Canadian Journal of Nursing Research, Vol. 27, No. 1

Methodological Commentary

Characteristics of Families– Implications for Statistical Analysis in Family Nursing Research

Sean P. Clarke

Research dealing with family phenomena of concern to nursing carries inherent methodological challenges, due in large part to special characteristics of families as social units (Copeland & White, 1991; Daly, 1992; Larzelere & Klein, 1987). Shared history and close contact between family members result in measures obtained from more than one member of the same family being statistically dependent. However, families are composed of distinct individuals and measures from different family members cannot be collapsed or "averaged" without losing important data about each person's unique experience. Perceptions and beliefs of family members are strongly influenced by individual values and by roles in families (determined by age, gender, generation, and family history). Also, families adapt and evolve over time. Describing the shape and determinants of change in families is a priority, but measuring change remains very difficult. Since families are embedded in larger social systems, complex methods of data collection and analysis are often required to capture the reciprocal influences of the environment, the family, and the individual on each other. While these aspects of families have implications for all phases of the research process, many of these characteristics pose major challenges for statistical analysis (Schumm, Barnes, Bollman, Jurich, & Milliken, 1985).

Some family researchers have turned to special analysis techniques, for example path analysis and structural equation modeling (LISREL), to assist them in making causal links and verifying theories (Biddle & Marlin, 1987). Ultimately, clinicians and policymakers look to nursing research for answers about which factors are associated with (and perhaps even causally related to) desired or undesired outcomes for families. Testing propositions of emerging theories of family adaptation is a major goal in family nursing research (McCubbin, 1991). The "softness" of the variables many family researchers work with, and the impossibility of manipulating most family conditions experimentally, interfere with conclusively linking variables, yet methods like structural equation

Sean P. Clarke, N., M.Sc.(A.), is a Doctoral Candidate in the School of Nursing at McGill University, Montreal, Quebec.

modeling are far from simple solutions to these problems (Biddle & Marlin, 1987; Cliff, 1983).

Statistical analysis is often seen as a matter of applying mathematical manipulations to data sets in a objective manner in order to arrive at unambiguous answers to research questions (Hornstein, 1988). The truth, however, is that data analysis involves an extensive set of subjective decisions regarding which statistical techniques to use, which variables to analyze, how these variables should be selected and possibly transformed for analysis, and how the results should be interpreted. Making sound choices in data analysis requires a familiarity not only with the practical aspects of statistics, but also an understanding of common issues in data analysis in one's field. Equipped with both of these types of knowledge, the nurse researcher can make informed choices about statistics throughout the planning and execution of a study. In this paper, some statistical implications of research involving three aspects of family phenomena will be discussed.

To provide a running example of choices in statistical analysis available in family research, reference will be made to Stewart's (1990) longitudinal study of family adaptation prior to and following the birth of a second child. Stewart and his colleagues followed 41 families over a 15-month period. Interview questions, pencil-and-paper questionnaires, and observations of family interactions were used to assess various dimensions of family adjustment including parental stress and responsibility for household and child-care tasks. Stewart's work is notable for its strong theoretical base (i.e. family systems theory and McCubbin and Patterson's Double ABC-X model of family adaptation) and innovative data collection techniques, as well as the proficient use of statistical analysis. The study is also of particular interest to family nursing research because Stewart's aim was to capture healthy family functioning and adaption and because Stewart, like many family nursing researchers, chose to gather a large amount of data from a relatively small sample of families.

Statistics and Selected Features of Families

Complexity of Family Phenomena

Family phenomena, like many objects of study in the social sciences, have a wide range of possible inputs, outputs, causes, and effects (Blalock, 1979). An obvious example is family adaptation to normative events and stressors, which research and clinical experience has demonstrated to be influenced by a wide range of factors (McCubbin, 1991). Often, family researchers measure a multiplicity of variables and attempt to determine the effects of these variables on a given individual or family outcome. Indeed, Abell (1990) mentions that some researchers and critics believe that family research studies are only worthwhile

when all aspects of families, family members and the environment are taken into account.

When the research question involves testing a set of relationships amongst a large number of the variables simultaneously, path analysis and structural equation modeling are often used. Outcome variables and variables believed to represent determinants of those outcomes are clearly set out, and the literature and clinical experience are used to justify all the hypothesized connections between variables. Unfortunately, when these approaches are chosen, there are serious consequences to not including variables known to influence a given phenomenon: "...without an attempt to include every relevant cause of an endogenous variable in the model ..., the estimates of relationships that are included in the model are biased [erroneously estimated]." (Godwin, 1988, p. 919) Specifically, this means that the magnitude and direction of relationships between variables may be incorrectly depicted if important variables are left out (Schumm, Southerly, & Figley, 1980). Clearly, the researcher must make a careful compromise between including "all" variables and selecting an insufficient range of variables, if and when complex family phenomena are studied using these statistical techniques.

A major drawback of selecting multivariable methods is that sample sizes required for analyses to provide stable and meaningful results increase with the number of variables included, and tend to be quite large (Tabachnick & Fidell, 1989). In structural equation modeling, usually data from a minimum of 150 to 200 subjects must be included, depending on the exact model being tested (Lavee, 1988). Clearly, considerable planning and expense are involved in data collection on this scale, and samples of this size may not be available to the nurse researcher. There are also other built-in limits to using these techniques to expand the number of variables that can be tied together in any single analysis. These limits are created at least partially by statistical aspects of the methods themselves, which restrict the kinds of variables (level of measurement, statistical distributions, and measurement assumptions) that the methods can accommodate, and (even in 1995) by limitations on computing power available to the researcher (Copeland & White, 1991; Teachman & Neustadtl, 1993).

LISREL and related methods are not the only statistical means of capturing complex relationships in family research. In Stewart's study, at least three major areas of family adaptation preceding and following the birth of a second child were of interest: parental stress and use of social support, household and child-care responsibilities, temperament of all members of the family, and adjustment of the firstborn child. Measures tapped a wide range of family unit characteristics and family member experiences: for instance, mothers' stress levels were believed to be related to demographic and overall family situation

characteristics, stressful life events, stressors related to the child and the infant, and social support. Stewart obtained many of these measures five times from as many as three members of each of the 41 families studied. With a sample size of 41, only a handful of variables could have been processed together in a multivariate regression analysis to determine their associations with an outcome variable.

However, methods that predict outcomes from background variables would have been inconsistent with the family systems theoretical framework grounding Stewart's study. Stewart sought to detect patterns of stability and change in family variables. Differences in measures between parents and across time, for instance, were analyzed using repeated-measures univariate analysis of variance (ANOVA) on a variable-by-variable basis. When possible associations between social support and reported stress were investigated, Stewart used canonical correlation, a technique that reveals overlap in statistical variation between groups of variables. In this particular case, one group was a set of measures of stress related to oneself, the firstborn child, and the infant, and the second set contained measures of social support. Handling the many variables included in the study, either one at a time or in groups, provided Stewart with the means to paint a very detailed picture of stability and change in the families he studied. Stewart selected subsets of variables and did not attempt to enter all variables in statistical models simultaneously. Cohen (1990) recommends a similar scaling back of the number of variables used in any given analysis, and cites clearer, more readily interpretable findings as a strength of such an approach.

If determining the predictive power of background variables had been considered important, there would have been at least one other analytic possibility to accommodate large numbers of variables. At least some of many variables in Stewart's study were highly intercorrelated (as we shall discuss next). The number of variables from the data set used in individual analyses could probably have been reduced by the analyst (using knowledge of the subject area) or through various statistical techniques (including factor analysis) (Fisher, Terry, & Ransom, 1990). Subsequently, small groups of independent and dependent variables could have been run through multiple regression or other statistical procedures.

Reciprocity of Family Relationships

Because family members have sustained involvement with each other and often share a common history and environment, measures using the same instrument taken from members of the same family are often highly intercorrelated. This creates at least two statistical issues. When two or more scores on the same instrument from different family members (for example mother's and

father's mood) are used as independent variables in a model to predict an outcome, if these variables are highly intercorrelated (bivariate *r*'s above .80-.90), the results of many multivariable methods will be misleading and statistical packages may not be able to compute the parameters in the requested models at all (Tabachnick & Fidell, 1989). This problem is known as multicollinearity.

Secondly, shared variance in scores between members of the same family complicates the issue of trying to detect differences in these scores. If, for example, the means of statistically-dependent variables are compared using analysis methods that assume the variables are statistically independent, the chances of correctly identifying differences in responses between family members are reduced. This is due in part to the fact that independent groups comparison methods partition the variance in values differently than repeated-measures methods. Ball, McKenry, and Price-Bonham (1983) emphasize that values on the same instrument from members of the same family are indeed repeated measures of the same "subjects" and that repeated-measures (within-subjects) approaches are the appropriate techniques for comparing variables in within-family designs.

In Stewart's (1990) study, one of the major goals was to determine whether the experience of stress and of social support was comparable in husbands and wives, whether the sharing of household tasks changed over time, and whether behaviours engaged in by mother and father in interacting with the firstborn child changed over time. A wealth of data was collected from both mother and father using the same instruments around these areas. Stewart's main analytic approach was, consistent with Ball, McKenry, and Price-Bonham's recommendations, repeated-measures ANOVA using not only time of measurement, but also parent (mother versus father), as within-subjects factors.

Again, Stewart's aim was not to demonstrate which variables were statistically-significant predictors of individual or family outcomes. If prediction had been a goal in his study, highly-intercorrelated measures could not have been used as independent variables in the statistical models tested. Measures from different family members would need to be aggregated or discrepancy scores would need to be computed (Ferketich & Mercer, 1992), or a single member's score on the measure for each family would have to be selected. The technique chosen would be determined largely by considerations about which of these methods would use the information from the set of variables in the most meaningful way (Copeland & White, 1991).

Circularity of relationships between variables: Determining the direction of relationships

Systems theory is a major force in the development of theory and clinical practice in family nursing. Major elements of family systems theory include

propositions that relationships between family variables are complex, intertwined, and characterized by multiple feedback loops. According to family systems theory, families are more than the sum of their parts, and there are no simple relationships where one or more family variables relate to each other through a single line of influence (Fisher, Terry, & Ransom, 1990; Stewart, 1990). Ecological perspectives proposing that families are embedded in larger social systems, and that families and the environment have reciprocal influences on each other, have also been very influential in family nursing (Bronfenbrenner, 1986).

Feetham (1990) is among the many authors who have commented that linear, additive statistical models cannot address the complex, multidirectional relationships that emerge from family systems theory. Since the researcher's analytic plan must be consistent with the theoretical framework guiding a particular study (Schumm, 1982), family nursing researchers must give considerable thought to the relationship of statistical analysis to the theoretical assumptions driving their studies.

As was stated previously, determining causes and consequences of various aspects of family adjustment is a major thrust in family nursing research. However, as numerous philosophers of science have shown, causation is extremely difficult to prove (Baumrind, 1983; Cliff, 1983). Baumrind (1983), borrowing from other authors, defines minimal criteria for causation: statistical association between variables, an appropriate temporal relationship between them (events cannot precede their putative causes), and, most importantly, the successive elimination of competing causes for the relationship between the variables. Certainty regarding the effect of variables on a family phenomenon is derived from a coherent pattern of results. In the words of John Tukey, the distinguished statistician:

> Causation can only be established as a theoretically inevitable consequence of empirical observations. Failure to recognize this dual requirement leads to asking too much of statistics and to consequent dissatisfaction. (Tukey, 1986, p.309).

Stewart's study was explicitly guided by family systems theory. He addressed the issues of circular relationships and causality in presenting both the framework for his study and his results. Distinctions between causes and effects were not made, and as noted before, shared variance between variables and differences in variables across spouses and over time were the major trends evaluated. Scores on the wide range of instruments used were first treated as independent variables, and then were subsequently treated as dependent variables in later analyses. Based on similarities and differences in these family variables, as well as changes in these variables over time, Stewart demonstrated that the birth of a second child is experienced differently by

husband and wife and proposed mechanisms that explain both the use of social support and possible changes in family roles during this transition.

Although Stewart provides an excellent example of how complex family data sets can be analyzed in a manner congruent with systems theory, it should be noted that other researchers have been able to reconcile testing linear models involving dependent and independent variables with a family systems approach. Accepting the inherently reductionist nature of quantitative research, they believe that one can examine small portions of the multidirectional webs linking family variables while acknowledging that the reality of family life is far more complicated (Abell, 1990; Fisher, Terry, & Ransom, 1990). Tests of linear relationships (for example, using multiple regression procedures), interpreted in the proper context, can provide the researcher with the pieces needed to construct richly-textured and highly complex pictures of family phenomena.

Path analysis, and later structural equation modeling (LISREL and related techniques), were at one time widely thought to provide researchers with a means of confirming theoretical models involving complex interrelationships and drawing conclusions about causality. Mulaik (1993), among other writers, reminds us that the results of any multivariate analyses can be read in many different ways. The consistency of a set of parameters with a particular theoretical interpretation does not prove that this interpretation reflects an underlying truth. Plausible alternative explanations for the connections between variables are rarely ruled out by any single analysis, whether a path analysis or not. Therefore, to speak of the results of any path analysis, or LISREL model as anything more than consistent with a given theory is highly misleading (Baumrind, 1983; Biddle & Marlin, 1987; Cliff, 1983; Lavee, 1988). Harsh criticism has led to a justified hesitation to associate the terms "causal" and "confirmatory" with such forms of analysis (Baumrind, 1983).

Nonetheless, path analysis and LISREL modeling remain powerful tools for knowledge-building since they allow probing of interconnections in data sets as well as the possibility, when properly used, of illuminating the directions of relationships between variables. These approaches are also often consistent with systems theory. For instance, Stewart himself recommended that some form of path or causal analysis be used to further assess some of the relationships brought out in his study. The family nursing researcher needs to be aware of the strengths and limitations of these methods.

Conclusion

The nature of families and family variables must be taken into careful account when researchers plan the analysis phase of their studies. When considering using multivariable techniques to capture the complexity of family phenom-

ena, it should be remembered that simple statistical methods that handle variables individually or in pairs, instead of in large groups simultaneously, can often portray this complexity quite well (Schumm, 1982). Handling variables that are measured in more than one family member requires careful thought, due to the statistical dependence of such measures and the problem of multicollinearity if these variables are entered into multivariable models together. Increasing the number of useful, supported theories in family nursing will require more than highly technical statistical methods that provide global tests of complicated interrelationships and verify causality only in a very weak sense. Families are not simply another type of social group. Family nurse researchers should be familiar with statistical aspects of describing family form and functioning, and should strive to understand the assumptions, strengths and limitations of the breadth of existing statistical techniques well enough to make informed decisions and choices in their studies (Brogan, 1989; Teachman & Neustadtl, 1993).

References

Abell, T.D. (1990). Description, inference, prescription: Data analysis, scientific inference, and epistemologic reasoning in clinical research. In C.N. Ramsey (Ed.), *Family systems in medicine* (pp. 539-560). New York: Guilford.

Baumrind, D. (1983). Specious causal attributions in the social sciences: The reformulated stepping-stone theory of heroin use as exemplar. *Journal of Personality and Social Psychology, 45,* 1289-1298.

Biddle, B.J., & Marlin, M.M. (1987). Causality, confirmation, credulity, and structural equation modeling. *Child Development, 58,* 4-17.

Blalock, HM. (1979). Measurement and conceptualization problems: The major obstacle to integrating theory and research. *American Sociological Review, 44,* 881-894.

Brogan, D. (1989). A statistician's view on statistics, quantitative methods, and nursing. In I.L. Abraham, D.M. Nadzam, & J.J. Fitzpatrick (Eds.), *Statistics and quantitative methods in nursing: Issues and strategies for research and education* (pp. 220-232). Philadelphia: Saunders.

Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives. *Developmental Psychology, 22,* 723-742.

Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research, 18,* 115-126.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45,* 1304-1312.

Copeland, A.P., & White, K.M. (1991). *Studying families.* Newbury Park, CA: Sage.

Daly, K. (1992). The fit between qualitative research and characteristics of families. In J.F. Gilgun, K. Daly, & G. Handel (Eds.), *Qualitative methods in family research* (pp. 3-11). Newbury Park, CA: Sage.

Feetham, S.L. (1990). Conceptual and methodological issues in research of families. In J. Bell, W. Watson, & M. Wright (Eds.), *The cutting edge of family nursing* (pp. 35-49). Calgary, AB: Family Nursing Unit Publications.

Ferketich, S.L., & Mercer, R.T. (1992). Focus on psychometrics: Aggregating family data. *Research in Nursing and Health, 15,* 313-317.

Fisher, L., Terry, H.E., & Ransom, D.C. (1990). Advancing a family perspective in health research: Models and methods. *Family Process, 29,* 177-189.

Godwin, D.D. (1988). Causal modeling in family research. *Journal of Marriage and the Family, 50,* 917-927.

Hornstein, G.A. (1988). Quantifying psychological phenomena: Debates, dilemmas, and implications. In J.G. Morawski (Ed.), *The rise of experimentation in American psychology* (pp. 1-34). New Haven: Yale University Press.

Larzelere, R.E., & Klein, D.M. (1987). Methodology. In M.B. Sussman, & S.K. Steinmetz (Eds.), *Handbook of marriage and the family* (pp. 125-155). New York: Plenum.

Lavee, Y. (1988). Linear structural relationships (LISREL) in family research. *Journal of Marriage and the Family, 50,* 937-948.

McCubbin, M.A. (1991). Family stress theory and the development of nursing knowledge about family adaptation. In S.L. Feetham, S.B. Meister, J.M. Bell, & Gilliss, C.L. (Eds.), *The nursing of families* (pp. 46-58). Newbury Park, CA: Sage.

Mulaik, S.A. (1993). Objectivity and multivariate statistics. *Multivariate Behavioral Research, 28* (2), 171-203.

Schumm, W.R. (1982). Integration of theory, measurement, and data analysis in family studies survey research. *Journal of Marriage and the Family, 44,* 983-998.

Schumm, W.R., Barnes, H.L., Bollman, S.R., Jurich, A.P., & Milliken, G.A. (1985). Approaches to the statistical analysis of family data. *Home Economics Research Journal, 14* (1), 112-122.

Schumm, W.R., Southerly, W.T., & Figley, C.R. (1980). Stumbling block or stepping stone: Path analysis in family studies. *Journal of Marriage and the Family, 42,* 251-262.

Stewart, R.B. (1990). *The second child: Family transition and adjustment.* Newbury Park, CA: Sage.

Tabachnick, B.G., & Fidell, L.S. (1989). *Using multivariate statistics* (2nd ed.). New York: HarperCollins.

Teachman, J.D., & Neustadtl, A. (1993). Emerging methods. In P.G. Boss, W.J. Doherty, R. LaRossa, W.R. Schumm, & S.K. Steinmetz (Eds.), *Sourcebook of family theories and methods: A contextual approach* (pp. 715-727). New York: Plenum.

Tukey, J.W. (1986). Data analysis and behavioral science or learning to bear the quantitative man's burden by shunning badmandments. In L.V. Jones (Ed.), *The collected works of John W. Tukey* (Vol. 3) (pp. 187-389). Belmont, CA: Wadsworth.

Acknowledgement

Preparation of this paper was assisted in part by an NHRDP Ph.D. Training Fellowship, Health Canada.