from 211 families were enrolled in CRISP and followed up to 3

Table 1: Measurements of log-transformed KVS, CVS and GFR in CRISP

VISIT	KVS (X)			CVS (Y)			GFR (Z)		
	N	mean	SD	N	mean	SD	N	mean	SD
0	241	6.81	0.56	240	5.74	1.16	236	4.55	0.25
1	228	6.85	0.57	223	5.89	1.14	227	4.55	0.28
2	214	6.94	0.58	205	5.99	1.18	217	4.51	0.33
3	203	6.98	0.60	199	6.11	1.15	210	4.48	0.34

years, using MRI to prospectively monitor the change of the renal structure. The study population was described elsewhere (Chapman et al 2003). In this paper, we only considered two markers for renal structure, kidney volume (KVS) and cyst volume (CVS), as well as one marker for renal function -- glomerular filtration rate (GFR) as estimated by observed iothalamate clearance. Our primary interest is to assess the interrelationships among these markers. Table 1 shows the log-transformed values for each variable. For an easy notation, throughout this paper we let X, Y, and Z represent KVS, CVS and GFR respectively. We note that, as ubiquitous in longitudinal studies, not all markers

are measured at all occasions in CRISP data.

Multivariate repeated measurement models with a Kronecker product covariance

The first approach considered is to fit a model with a Kronecker product covariance structure. This model allows investigators to specify the variance-covariance matrix of the measurement error within each study unit (i.e., subject) and thus to examine the intra- and inter-marker correlations of the measurement errors. The data layout for this model has the following presentation, where the variable VISIT indicates that all the assessments are equally spaced while the variable TIME gives the exact measurement time (in month). To assure normality and homoskedasticity of the residual distribution, the response variable is defined as the change in value of marker at time t since the initial visit, i.e., $X_i(t) = \log_e KVS_i(t) - \log_e KVS_i(0)$, $Y_i(t) = \log_e CVS_i(t) - \log_e CVS_i(0)$, and $Z_i(t) = \log_e GFR_i(t) - \log_e GFR_i(0)$, for i=1, 2, …, 241.

PKDID	VNAME	VISIT	TIME	VAL
100505	KVS	1	10.6	0.07001
100505	KVS	2	25.5	0.09635
100505	KVS	3	39.0	0.09666
100505	CVS	1	10.6	0.03182
100505	CVS	2	25.5	0.19006
100505	CVS	3	39.0	0.27320
100505	GFR	1	10.6	0.12562
100505	GFR	2	25.5	-0.08201
100505	GFR	3	39.0	-0.03200

For models with a Kronecker product covariance, SAS only provides the possibility to fit bivariate models. The model is implemented by the *REPEATED* statement in the PROC MIXED and currently SAS only provides 3 alternative covariance structures: UN@CS, UN@AR(1) and UN@UN. For example, the following SAS codes fit a model with a covariance of UN@AR(1) for the relationship between KVS and CVS, where the *COVTEST* option requests significance tests for the covariance estimates (i.e., the random effects portion of the model), the *NOCLPRINT* option suppresses the display of the "Class Level Information" table, and the *NOINT* option in the *MODEL* statement suppresses the intercept terms because we have $X_i(0) = Y_i(0) = 0$ for each subject i. Note that the variable VISIT rather than TIME is used in the *MODEL* statement, and thus assumes that all patients have equally spaced assessments. The variable VISITC in the *REPEATED* statement takes exactly the same values as in VISIT, but VISITC is treated as a class variable.

```
proc mixed data=CRISP_MIX covtest noclprint;
  title1  "Model 1.1 Mixed model with a Kronoker product covariance";
  title2  "KVS (X) versus CVS (Y)";
  class pkdid vname visitc;
  model val=vname*visit/s noint;
  repeated vname visitc /type=un@ar(1) subject=pkdid r rcorr;
  where vname="KVS" or vname="CVS";
run;
```

The option *TYPE=UN@AR(1)* in the *REPEATED* statement specifies that the covariance matrix within an individual has the following structure:

$$V \otimes \Sigma = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} \otimes \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix} = \begin{pmatrix} \sigma_x^2 * \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix} & \sigma_{xy} * \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix} \\ \sigma_{xy} * \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix} & \sigma_y^2 * \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix} \end{pmatrix}.$$

Thus, the above model assumes that a) the two markers share a common intra-marker correlation as measured by $\Sigma = \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}$, and b) the inter-maker correlations (as measured by $\rho_{xy} = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2 \sigma_y^2}}$) are the same for two markers measured at the same time point. The two markers are independent if the matrix V has the form of

$V = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}$, and the hypothesis H_0: $\sigma_{xy} = 0$ will be tested by the *COVTEST* option. Results of the fitted mixed models are shown in Table 2, where the slope parameter reflects the average annual change of the marker over time. Since our primary interest focuses on the interrelationships among these markers, for simplicity, only the estimated correlation coefficients out of the random effect portion are listed in Table 2. Based on the fit

Table 2: Bivariate mixed models with a Kronoker product covariance

Features	KVS (X) and CVS (Y)		KVS (X) and GFR (Z)		CVS (Y) and GFR (Z)	
	UN@AR(1)	UN@UN	UN@AR(1)	UN@UN	UN@AR(1)	UN@UN
Fixed effects						
$Slope_x$	0.0513**	0.0529**	0.0510**	0.0522**	--	--
$Slope_y$	0.1133**	0.1128**	--	--	0.1144**	0.1125**
$Slope_z$	--	--	-0.0179*	-0.0171*	-0.0182*	-0.0182*
Random effects						
ρ_{xy}	0.377**	0.374**	--	--	--	--
ρ_{xz}	--	--	0.018	0.058	--	--
ρ_{yz}	--	--	--	--	-0.049	-0.044
Fit statistics						
-2 LnL	-2278	-2348	-1716	-1851	-798	-860
AIC	-2270	-2332	-1708	-1835	-790	-844
BIC	-2256	-2304	-1694	-1807	-776	-816

*p<0.05, **p<0.001

statistics (smaller indicating a better fit), those models with UN@UN covariance structure provide a better fit to the data. The results show that, in patients with AKPKD, the values of KVS and CVS increase over time while GFR decreases over time. All these changes are significantly different from zero (with p<0.001 for KVS and CVS, p<0.05 for GFR respectively). The results also reveal a strong positive correlation between KVS and CVS (ρ_{xy}=0.374 with p<0.001), but the correlation between KVS and GFR (ρ_{xz}=0.058, p=0.159) as well as the correlation between CVS and GFR (ρ_{yz}=-0.044, p=0.305) are not statistically significant. We note that these correlations index the extra associations among biomarkers after removing the effect of involution process over time.

Random coefficient mixed models

Instead of modeling the variation *within* study unit as in the repeated measurement models, the random coefficient mixed models assume that the regression coefficients are a random sample from some population of possible coefficient and allow one to model variations *between* study units (Littell et al 1996). In the presence of multiple response variables, a separate set of regression coefficients will be fitted for each response variable and the correlations among these random coefficients can be examined. In SAS this model is also implemented with PROC MIXED and the data layout of the model is exactly the same as in previous section. The following SAS codes fit a random coefficient mixed model for KVS and CVS. We note that the variable TIME rather than VISIT is used in the model, indicating its capability to handle unequally spaced measurements.

```
proc mixed data=CRISP_MIX covtest noclprint;
  title1  "Model 2.1 Mixed model with random coefficients";
  title2  "KVS (X) versus CVS (Y)";
  class pkdid vname;
  model val=vname*time/s noint;
  random vname*time/type=un subject=pkdid g gcorr;
  repeated /type=VC group=vname subject=pkdid;
  where vname="KVS" or vname="CVS";
run;
```

The *RANDOM* statement requests that two random slopes (one for KVS and one for CVS) be fitted for each individual and the *GROUP* option in the *REPEATED* statement specifies that the variances of measurement errors are different for different markers. The *G* and *GCORR* options in the *RANDOM* statement require the display of covariance and correlation matrix for the random slopes respectively, with $G = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}$, and the two slopes are independent

if $\sigma_{xy} = 0$. The results of mixed models with random coefficients are listed in Table 3 where the slope parameter represents the average change per month for each marker over time. For simplicity, only the estimated correlation coefficients out of the random effect portion are listed in Table 3. Results show that KVS and CVS increase over time (p<0.001) while GFR decreases (p<0.05). There exists a strong positive correlation between KVS and CVS (ρ_{xy}=0.647 with p<0.001), but the correlation between KVS and GFR (ρ_{xz}=-0.063, p=0.446) as well as the correlation between CVS and GFR (ρ_{yz}=-0.036, p=0.694) are not statistically significant. However, it should be pointed out that the correlation coefficients in Table 3 represent the associations among random slopes (i.e., the trajectory of markers over time) rather than measurement errors. We also note that this approach is easily extendable to multivariate longitudinal data with more than two response variables.

Table 3: Heterogeneous mixed models with random coefficients

Features	(X, Y)	(X, Z)	(Y, Z)	(X, Y, Z)
Fixed effects				
$Slope_x$	0.0042**	0.0042**	--	0.0042**
$Slope_y$	0.0099**	--	0.0098**	0.0099**
$Slope_z$	--	-0.0011*	-0.0011*	-0.0011*
Random effects				
ρ_{xy}	0.647**	--	--	0.647**
ρ_{xz}	--	-0.063	--	-0.063
ρ_{yz}	--	--	-0.036	-0.024
Fit statistics				
-2 LnL	-2256	-1832	-722	-2443
AIC	-2246	-1822	-712	-2423
BIC	-2229	-1804	-694	-2393

* p<0.05, ** p<0.001

Structural equation models

A structural equation model (SEM) is a hypothesized pattern of directional and non- directional relationships among a set of observed and unobserved variables (MacCallum and Austin 2000). SAS usually works on covariance or correlation matrix to fit a structural equation model. Since the two renal structural markers (KVS and CVS) show an almost perfect correlation (with a correlation coefficient $r_{xy} \geq 0.93$ at every time point), in this section our interest only focuses on the relationships between renal structure and renal function. A set of structural equation models will be fitted for KVS versus GFR and CVS versus GFR separately. The following SAS codes present the correlation matrix that is used as the input data. Note that the *NOMISS* option in PROC CORR excludes those subjects with missing assessments and thus assures a non-singular covariance matrix (Hatcher 1998). As a consequence, only 184 subjects are left to fit the structural equation models for KVS and GFR.

```
proc corr data=CRISP_SEM nomiss out=SEM_XZ noprint;
  var x0 x1 x2 x3 z0 z1 z2 z3;
 run;
 run;
 proc print data=SEM_XZ;
  title "Correlation matrix of KVS versus GFR";
 run;
```

TYPE	_NAME_	x0	x1	x2	x3	z0	z1	z2	z3
MEAN		6.854	6.891	6.965	7.001	4.546	4.553	4.519	4.482
STD		0.563	0.576	0.595	0.614	0.232	0.266	0.315	0.328
N		184	184	184	184	184	184	184	184
CORR	x0	1.000	0.994	0.991	0.985	-0.364	-0.450	-0.481	-0.540
CORR	x1	0.994	1.000	0.996	0.990	-0.368	-0.440	-0.475	-0.534
CORR	x2	0.991	0.996	1.000	0.994	-0.365	-0.440	-0.473	-0.542
CORR	x3	0.985	0.990	0.994	1.000	-0.371	-0.446	-0.478	-0.540
CORR	z0	-0.364	-0.368	-0.365	-0.371	1.000	0.659	0.683	0.641
CORR	z1	-0.450	-0.440	-0.440	-0.446	0.659	1.000	0.717	0.757
CORR	z2	-0.481	-0.475	-0.473	-0.478	0.683	0.717	1.000	0.815
CORR	z3	-0.540	-0.534	-0.542	-0.540	0.641	0.757	0.815	1.000

Figure 1: Path diagram of a cross-lagged model to describe the interrelationship between KVS and GFR in CRISP data

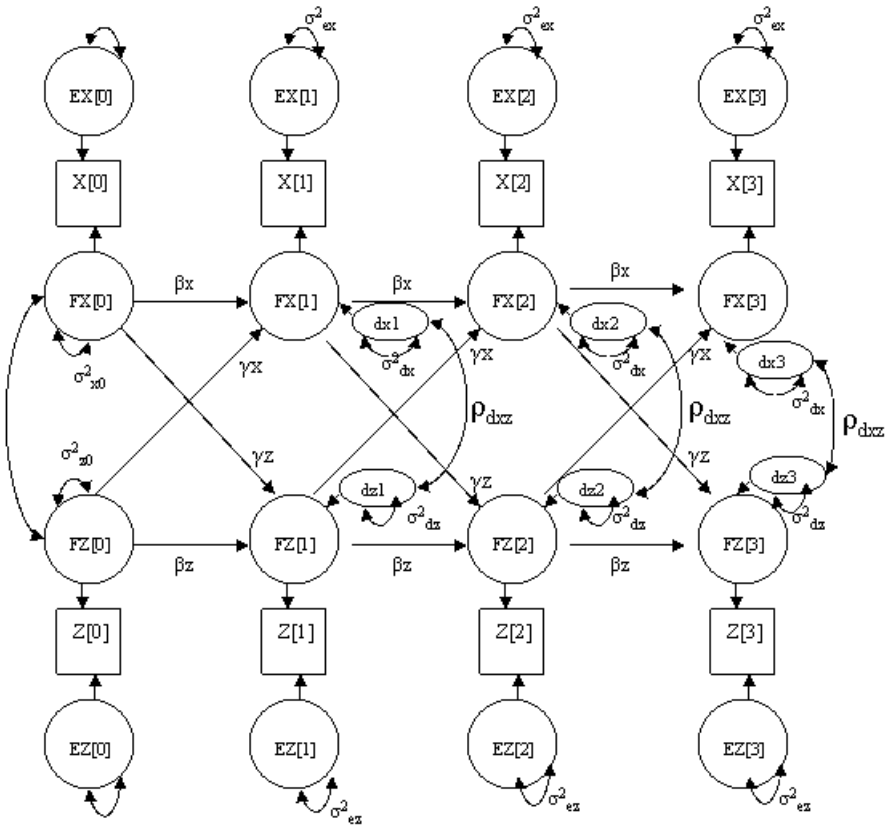

Figure 1 depicts the model we used to assess the relationship between KVS and GFR where, following the SEM notation practice, a square is used to represent an observed (measured) variable while a circle or an oval is used for an unobserved (latent) variable. This model is a classical SEM for bivaraite longitudinal data (Ferrer and McArdle 2003). Let X[t] and Z[t] denote the measured variables at time t, let FX[t] and FZ[t] be the latent variables reflecting the corresponding (unobserved) true values, let EX[t] and EZ[t] represent the measurement errors, and finally let DX[t]

and DZ[t] denote the residual terms. Then, following the classical test theory, we assume that the observed variables are a function of the true variables and the measurement errors, i.e.,

X[t] = FX[t] + EX[t] and Z[t] = FZ[t] + EZ[t], for t = 0, 1, 2, 3.

Next, we specify the interrelations between the two response variables at the level of the (unobserved) true values. We assume that the true value at time t is a function of three components: a) an auto-regression parameter (β_x or β_z) that describes the effect of the same variable at the previous measurement, b) a cross-lagged regression parameter (γ_x or γ_z) which is the effect of the other variable at previous time, and c) a residual term (DX[t] or DZ[t]) at time t, i.e.,

FX[t] = β_x * FX[t-1] + γ_x * FZ[t-1] + DX[t] and

FZ[t] = β_z * FZ[t-1] + γ_z * FX[t-1] + DZ[t], for t = 1, 2, 3.

The following SAS codes fit this structural equation model using PROC CALIS. The *LINEQS* statement gives the linear equations described above, the *STD* statement identifies the variables whose variances are to be estimated, the *COV* statement identifies pairs of variables that are expected to be correlated, and the *VAR* statement presents these observed variables. Note that this classical model assumes a) all regression parameters (β_x, β_z, γ_x and γ_z) are constant over lags, b) the variances of the measurement errors (σ^2_{ex} and σ^2_{ez}) and the variances of the residuals (σ^2_{dx} and σ^2_{dz}) are constant over time, and c) the correlations between the two residual terms (ρ_{dxz}) are also constant over time.

```
proc calis data=SEM_XZ;
  title1  "Model 3.1 Structural Equation Modeling";
  title2  "KVS (X) versus GFR (Z): with cross lag-effects";
  lineqs
   x0=fx0+ex0,   x1=fx1+ex1, x2=fx2+ex2, x3=fx3+ex3,
   z0=fz0+ez0,   z1=fz1+ez1, z2=fz2+ez2, z3=fz3+ez3,
   fx1=betax fx0+ gammax fz0 + dx1,
   fx2=betax fx1+ gammax fz1 + dx2,
   fx3=betax fx2+ gammax fz2 + dx3,
   fz1=betaz fz0+ gammaz fx0 + dz1,
   fz2=betaz fz1+ gammaz fx1 + dz2,
   fz3=betaz fz2+ gammaz fx2 + dz3;
  std
   fx0 fz0=sigmaex0 sigmaez0,
   ex1-ex3=3*sigmaex,
   ez1-ez3=3*sigmaez,
   dx1-dx3=3*sigmadx,
   dz1-dz3=3*sigmadz;
  cov
   dx1 dz1 = rhoxz,
   dx2 dz2 = rhoxz,
   dx3 dz3 = rhoxz;
  var x0 x1 x2 x3 z0 z1 z2 z3;
run;
```

Table 4 lists the results of 4 alternative SEMs for the relationship between KVS and GFR: a) a model with cross-lagged effects for both markers ($\gamma_x \neq 0, \gamma_z \neq 0$), b) a model with only a cross-lagged effect from KVS to GFR ($\gamma_x = 0, \gamma_z \neq 0$), c) a model with only a cross-lagged effect from GFR to KVS ($\gamma_x \neq 0, \gamma_z = 0$), and d) a model without any cross-lagged effects ($\gamma_x = \gamma_z = 0$). Unlike traditional statistical methods which usually use one statistical test to determine the adequacy of overall model fit, SEM relies on several fit statistics. Comparative Fit Index (CFI) and Root Mean Square Error of Approximation (RMSEA) are most frequently used for such a purpose. Both CFI and RMSEA range from 0 to 1. A model with CFI of 0.9 or more would be considered acceptable while a RMSEA of 0.06 or less indicates an adequate fit (Hu and Bentler 1999). The examination of the fit statistics indicates that the model with $\gamma_x = 0$ and $\gamma_z \neq 0$ fits the data best though there are no large differences among the 4 alternative models. The CFI values indicate an adequate fit to the data but the corresponding RMSEA values are all above the 0.06 critical value. The (latent) true measurements of KVS and GFR at time t=1, 2, 3 can be described as,

```
fx1 =  0.9948 * fx0                    + 1.0000 dx1
fx2 =  0.9948 * fx1                    + 1.0000 dx2
fx3 =  0.9948 * fx2                    + 1.0000 dx3
fz1 = -0.1700 * fx0  + 0.7445 * fz0    + 1.0000 dz1
fz2 = -0.1700 * fx1  + 0.7445 * fz1    + 1.0000 dz2
```

fz3 = -0.1700 * fx2 + 0.7445 * fz2 + 1.0000 dz3.

The result shows that a) the auto-regression parameters are significant for both KVS and GFR, and b) the cross-lag effect of KVS to GFR is significant but the cross-lag effect of GFR to KVS is not. That means the measurement of GFR at time t tends to be small if a large value of KVS is observed at time t-1. In average, for example, 1 unit increase in $\log_e$(KVS) at time t-1 is associated with 0.17 unit decrease of $\log_e$ (GFR) at time t. The results also show that the correlation between the measurement errors in KVS and GFR is not significant. A similar conclusion is also obtained regarding the relationship between CVS and GFR (Table 5).

Table 4: Alternative structural equation models for KVS and GFR (N=184)

Features	$\gamma_x \neq 0, \gamma_z \neq 0$	$\gamma_x = 0, \gamma_z \neq 0$	$\gamma_x \neq 0, \gamma_z = 0$	$\gamma_x = \gamma_z = 0$
Regression effects				
β_x	0.9916**	0.9948**	0.9948**	0.9973**
β_z	0.7410**	0.7445**	0.8217**	0.8242**
γ_x	-0.0074	--	-0.0059	--
γ_z	-0.1716**	-0.1700**	--	--
Variances				
ρ_{dxz}	0.1101	0.1030	0.1059	0.1092
Fit statistics				
Parameters	11	10	10	9
Degree of freedom	25	26	26	27
CFI	0.9731	0.9728	0.9601	0.9591
RMSEA	0.1339	0.1320	0.1617	0.1588
AIC	57	57	98	98
X^2	107	109	150	152

* p<0.05, **p<0.001

Table 5: Alternative structural equation models for CVS and GFR (N=167)

Features	$\gamma_y \neq 0, \gamma_z \neq 0$	$\gamma_y = 0, \gamma_z \neq 0$	$\gamma_y \neq 0, \gamma_z = 0$	$\gamma_y = \gamma_z = 0$
Regression effects				
β_y	0.9965**	0.9949**	0.9942**	0.9931**
β_z	0.7529**	0.7543**	0.8318**	0.8328**
γ_y	-0.0036	--	-0.0025	--
γ_z	-0.1535**	-0.1530**	--	--
Variances				
ρ_{dyz}	-0.0741	-0.0673	-0.0895	-0.0894
Fit statistics				
Parameters	11	10	10	9
Degree of freedom	25	26	26	27
CFI	0.9553	0.9555	0.9436	0.9439
RMSEA	0.1692	0.1655	0.1862	0.1822
AIC	94	92	124	122
X^2	144	144	176	176

* p<0.0.5, **p<0.001

Comparison of the 3 approaches for multivariate longitudinal data modeling

The 3 different approaches considered in this paper explore different aspects of the joint evolution of multiple response variables, and thus each approach has its own advantages and limitations. Models with Kronecker product covariance provide a convenient way to fit bivariate data and enable one to examine both inter- and intra-marker correlations of the measurement errors. However, this model possesses several apparent limitations, namely, a common intra-marker correlation for different markers, a constant inter-marker correlation for markers measured at the same time point, as well as the demanding for equally spaced measurements. Furthermore, SAS only provides the possibility to fit bivariate mixed models.

Models with random coefficients provide an opportunity to examine correlations among the trajectories of markers over time and to capture the growth of these response variables. As exemplified by the CRISP data, this approach is capable of handling unequally spaced measurements and is easily extendable to multivariate models with more than two response variables. Another advantage of these mixed-model approaches is that they allow incomplete observations and thus can use information more efficiently. As long as the missing data are at random, incomplete observations have little influence on mixed models (Littell et al 1996).

The third approach, structural equation modeling (SEM), is a powerful and flexible tool to explore the interrelationships among a set of response variables. SEM treats several traditional multivariate procedures (including regression analysis, factor analysis, discriminant analysis, etc.) as special cases. SEM can be readily applied to nested longitudinal data as commonly seen in ophthalmology studies. It can assess the directional and non-directional relationships among observed and unobserved variables. SEM recognizes the imperfect nature of the measurement and explicitly specifies error terms for all measurements. Since SEM is a multivariate technique in nature, this approach can be easily extended to models with more than two response variables. SEM offers no default model structure and the hypothesized interrelationships among the variables need to be explicitly specified by a researcher. Therefore, it provides more flexibility and allows investigators to examine different aspects of the interrelationships such as cross-lagged effects, the latent growth, or both (Ferrer and McArdle 2003). However, sometimes it is not an easy task to construct an appropriate model, especially for data with complicated structures. One limitation of PROC CALIS is that observations with missing values for any variables in the analysis are excluded from the computations. This will not only result in a considerable loss of information but may also introduce potential bias for interpretation. For instance, only 184 out of 241 subjects are used to explore the relationship between KVS and GFR while only 167 out of 241 subjects are used for the modeling of CVS and GFR.

CONCLUSION

For patients in the early stage of ADPKD, kidney volumes and cyst volumes steadily increase over time while the values of GFR decrease over time. There exists a strong positive correlation between the two renal structural markers. The correlations between the structural markers (KVS and CVS) and renal functional marker (GFR) are not significant in either measurement errors or random slopes, but KVS and CVS show a negative cross-lagged effect on GFR, i.e., the higher the value of KVS (CVS) at time t-1, the lower the GFR measurement at time t.

REFERENCES

Chapman AB, Guay-Woodford LM, Grantham JJ, Torres VE, Bae KT, Baumgarten DA, Kenney PJ, King BF, Glockner JF, Wetzel LH, Brummer ME, O'Neill WC, Robbin ML, Bennett WM, Klahr S, Hirschman GH, Kimmel PL, Thompson PA, Miller JP (2003). Renal structure in early autosomal-dominant polycystic kidney disease (ADPKD): the Consortium for Radiology Imaging Study of Polycystic Kidney Disease (CRISP) cohort. *Kidney International*, 64:1035-1045.

Ferrer E and McArdle JJ (2003). Alternative structural models for multivariate longitudinal data analysis. *Structural Equation Modeling*, 10:493-524

Galecki, AT (1994). General class of covariance structures for two or more repeated factors in longitudinal data analysis. *Communications in Statistics - Theory and Methods*, 23:3105-3119

Hu L and Bentler PM (1999). Cutoff criteria for fit index in covariance structural analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6:1-55.

Hatcher L (1998). *A step-by step approach to using the SAS System for factor analysis and structural equation modeling,* Cary, NC: SAS Institute Inc

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

MacCallum R and Austin JT (2000). Application of Structural Equation Modeling in psychological research. *Annual Review of Psychology*, 51:201-226

Newsom JT (2002). A multilevel structural equation model for dyadic data. *Structural Equation Modeling*, 9:431-447

Thiebaut R, Jacqmin-Gadda H, Chene G, Leport C and Commenges D (2002).
Bivariate linear mixed models using SAS proc MIXED. *Computer Methods and Programs in Biomedicine*, 69: 249-256

ACKNOWLEDGMENTS
This study is partially supported by the National Institute of Health grants 5U01DK6240104 and U01-DK56956 for J. Philip Miller and 5U10 EY09341 for Mae Gordon

CONTACT INFORMATION
Your comments and questions are valued and encouraged. Address all correspondences to:

Feng Gao, Ph.D.
Division of Biostatistics, Campus Box 8067
Washington University in St. Louis
St. Louis, MO 63110, U.S.A
Phone: (314)-362-3682
Fax: (314)-362-3728
Email: feng@wustl.edu