

# Quantitative Family Data Analysis

International Family Nursing Conference 2009

George J. Knaf1, PhD

School of Nursing

University of North Carolina - Chapel Hill

[gknaf1@unc.edu](mailto:gknaf1@unc.edu)



# Summary

- I. overview of different types of family data and alternative family data analysis methods
- II. selected analyses of actual family data including
  - summary statistics
  - regression
  - analysis of variance
  - cluster analysis
  - factor analysis
- III. detailed analyses using SPSS and SAS
  - in the breakout session

# Quantitative Family Data Analysis

## Part I

### Overview of Family Data Analysis

# Number of Measurements per Family

- family data are often dyadic
  - e.g., mother's and father's assessments of family functioning
  - e.g., mother's assessments of behavioral problems for a chronically ill child and a healthy sibling
- but can involve more than two family members
  - e.g., adolescents assessments of family functioning along with assessments from mothers and fathers
- or multiple family dimensions
  - e.g., mother's and father's assessments of behavioral problems for an ill child and a healthy sibling

# Number of Members per Family

- family data can involve different numbers of family members for different families
- e.g., a survey of parents of a chronically ill child can involve
  - single-mother, single-father, and two-parent families
  - the two-parent families can have
    - only mothers participating, only fathers participating, or both mothers and fathers participating

# Family Outcomes

- outcome (dependent, response,  $y$ ) variables can be the same for all family members
  - e.g., a child's behavioral problems from both the mother and the father
- or different for different family members
  - e.g., the mother's depressive symptoms and a child's behavioral problems

# Family Data Analyses

- family data need standard kinds of analyses
  - e.g., summary statistics, regression, analysis of variance, analysis of covariance
- standard methods inappropriate for analyzing combined data for all family members
  - they assume that outcome measurements for different subjects are independent
  - but outcome measurements for members of the same family are usually correlated

# Family Data Analyses

- can be more complex than for individuals
- predictor (independent, explanatory, x) variables can provide both actor and partner effects on outcomes
  - e.g., a parent's assessment of family functioning provides an actor effect for predicting that parent's assessment of a child's behavioral problems and a partner effect for predicting the spouse's assessment of those behavioral problems
  - part of the Actor Partner Interdependence Model



# Number of Time Points

- family data can be cross-sectional
  - i.e., measured at one point in time
  - analyses then need to account for intra-familial correlation (IFC) between outcome measurements for members of the same family
- or longitudinal
  - i.e., measured at multiple points in time
  - analyses then need to account for temporal correlation as well as for IFC

# Simplified Outcomes

- outcomes for each family can be aggregated and the aggregated data analyzed with standard methods
  - e.g., averaging mother's and father's assessments of a child's behavioral problems
- can result in substantial loss of information on variability within families
- can produce non-comparable data
  - e.g., averages for different numbers of family members have different variances and so should not be analyzed with standard methods that assume constant variances

# Separate Analyses

- standard methods can be applied to the data for each family member separately
  - e.g., separate averages of child behavioral problems as assessed by mothers and by fathers
- ignores the IFC, an important property of the family as a unit
  - the family unit is more than just the aggregate of its individual family members
  - family data analyses should reflect this by addressing the combined data for all family members
- can also lose power for testing hypotheses

# Combined Analyses

- to test for differences in means for  $Y$  = child behavioral problems using combined data for mothers and fathers
- could try a standard regression model

$$Y = a + b \cdot \text{FATHER} + e$$

- where FATHER is an indicator for being a father
  - i.e., it equals 1 for fathers and 0 for mothers
- the intercept  $a$  is the mean for mothers
- the slope  $b$  is how much the mean for fathers differs from the mean for mothers
  - so you want to test for  $b = 0$
- $e$  is a mean zero random error term
- problem: the errors  $e$  are independent for different subjects, which usually does not hold for mothers and fathers from the same families

# Adjusting the Errors

- start with a standard regression model

$$Y = a + b \cdot \text{FATHER} + e$$

- the intercept  $a$  is the mean for mothers
  - the slope  $b$  is how much the mean for fathers differs from the mean for mothers
- add in a mean zero random term  $u$ , one for each family and independent for different families

$$Y = a + b \cdot \text{FATHER} + e + u$$

- the new error terms  $e' = e + u$  are correlated within families and the correlation is the IFC
- use this model to test for  $b = 0$

# Multilevel (Hierarchical Linear) Model

- first level is standard regression model

$$Y = a + b \cdot \text{FATHER} + e$$

– intercept  $a$  and slope  $b$  are fixed coefficients

- second level changes one or more of the fixed coefficients to random coefficients

- in this case, change fixed intercept  $a$  to random intercept  $a + u$  (so  $a$  is average intercept) giving

$$Y = \mathbf{a + u} + b \cdot \text{FATHER} + e = a + b \cdot \text{FATHER} + \mathbf{e + u}$$

- correlation for  $e' = e + u$  called the intraclass correlation coefficient in general but is the IFC for family data

# Linear Mixed Model

$$Y = a + b \cdot \text{FATHER} + e'$$

- has fixed component  $a + b \cdot \text{FATHER}$  modeling means and random component  $e'$  modeling variances/correlations (or equivalently, covariances)
- random errors  $e'$  are independent for different families but correlated for different members of the same family
- $e'$  generated by a random intercept works for any number of family members and produces
  - same variance for all family members and
  - same correlation for all pairs of distinct family members
  - so it is called compound symmetry

# Adjustments to Random Component

- compound symmetry is the model used in standard repeated measures analyses
- same variances for all family members can be too simplistic for family data
  - can be changed to heterogeneous variances
- same correlations can be too simplistic for data based on more than two family members
  - can be changed to unstructured correlations with different IFCs for each pair of family members
- unstructured correlations with heterogeneous variances is model used in multivariate ANOVA



# Adjustments to Fixed Component

- should account for differences in means for
  - different types of family members
    - e.g., fathers and mothers
  - different types of families
    - e.g., one-parent versus two-parent families
  - different numbers of participating family members
    - e.g., only mother, only father, or both participating
  - can do this with indicator variables like FATHER
- can also add predictor variable(s) X

# Types of Outcomes

- linear mixed models apply to continuous family outcomes that can be treated as normally distributed
  - e.g., depressive symptoms for family care givers/receivers
  - in SAS use PROC MIXED
- categorical family outcomes require generalized linear modeling
  - e.g., a high level of depressive symptoms or not for family care givers/receivers
  - can use generalized estimating equations (GEE) techniques to account for IFC
  - in SAS use PROC GENMOD

# Data Structure

- family data are often stored as one record per family with separate variables containing measurements for different family members
- then they need to be restructured for analysis

**from**

**to**

FAMID	YMOTHER	YFATHER	FAMID	FATHER	Y
1	10	12	1	0	10
2	15	10	1	1	12
...			2	0	15
			2	1	10
			...		

# Handling Longitudinal Data

- to understand how to handle longitudinal family data, need to first understand how to handle non-family longitudinal data
  - e.g., a mother's assessments of a child's behavioral problems  $Y$  at multiple time points  $T$

MOMID	T	Y
1	1	80
1	2	82
1	3	83
...		

# Longitudinal Data

- the multilevel model starts with a standard regression model in time  $T$

$$Y = a + b \cdot T + e$$

- and replaces fixed slope  $b$  by a random slope  $b + v$  (so  $b$  is the average slope) where  $v$  has mean zero and is independent across subjects but correlated within times for each subject

$$Y = a + (\mathbf{b + v}) \cdot T + e = a + b \cdot t + \mathbf{e + v \cdot T}$$

- new error terms  $e' = e + v \cdot T$  are correlated within times for the same subject

# Linear Mixed Model

$$Y = a + b \cdot T + e'$$

- where the error terms  $e'$  have mean zero, are independent across subjects and correlated within times for the same subject
- $e' = v \cdot T + e$  is the case generated by the multilevel model
- autoregression is a common alternative for  $e'$ 
  - treats correlations as weakening the farther apart outcomes are in time
    - i.e., the correlation for outcomes at times  $t$  and  $t'$  for the same subject equal  $\rho^{|t - t'|}$  where  $\rho$  is called the autocorrelation

# Handling Longitudinal Family Data

- e.g., mother's and father's assessments of child's behavioral problems  $Y$  at multiple time points  $T$

FAMID	FATHER	T	Y
1	0	1	80
1	0	2	82
1	0	3	83
1	1	1	82
1	1	2	83
1	1	3	84
...			

# Longitudinal Family Data

- multilevel model starts with standard regression model in time T and FATHER

$$Y = a + b_1 \cdot T + b_2 \cdot \text{FATHER} + e$$

- and replaces fixed intercept a and slope  $b_1$  for T by random coefficients  $a + u$  and  $b_1 + v$

$$\begin{aligned} Y &= \mathbf{a + u} + (\mathbf{b_1 + v}) \cdot T + b_2 \cdot \text{FATHER} + e \\ &= a + b_1 \cdot T + b_2 \cdot \text{FATHER} + \mathbf{e + u + v \cdot T} \end{aligned}$$

- new error terms  $e' = e + u + v \cdot T$  are correlated within members of same family and within times for the same family member



# Linear Mixed Model

$$Y = a + b \cdot T + b_2 \cdot \text{FATHER} + e'$$

- the error terms  $e'$  are independent across families and correlated within members of the same family and within times for the same family member
- $e' = u + v \cdot T + e$  is the case generated by the multilevel model
- two-dimensional correlation structures are alternatives for  $e'$ 
  - e.g., unstructured correlations within family members and autoregressive correlations within times
  - supported by SAS but not by SPSS

# Types of Outcomes

- linear mixed models apply to continuous longitudinal family outcomes that can be treated as normally distributed
  - e.g., depressive symptoms for family care givers/receivers over multiple time points
  - in SAS use PROC MIXED
- categorical longitudinal family outcomes require generalized linear modeling
  - e.g., high levels of depressive symptoms or not for family care givers/receivers over multiple time points
  - can use GEE techniques to account for both IFC and temporal correlation
  - but it can be complicated to run the analyses
  - in SAS use PROC NLMIXED or PROC GLIMMIX

# Summary for Part I

- family data
  - can involve different numbers of family members and different types of families
  - can be cross-sectional or longitudinal
- family outcomes
  - can be the same or different for different family members
  - can be continuous or categorical
  - can be analyzed with mixed models accounting for
    - differences in means and standard deviations
    - IFC and possibly also temporal correlation
    - dependence on one or more predictor variables

# Quantitative Family Data Analysis

## Part II

### Selected Family Data Analyses

# Child Adaptation Data

- cross-sectional study of parents of a chronically ill child
- available measures
  - Y = child adaptation in intensity of behavioral problems (Eyberg Child Behavioral Inventory)
    - larger values mean more problems and so worse child adaptation
  - X = family functioning (McMaster Family Assessment Device)
    - larger values mean poorer levels of family functioning
- 324 two-parent families
  - with mothers participating for all families
  - with fathers also participating for 145 (44.8%) of the families

# Summary Statistics - Child Adaptation

$$Y = a + b_1 \cdot \text{FATHER} + b_2 \cdot \text{MOMONLY} + e'$$

- $a$  is mean adaptation for mothers with participating fathers
- $b_1$  is change in mean for fathers compared to mothers of the same families
  - FATHER is 1 if a father and 0 if a mother
- $b_2$  is change in mean for mothers without participating fathers compared to mothers with participating fathers
  - MOMONLY is 1 if a mother without a participating father and 0 if a mother with a participating father or a father
- $e'$  = errors independent across families, correlated within families, with different variances for mothers/fathers

# Summary Statistics - Child Adaptation

$$Y = a + b_1 \cdot \text{FATHER} + b_2 \cdot \text{MOMONLY} + e'$$

- estimated mean adaptation  $a$  for mothers with participating fathers was 87.5
- estimated change in mean  $b_1$  for fathers compared to mothers of the same families was 2.8 and was not significant ( $p=.16$ )
- estimated change in mean  $b_2$  for mothers without participating fathers compared to mothers with participating fathers was  $-1.5$  and was not significant ( $p=.65$ )
- estimated standard deviations were close at 29.0 and 28.4 for mothers and fathers, respectively
- estimated IFC was .65 and significant ( $p<.01$ )
- mean child adaptation did not change with type of family and type of family member but the IFC was substantial

# Child Adaptation vs. Family Functioning

$$Y = a + b_1 \cdot \text{FATHER} + b_2 \cdot \text{MOMONLY} + b_3 \cdot X + e'$$

- $X$  = family functioning
- $a + b_3 \cdot X$  is how mean child adaptation changes with family functioning for mothers with participating fathers
- $b_1$  is how much this is shifted for fathers compared to mothers of the same families
- $b_2$  is how much this is shifted for mothers without participating fathers compared to mothers with participating fathers
- $e'$  = errors independent across families, correlated within families, with different variances for mothers/fathers



# Child Adaptation vs. Family Functioning

$$Y = a + b_1 \cdot \text{FATHER} + b_2 \cdot \text{MOMONLY} + b_3 \cdot X + e'$$

- the estimated slope  $b_3$  for family functioning  $X$  was 16.4 and was significant ( $p < .01$ )
  - mean child adaptation got worse (larger values or more behavioral problems) with poorer family functioning (larger values)
- this relationship was not significantly shifted for fathers ( $p = .08$ ) or for mothers without participating fathers ( $p = .57$ )
- estimated standard deviations were close (27.8 for mothers and 27.2 for fathers)
- estimated IFC was substantial at .63 ( $p < .01$ )

# Summary of Adaptation Analyses

- as expected from the literature better/worse family functioning corresponded to better/worse adaptation of a child to a chronic illness
- this relationship as well as overall means were the same for mothers and fathers within the same families and for mothers with and without participating fathers
  - but in general this will not be the case
- variances did not differ much with family member
  - but in general this will not be the case
- the IFC was substantial
  - it was important to account for this correlation
  - treating it as zero would have been very inappropriate

# Other Possible Analyses

- if single mothers are also included in the study

$$Y = a + b_1 \cdot \text{FATHER} + b_2 \cdot \text{MOMONLY} + b_3 \cdot X + b_4 \cdot \text{SINGLE} + e'$$

- where SINGLE is 1 if single mother and 0 otherwise
- single fathers handled in the same way

- mediation of effect of predictor X by a mediator M

- X should have an effect on M

$$M = a + b_1 \cdot \text{FATHER} + b_2 \cdot \text{MOMONLY} + b_3 \cdot X + e'$$

- the effect  $b_3$  for X on Y should be reduced by adding in M

$$Y = a + b_1 \cdot \text{FATHER} + b_2 \cdot \text{MOMONLY} + b_3 \cdot X + b_4 \cdot M + e'$$

# Other Possible Analyses

- moderation of the effect of X for fathers compared to mothers

$$Y = a + b_1 \cdot \text{FATHER} + b_2 \cdot \text{MOMONLY} + b_3 \cdot X + b_4 \cdot \text{FATHER} \cdot X + e'$$

$$Y = a + b_1 \cdot \text{FATHER} + b_2 \cdot \text{MOMONLY} + (b_3 + b_4 \cdot \text{FATHER}) \cdot X + e'$$

- introduce an interaction term  $\text{FATHER} \cdot X$
- $b_4$  is how much the slope for X changes for fathers compared to mothers
- moderation of the effect of X for mothers without participating fathers handled in the same way
  - separately from or together with moderation of the effect of X for fathers

# Family Functioning Data

- cross-sectional study of parents of child with genetic condition
  - including PKU, CF, neurofibromatosis, sickle cell disease, thalassemia, hemophilia, and Marfan's syndrome
- two available measures of family functioning
  - $Y_1$  = satisfaction (Family APGAR)
    - larger values mean more satisfied with family life
  - $Y_2$  = hardiness (Family Hardiness Index)
    - larger values mean higher levels of hardiness (strength and durability)
- 52 two-parent families with both parents participating

# Family Functioning Clusters

- wanted to classify family functioning for these 52 families
  - using both  $Y_1$  = satisfaction and  $Y_2$  = hardiness for both mothers and fathers
- should not cluster the  $2 \cdot 52 = 104$  vectors  $(Y_1, Y_2)$  for mothers and fathers combined
  - clustering methods allow for correlation within a vector for the same subject, but assume vectors independent across subjects
- can cluster the 52 vectors  $(Y_{1,mother}, Y_{2,mother})$  for mothers separately from the 52 vectors  $(Y_{1,father}, Y_{2,father})$  for fathers
  - but this will not take into account similarities and differences for mothers and fathers of the same families
- clustered the 52 combined vectors for both parents
  - i.e.,  $(Y_{1,mother}, Y_{2,mother}, Y_{1,father}, Y_{2,father})$

# Generated Clusters

- Ward's method generated 5 family functioning types
  - well-adapted families (21 or 40.4%) with both parents tending to rate both satisfaction and hardiness as high
  - discrepant families (10 or 19.2%) with mothers tending to rate satisfaction and hardiness as high but with fathers tending to rate them both as moderate
  - diminished families with both parents tending to rate satisfaction and/or hardiness as moderate
    - with mothers diminished more in satisfaction (8 or 15.4%)
    - or with mothers diminished more in hardiness (11 or 21.2%)
  - compromised families (2 or 3.8%) with both parents tending to rate satisfaction and hardiness as low

# Validity of the Clusters

- validated the clusters using
  - parental quality of life (QOL) (Quality of Life Index)
  - child functional status (Functional Status II)
  - reduced clusters to 3 so no small clusters
    - by combining the diminished and compromised clusters into a single diminished/compromised cluster
- considered ANOVA models for these variables
  - with main effects to type of parent and to cluster as well as an interaction effect to type of parent and cluster
    - expected means to change with clusters
  - using linear mixed models with error terms correlated within families and independent across families



# Cluster Validity Results

- mean QOL changed significantly with cluster ( $p < .01$ ) and with interaction ( $p = .01$ ) but not with type of parent ( $p = .06$ )
  - a post hoc analysis revealed that mean QOL was lower for diminished/compromised parents than for well-adapted parents with the difference being greater for diminished/compromised mothers than for diminished/compromised fathers (joint  $p < .05$ )
- mean child functional status changed significantly with cluster ( $p = .01$ ) but not with type of parent ( $p = .37$ ) or with the interaction ( $p = .47$ )
  - a post hoc analysis revealed that child functional status was lower for diminished/compromised families than for the well-adapted and the discrepant families
- these results supported the validity of the clusters

# Issues in Clustering Family Data

- these analyses used data from families for which there were two participating parents
  - data from families with only mothers participating were not used in the analyses
  - since hierarchical clustering procedures (e.g., SAS PROC CLUSTER) supporting Ward's method drop partial vectors
- possible to include families with partial vectors
  - k-means clustering (e.g., SAS PROC FASTCLUS) automatically adjusts the computation of distances between vectors of different sizes
  - hierarchical clustering procedures (e.g., SAS PROC CLUSTER) can use predetermined distances (e.g., computed with SAS PROC DISTANCE)

# Family Instrument Development Data

- cross-sectional study of parents of a child with a chronic condition to develop a survey instrument for measuring aspects of family management of the chronic condition
- 65 items were developed based on the Family Management Style Framework
- 579 parents from 417 families were surveyed including
  - 414 mothers with 349 partnered and 65 single
  - 165 fathers with mothers participating for all but 3 fathers
- only 1% (324 out of the  $579 \cdot 65 = 37,635$ ) of the item values were missing, so these were imputed

# Factor Analysis

- factor analysis models assume independence for item responses for different subjects
  - and so should not be applied to combined item responses for mothers and fathers
- 8 of the 65 items addressed parental mutuality which only applies to partnered parents
  - so single mothers responded to only 57 of the items
  - factor analysis procedures would drop such partial sets of item responses
  - imputation of parental mutuality items not sensible for single mothers
- so we factor analyzed only the item responses for the 349 partnered mothers
  - we considered mothers the primary family member

# Family Management Measure

- 6 scales were produced based on 53 of the 65 items
  - child's daily life
  - condition management ability
  - condition management effort
  - family life difficulty
  - parental mutuality (only for partnered parents)
  - view of condition impact
- special methods were used to generate the scales that also allowed us to justify that those scales were appropriate for use by fathers and by single mothers
- internal consistency reliability, test-retest reliability, and construct validity were based on data for all 579 parents using specialized linear mixed models accounting for IFC and differences for family members and for types of families

# Summary of Parts I-II

- have provided an overview of family data and the kinds of analyses they require
- have also provided examples of actual analyses of selected family data
- methods are available for analyzing combined data for family members that account for IFC in most situations
  - factor analysis is an exception
- in many situations, these analyses require only fairly straightforward linear mixed models
  - but some cases require more sophisticated techniques and the assistance of a statistician
- however, it is important to address IFC in analyses of family data whenever possible
- the breakout session will cover the details of generating some of these analyses in SPSS and SAS

# Quantitative Family Data Analysis

## Part III

### Detailed Family Data Analyses

## Overview of Part III

- linear mixed modeling in SPSS 17.0
  - with demonstration of how to run analyses
- linear mixed modeling in SAS 9.2
  - with demonstration of how to run analyses
- comparison of SPSS and SAS for family data analysis



# Linear Mixed Modeling in SPSS

# Example Data with Alternate Structures

Wide Structure				Long Structure			
FAMID	Y0	Y1	MOMONLY	FAMID	FATHER	Y	MOMONLY
1	3.5	4.5	0	1	0	3.5	0
2	2.1	3.6	0	1	1	4.5	0
3	2.5	.	1	2	0	2.1	0
4	4.5	4.9	0	2	1	3.6	0
5	2.3	3.0	0	3	0	2.5	1
6	2.8	.	1	3	1	.	1
7	2.0	2.6	0	4	0	4.5	0
8	2.8	.	1	4	1	4.9	0
9	2.6	2.4	0	5	0	2.3	0
10	2.1	.	1	5	1	3.0	0
				6	0	2.8	1
				6	1	.	1
				7	0	2.0	0
				7	1	2.6	0
				8	0	2.8	1
				8	1	.	1
				9	0	2.6	0
				9	1	2.4	0
				10	0	2.1	1
				10	1	.	1

Note. Y0 contains outcome values for mothers and Y1 contains outcome values for fathers; FATHER=0 indicates a mother; FATHER=1 indicates a father; MOMONLY=0 indicates either a mother with a participating father or a father; MOMONLY=1 indicates a mother without a participating father; "." denotes a missing value.

# Alternate Data Structures

- the wide (or broad) data structure
  - one record (row, case, observation) per family
  - one outcome variable (column) for each type of family member
  - other variables for predictors
  - common way family data are stored
- the long data structure
  - one record per family member within each family
  - an extra variable to indicate the family member
  - one outcome variable for all family members
  - other variables for predictors
  - needed for linear mixed modeling

# Restructuring Data in SPSS

- use Data/Restructure... option to convert data from wide to long structure if necessary
  - click on "Next >" for default of "Restructure selected variables into cases"
  - click on "Next >" for default of "One" for "How many variable groups do you want to restructure?"
  - set "Case Group Identification" to "Use selected variable" and set "Variable:" to "FAMID"; set "Target Variable:" to "Y", "Variables to be Transposed" to Y0 and Y1, and "Fixed Variable(s):" to MOMONLY; and click on "Next >"
  - click on "Next >" for default of "One" for "How many index variables do you want to create?"; leave "What kind of index values?" at the default of "Sequential numbers"; set name of Index Variable to "FATHER" and its label to "Father"; then click on "Next >"
  - click on "Finish" (no need for the extra steps); give variable "Y" the label "Outcome"; save data under a new file name like "long.sav" (use File/Save As not File/Save since that writes over the source file)
  - can delete all cases with missing Y values since they will be dropped from analyses anyway

# SPSS Analysis of Example Data

- use Analyze/Mixed Models/Linear... option to generate the model for data in long structure
  - "Subjects:" in this case are families, so set to the FAMID variable
  - "Repeated:" is set to the FATHER variable
  - **choose "Compound Symmetry: Heterogeneous" (CSH) for "Repeated Covariance Type:"**
  - **set "Dependent Variable:" to Y**
  - **set "Covariates:" to FATHER and MOMONLY (not "Factor(s):")**
  - click on "Fixed...", change default of "Factorial" to "Main Effects", mark FATHER and MOMONLY and click on "Add" (one at a time of both), then click on "Continue"
  - click on "Statistics..." and request "Parameter estimates" and "Tests for covariance parameters", then click on "Continue"
  - click on "OK" to run the model

# CSH Fixed Component Estimates

$$Y = a + b_1 \cdot \text{FATHER} + b_2 \cdot \text{MOMONLY} + e'$$

- a (Intercept) is the mean for mothers with fathers participating
  - estimate is 2.83
- $b_1$  (FATHER) is how much the mean changes for fathers
  - estimate is 0.67 and is significant ( $p=.033$ )
- $b_2$  (MOMONLY) is how much the mean changes for mothers without participating fathers
  - estimate is  $-0.28$  and is nonsignificant ( $p=.597$ )

Estimates of Fixed Effects<sup>a</sup>

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	2.833333	.326705	8.509	8.672	.000	2.087727	3.578940
FATHER	.666667	.231585	5.205	2.879	.033	.078357	1.254976
MOMONLY	-.283333	.516566	8.509	-.548	.597	-1.462241	.895574

a. Dependent Variable: Y.

# CSH Random Component Estimates

$$Y = a + b_1 \cdot \text{FATHER} + b_2 \cdot \text{MOMONLY} + e'$$

- error terms  $e'$ 
  - have mean zero and are independent across families
  - are correlated within families
    - estimate of IFC is .78 (CSH rho) and is significant ( $p < .001$ )
  - have different variances for different family members
    - estimates are 0.64/0.80 for mothers/fathers (FATHER = 0/1)
    - take square roots to get standard deviations (0.80/0.90)

Estimates of Covariance Parameters<sup>a</sup>

Parameter	Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval		
					Lower Bound	Upper Bound	
Repeated Measures	Var: [FATHER=0]	.640417	.310479	2.063	.039	.247624	1.656272
	Var: [FATHER=1]	.803662	.385724	2.084	.037	.313713	2.058801
	CSH rho	.782179	.148921	5.252	.000	.290463	.947100

a. Dependent Variable: Y.

# Impact of IFC on Fixed Component

- using "Diagonal", the default for "Repeated Covariance Type:"
  - variances change with family member
  - but IFC equals zero
  - also called variance components
- $b_1$  (FATHER) has the same estimate (0.67) but is no longer significant ( $p=.236$ )
- ignoring IFC can result in loss of power
- so change back to the CSH model

Estimates of Fixed Effects<sup>a</sup>

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	2.833333	.326705	8	8.672	.000	2.079950	3.586716
FATHER	.666667	.530474	10.522	1.257	.236	-.507400	1.840734
MOMONLY	-.283333	.516566	8	-.548	.598	-1.474536	.907870

a. Dependent Variable: Y.



# ANOVA versus Regression

- entering FATHER and MOMONLY as covariates generates a regression model
- an equivalent ANOVA model is generated by entering them as factors instead
  - remove them from "Covariate(s)", set them to "Factor(s)", click on the "Fixed..." button, and reset the fixed component
- but has a different fixed component

# ANOVA Fixed Component Estimates

- 0's/1's handled oppositely than in regression model
- intercept has different estimate (3.22)
  - corresponds to being a father (FATHER = 1) and being a mother without a participating father (MOMONLY = 1) which makes no sense for these data
- signs reversed for other 2 effects but p-values same
- do get same random component estimates

Estimates of Fixed Effects<sup>b</sup>

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	3.216667	.462316	12.913	6.958	.000	2.217212	4.216122
[FATHER=0]	-.666667	.231585	5.205	-2.879	.033	-1.254976	-.078357
[FATHER=1]	0 <sup>a</sup>	0	.	.	.	.	.
[MOMONLY=0]	.283333	.516566	8.509	.548	.597	-.895574	1.462241
[MOMONLY=1]	0 <sup>a</sup>	0	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

b. Dependent Variable: Y.

# Alternative ANOVA Model

- to get exactly the same model
  - create  $MOTHER = 1 - FATHER$  and  $TWOPRNTS = 1 - MOMONLY$ 
    - use Transform/Compute Variable...
  - change "Factor(s)" to these two variables with 0's/1's reversed, click on the "Fixed..." button, and reset the fixed component
- now get same estimates and p-values
  - since  $MOTHERS = 1$  and  $TWOPRNTS = 1$  means mothers with participating fathers and so intercept is mean for that case

Estimates of Fixed Effects<sup>b</sup>

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	2.833333	.326705	8.509	8.672	.000	2.087727	3.578940
[MOTHER=0]	.666667	.231585	5.205	2.879	.033	.078357	1.254976
[MOTHER=1]	0 <sup>a</sup>	0	.	.	.	.	.
[TWOPRNTS=0]	-.283333	.516566	8.509	-.548	.597	-1.462241	.895574
[TWOPRNTS=1]	0 <sup>a</sup>	0	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

b. Dependent Variable: Y.

# Alternative Random Components

- are the estimates of variance (0.64/0.80) distinctly different or not?
  - if not, can change to "Compound Symmetry: Correlation Metric" (CSR) to request homogeneous variances
- mixed modeling procedure does not provide a test
- can address this issue with a likelihood ratio test (LRT)
  - using likelihoods ( $L_{CSH}$ ,  $L_{CSR}$ ) for the 2 cases
- compare the following difference
$$-2 \cdot \log(L_{CSR}) - (-2 \cdot \log(L_{CSH}))$$

to 3.841, the cutoff for 5% significance for the  $\chi^2$  test with 1 degree of freedom (DF)

  - use DF = 1, since models differ by 1 parameter: 1 vs. 2 variances
  - increase DF if models differ by more parameters

# Variance Alternatives

- compare the following difference  

$$-2 \cdot \log(L_{CSR}) - (-2 \cdot \log(L_{CSH})) = 32.708 - 32.473 = 0.235$$
 to 3.841, the cutoff for 5% significance for the  $\chi^2$  test with 1 degree of freedom (DF)
- nonsignificant since  $0.235 < 3.841$ 
  - not a large enough increase to be of significance
- there is no need for these data to treat variances as different for mothers and fathers

## CSH

<b>-2 Restricted Log Likelihood</b>	<b>32.473</b>
Akaike's Information Criterion (AIC)	38.473
Hurvich and Tsai's Criterion (AICC)	41.140
Bozdogan's Criterion (CAIC)	43.168
Schwarz's Bayesian Criterion (BIC)	40.168

The information criteria are displayed in smaller-is-better forms.

a. Dependent Variable: Y.

## CSR

<b>-2 Restricted Log Likelihood</b>	<b>32.708</b>
Akaike's Information Criterion (AIC)	36.708
Hurvich and Tsai's Criterion (AICC)	37.908
Bozdogan's Criterion (CAIC)	39.838
Schwarz's Bayesian Criterion (BIC)	37.838

The information criteria are displayed in smaller-is-better forms.

a. Dependent Variable: Y.

# CSR Fixed Component Estimates

$$Y = a + b_1 \cdot \text{FATHER} + b_2 \cdot \text{MOMONLY} + e'$$

- $a$  = mean for mothers with fathers participating
  - estimate is 2.83
- $b_1$  = change in mean for fathers
  - estimate is 0.67 and is significant ( $p=.029$ )
- $b_2$  = change in mean for mothers without participating fathers
  - estimate is  $-0.28$  and is nonsignificant ( $p=.615$ )
- conclusions the same
  - estimates also same (won't always happen) but p-values not

Estimates of Fixed Effects<sup>a</sup>

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	2.833333	.344500	9.614	8.224	.000	2.061547	3.605120
FATHER	.666667	.227155	5.522	2.935	.029	.098960	1.234373
MOMONLY	-.283333	.544702	9.614	-.520	.615	-1.503635	.936969

a. Dependent Variable: Y.

# CSR Random Component Estimates

$$Y = a + b_1 \cdot \text{FATHER} + b_2 \cdot \text{MOMONLY} + e'$$

- error terms  $e'$ 
  - have mean zero and are independent across families
  - are correlated within families
    - estimate of IFC is .78 (CSR rho) and significant ( $p < .001$ )
  - have same variances for different family members
    - estimate is 0.71 (CSR diagonal)
    - take square root to get standard deviation (0.84)

Estimates of Covariance Parameters<sup>a</sup>

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Repeated Measures	CSR diagonal	.712080	.324775	2.193	.028	.291274	1.740831
	CSR rho	.782612	.149827	5.223	.000	.286080	.947818

a. Dependent Variable: Y.

# Other Compound Symmetry Model

- change "Repeated Covariance Type:" to "Compound Symmetry" (CS) and rerun model
  - the model induced by a random intercept
  - equivalent to "Compound Symmetry: Correlation Metric" (CSR)
- get same fixed component estimates and p-values

Estimates of Fixed Effects<sup>a</sup>

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	2.833333	.344500	9.614	8.224	.000	2.061547	3.605120
FATHER	.666667	.227155	5.522	2.935	.029	.098960	1.234373
MOMONLY	-.283333	.544702	9.614	-.520	.615	-1.503635	.936969

a. Dependent Variable: Y.



# CS Random Component Estimates

- get equivalent random component estimates
- to get the variance estimate add two estimates (CS diagonal offset + CS covariance)
  - =  $.154798 + .557283 = .712081$  (.712080 for CSR)
- to get the IFC divide the second estimate by the variance estimate (CS covariance / variance)
  - =  $.557283 / .712081 = .782612$  (.782612 for CSR)
- more convenient to use CSR than CS

Estimates of Covariance Parameters<sup>a</sup>

Parameter	Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval		
					Lower Bound	Upper Bound	
Repeated Measures	CS diagonal offset	.154798	.093162	1.662	.097	.047587	.503551
	CS covariance	.557283	.321969	1.731	.083	-.073764	1.188330

a. Dependent Variable: Y.

# Random Intercept Model

- to generate the random intercept model directly
  - change "Repeated Covariance Type:" to "Scaled Identity"
    - for independent errors  $e$  as used in standard regression models
  - click on "Random...", request "Intercept included" to add a random intercept  $u$ , and set "Combinations:" to FAMID
    - errors are now  $e' = e + u$
- get same model as "Compound Symmetry" (CS)
- with same fixed component estimates/p-values

Estimates of Fixed Effects<sup>a</sup>

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	2.833333	.344500	9.614	8.224	.000	2.061547	3.605120
FATHER	.666667	.227155	5.522	2.935	.029	.098961	1.234372
MOMONLY	-.283333	.544702	9.614	-.520	.615	-1.503635	.936969

a. Dependent Variable: Y.

# Random Intercept Model

- also get equivalent random component estimates
- use same computations as for CS
  - to get the variance estimate add two estimates  
(Repeated Measures Variance + Intercept Variance)  
=  $.154798 + .557283 = .712081$
  - to get the IFC divide the second estimate by the variance estimate (Intercept Variance / variance)  
=  $.557283 / .712081 = .782612$
- as before, more convenient to use CSR

Estimates of Covariance Parameters<sup>a</sup>

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Repeated Measures	Variance	.154798	.093162	1.662	.097	.047587	.503550
Intercept [subject = FAMID]	Variance	.557283	.321969	1.731	.083	.179596	1.729239

a. Dependent Variable: Y.

# Other Repeated Covariance Types

- Unstructured
  - heterogeneous variances and unstructured correlations
  - same as CSH for two family members
  - for more than two family members get different IFC for each pair of family members
- Unstructured: Correlation Metric
  - same as Unstructured, but output in terms of correlations rather than covariances, so more readily interpreted
- AR(1)
  - 1st order autoregression with homogeneous variances
  - useful for non-family longitudinal data
- AR(1): Heterogeneous
  - AR(1) with heterogeneous variances

# Maximum Likelihood (ML) Estimation

- linear mixed modeling procedures use ML to estimate parameters
- they use the full information version of ML
  - meaning that all available outcome values are used without loss of data
  - and without imputation of missing outcome values
- very important for analyzing family data
  - which usually have different numbers of outcomes for different families
- data still lost when predictors have missing values

# Restricted ML (REML) vs. ML

- ML generates biased estimates of covariances
- default procedure in SPSS uses REML that generates unbiased estimates of covariances
  - can be changed to ML in SPSS by clicking on "Estimation..."
  - REML more important for small sample sizes
  - should not be much difference for large sample sizes
- okay to base LRTs on REML likelihoods when comparing models with same fixed components and different random components
- should base LRTs on ML likelihoods for comparing models with different fixed components and either same or different random components

# SPSS Syntax

- SPSS has a windows-based user interface but it also has an underlying programming language
  - SPSS calls code syntax
- once a model has been fully specified, click on the "Paste" button, and save the generated code to a syntax file
- when opened, the code is read into the syntax editor
  - can be changed and related models run with less effort
- the following is code for the CSH model with covariates

```
DATASET ACTIVATE DataSet1.
MIXED Y WITH FATHER MOMONLY
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(5) SCORING(1)
    SINGULAR(0.00000000001) HCONVERGE(0,ABSOLUTE)
    LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=FATHER MOMONLY | SSTYPE(3)
  /METHOD=REML
  /PRINT=SOLUTION TESTCOV
  /REPEATED=FATHER | SUBJECT(FAMID) COVTYPE(CSH).
```

# Linear Mixed Modeling in SAS



# Creating a Long Data Set in SAS

- assume that a data set has been created in the SAS work library called WIDEDATA
- the following code creates the data set LONGDATA from the WIDEDATA data set
  - with two observations (SAS name for rows) for each family, one each for the mother and father
  - a variable FATHER and a copy called FAMMEMBR
  - but only for fathers with non-missing Y1 values

```
data longdata;
  set widedata;
  Y=Y0; FATHER=0; FAMMEMBR=FATHER; output;
  if Y1^=. then do;
    Y=Y1; FATHER=1; FAMMEMBR=FATHER; output;
  end;
  drop Y0 Y1;
run;
```

# Invoking PROC MIXED

- use the following code to request a CSH model

```
proc mixed data=LONGDATA ...;  
  class FAMID FAMMEMBR;  
  model Y=FATHER MOMONLY ...;  
  repeated FAMMEMBR / subject=FAMID type=CSH;  
run;
```
- the variable (FAMMEMBR) listed in the REPEATED statement must also be listed in CLASS statement
  - matches family members of the same type across families
- variables listed in the CLASS statement are treated in the MODEL statement as ANOVA factors
- use FATHER in the MODEL statement instead of FAMMEMBR to request a regression model

# Requesting Parameter Estimates

- adjust the code as follows

```
proc mixed data=LONGDATA covtest;
```

```
  class FAMID FAMMEMBR;
```

```
  model Y=FATHER MOMONLY / solution;
```

```
  repeated FAMMEMBR / subject=FAMID type=CSH;
```

```
run;
```

- "COVTEST" in PROC MIXED statement requests estimates and tests of random component parameters
- "/ SOLUTION" in MODEL statement request estimates and tests for fixed component parameters
- SAS also uses REML by default
  - add "METHOD=ML" to PROC MIXED statement to change to standard ML estimation

# CSH Fixed Component Estimates

$$Y = a + b_1 \cdot \text{FATHER} + b_2 \cdot \text{MOMONLY} + e'$$

- a (Intercept) is mean for mothers with fathers participating
  - estimate is 2.83
- $b_1$  (FATHER) is how much mean changes for fathers
  - estimate is 0.67 and is significant ( $p=.035$ )
- $b_2$  (MOMONLY) is how much mean changes for mothers without participating fathers
  - estimate is  $-0.28$  and is nonsignificant ( $p=.598$ )
- same estimates and conclusions but p-values different
  - SAS uses different degrees of freedom (DFs) values than SPSS

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr >  t
<b>Intercept</b>	<b>2.8333</b>	0.3267	8	8.67	<.0001
<b>FATHER</b>	<b>0.6667</b>	0.2316	5	2.88	<b>0.0346</b>
<b>MOMONLY</b>	<b>-0.2833</b>	0.5166	8	-0.55	<b>0.5983</b>

# CSH Random Component Estimates

$$Y = a + b_1 \cdot \text{FATHER} + b_2 \cdot \text{MOMONLY} + e'$$

- error terms  $e'$ 
  - have mean zero and are independent across families
  - are correlated within families
    - estimate of IFC is .78 (labeled CSH)
  - have different variances for different family members
    - estimates are 0.64/0.80 for mothers/fathers (labeled Var(1)/Var(2) for FAMMEMBR = 0/1)
    - take square roots to get standard deviations
- same estimates as SPSS but different p-values

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
<b>Var(1)</b>	FAMID	<b>0.6404</b>	0.3202	2.00	<b>0.0228</b>
<b>Var(2)</b>	FAMID	<b>0.8037</b>	0.4258	1.89	<b>0.0296</b>
<b>CSH</b>	FAMID	<b>0.7822</b>	0.1510	5.18	<b>&lt;.0001</b>

# Differences in P-Values

- SPSS always uses Satterthwaite's method for computing the denominator DFs for model (DDFM) parameters
- SAS uses different approach
  - allocates DFs to between-subjects and within-subjects effects
  - the option "DDFM=BETWITHIN" for the MODEL statement
- to get same p-values as SPSS, add "DDFM=SATTERTH" to the MODEL statement
- SAS supports 3 other DDFM options
- literature does not agree on how to compute p-values
- in most cases, the conclusions should be the same
  - but if not, use a LRT instead

# Homogeneous Variance Model

- change the repeated statement to  
repeated FAMMEMBR / subject=FAMID type=**CS**;
- "TYPE=CS" requests homogeneous variances
- estimates of fixed parameters readily identified in output, but not estimates of random parameters  
– need same computations as for CS option in SPSS

## Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr >  t
<b>Intercept</b>	<b>2.8333</b>	0.3445	8	8.22	<.0001
<b>FATHER</b>	<b>0.6667</b>	0.2272	5	2.93	<b>0.0324</b>
<b>MOMONLY</b>	<b>-0.2833</b>	0.5447	8	-0.52	<b>0.6170</b>

## Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
<b>CS</b>	FAMID	<b>0.5573</b>	0.3220	1.73	<b>0.0835</b>
<b>Residual</b>		<b>0.1548</b>	0.09316	1.66	<b>0.0483</b>

# Homogeneous Variance Model

- change the repeated statement to  
repeated FAMMEMBR / subject=FAMID type=CS **r=1 rcorr=1;**
- use "R=1 RCORR=1" to have SAS compute estimates of the variance and of the IFC
  - "R=1" denotes the covariance matrix for the 1st family
    - estimate of common variance on the diagonal (0.7121)
  - "RCORR=1" denotes correlation matrix for the 1st family
    - estimate of IFC on the off-diagonal (0.7826)
  - use 1 when the 1st family has two participating parents
    - change it if necessary

Estimated R Matrix for FAMID 1			Estimated R Correlation Matrix for FAMID 1		
Row	Col1	Col2	Row	Col1	Col2
1	<b>0.7121</b>	0.5573	1	1.0000	<b>0.7826</b>
2	0.5573	<b>0.7121</b>	2	<b>0.7826</b>	1.0000



# Random Intercept Model

- change from a repeated to a random statement  
random intercept / subject=FAMID v=1 vcorr=1;
- get equivalent model based on a random intercept
  - with same estimates for fixed and random components
  - but different p-values
- uses "V=1 VCORR=1" like "R=1 RCORR=1" to have SAS compute the variance and IFC estimates

## Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr >  t
<b>Intercept</b>	<b>2.8333</b>	0.3445	8	8.22	<.0001
<b>FATHER</b>	<b>0.6667</b>	0.2272	5	2.93	<b>0.0324</b>
<b>MOMONLY</b>	<b>-0.2833</b>	0.5447	5	-0.52	<b>0.6252</b>

## Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
<b>Intercept</b>	FAMID	<b>0.5573</b>	0.3220	1.73	<b>0.0417</b>
<b>Residual</b>		<b>0.1548</b>	0.09316	1.66	<b>0.0483</b>

# Differences in P-Values

- SAS uses a different approach for computing denominator degrees of freedom for models based on the RANDOM statement than for models based on the REPEATED statement
- to get the same results
  - add the option "DDFM=BETWITHIN" to the MODEL statement when using a RANDOM statement
- in most cases, the conclusions should be the same
  - if not, use an LRT instead

# Comparison of SPSS and SAS

# Other Covariance Types

- SAS "TYPE=UN" same as SPSS Unstructured
- SAS "TYPE=UNR" same as SPSS Unstructured: Correlation Metric
- SAS "TYPE=AR(1)" same as SPSS AR(1)
- SAS "TYPE=ARH" same as SPSS AR(1): Heterogeneous
- SAS "TYPE=VC" (variance components) same as SPSS Diagonal
- SAS supports other choices which SPSS also supports
- SAS supports other choices not supported by SPSS
- SAS also supports different covariance estimates for different groups of families but not SPSS
  - e.g., families with a male vs. a female chronically ill child

# Non-Family Longitudinal Data

- assume there is a variable  $T$  with values indicating times at which outcomes were measured
- can use a random coefficient model
  - include a random slope for  $T$  but a not random intercept
- can also use autoregressive models
  - AR(1) or ARH
    - with homogeneous or heterogeneous variances
- both approaches supported by SPSS and SAS

# Longitudinal Family Data

- assume there are variables
  - T with values indicating times outcomes were measured
  - FAMMEMBR with values for different family members
- can use a random coefficient model
  - include random slope for T and a random intercept
  - supported by SPSS and SAS
- SAS supports 2-dimensional structures but not SPSS
  - e.g., "TYPE=UN@AR" combines unstructured correlations for FAMMEMBR and AR(1) correlations for T
- SAS supports spatial autoregression but not SPSS
  - one dimension for time plus multiple family member dimensions like mother/father and ill/healthy child

Knafli, Knafli, & McCorkle (2005)

# Random Coefficient Models

- supported by both SPSS and SAS for multilevel modeling
- standard case starts with independent errors  $e$  as in standard regression models
  - called "Scale Identity" in SPSS
  - default in SAS when there is only a RANDOM statement
- can also combine random coefficients with non-independent errors  $e'$ 
  - supported in SPSS by changing from Scaled Identity
  - supported in SAS by including both a RANDOM and a REPEATED statement
  - but more alternatives for  $e'$  in SAS than in SPSS

# Categorical Cross-Sectional Family Data

- in SPSS, use Analyze/Generalized Linear Models/Generalized Estimating Equations...  
option
  - can use "Exchangeable" (constant correlations) for the same IFC for all pairs of family members
  - "Unstructured" for different IFCs for each pair of family members
  - AR(1) for categorical non-family longitudinal data
- in SAS, use PROC GENMOD
  - same three options available for correlations



# Categorical Longitudinal Family Data

- not fully supported by SPSS
  - could use 1-dimensional unstructured correlations for all combinations of times and family members
  - but this ignores 2-dimensionality of data
- SAS supports 2-dimensional correlations
  - use PROC NLMIXED
    - for nonlinear mixed modeling
  - or PROC GLIMMIX
    - for generalized linear mixed modeling
    - have to download this from SAS website

# Summary of Part III

- have demonstrated the use of SPSS and SAS for conducting a variety of analyses of cross-sectional dyadic family outcomes
- these analyses cover important cases that every family researcher ought to be familiar with
- other important types of analyses have not been demonstrated including
  - analyses of longitudinal family outcomes
  - analyses of categorical family outcomes
- these are more complicated and may require the assistance of a statistician

# Selected References

- Campbell, L., & Kashy, D. A. (2002). Estimating actor, partner, and interaction effects for dyadic data using PROC MIXED and HLM: A user-friendly guide. *Personal Relationships, 9*, 327-342.
- Cook, W. L., & Kenny, D. A. (2005). The Actor-Partner Interdependence Model: A model of bidirectional effects in developmental studies. *International Journal of Behavioral Development, 29*, 101-109.
- Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York: The Guilford Press.
- Knafl, G. J., Dixon, J. K., O'Malley, J. P., Grey, M., Deatrck, J. A., Gallo, A., & Knafl, K. A. (2009). Analysis of cross-sectional univariate measurements for family dyads using linear mixed models. *Journal of Family Nursing, 15*, 130-151. DOI: 10.1177/1074840709331641.
- Knafl, G. J., Knafl, K. A., & McCorkle, R. (2005). Mixed models incorporating intra-familial correlation through spatial autoregressive methods. *Research in Nursing and Health, 28*, 348-356.
- Knafl, K., Knafl, G., Gallo, A., & Angst, D. (2007). Parents' perceptions of functioning in families having a child with a genetic condition. *Journal of Genetic Counseling, 16*, 481-492.
- Knafl, K., Deatrck, J. A., Gallo, A., Dixon, J. D., Grey, M., Knafl, G., & O'Malley, J. (in press). Assessment of the psychometric properties of the Family Management Measure. *Journal of Pediatric Psychology*. DOI: 10.1093/jpepsy/jsp034.
- Lyons, K. S., & Sayer, A. G. (2005). Longitudinal dyad models in family research. *Journal of Marriage and Family, 67*, 1048-1060.
- Rayens, M. K., & Svavarsdottir, E. K. (2003). A new methodological approach in nursing research: An actor, partner, and interaction effect model for family outcomes. *Research in Nursing and Health, 26*, 409-419.
- Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth curves. *Journal of Educational and Behavioral Statistics, 24*, 323-355.