Multiple Linear Regression Viewpoints

A Publication sponsored by the American Educational Research Association's Special Interest Group on Multiple Linear Regression: The General Linear Model

MLRV

Volume 32 • Number 1 • Fall 2006

Table of Contents

Multiple Linear Regression Viewpoints

Editorial Board

Multiple Linear Regression Viewpoints (ISSN 0195-7171) is published by the AERA Special Interest Group on Multiple Linear Regression: General Linear Model through the **University of Alabama at Birmingham**.

Subscription and SIG membership information can be obtained from:

Jeffrey B. Hecht, MLR:GLM/SIG Executive Secretary
Department of Educational Technology, Research & Assessment
Northern Illinois University
DeKalb, IL   60115-2854.
jbhecht@niu.edu

MLRV abstracts appear in CIJE, the ERIC system, and microform copies are available from University Microfilms International, 300 North Zeeb Road, Ann Arbor, MI 48106. *MLRV* is listed in the *EBSCO Librarians Handbook.*

Bonferroni Adjustments in Tests for Regression Coefficients

Daniel J. Mundfrom **Jamis J. Perrett** **Jay Schaffer**
Adam Piccone **Michelle Roozeboom**
University of Northern Colorado

A common application of multiple linear regression is to build a model that contains only those predictors that are significantly related to the response. In so doing, tests regarding the unique contribution of individual predictors to the model are often performed. It is not uncommon for practitioners to conduct each of these tests at the nominal $\alpha = 0.05$ level, without regard to the effect that this practice may have on the overall Type I error rate. This research investigated the utility of making a Bonferroni adjustment when conducting these tests of the partial regression coefficients. Simulated multivariate normal populations with various correlational structures, different numbers of predictors in the model, and differing numbers of "significant" predictors in the model were generated. Ten thousand samples, 5000 each of sizes 50 and 300, were drawn from each population condition and a multiple regression analysis was performed on each sample. In every case, the observed significance levels for the Bonferroni-adjusted tests were controlled below the nominal 0.05 level as expected, and in most cases substantially lower than the observed significance levels for the unadjusted tests.

Multiple Linear Regression (MLR) is a popular statistical procedure for investigating the nature of the relationships among several numerical characteristics. Typically, one of the characteristics is identified as the dependent or response variable and the remainder of the characteristics are called independent or predictor variables. Most introductory level statistics texts identify regression analyses as having two uses: 1) to estimate the average response for a sample of individuals having various values for each variable in a set of predictors, and 2) to predict the response for a "new" individual for whom only values of the predictors are measured/observed. In either case, a linear model, based on observed data is used to make the estimation or prediction.

In some applications, the researcher knows which variables should be used as predictors in the model and the purpose of the analysis is to predict the value of the response using previous information regarding the nature of the variables' relationships with each other. Data are collected on the predictor variables and the model is used to predict the value of the response variable for one or more "new" individuals. In other situations, the researcher is interested in determining which, if any, of several numerical characteristics are significantly related to a specific outcome. Data are collected on all the variables of interest—the dependent variable and all the independent variables—and an MLR analysis is performed to build a model that may later be used for prediction, i.e., the researcher determines which of these predictor variables displays a significant unique ability to explain variation in the response variable. While both of these applications of MLR are useful and appropriate, it is the latter situation which is the focus of this research.

When the purpose of the regression analysis is to determine which independent variables are unique contributors to the model, it is typical for the researcher to perform separate tests of the partial regression coefficients (i.e., the beta coefficients) for each predictor. Those predictors for which the test of the beta coefficient has a p-value that is less than the specified α-level are deemed to be making a unique contribution to the model and will be retained in the model as a predictor variable. On the other hand, those variables for which the test of the beta coefficient has a p-value that is larger than that specified α level are not identified as useful predictors and may be dropped from the model in the interest of parsimony. It is not an uncommon practice for each of these separate tests to be conducted at the nominal 5% significance level. The purpose of this research is to investigate whether conducting each of these tests at $\alpha = 0.05$ inflates the overall Type I error rate for the collection of all these tests and if a Bonferroni-type adjustment to the α level for each test would be appropriate to control the overall α-level closer to that nominal level.

Making adjustments to the significance level of a statistical test when multiple tests are conducted on the same data is a common statistical practice. Many procedures have been developed for making such adjustments. One of these procedures is the Bonferroni adjustment.

The Bonferroni adjustment is based on an inequality in probability theory that was derived by C. E. Bonferroni. The inequality states that if $A_1, A_2, \ldots, A_k$ represent k events and $A_1', A_2', \ldots, A_k'$ represent the corresponding complements, then

$$P\left(\bigcap_{i=1}^{k} A_i'\right) = 1 - \sum_{i=1}^{k} P(A_i).$$

An application of the Bonferroni inequality, the Bonferroni adjustment, is one of the commonly used methods for adjusting the significance levels of individual tests when multiple tests are performed on the same data. For example, consider three statistical tests being performed simultaneously, each at level α, such that

A = {a Type I error occurred in test 1}
B = {a Type I error occurred in test 2}
C = {a Type I error occurred in test 3}

so that $P(A) = P(B) = P(C) = \alpha$. Under these conditions, the probability that at least one Type I error occurs in the three tests, i.e., the overall significance level of the three tests, is inflated. The Bonferroni inequality provides an upper bound for the overall level of significance such that,

$$\begin{aligned} P(\text{at least one Type I error occurs}) &= 1 - P(\text{no Type I errors occur}) \\ &= 1 - P(A' \cap B' \cap C') \\ &< 1 - \{1 - [P(A) + P(B) + P(C)]\} \\ &= 1 - [1 - 3\alpha] \\ &= 3\alpha. \end{aligned}$$

The Bonferroni adjustment divides the nominal significance level, α, by the number of tests being performed simultaneously to prevent the overall level of significance from exceeding the nominal level, α. The adjusted level of significance, in general α/k for k tests, is used to conduct each of the k individual tests.

Virtually every statistics textbook recommends some type of adjustment when pairwise comparisons of means are performed as a follow-up to a significant ANOVA (see, for example, Glass and Hopkins, 1996; Hinkle, Wiersma, & Jurs, 1998; Agresti and Finlay, 1997). It is rare, if ever however, that these same textbooks would recommend these same types of adjustments when conducting tests of main effects and interactions in a factorial ANOVA design or tests of the partial regression coefficients in a MLR analysis, yet these two situations also consist of multiple tests being conducted on the same sample of data. Hinkle, Wiersma & Jurs (1998), for example, provide a multiple regression example with four predictors in which each partial regression coefficient is tested at $\alpha = 0.05$ to determine if the corresponding predictor variable should be retained in the model. It would not be surprising to find similar examples in just about any text covering basic regression analysis.

Galambos and Simonelli (1996) discuss additional applications of the Bonferroni inequality beyond pair-wise comparisons of means as a post-hoc or a priori follow-up to a significant ANOVA. Simultaneous confidence intervals are presented for differences between pairs of means or variances, along with joint confidence intervals for partial regression coefficients, joint prediction intervals for n new observations, and the detection of outliers in multiple regression. Although Galambos and Simonelli (1996) discuss applications of the Bonferroni inequality in various fields of application, including number theory, extreme value theory, linear programming, and statistical methods, they do not discuss its use for variable selection in a MLR analysis.

For a multiple regression analysis with, say, seven predictor variables, it doesn't appear to be difficult to see that conducting seven separate tests, each at $\alpha = .05$, would inflate the overall Type I error rate for that analysis. Consequently, it also would not appear to be surprising that using a Bonferroni-type adjustment that would conduct each of those tests at $\alpha/7$, i.e., $0.05/7 = 0.0071$, would provide some protection for that overall error rate. We see no compelling philosophical perspective that would preclude such adjustments from this type of application. Indeed, Korn and Graubard (1990) compared the power of Bonferroni-adjusted t-tests of the partial regression coefficients and the Wald statistic. Their results were mixed with each method outperforming the other in certain situations, leading them to conclude that, "One possible interpretation . . . is not that the Bonferroni procedure works well, but that the Wald statistic works poorly . . ." (Korn and Graubard, 1990, p.274). More recently, Foster and Stine (2004) used a modified stepwise selection procedure incorporating a Bonferroni-type adjustment on tests of individual partial regression coefficients in an application using bankruptcy data. Using their method resulted in an earlier end to the stepwise variable selection procedure and that the resulting prediction model performed better than the conventional data mining techniques that were more typically used. Both of these studies indicate that the use of a Bonferroni adjustment to the significance level for tests of individual partial regression coefficients in variable selection when building a prediction model using a multiple regression analysis is a reasonable approach for providing some control over the overall Type I error rate.

Method

In this study, simulated data were used to create populations with known characteristics so that the effect on the overall Type I error rate from varying those characteristics could be determined. Data were generated using PROC IML in SAS. Populations were created that contained data from multivariate normal distributions and varied according to the number of predictor variables in the model, the magnitude of the zero-order correlations among the predictors, the magnitude of the zero-order correlations between the predictors and the response variable, and the proportion of predictors having non-zero correlations with the response variable.

The number of predictor variables in each population was varied across the values 2, 4, 6 and 8. Zero-order correlations among the predictor variables were varied across the values 0.1, 0.3, and 0.5, keeping these correlations low to eliminate any potential multicollinearity problems which could influence the results of individual tests regarding the partial regression coefficients. Correlations between the predictor variables and the response variable were varied across the values 0, 0.4, and 0.8. Because a Type I error can occur only when a non-significant predictor is identified as significant, only models that contained at least one non-significant predictor were investigated, i.e., all of the models investigated in this study had at least one predictor variable that had a correlation of 0 with the response variable. The proportion of predictor variables that had non-zero correlations with the response was varied across the values 0, 1/6, 1/4, 1/3, and 1/2. Only proportions that resulted in an integral value for the number of predictors with non-zero correlations were used in any particular model. For example, with two predictors, only the proportions 0 and 1/2 were used, i.e., we investigated cases with both variables uncorrelated with the response and with only one variable correlated with the response. With four predictors, only the proportions 0, 1/4, and 1/2, were used, i.e., we considered cases with all four predictors, one, and two predictors respectively being uncorrelated with the response. Similar restrictions were used in the scenarios with 6 and 8 predictor-variable models. In all the population conditions investigated, the proportion of predictor variables that had non-zero correlations with the response was restricted to no more than half of the variables in the model. This restriction was used so that the possible number of Type I errors that could occur in any sample did not become too small. We considered two different sample sizes—50 and 300—to see if the amount of data had any influence on the unadjusted and Bonferroni-adjusted tests of the individual partial regression coefficients.

Initially, 10 populations of size 100,000 were generated for each of the population conditions identified above. A multiple regression analysis was performed on each population to determine which predictors were actually significantly related to the response. For each sample size (i.e., 50 and 300), 500 samples were generated from each of the 10 populations in each of the population conditions, for a total of 5000 samples of size 50 and 5000 samples of size 300 for each condition. A multiple regression analysis was performed on each sample and each of the partial regression coefficients was tested at both the unadjusted α-level, 0.05, and the adjusted α-level, $0.05/k$, where k is the number of predictors in the model. A Type I error was defined as interpreting one predictor as significant in the sample that was not identified as significant in the parent population of that sample. The proportion of samples that lead to a Type-I error for any of the predictors in the model was computed as the actual significance level of the test. Actual significance levels for both the adjusted tests and the unadjusted tests were computed and compared.

Table 1. *Actual significance levels for Bonferroni-adjusted and unadjusted tests of the partial regression coefficients for various numbers of predictors, various*

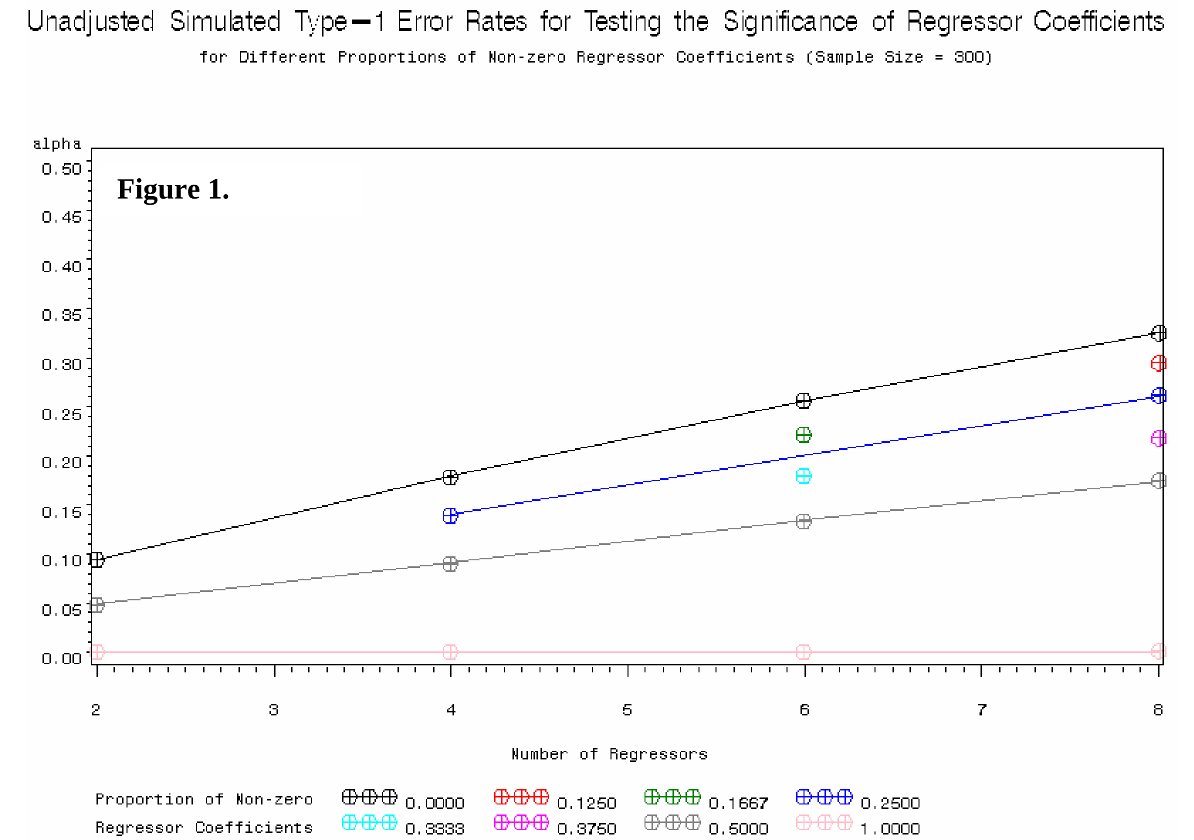
numbers of correlated predictors, and two different sample sizes

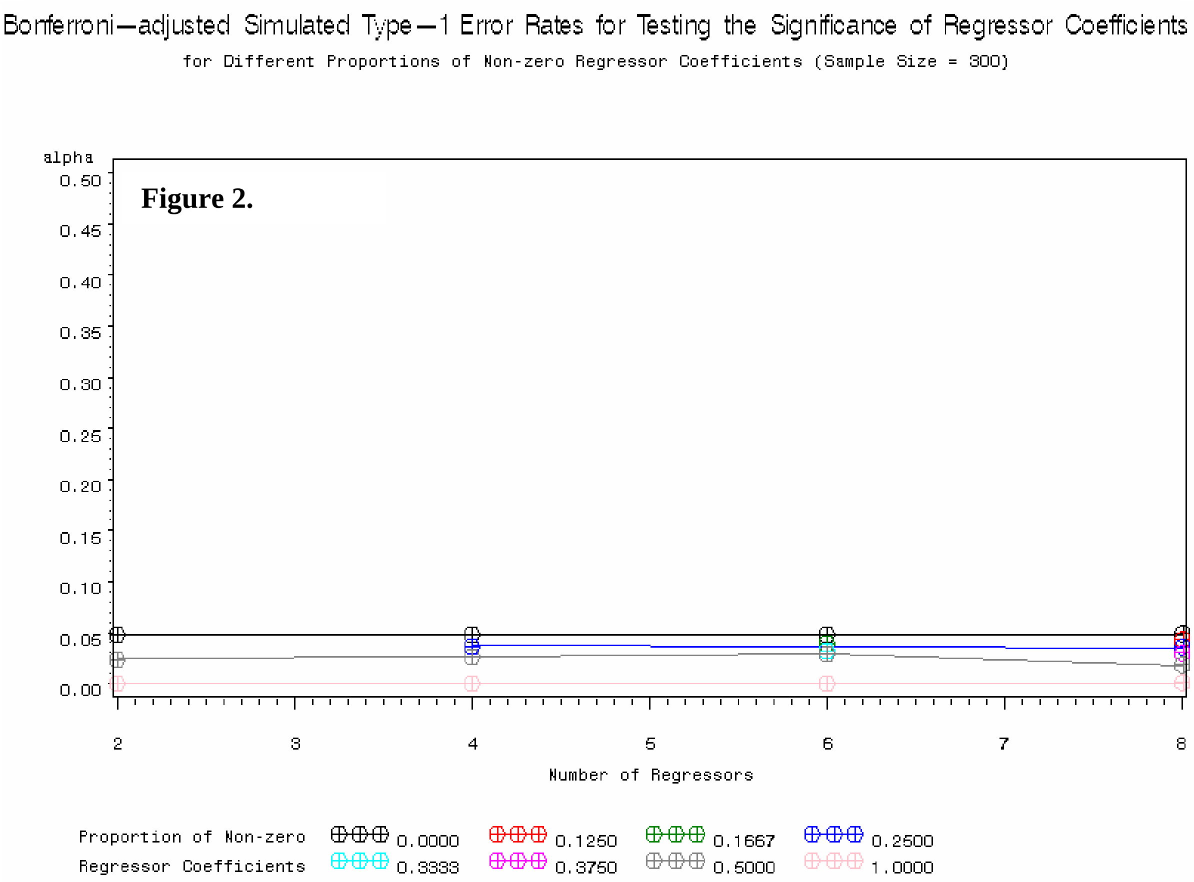
number of predictors	number of non-zero correlations with response	sample size	actual significance levels unadjusted	Bonferroni-adjusted
2	0	50	0.0930	0.0476
2	0	300	0.0940	0.0481
2	1	50	0.0511	0.0261
2	1	300	0.0489	0.0244
4	0	50	0.1744	0.0476
4	0	300	0.1792	0.0479
4	1	50	0.1388	0.0364
4	1	300	0.1398	0.0373
4	2	50	0.0993	0.0252
4	2	300	0.0908	0.0264
6	0	50	0.2450	0.0462
6	0	300	0.2565	0.0479
6	1	50	0.2131	0.0381
6	1	300	0.2214	0.0403
6	2	50	0.1827	0.0322
6	2	300	0.1794	0.0324
6	3	50	0.1447	0.0193
6	3	300	0.1340	0.0293
8	0	50	0.3065	0.0468
8	0	300	0.3249	0.0479
8	1	50	0.2810	0.0415
8	1	300	0.2944	0.0425
8	2	50	0.2494	0.0374
8	2	300	0.2603	0.0356
8	3	50	0.2092	0.0312
8	3	300	0.2178	0.0289
8	4	50	0.1732	0.0176
8	4	300	0.1740	0.0180

Results

The actual significance levels for both the unadjusted tests and the Bonferroni-adjusted tests are displayed in Table 1, for both of the sample sizes investigated. Separate values are reported for each of the four values for the number of predictors in the model and for the separate values of the number of predictors with non-zero correlations with the response. The actual significance levels are aggregated over the three different values of the correlations between the predictor variables and over the two values for the non-zero correlation between the predictors and the response. In every case, the actual significance levels for the Bonferroni-adjusted tests were less than the nominal 0.05 level as expected. In every case except the one with only two predictors and only one of them having a non-zero correlation with the response, the unadjusted significance levels were substantially larger than the nominal 0.05 level. It is clear from the table that as the number of predictors in the model increases, the unadjusted actual significance level also increased, with the increase being smaller with a larger number of the predictors having a non-zero correlation with the response.

For the Bonferroni-adjusted significance levels, in every case with all of the predictors having no correlation with the response, the actual significance level was close to, but slightly smaller than, the nominal 0.05 level. As the number of predictors with a non-zero correlation with the response increases, the Bonferroni-adjusted significance levels get smaller, indicating, as expected, that the Bonferroni adjustment over-corrects so that the observed significance level is less than the nominal value. Although variations exist between the actual significance levels for both the unadjusted and adjusted cases for the


Unadjusted Simulated Type-1 Error Rates for Testing the Significance of Regressor Coefficients
for Different Proportions of Non-zero Regressor Coefficients (Sample Size = 300)
Figure 1.
alpha
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
2
3
4
5
6
7
8
Number of Regressors
Proportion of Non-zero
Regressor Coefficients
0.0000
0.1250
0.1667
0.2500
0.3333
0.3750
0.5000
1.0000


Bonferroni-adjusted Simulated Type-1 Error Rates for Testing the Significance of Regressor Coefficients
for Different Proportions of Non-zero Regressor Coefficients (Sample Size = 300)
Figure 2.
alpha
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
2
3
4
5
6
7
8
Number of Regressors
Proportion of Non-zero
Regressor Coefficients
0.0000
0.1250
0.1667
0.2500
0.3333
0.3750
0.5000
1.0000

two sample sizes, these variations are small, indicating that the size of the sample had little meaningful effect on the observed significance levels.

Figures 1 and 2 display the information in Table 1 graphically for the unadjusted actual significance levels and the Bonferroni-adjusted actual significance levels, respectively, for the sample size = 300 cases. Similar graphs for sample size = 50 were nearly identical to those in figures 1 and 2, indicating no differences in these observed significance levels due to sample size. The graphs corresponding to sample size = 50 have been omitted. In both figures, the data are aggregated according to the proportion of predictors having non-zero correlations with the response. Actual data values occurred only at the discrete values of the number of predictors equal to 2, 4, 6, and 8. The lines are drawn to indicate the linear trends that correspond to the observed significance levels for both the adjusted and unadjusted tests as the number of predictors increases and the number of predictors having non-zero correlations with the response decreases.

Conclusion

It is not common practice among applied researchers to use adjusted t-tests for variable selection in regression analysis. Rather, it is much more common to see each individual test conducted at the nominal level, usually using $\alpha = 0.05$. This research indicates that when unadjusted t-tests are used for individual variable selection, the associated overall Type-I error rate may be inflated by as much as 2 to 6 times the nominal α-level depending upon the number of predictors in the model and the number of predictors that have a non-zero correlation with the response. Consequently, one or more variables are identified as "significant" predictors of the response that are not actually needed in the model, i.e., the amount of unique variance in the response explained by these variables is negligible. A more conservative approach, one that controls the overall α-level, would be to use the Bonferroni-adjusted approach to conduct these tests. As shown here, tests based on this adjustment are overly conservative, especially as the number of predictors having non-zero correlations with the response increases. However, if the goal of the research is to identify only those predictors that are actually related to the response, these results seem to indicate that using the Bonferroni adjustment would be preferred.

References

Agresti, A. & Finlay, B. (1997). *Statistical methods for the social sciences* (3rd Ed.). Upper Saddle River, NJ: Prentice Hall.

Foster, D. P. & Stine, R. A. (2004). Variable selection in data mining: Building a predictive model for bankruptcy. *Journal of the American Statistical Association, 99,* 303-313.

Galambos, J. & Simonelli, I. (1996). *Bonferroni-type inequalities with applications*. New York: Springer-Verlag New York, Inc.

Glass, G. V & Hopkins, K. D. (1996). *Statistical methods in education and psychology* (3rd Ed.) Needham Heights, MA: Allyn & Bacon.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1998). *Applied statistics for the behavioral sciences* (4th Ed.). Boston, MS: Houghton Mifflin.

Korn, E. L. & Graubard, B. I. (1990). Simultaneous testing of regression coefficients with complex survey data: Use of Bonferroni t statistics. *The American Statistician, 44,* 270-276.

Send correspondence to: Daniel J. Mundfrom
University of Northern Colorado
Email: daniel.mundfrom@unco.edu

Protective Factors and Risk Factors in Preschool Aged Children

Monica R. Geist
University of Northern Colorado

P. Antonio Olmos Gallo
Mental Health Corporation of Denver

Mary Grimmer
Mental Health Corporation of Denver

Daniel J. Mundfrom
University of Northern Colorado

Do resiliency or protective factors moderate risk factors in preschool aged children? This study looks at pre- and post-treatment data from 49 preschool children whose average age was 4 years old. The treatment included prevention and early intervention programs that promote emotional well-being and the development of healthy interpersonal relationships in children, ages birth to eight, and their caregivers. The Devereux Early Childhood Assessment Scale (DECA) was used to measure protective factors as well as behavior concerns by parents and teachers. There was no significant interaction effect between the number of risk factors and the post-treatment protective factors scores. This result does not support prior research on protective and risk factors. Cluster analysis was used to identify subgroups.

Children who are exposed to many risks growing up, are in danger of becoming a burden to society. In order to develop and improve preventions and interventions for children at risk, research in the resiliency of children is important. Resiliency can be defined as *good outcomes in spite of serious threats to adaptation or development* (Masten, 2001). Research in resiliency grew out of risk research (Masten, Best, & Garmezy, 1990). Risk factors include characteristics such as poverty, low maternal education, low socioeconomic status, low birth-weight, family instability, family violence, divorce, birth to a single parent, child abuse, homelessness, substance abuse, natural disasters and war, (Masten & Coatsworth, 1998, Masten, Best, Garmezy, 1990).

Resilient individuals seem to possess protective factors (Masters & Coatsworth, 1998). Protective factors moderate individual vulnerabilities or environmental hazards to increase the likelihood of success for a child (Baldo, 2000). Examples of protective factors include good intellectual functioning; an appealing, sociable, and easy going disposition; self-efficacy, self-confidence, initiative, and high self-esteem; talents; and faith. Some protective factors are external and include close relationships with caring parent figure(s), socioeconomic advantages, connections to extended supportive family networks, bonding with pro-social adults outside the family, connections to pro-social organizations, and attending effective schools (Baldo, 2000).

Researchers do not generally agree on how to measure resiliency. Many researchers study resilience in terms of *an observable track record of meeting the major expectations of a given society or culture in historical context for the behavior of children of that age and situation* (Masten, 2001). Other researchers focus on the absence of psychopathology. Still others look at both kinds of criteria (Masten, 2001). Three constructs that commonly occur in resiliency research are attachment, initiative, and self-control.

Attachment, as defined by LeBuffe and Naglieri (1999), is “a mutual, strong, and lasting relationship between a child and significant adult such as parents, family members, and teachers” (p. 4). Researchers have found empirical evidence that infants can be classified into one of three categories: (1) secure, (2) anxious-ambivalent, and (3) anxious-avoidant (Ainsworth & Bell, 1970). Recently Main and Solomon (1990) have found a fourth category called disorganized-disoriented. Current research suggests a strong relation between a child’s early attachment classification and later social, emotional, behavioral, and academic outcomes (Jacobsen & Hofmann, 1997). Not forming secure attachments as infants has been linked with behavioral problems (Kennedy & Kennedy, 2004). Boys with insecure attachments have been shown to be more aggressive, disruptive, assertive, controlling, and attention-seeking than boys with secure attachments (Turner, 1991). Girls with insecure attachments show more dependent behavior than girls with secure attachments (Turner, 1991). Infants who were securely attached at 18 months were found to be more enthusiastic, persistent, and cooperative than insecurely attached infants (Matas, Arend, & Sroufe, 1978). Waters, Wippman, & Sroufe (1979) suggest that secure attachment is not merely the absence of negative behavior. In fact, they found that securely attached infants display positive affective sharing while their anxiously attached counterparts do not.

LeBuffe and Naglieri (1999) define *initiative* as “the child’s ability to use independent thought and action to meet his or her needs” (p. 4). Hoehne (1990) found that motivation and initiative are related but are different concepts in that motivation’s activating force is the achievement of a specific objective while initiative is rather a self-starting, self-activating, self-reliant urge or drive to act, question, search, probe and persevere (Hoehne, 1990). Much research has been conducted on motivation. On the other hand, no empirical research on initiative was found.

LeBuffe and Naglieri (1999) define *self-control* as “the child’s ability to experience a range of feelings and express them using the words and actions that society considers appropriate” (p. 4). Self-control and self-regulation are fundamental to successful functioning in society (Masten & Coatsworth, 1998). Failure to develop self-control in the early years has been shown to set the stage for aggressive and disruptive behavior (Patterson, 1986).

Need for the Study

Protective factors such as attachment, initiative, and self-control have been shown to moderate, or diminish, the effects of risk factors (Baldo, 2000). It is important to try to duplicate those results, as well as, attempt to understand the characteristics of preschoolers with similar protective and behavior scores. Understanding the relationship between risk factors and protective factors will enable mental health professionals to screen and identify young children at risk, as well as, raise awareness for prevention of risk factors.

Purpose of the Study

The purpose of this study was two-fold. One purpose was to determine the relationship between risk factors and protective factors on children’s behavior as measured by the Devereux Early Childhood Assessment scale (DECA). A second purpose was to determine what subgroups of at-risk preschoolers who had similar protective and behavior scores had in common. The research questions that guide this study are:

1. Do protective factors moderate risk factors?
2. Are there identifiable subgroups of participants wherein the members within any group are similar and the subgroups are different from one another?

Methods

Procedures

Children enrolled in several preschools in an urban area of a large metropolitan Midwestern city were identified as having one or more risk factors. Once they were identified, their parents were invited to participate in the study. Participation in the research was voluntary. Researchers interviewed the parents and filled out an Early Intervention Child Data Sheet (see Appendix B) on each child in the study. Specific risk factors were identified in that interview. One parent of each child completed a Devereux Early Childhood Assessment (DECA) specifically regarding the child. Each child’s preschool teacher also completed a DECA.

Each set of children and parents then received treatment. The treatment included prevention and early intervention programs that promote emotional well-being and the development of healthy interpersonal relationships in children, ages birth to eight, and their caregivers. There were two categories of treatment: center-based services and child-specific services. Center-based services involved consultation to center directors and staff regarding child development, strategies to handle behavioral problems, overall classroom environment and quality improvements. Center-based activities could also include educational programs to which all parents were invited or classroom instruction in which all children participated. All children enrolled in the contract-center programs benefited from these services. It is through these more general center-based services that children needing specific intervention were often identified.

Child-specific services included direct involvement with specific children and consultation with the parents and teachers of those children. The consultation and involvement focused on reducing problem behaviors and encouraging social and emotional competence in that particular child. Child-specific services also included direct involvement with the child individually or in a small group. Referral for evaluation and treatment of related problems often occurred. The child was in contact with some service category at least once a week. Treatment lasted one academic year (9 months). Parents and teachers then filled out a DECA for each child after the treatment.

Participants

The participants were 250 preschool age children enrolled in several preschools. There were only 49 useable participants due to missing data on the other 201. Of the 49 useable participants there were 27 males and 22 females. There were 14 African-Americans, 25 Hispanics, 2 mixed African-American and Hispanic, and 8 Whites. The average family annual income was $8,045. The average age was 4 years old.

Instrumentation

The Devereux Early Childhood Assessment scale (DECA; LeBuffe & Naglieri, 1999) is a nationally normed instrument designed to measure protective factors. It is the first instrument of its kind. The Devereux Early Childhood Assessment (DECA) will be used to study the relationship between risk factors, protective factors and behaviors. The DECA contains 37 items (see Appendix A). The items asked how often in the last four weeks did the child do a specific behavior. There were 5 choices for answering each item: never, rarely, occasionally, frequently or very frequently. The DECA has three subscales scores (attachment, initiative, and self-control) and a challenging behaviors score.
In order to get the initiative sub-score, items 2, 3, 7, 12, 16, 19, 20, 24, 28, 32, and 36 were added together and then made into a T-score. In order to get the self-control sub-score, items 4, 5, 13, 21, 25, 30, 33, and 34 were used to obtain a T-score. In order to get the attachment sub-score, items 1, 6, 10, 17, 29, 31, and 37 were used. The total protective factors added all the items from the three sub-scores together for another T-score. The behavior concerns score was a result of combining items 8, 9, 11, 14, 15, 18, 23, 26, 27, and 35. A high behavior score indicates the behavior is a problem. A low behavior score indicates the child does not have a behavior problem out of the ordinary average preschooler. A high protective score means the child has more protective factors than a child with a low protective score. The risk factors were determined by interviewing the parents. The data was recorded on the Early Intervention Child Data Sheet (see Appendix B).

Data Analysis

In order to address the first research question, regression analysis was used. Two different models were created, one for teacher and one for parent responses. They were separated because the teacher and parent responses were not significantly correlated (Table 1).

Table 1. Pearson Correlations between Teacher and Parent Responses (N = 49)..

	R	Prob > $\|r\|$
Pre-treatment total protective score by Teacher & Pre-treatment total protective score by Parent	-0.33636	0.8257
Post-treatment total protective score by Teacher & Post-treatment total protective score by Parent	-0.09347	0.5229
Pre-treatment behavior concern score by Teacher & Pre-treatment behavior concern score by Parent	0.15714	0.2809
Post-treatment behavior concern score by Teacher & Post-treatment behavior concern score by Parent	0.16989	0.2432

Regression was used to determine if protective factors moderated risk factors. The first model included scores by the teachers.

Behavior_Teacher_Post = B_0
+ B_1 * Total Number of Risk Factors
+ B_2 * Protective_Teacher_Post
+ B_3 * Total Number of Risk Factors * Protect_Teacher_Post
+ B_4 * Behavior_Teacher_Pre

$$\hat{Y} = B_0 + B_1X_1 + B_2X_2 + B_3X_1X_2 + B_4X_4$$

This model included the total number of risk factors, the post-treatment total protective scores by the teacher, the interaction term between the total number of risk factors and the post-treatment protective scores by the teacher, and the pre-treatment behavior concern scores by the teacher as independent variables. The post-treatment behavior concern scores by the teacher was the dependent variable. R-Square for the model was 0.680672. The post-treatment protective scores by the teacher were significant predictors in the model (*F* = 6.44, *df* = 1, *p*= 0.0148,). The pre-treatment behavior concern scores by the teachers were also significant predictors in the model (*F*=42.25, *df* = 1, *p*<0.0001). The total number of risk factors was not a significant predictor in the model (*F* = 0.01, *df* = 1, *p* = 0.9180). There was no significant interaction effect between total number of risk factors and the post-treatment protective scores by the teacher (*F* = 0.05, *df* = 1, *p* = 0.8265).The resulting model was: $\hat{Y} = 33.05 - 0.32X_1 - 0.35X_2 + 0.1X_1X_2 + 0.68X_4$.

The second model included scores by the parents.

Behavior_Parent_Post = B_0
+ B_1 * Total Number of Risk Factors
+ B_2 * Protective_Parent_Post
+ B_3 * Total Number of Risk Factors * Protect_Parent_Post
+ B_4 * Behavior_Parent_Pre

$$\hat{Y} = B_0 + B_1X_1 + B_2X_2 + B_3X_1X_2 + B_4X_4$$

This model included the total number of risk factors, the post-treatment total protective scores by the parent, the interaction term between the total number of risk factors and the post-treatment protective scores by the parent, and the pre-treatment behavior concern scores by the parent as independent variables. The post-treatment behavior concern scores by the parent was the dependent variable. *R*-Square for the model was 0.297152. The total number of risk factors was not a significant predictor in the model (*F* = 0.09, *df* = 1, *p* = 0.7639). The post-treatment protective scores by the parent were not significant predictors in the model (*F* = 4.05, *df* = 1, *p* = 0.0504). There was no significant interaction effect between total number of risk factors and the post-treatment protective scores by the parent (*F* = 0.00, *df* = 1, *p* = 0.9612,). The pre-treatment behavior concern scores by the parent were also not significant predictors in the model (*F* = 1.33, *df* = 1, *p* = 0.2553,). This "Parent's Model" is believed to have multicolinearity because none of the predictors were significant, however the overall model was significant (*F* = 4.65, *df* = 4, *p* = 0.0032). The resulting model was: $\hat{Y} = 74.24 - 1.2X_1 - 0.43X_2 + 0.004X_1X_2 + 0.15X_4$.

In order to deal with the multicollinearity in the Parent's Model, the interaction term between risk factors and the post-treatment protective score by the parent was removed to create a third model.

Behavior_Parent_Post = B_0
+ B_1 * Total Number of Risk Factors
+ B_2 * Protect_Parent_Post
+ B_4 * Behavior_Parent_Pre

$$\hat{Y} = B_0 + B_1X_1 + B_2X_2 + B_4X_4$$

This model was significant (*F*=6.34, *df* = 3, *p* = 0.0011). Neither the total number of risk factors nor the pre-treatment behavior concern scores by the parents were significant predictors in this model. The post-treatment protective scores by the parents, however, was significant (*F* = 16.71, *df* = 1, *p* = 0.0002). The resulting model was: $\hat{Y} = 73.80 - 1.01X_1 - 0.42X_2 + 0.15X_4$.

Cluster analysis was used to address the second research question. The cluster analysis identified four unique subgroups of children. The four identified clusters accounted for an R-square of 0.428. Increasing the number of clusters only increased R-square to 0.488. Also when five clusters were used the cluster sizes began to be too small since the sample size was small (*N*=49).

The variables considered were gender, ethnicity, income, single parent, parent health problem, parent unemployed, marital instability, substance abuse, domestic violence, and history of child abuse. Low income was not a defining feature of any of the clusters, as most participants had a low income family.

Cluster 1 *n*=13. Mostly male (85%); 85% African-American or Hispanic; 15% White; 92% had single parent home; 62% had an unemployed parent; 46% had some sort of substance abuse, domestic violence or child abuse in the home; highest post-treatment behavior concerns score as scored by the teacher (mean of 64.23); highest post-treatment behavior concerns score as scored by the parent (mean of 68.46); pre- and post-treatment protective factor cores as scored by the teacher were very close (pre 36.2 and post 37.4); pre- and post-treatment behavior concerns scores as scored by the teacher were also very close (pre 6.4 and post 64.2); pre- and post-treatment protective factor scores as scored by the parents decreased slightly (pre 42.5 and post 39.9); pre- and post-treatment behavior concerns scores as scored by the parent increased (pre 64.3 and post 68.5).

Cluster 2 *n*=18. 28% African-American; 55% Hispanic; 17% White; only one participant had a single parent home (5%); 17% had an unemployed parent; 39% had some sort of substance abuse, domestic violence or child abuse in the home; post-treatment behavior concerns score as scored by the teacher mean of 62.39; post-treatment behavior concerns score as scored by the parent mean of 60.28; pre- and post-treatment protective factor scores as scored by the teacher were very close (pre 42.5 and post 42.3); pre- and post-treatment behavior concerns scores as scored by the teacher were also very close (pre 62.1 and post 62.4); pre- and post-treatment protective factor scores as scored by the parents increased (pre 45.3 and post 53.8); pre- and post-treatment behavior concerns scores as scored by the parent decreased (pre 63.6 and post 60.3).

Cluster 3 *n*=10. Mostly Hispanic (90%); 10% White; mostly female (70%); 50% had single parent home; 30% had an unemployed parent; 20% had some sort of domestic violence or child abuse in the home; lowest post-treatment behavior concerns score as scored by the teacher (mean of 44.8); a high post-treatment behavior concerns score as scored by the parent (mean of 64.8); pre- and post-treatment protective factor scores as scored by the teacher increased (pre 48.5 and post 55.0); pre- and post-treatment behavior concerns scores as scored by the teacher decreased (pre 50.6 and post 44.8); pre- and post-treatment protective factor scores as scored by the parents were very close (pre 35.5 and post 35.8); pre- and post-treatment behavior concerns scores as scored by the parent decreased (pre 67.8 and post 64.8).

Cluster 4 *n*=8. 50% African-American; 25% Hispanic; 25% White; 75% had single parent home; 50% had an unemployed parent; 12.5% had some sort of domestic violence in the home; post-treatment behavior concerns score as scored by the teacher (mean of 56.75); post-treatment behavior concerns score as scored by the parent (mean of 55); pre- and post-treatment protective factor scores as scored by the teacher increased (pre 55.25 and post 57.25); pre- and post-treatment behavior concerns scores as scored by the teacher decreased (pre 58.875 and post 56.75); pre- and post-treatment protective factor scores as scored by the parents increased (pre 46.6 and post 50.5); pre- and post-treatment behavior concerns scores as scored by the parent decreased (pre 64.9 and post 5).

Discussion

Initially the answer to the first research question, "Do protective factors moderate risk factors?" appears to be no. In both of the models (teacher scores and parent scores), there was no significant interaction effect between the number of risk factors and the post-treatment protective factors score. Therefore, protective factors did not moderate risk factors. This does not support what previous research found (Baldo, 2000; Jacobsen & Hofmann, 1997; Patterson, 1986). In the third model, when the interaction term was removed from the parent model in order to deal with the multicolinearity, the risk factors were not significant predictors of behavior.

In the second research question, "Are there identifiable subgroups of participants wherein the members within any group are similar and the subgroups are different from one another?" was affirmative. Four clusters were identified (see Table 2).

Cluster 1 was made up of mostly African-American or Hispanic males and had the highest behavior scores as scored by the teacher. These higher scores mean the behavior was more challenging in this

Table 2. Cluster Demographics, Protective and Behavior Scores

	Cluster 1 (n=13)	Cluster 2 (n=18)	Cluster 3 (n=10)	Cluster 4 (n=8)
Gender	11 male (85%) 2 female (15%)	8 male (44%) 10 female (56%)	3 male (30%) 7 female (70%)	5 male (62.5%) 3 female (37.5%)
Ethnicity	5 African-Am (38%) 4 Hispanic (31%) 2 Af-Am/Hisp (15%) 2 White (15%)	5 African-Am (28%) 10 Hispanic (56%) 3 White (17%)	0 African-Am (0%) 9 Hispanic (90%) 1 White (10%)	4 African-Am (50%) 2 Hispanic (25%) 2 White (25%)
Single Parent Home	12 (92%)	1 (5%)	5 (50%)	6 (75%)
Unemployed Parent	8 (62%)	3 (17%)	3 (30%)	4 (50%)
Substance abuse, domestic violence, or history of child abuse	6 (46%)	7 (39%)	2 (20%)	1 (12.5%)
Protective Scores by Teacher	Pre = 36.15 Post = 37.38	Pre = 42.56 Post = 42.28	Pre = 48.50 Post = 55.00	Pre = 55.25 Post = 57.25
Behavior Scores by Teacher	Pre = 66.38 Post = 64.23	Pre = 62.06 Post = 62.39	Pre = 50.60 Post = 44.80	Pre = 58.86 Post = 56.75
Protective Scores by Parent	Pre = 42.54 Post = 39.92	Pre = 45.33 Post = 53.83	Pre = 35.50 Post = 35.80	Pre = 45.63 Post = 50.50
Behavior Scores by Parent	Pre = 64.31 Post = 68.46	Pre = 63.61 Post = 60.28	Pre = 67.80 Post = 64.80	Pre = 64.88 Post = 55.00

cluster according to the teachers. This cluster also had the lowest protective scores as scored by the teachers. This means that these children had fewer protective factors. The parents viewed these children with more protective factors than did the teachers. Cluster 1 had the largest percentage of single family homes and the largest percentage of unemployed parents.

Cluster 2 had a mix of ethnicities and genders. This group had the most number of instances of substance abuse, domestic violence or child abuse. Cluster 2 had the second to lowest number of protective factors with the second to highest score of behavior problems as reported by the teachers. This subgroup had the least number of single family homes.

Cluster 3 was made up of mostly Hispanic females and had the lowest behavior scores as scored by the teachers. In other words, these children had the least amount of behavior problems according to the teachers. This cluster had the second to highest protective scores as scored by the teacher, but the lowest protective scores as scored by the parents.

Cluster 4 had a mix of ethnicities and genders. They had the fewest number of substance abuse, domestic violence or child abuse incidences. Their protective scores by the teachers were the highest of the four subgroups. Only cluster 3 had better behavior than this subgroup as reported by the teachers.

Caution should be taken when considering the results of this study. The main limitation of this research was the small useable sample size of 49. Even though the original data included 250 participants, there was not enough data collected for each child to construct a complete picture. Further research needs to be conducted with larger useable sample sizes.

References

Ainsworth, M.D.S. & Bell, S. (1970). Attachment, exploration, and separation: Illustrated by the behavior of one-year-olds in strange situation. *Child Development*, *41*(1), 49-67.

Baldo, A.J. (2000). Risk factors, protective factors, and challenging behaviors in at risk preschool aged children: An exploratory analysis (Doctoral dissertation, University of Northern Colorado, 2000). *Dissertation Abstracts International*, (*61*)12, 6733.

Hoehne, K.A.K. (1990). Initiative – a neglected psychosocial dimension. *Social Psychiatry and Psychiatric Epidemiology*, *25*, 101-107.

Jacobsen, T., & Hofmann, V. (1997). Children's attachment representations: Longitudinal relations to school behavior and academic competency in middle childhood and adolescence. *Developmental Psychology*, *33*(4), 703-710.

Kennedy, J.H., & Kennedy, C.E. (2004) . Attachment theory: Implications for school psychology. *Psychology in the Schools*, *41*(2), 247-259.

LeBuffe, P.A. & Naglieri, J.A. (1999). *Devereux Early Childhood Assessment User's Guide*. Lewisville, NC: Kaplan Press.

Main, M., & Solomon, J. (1990). Procedures for identifying infants as disorganized-disoriented during the Ainsworth Strange Situation. In M. Greenberg, D. Cicchetti, & E.M. Cummings (Eds.), *Attachment in the preschool years: Theory, research and intervention* (pp. 121-160). Chicago: University of Chicago Press.

Masten, A.S. (2001). Ordinary magic: Resilience processes in development. *American Psychologist*, *56*(3), 227-238.

Masten, A.S., Best, K.M., & Garmezy, N. (1990). Resilience and development: Contributions
from the study of children who overcome adversity. *Development and Psychopathology*, *2*, 425-444.

Masten, A.S. & Coatsworth, J.D. (1998). The development of competence in favorable and unfavorable environments: Lessons from research on successful children. *American Psychologists*, *53*, 205-220.

Matas, L., Arend, R.A., & Sroufe, L.A. (1978). Continuity of adaptation in the second year: The relationship between quality of attachment and later competence. *Child Development*, *49*, 547-556.

Mental Health Corporation of Denver (1999/2000). *Parent Empowerment Alternatives with Resources and Learning program*.

Patterson, G. (1986). Performance models for antisocial boys. *American Psychologist*, *44*, 432-444.

Turner, P.J. (1991). Relations between attachment, gender, and behavior with peers in preschool. *Child Development*, *62*, 1475-1488.

Send correspondence to: Monica R. Geist
University of Northern Colorado
Email: jfraas@ashland.edu

Application of a Structure Coefficient Rule of Thumb For Two-Group Descriptive Discriminant Analysis

Mercedes K. Schneider
Ball State University

This simulation study considered the rule of thumb as noted in Pedhazur (1997) for judging the usefulness of continuous (in a previous MANOVA, dependent) variables at determining group separation in descriptive discriminant analysis; namely, that a structure coefficient value equal to or greater than .3 identifies a useful continuous variable. No research to date has tested this rule. Results indicate that the rule is generally useful for identifying variables with medium to large effects but not small effects.

In an effort to easily interpret results of complex statistical analyses, practitioners often consult the literature for guidelines, or "rules of thumb," to aid in understanding their results. Although statisticians themselves might hesitate to distill complex results into a few general criteria, they also wish to offer some means of helping researchers from other fields utilize complex results (e.g., Cohen's work with effect size guidelines, 1992). Thus, they offer "rules of thumb." In Cohen's work, the rules of thumb for identifying small, medium, and large effect sizes is based upon extensive research. However, some rules of thumb are not so obviously supported via prior research. Such is the case of interpreting structure coefficient (SC) values in discriminant analysis (DA), specifically, descriptive discriminant analysis (DDA). Before continuing about the issues surrounding SC interpretation, a more detailed discussion of DA might prove helpful.

Cooley and Lohnes (1971) describe DA as the search for the best reduced-rank linear model to account for differences among groups as such differences have been measured on a vector of p continuous variables. Though mathematically it makes no difference whether the continuous variables are viewed as independent and the grouping variable, dependent, often in DA, the grouping variable is considered the outcome variable, with orthogonal linear discriminant functions derived such that the resulting coefficients associated with the vector of p continuous, independent variables maximize group differences. The number of possible linear functions is the lesser of p and the number of groups, k, minus one. In the case of two-group MANOVA, only one linear function is possible. With slight modification in notation, this function, Z, may be written as:

$$Z = X_1v_1 + X_2v_2 + \ldots + X_pv_p = \mathbf{Xv} \tag{1}$$

as noted in Schneider, 2002 (see Tatsuoka, 1988a, for further explanation). In Equation 1, X_p is the *p*th continuous variable and v_p, the raw weight associated with the *p*th variable. The raw weights are not readily interpretable and must be converted into other coefficients. One vector of coefficients commonly used for lending meaning to the linear function is the vector of structure coefficients (SCs). In the case of two groups, only one linear function is possible. Therefore, the SSCP matrix for the total sample is reduced to a scalar, T. Using notation from Tatsuoka (1988b), if **D(.)** represents the diagonal elements of a given matrix, then the matrix of SCs based on total variance, **A**, can be written as:

$$\mathbf{A} = [\mathbf{D}(\mathrm{T})]^{-1/2}\,(\mathrm{T}\mathbf{V})\,[\mathbf{D}(\mathbf{V'}\mathrm{T}\mathbf{V})]^{-1/2}. \tag{2}$$

The elements of **A** are the SCs , a_1 thru a_p, associated with the single linear function, *Z*. Using the above algorithm, in order to calculate SCs based on pooled within-group variance as opposed to total variance, one need only substitute the scalar T with the scalar for pooled within-group variance, W. In the case of Equation 2, SCs are normalized because calculation of **A** includes multiplying the square root of the inverse matrix, $[\mathbf{D}(\mathbf{V'}\mathrm{T}\mathbf{V})]^{-1}$. Thus, any given vector of SCs is restricted to a length of one. However, in popular statistical computer packages (SAS and SPSS), SCs are not normalized. Furthermore, DA output in SAS includes SCs based on both total variance and pooled within-group variance; concordant output in SPSS includes only SCs based on within-group variance. Because SCs are not normalized in contemporary statistical computer software, as it applies to non-normalized SCs, Equation 3 may be rewritten as a modification of Tatsuoka's algorithm from Equation 2:

$$\mathbf{A}_{\mathbf{non}} = [\mathbf{D}(\mathrm{T})]^{-1/2}\,(\mathrm{T}\mathbf{V})\ , \tag{3}$$

where the resulting matrix on SCs, $\mathbf{A}_{\mathbf{non}}$, is not normalized.

Researchers utilize DA for two primary purposes: prediction or description (Hubery, 1994). The first purpose involves predicting group membership based on the vector of *p* continuous variables. In the second use of DA, instead of predicting group membership, the researcher is interested utilizing DA as a *post hoc* procedure following a significant MANOVA. The focus of this paper is on the latter use of DA, commonly known as descriptive discriminant analysis (DDA). In DDA, the researcher interprets the vector of *p* coefficients to understand, for example, which of the *p* variables contributed to separation on the grouping variable and which did not. As previously mentioned, the vector of raw weights must be converted into interpretable coefficients. This paper investigates two types of SCs: Those based on total correlation matrix, $\mathbf{s}_T$, and those based upon the pooled within-group variance, $\mathbf{s}_W$ (Dalgleish, 1994). As previously noted in reference to Equations 2 and 3, these two types of SCs are available in SAS, and only $\mathbf{s}_W$ is available in SPSS. Both $\mathbf{s}_T$ and $\mathbf{s}_W$ will be examined in this study. If the researcher's goal is to identify which among the *p* continuous variables are contributing to group separation by consulting SC values, then a rule of thumb would certainly be helpful. Regarding interpretation of SC values, Pedhazur (1997) notes that SC values $\geq .3$ "are treated as meaningful" (p. 910). Pedhazur also notes that rules of thumb might be problematic and refers the reader to Dalgleish regarding testing SC significance. However, tests of the significance of SCs are not readily available for researcher use, which may be why the researcher seeks a rule of thumb in the first place. Because tests of SC significance are not readily available, testing the rule of thumb might yield useful information regarding its application to SC interpretation.

Purpose of the Study

To date, no simulation study has examined the usefulness of the rule of thumb that SCs with values of .3 or greater might meaningfully identify continuous variables influential upon group separation in DA. The primary goal of the current study was to investigate conditions under which the rule might identify "meaningful" (formerly MANOVA dependent) variables when DA is used as a *post hoc* test following a significant MANOVA, known as *descriptive discriminant analysis* (DDA) (Huberty, 1994). In this work, operationalizing of "meaningful" variables is addressed in the Procedures section. In the presence of significant differences among group means, $\mathbf{s}_W$ might be preferred to $\mathbf{s}_T$ (Dalgleish, 1994; Huberty, 1975). However, because DDA research is inconclusive regarding the utility of SCs based on both the total matrix versus those based upon the within matrix (Schneider, 2004), both types of SCs will be compared in this study.

Procedures

SAS PROC IML was employed for the current Monte Carlo, two-group simulation, with two *p*-dimensional, multivariate population matrices generated, each being $\mathcal{N}(\mu, \Sigma)$ (SAS Institute, 1999). The general procedure was employed in Schneider, 2004: In all cells, μ_1 was a *p* x 1 null vector, and μ_2, a *p* x 1 vector of effects of some combination such that $\mu_1 \neq \mu_2$. A sample of dimension *n* x *p* was then drawn from each population ($n_1 = n_2$; $p_1 = p_2$) and analyzed as a two-way MANOVA using Wilks' Λ and a special case of Bartlett's V as a test of significance:

$$V = -[N - 1 - (p + 2) / 2] \ln \Lambda \quad (3)$$

where Λ is also calculated using a modified formula

$$\Lambda = 1 / (1 + \lambda) \quad (4)$$

Based on Tatsuoka (1988a, 1988b), V is approximately a χ^2 distribution with *p* degrees of freedom. In the current work, the variables and corresponding levels manipulated were as follows:

1. Continuous variable levels $p = 2, 3, 4,$ and 5.
2. Group sample size $n = 50, 100, 150$ and 200.
3. The population correlation matrices, $\mathbf{P}_1$ and $\mathbf{P}_2$. Five levels were used, reflecting five possible ranges of *p* intercorrelation (hereby denoted as **P** for population and **R**, for sample): 0 - .20; .21 - .40; .41 - .60; .61 - .80, and .81 – 1.00. For a given experiment, the exact correlation for the two groups between continuous variables *p* and *p'* (where $p \neq p'$) was randomly generated within any one of these five ranges. The two most highly correlated ranges were included to consider potential effects of collinearity upon the rule of thumb.

4. Population mean vector, μ_2. As previously mentioned, μ_1 was held constant as a null vector. Thus, μ_2 was manipulated as the vector of effects. The *p* elements of a given μ_2 were some combination of effects, with three possible levels of effect size: .2, .5 and .8. These levels were based upon Cohen's (1992) determination of small, medium, and large effects, respectively, for two independent means. For the purposes of this study, all three levels of effect were considered "meaningful," as all three could contribute to MANOVA significance. All combinations of .2, .5, and .8 were investigated, for a total of 45 *p* x 1 mean vector pairs: 6 for *p* = 2; 10 for *p* = 3; 13 for *p* = 4, and 16 for *p* = 5.

Each *n* x *p* cell was replicated 5,000 times. For the replications where the MANOVA null hypothesis $\mathbf{H_0}$: $\mu_1 = \mu_2$ was correctly rejected within each cell, the *p* x 1 vectors of total and within SCs, $\mathbf{s_T}$ and $\mathbf{s_W}$, representing the first discriminant function, were calculated, and the proportions of $\mathbf{s_T}$ and $\mathbf{s_W}$ vectors conforming to rule of thumb (SC element values $\geq$ .3) were subsequently calculated. For all vectors, regardless of MANOVA significance, information on the proportion of individual elements conforming to the rule of thumb was also tabulated.

Results

In general, the rule of thumb that an SC value $\geq$ .3 indicates a continuous variable contributing to group separation works best for vectors involving medium and large effects across all *n* x *p* cells. The exceptions appear to be cells where *p* variable intercorrelation was highest (**R** = .81 - .99). In the case of highest intercorrelation, the proportion of sample vectors fitting the rule dropped notably, but only where the elements were some combination of medium and large effects, not where entire vectors contained either medium or large effects. As for vectors with small effects, the rule fit best where the entire vector was comprised of small effects. For remaining vectors containing at least one small effect, the rule did not fit vectors well at all. However, when the proportion of elements fitting the rule was examined (as opposed to proportions of entire vectors), it is clear that the elements with small effects are responsible for entire vector ill fit. A final notable finding is that overall, $\mathbf{s_T}$ outperformed $\mathbf{s_W}$.

In subsequent sections, discussion focuses first on vectors with either all medium or all large effects. Next are results for vectors with combined medium and large effects. Final discussion is on vectors containing 1) all small effects and 2) at least one small effect.

Vectors with All Medium or All Large Effects

Table 1 includes the proportions of both $\mathbf{s_T}$ and $\mathbf{s_W}$ conforming to the rule for *p* = 4 where all elements have medium effects (.5, as noted in Cohen, 1992). Table 2 contains the same information for *p* = 5 for all large effects (.8, as noted in Cohen). If the proportion of sample vectors conforming to the rule equaled or exceeded .8, then the result is reported in bold in all tables, for this result is deemed as indicating the rule worked well for such a cell.

For the two examples above where all effects were either medium or large, the rule of thumb worked well for almost all cells, even those with highest *p* variable intercorrelation. Note also that there appears to be little difference in the proportions of both $\mathbf{s_T}$ and $\mathbf{s_W}$ conforming to the rule, with one exception: the cell with the smallest group sample size *n* = 50 and lowest intercorrelation, **R** = .00 - .21 in Table 2.

Table 1. Proportions of SCs for *p* = 4 Vector with all medium effects

n	*R*	.00 - .20	.21 - .40	.41 - .60	.61 - .80	.81 - .99
50		.6919	**.8222**	**.9049**	**.9584**	**.9790**
		.6190	*.7886*	***.8900***	***.9488***	***.9766***
100		**.9047**	**.9548**	**.9819**	**.9914**	**.9976**
		.8718	***.9403***	***.9801***	***.9903***	***.9976***
150		**.9714**	**.9886**	**.9956**	**.9988**	**1.000**
		.9526	***.9858***	***.9942***	***.9988***	***1.000***
200		**.9772**	**.9956**	**.9988**	**.9998**	**.9984**
		.9610	***.9930***	***.9986***	***.9998***	***.9978***

Note: S_T are in Roman font, and S_W are italicized. Proportions $\geq$ .8 are in bold.

Table 2. Proportions of SCs for *p* = 5 Vector with all large effects

n	*R*	.00 - .20	.21 - .40	.41 - .60	.61 - .80	.81 - .99
50		**.9208**	**.9738**	**.9910**	**.9969**	**.9976**
		.7288	***.9426***	***.9841***	***.9950***	***.9976***
100		**.9956**	**.9994**	**1.000**	**1.000**	**1.000**
		.9636	***.9990***	***1.000***	***1.000***	***1.000***
150		**.9996**	**1.000**	**1.000**	**1.000**	**1.000**
		.9978	***1.000***	***.9998***	***1.000***	***1.000***
200		**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
		.9998	***1.000***	***1.000***	***1.000***	***1.000***

Note: S_T are in Roman font, and S_W are italicized. Proportions ≥ .8 are in bold.

Table 3. Proportions of SCs for *p* = 5 Vector with all medium effects

n	*R*	.00 - .20	.21 - .40	.41 - .60	.61 - .80	.81 - .99
50		.5525*	.7494**	**.8612******	**.9288**	**.9744**
		.4341	*.6875*	***.8338***	***.9170***	***.9694***
100		**.8322**	**.9176**	**.9691**	**.9906**	**.9966**
		.7443	***.8961***	***.9628***	***.9886***	***.9959***
150		**.9502**	**.9832**	**.9950**	**.9990**	**.9996**
		.9076	***.9774***	***.9930***	***.9986***	***.9996***
200		**.9700**	**.9954**	**.9978**	**1.000**	**.9998**
		.9364	***.9926***	***.9972***	***1.000***	***.9998***

Note: SC_T are in Roman font, and SC_W are italicized. Proportions ≥ .8 are in bold.
*d_1= .5525 - .4341 = **.1184** **d_2= .7494 - .6875 = **.0619** ***d_3= .7494 - .6875 = **.0619**

In this case, $\mathbf{s_T}$ noticeably outperforms $\mathbf{s_W}$ in identifying correct contribution to group separation based upon the rule of thumb ($\mathbf{s_T}$ = .9208; $\mathbf{s_W}$ = .7288). For other *p* vectors with either all medium or all large effects, $\mathbf{s_W}$ did not fare as well as $\mathbf{s_T}$ for lower levels of *n* and/or **R**, with the most notable difference evident where *p* = 5 and all effects were medium (Table 3).

As is evident in Table 3, the difference in performance of $\mathbf{s_T}$ versus $\mathbf{s_W}$ is less noticeable as proportion of sample vectors conforming to the rule increases. Furthermore, for cells where all effects are either medium or large, an *n* x **R** interaction is evident. As for the difference in performance of $\mathbf{s_T}$ and $\mathbf{s_W}$, where *n* = 50, this difference is most noticeable for the cell with the lowest intercorrelation (d_1 = .5525 - .4341 = **.1184**) and less noticeable as intercorrelation increases (d_2 = .7494 - .6875 = **.0619**; d_3 = .8612 - .8338 = **.0274**).

Vectors with a Combination or Medium and Large Effects

As previously mentioned, the rule of thumb fit well for vectors with a combination of medium and large effects, with the exception of cells with the highest *p* variable intercorrelation. Tables 4 thru 6 provide representative examples of the drop in the proportion of conforming cells at highest intercorrelation for cells where *p* = 3, 4, and 5, respectively. As was true for vectors with all medium or all large effects, an *n* x **R** interaction is present, this time for all cells excepting those of the highest intercorrelation. In this interaction, fewer cells fit the rule of thumb as *n* and **R** decreased. Furthermore, regardless of the number of *p* variables, the rule worked well consistently for cells in the center of the tables, with sample sizes *n* ≥ 100 and intercorrelation **R** = .21 - .80.

As the results in Table 5 show, not all cells for the highest level of intercorrelation yielded poor results. For cells with highest intercorrelation where group sample size *n* = 100 and 200, the rule fit well (Table 5) (*n* = 100: $\mathbf{s_T}$ = .9194; $\mathbf{s_W}$ = .8562, and *n* = 200: $\mathbf{s_T}$ = .9940; $\mathbf{s_W}$ = .9882). However, even where the rule fit erratically for cells with highest correlation (Table 5) as opposed to not fitting well at all (Tables 4 and 6), no pattern was evident except that $\mathbf{s_T}$ consistently outperformed $\mathbf{s_W}$.

Table 4. Proportions of SCs for *p* = 3 Vector with two medium and one large effect

n	*R*	.00 - .20	.21 - .40	.41 - .60	.61 - .80	.81 - .99
50		.7787	**.8431**	**.8707**	**.8914**	.7708
		.7019	***.8015***	***.8476***	***.8678***	*.6516*
100		**.9298**	**.9446**	**.9608**	**.9716**	.4500
		.8910	***.9196***	***.9505***	***.9590***	*.0074*
150		**.9698**	**.9848**	**.9894**	**.9896**	.6174
		.9518	***.9722***	***.9826***	***.9826***	*.0860*
200		**.9886**	**.9940**	**.9958**	**.9960**	.6844
		.9746	***.9896***	***.9928***	***.9926***	*.1196*

Note: S_T are in Roman font, and S_W are italicized. Proportions $\geq$ .8 are in bold.

Table 5. Proportions of SCs for *p* = 4 Vector with one medium and three large effects

n	*R*	.00 - .20	.21 - .40	.41 - .60	.61 - .80	.81 - .99
50		.7666	**.8484**	**.8890**	**.9012**	.7246
		.6286	*.7878*	***.8493***	***.8732***	*.5310*
100		**.8988**	**.9330**	**.9548**	**.9652**	**.9194**
		.7958	***.8802***	***.9318***	***.9476***	***.8562***
150		**.9424**	**.9664**	**.9838**	**.9816**	.3008
		.8392	***.9360***	***.9728***	***.9702***	*0*
200		**.9562**	**.9840**	**.9920**	**.9936**	**.9940**
		.8640	***.9618***	***.9778***	***.9882***	***.9882***

Note: S_T are in Roman font, and S_W are italicized. Proportions $\geq$ .8 are in bold.

Table 6. Proportions of SCs for *p* = 5 Vector with four medium and one large effect

n	*R*	.00 - .20	.21 - .40	.41 - .60	.61 - .80	.81 - .99
50		.5041	.6701	.7679	**.8094**	.4996
		.3466	*.5824*	*.7156*	*.7616*	*.1918*
100		.7828	**.8774**	**.9260**	**.9279**	.7004
		.6382	***.8252***	***.9014***	***.9033***	*.4170*
150		.8810	**.9606**	**.9736**	**.9828**	.3928
		.7662	***.9334***	***.9622***	***.9740***	*.0022*
200		**.9554**	**.9852**	**.9914**	**.9936**	.5884
		.8982	***.9718***	***.9832***	***.9894***	*.0400*

Note: SC_T are in Roman font, and SC_W are italicized. Proportions $\geq$ .8 are in bold.

Vectors with All Small Effects

Tables 7 though 10 contain the results for the four *p* x 1 mean vector pairs where all elements had small effects (.2, as noted in Cohen, 1992). In the case of vectors containing small effects, only those with all small effects had cells that fit well with the rule of thumb that an SC value $\geq$ .3 indicates a continuous variable contributing to group separation. As was true for previously reported results, if the proportion of sample vectors conforming to the rule equaled or exceeded .8, then the result is reported in bold in all tables, for this result is deemed as indicating the rule worked well for such a cell.

It is clear that as one reads Tables 7 through 10 in sequence, the proportion of SC values fitting the rule well (i.e., cells with the proportion equaling or exceeding .8 for SC vectors conforming to the rule) narrows as the number of *p* variables increases. Specifically, in the case of vectors with all small effects, as the number of *p* variables increases, both group sample size *n* and intercorrelation **R** must increase in order for the rule of thumb to work well. The vectors with all small effects present the clearest evidence of a three-way interaction, *p* x *n* x **R**, in this entire study. The most noticeable decrease in the number of *n* x **R** cells fitting the rule occurs when the number of *p* variables increases from 2 to 3. Seventeen cells show SC proportions fitting the rule as equaling or exceeding .8 where *p* = 2. However, this number drops to ten cells where *p* = 3. The drop is no so drastic when *p* = 4 (6 cells fit the rule) or *p* = 5 (3 cells).

Table 7. Proportions of SCs for *p* =2 Vector with both small effects

n	*R*	.00 - .20	.21 - .40	.41 - .60	.61 - .80	.81 - .99
50		.6859	.7728	**.8261**	**.8593**	**.8811**
		.6726	*.7546*	***.8152***	***.8552***	***.8811***
100		.7957	**.8298**	**.8738**	**.9176**	**.9382**
		.7874	***.8246***	***.8690***	***.9153***	***.9373***
150		**.8479**	**.8989**	**.9176**	**.9460**	**.9701**
		.8437	***.8984***	***.9166***	***.9455***	***.9701***
200		**.8658**	**.9169**	**.9422**	**.9816**	**.9878**
		.8616	***.9162***	***.9396***	***.9803***	***.9878***

Note: S_T are in Roman font, and S_W are italicized. Proportions $\geq$ .8 are in bold.

Table 8. Proportions of SCs for *p* = 3 Vector with all small effects

n	*R*	.00 - .20	.21 - .40	.41 - .60	.61 - .80	.81 - .99
50		.4670	.5347	.6466	.7469	**.8159**
		.4336	*.5197*	*.6248*	*.7362*	***.8035***
100		.5560	.6511	.7635	**.8219**	**.8523**
		.5429	*.6416*	*.7583*	***.8201***	***.8468***
150		.6638	.7673	**.8162**	**.8941**	**.8996**
		.6543	*.7595*	***.8111***	***.8911***	***.8980***
200		.7139	**.8280**	**.8720**	**.9284**	**.9733**
		.7060	***.8237***	***.8685***	***.9252***	***.9728***

Note: S_T are in Roman font, and S_W are italicized. Proportions $\geq$ .8 are in bold.

Table 9. Proportions of SCs for *p* = 4 Vector with all small effects

n	*R*	.00 - .20	.21 - .40	.41 - .60	.61 - .80	.81 - .99
50		.2638	.3811	.5202	.6331	.7382
		.2257	*.3585*	*.4933*	*.6071*	*.7236*
100		.3335	.4943	.6832	.7376	**.8466**
		.3126	*.4754*	*.6741*	*.7312*	***.8433***
150		.4891	.6134	.7491	**.8327**	**.8952**
		.4707	*.6049*	*.7448*	***.8301***	***.8931***
200		.4626	.6638	**.8125**	**.9086**	**.9339**
		.4446	*.6548*	***.8070***	***.9070***	***.9294***

Note: S_T are in Roman font, and S_W are italicized. Proportions $\geq$ .8 are in bold.

Table 10. Proportions of SCs for *p* = 5 Vector with all small effects

n	*R*	.00 - .20	.21 - .40	.41 - .60	.61 - .80	.81 - .99
50		.1168	.2471	.4252	.5507	.6601
		.0880	*.2096*	*.3896*	*.5217*	*.6444*
100		.1825	.3750	.5706	.7444	.7868
		.1557	*.3530*	*.5517*	*.7363*	*.7797*
150		.2229	.4943	.6636	.7991	**.8804**
		.2004	*.4731*	*.6512*	*.7937*	***.8755***
200		.2938	.5708	.7326	**.8276**	**.9222**
		.2757	*.5569*	*.7231*	***.8229***	***.9210***

Note: SC_T are in Roman font, and SC_W are italicized. Proportions $\geq$ .8 are in bold.

As was true for all previous *p* x 1 mean vectors of effects discussed in this paper, $\mathbf{s_T}$ consistently outperformed $\mathbf{s_W}$, with the differences between $\mathbf{s_T}$ and $\mathbf{s_W}$ proportions being less pronounced as proportions of SC vectors fitting the rule increased. Too, as was true for vectors containing either all medium or all large effects, the drop was not present in proportions conforming to the rule for cells with the highest intercorrelation.

Table 11. Proportions of SCs for *p* = 2 Vector with one small and one large effect

n	*R*	.00 - .20	.21 - .40	.41 - .60	.61 - .80	.81 - .99
50		.4594	.4551	.4254	.4266	.2118
		.4154	*.4145*	*.3757*	*.3685*	*.1130*
100		.4355	.4238	.3999	.2780	.2120
		.3896	*.3767*	*.3449*	*.1986*	*.1240*
150		.4180	.4246	.4006	.3226	.0112
		.3612	*.3606*	*.3390*	*.2532*	*0*
200		.3906	.3948	.3888	.3160	.0402
		.3280	*.3278*	*.3194*	*.2352*	*.0048*

Note: S_T are in Roman font, and S_W are italicized. Proportions $\geq$ .8 are in bold.
MANOVA power: $1 - \beta$ = .9504

Table 12. Proportions of SCs for *p* = 3 Vector with two small and one medium effect

n	*R*	.00 - .20	.21 - .40	.41 - .60	.61 - .80	.81 - .99
50		.3320	.4031	.4321	.4457	.3164
		.2978	*.3715*	*.3972*	*.4195*	*.2698*
100		.3866	.4627	.4790	.4329	.0116
		.3543	*.4377*	*.4560*	*.4063*	*0*
150		.4333	.4896	.5112	.4376	.0056
		.4063	*.4679*	*.4884*	*.4055*	*0*
200		.4544	.5199	.5235	.5529	.0148
		.4201	*.4964*	*.4950*	*.5255*	*.0006*

Note: S_T are in Roman font, and S_W are italicized. Proportions $\geq$ .8 are in bold.
MANOVA power: $1 - \beta$ = .5964

Table 13. Proportions of SCs for *p* = 4 Vector with one small and three large effects

n	*R*	.00 - .20	.21 - .40	.41 - .60	.61 - .80	.81 - .99
50		.2256	.2944	.3277	.2735	.1912
		.1262	*.2086*	*.2463*	*.1879*	*.0806*
100		.1874	.2412	.2280	.1752	.0072
		.0874	*.1620*	*.1540*	*.0900*	*0*
150		.1154	.1712	.2168	.1746	.0868
		.0418	*.0886*	*.1346*	*.0900*	*.0218*
200		.0998	.1128	.1870	.0816	.0004
		.0328	*.0374*	*.1006*	*.0202*	*0*

Note: S_T are in Roman font, and S_W are italicized. Proportions $\geq$ .8 are in bold.
MANOVA power: $1 - \beta$ = 1.000

Table 14. Proportions of SCs for *p* = 5 Vector with two small, one medium, two large effects

n	*R*	.00 - .20	.21 - .40	.41 - .60	.61 - .80	.81 - .99
50		.0953	.1329	.1281	.1200	.0338
		.0506	*.0831*	*.0750*	*.0476*	*0*
100		.0546	.0912	.1086	.1052	.0786
		.0192	*.0494*	*.0502*	*.0504*	*.0178*
150		.0240	.0824	.0836	.1162	.0016
		.0046	*.0402*	*.0350*	*.0638*	*0*
200		.0224	.0530	.0520	.0574	.0084
		.0054	*.0180*	*.0168*	*.0166*	*0*

Note: SC_T are in Roman font, and SC_W are italicized. Proportions $\geq$ .8 are in bold.
MANOVA power: $1 - \beta$ = .9928

Mixed Vectors Containing at Least One Small Effect

Tables 11 through 14 contain results representative of the 23 *p* x 1 vectors involving at least one element with a small effect. What is clear from examination of Tables 11 through 14 is that the rule of thumb that an SC value ≥ .3 indicates a continuous variable contributing group separation does not fit well at all for vectors containing even only one element with a small effect. Furthermore, there is no pattern evident across these tables to indicate that increased group sample size would remediate the problem. For example, in Table 12, for all except the highest level of *p* variable intercorrelation (**R** .00 - .80), it seems that as group sample size *n* increases, so does the proportion of cells fitting the rule. However, results of Table 13 show the opposite effect: For the same range of **R** intercorrelation, the proportion of cells fitting the rule decreases as group sample size *n* increases. Finally, Tables 11 and 14 yield results neither consistently increasing nor decreasing as group size *n* increases but instead fluctuates, showing yet another pattern for all except the highest level of intercorrelation (**R** .00 - .80).

As for the highest level of *p* variable intercorrelation (**R** = .81 - .99), proportions of cells fitting the rule drop in a similar fashion to the vectors containing mixed medium and large effects (Tables 4 – 6). In the case of vectors including at least one small effect (Tables 11 – 14), cells were likely to approach or reach zero proportions fitting the rule than was true for vectors containing only medium and large effects. Too, even though general results were poor for both types of SCs, $\mathbf{s_T}$ continued to outperform $\mathbf{s_W}$ as far as rule fit was concerned.

An interesting comparison involves Tables 5 and 13. The only difference between the *p* = 4 vectors for the two tables is that the single medium effect in Table 5 is replaced with a small effect in Table 13. In both vectors, the remaining three elements are large effects. What is noteworthy is the difference the change from medium to small effect has upon the fit of the entire vector to the rule of thumb. Tables 15 through 18 contain detailed information regarding rule fit for specific elements in a vector and correspond to Tables 11 though 14, respectively, where the information is on rule fit for entire vectors. As one can see, the presence of elements with small effects in these mixed vectors would reduce the fit of the entire vector. For example, the proportion of vectors conforming to the rule where *p* = 2 with one small and one large effect where the MANOVA was correctly rejected (MANOVA power: $1 - \beta = .9504$.) was $\mathbf{s_T}$ = .4594 and $\mathbf{s_W}$ = .4154 for *n* = 50 and **R** = .00 - .20 (Table 11). However, as one examines proportions of elements conforming to the rule for these same conditions across all 5000 replications (Table 15), one sees that the proportion fitting the rule was high for the element with the large effect (both $\mathbf{s_T}$ and $\mathbf{s_W}$ = .9502) and low for the element with the small effect ($\mathbf{s_T}$ = .4368 and $\mathbf{s_W}$ = .3950). Another example involves the proportion of vectors conforming to the rule where *p* = 5 with two small, one medium, and two large effect where the MANOVA was correctly rejected (MANOVA power: $1 - \beta = .9928$.) was $\mathbf{s_T}$ = .1052 and $\mathbf{s_W}$ = .0504 for *n* = 100 and **R** = .61 - .80 (Table 14). As one examines proportions of elements conforming to the rule for these same conditions across all 5000 replications (Table 18), one sees that the proportion fitting the rule was high for the elements with the one medium ($\mathbf{s_T}$ = .9182 and $\mathbf{s_W}$ = .8466) and two large effects (both elements $\mathbf{s_T}$ = .9998 and $\mathbf{s_W}$ = .9996) and low for the two elements with small effects (first small element: $\mathbf{s_T}$ = .2110 and $\mathbf{s_W}$ = .1186; second small element: $\mathbf{s_T}$ = .2104 and $\mathbf{s_W}$ = .1218).

Discussion

Research practitioners often search the literature for guidelines regarding interpretation of statistical analysis results. The researcher interested in interpreting structure coefficients (SCs) in discriminant analysis (DA) might use the rule of thumb as noted in Pedhazur (1997) that an SC value ≥ .3 indicates a continuous variable useful for contributing to separation on the grouping variable. However, this rule has apparently not been tested before this study. In the case of two-group MANOVA, results indicate that the rule of thumb works well for vectors with medium or large effects (.5 and .8, respectively, as noted in Cohen, 1992) but not well for small effects (.2, as noted in Cohen). The exception appears to be *p* = 2 continuous variables where both effects are small (Table 7). Because the most common effect size in many fields is the medium effect size (Cohen), the rule of thumb could prove useful for practitioners despite the apparent poor results for vectors involving small effects.

Table 15. Proportions of Elements Fitting the Rule for the *p* = 2 Vector in Table 11Population Effects: One Small and One Large Effect, Respectively

n	*R*	.00 - .20	.21 - .40	.41 - .60	.61 - .80	.81 - .99
50		.4368	.4320	.4142	.4192	.2140
		.9502	**.9492**	**.9736**	**.9812**	**.9896**
		.3950	*.3934*	*.3658*	*.3626*	*.1138*
		.9502	***.9492***	***.9736***	***.9810***	***.9762***
100		.4352	.4342	.3996	.2780	.2120
		.9994	**.9986**	**.9992**	**1.000**	**1.000**
		.3894	*.3762*	*.3446*	*.1986*	*.1240*
		.9994	***.9986***	***.9992***	***1.000***	***.9996***
150		.4180	.4246	.4006	.3226	.0112
		1.000	**1.000**	**1.000**	**1.000**	**.9970**
		.3612	*.3606*	*.3390*	*.2532*	*0*
		1.000	***1.000***	***1.000***	***1.000***	*.7200*
200		.3906	.3948	.3888	.3160	.0402
		1.000	**1.000**	**1.000**	**1.000**	**1.000**
		.3280	*.3278*	*.3194*	*.2352*	*.0048*
		1.000	***1.000***	***1.000***	***1.000***	***1.000***

Note: S_T are in Roman font, and S_W are italicized. Proportions ≥ .8 are in bold.

Table 16. Proportions of Elements Fitting the Rule for the *p* = 3 Vector in Table 12 Population Effects: Two Small and One Medium Effect, Respectively

n	*R*	.00 - .20	.21 - .40	.41 - .60	.61 - .80	.81 - .99
50		.3536	.3498	.3420	.3750	.3926
		.3508	.3458	.3356	.3708	.3934
		.5908	.5476	.5414	.6380	**.8838**
		.3374	*.3368*	*.3266*	*.3574*	*.3364*
		.3346	*.3312*	*.3198*	*.3526*	*.3416*
		.5902	*.5474*	*.5412*	*.6362*	***.8616***
100		.5610	.5604	.5676	.5564	.0216
		.5470	.5688	.5618	.5506	.0248
		.8906	**.8570**	**.8614**	**.9428**	.5264
		.5404	*.5428*	*.5502*	*.5332*	*0*
		.5252	*.5492*	*.5466*	*.5268*	*0*
		.8906	***.8570***	***.8614***	***.9426***	*.0238*
150		.6490	.6574	.6514	.5828	.0068
		.6478	.6462	.6560	.5826	.0058
		.9766	**.9674**	**.9624**	**.9970**	.5626
		.6288	*.6424*	*.6320*	*.5510*	*0*
		.6294	*.6304*	*.6402*	*.5502*	*0*
		.9764	***.9674***	***.9624***	***.9970***	*.0134*
200		.6688	.6980	.6872	.6562	.0296
		.6712	.7160	.6748	.6550	.0254
		.9974	**.9944**	**.9960**	**.9976**	**.8472**
		.6740	*.6802*	*.6660*	*.6342*	*.0018*
		.6454	*.6978*	*.6526*	*.6296*	*.0012*
		.9974	***.9944***	***.9960***	***.9976***	*.4308*

Note: S_T are in Roman font, and S_W are italicized. Proportions ≥ .8 are in bold.

Table 17. Proportions of Elements Fitting the Rule for the p = 4 Vector in Table 13 Population Effects: One Small and Three Large Effects, Respectively

n	*R*	.00 - .20	.21 - .40	.41 - .60	.61 - .80	.81 - .99
50		.2302	.2954	.3280	.2772	.1922
		.9906	**.9936**	**.9914**	**.9940**	**.9866**
		.9912	**.9942**	**.9918**	**.9940**	**.9856**
		.9912	**.9934**	**.9918**	**.9944**	**.9840**
		.1326	*.2106*	*.2466*	*.1906*	*.0810*
		.9796	***.9908***	***.9894***	***.9916***	***.9534***
		.9780	***.9902***	***.9892***	***.9904***	***.9560***
		.9804	***.9900***	***.9892***	***.9914***	***.9558***
100		.1874	.2412	.2280	.1754	.0072
		.9998	**.9998**	**.9998**	**.9992**	**.9564**
		1.000	**1.000**	**1.000**	**1.000**	**.9558**
		1.000	**1.000**	**.9998**	**.9998**	**.9590**
		.0876	*.1620*	*.1540*	*.0900*	*0*
		.9992	***.9998***	***.9998***	***.9994***	*.0084*
		.9986	***.9998***	***1.000***	***.9994***	*.0078*
		.9998	***1.000***	***.9998***	***.9992***	*.0080*
150		.1154	.1712	.2168	.1746	.0860
		1.000	**1.000**	**1.000**	**1.000**	**1.000**
		1.000	**1.000**	**1.000**	**1.000**	**1.000**
		1.000	**1.000**	**1.000**	**1.000**	**1.000**
		.0418	*.0886*	*.1346*	*.0900*	*.0218*
		1.000	***1.000***	***1.000***	***1.000***	***.9994***
		1.000	***.9998***	***1.000***	***1.000***	***.9996***
		.9998	***1.000***	***1.000***	***1.000***	***.9998***
200		.0998	.1128	.1870	.0816	.0004
		1.000	**1.000**	**1.000**	**1.000**	**.9956**
		1.000	**1.000**	**1.000**	**1.000**	**.9982**
		1.000	**1.000**	**1.000**	**1.000**	**.9976**
		.0328	*.0374*	*.1006*	*.0202*	*0*
		1.000	***1.000***	***1.000***	***1.000***	*.0302*
		1.000	***1.000***	***1.000***	***1.000***	*.0304*
		1.000	***1.000***	***1.000***	***1.000***	*.0308*

Note: S_T are in Roman font, and S_W are italicized. Proportions ≥ .8 are in bold.

As for the idea that SCs based upon pooled within group variance ($\mathbf{s_W}$) might outperform those based on total variance ($\mathbf{s_T}$) in the presence of significant differences among group means (Dalgliesh, 1994; Huberty, 1975), the results of this study indicate that the opposite is true. For two-group MANOVA where the MANOVA was significant and subsequent descriptive discriminant analysis (DDA) conducted, $\mathbf{s_T}$ consistently outperformed $\mathbf{s_W}$. However, when the rule fit well (SC vector proportions fitting the rule ≥ .8), differences between $\mathbf{s_T}$ and $\mathbf{s_W}$ proportions were minimal. This minimal difference when the rule fits well is important given that SPSS DDA output includes $\mathbf{s_W}$ coefficients and not $\mathbf{s_T}$. If the conditions the researcher is testing are conditions where SC rule of thumb fit is high, the researcher using SPSS for DDA need not be concerned about not having $\mathbf{s_T}$ coefficients available.

For vectors containing either all small, medium or large effects, there was a three-way interaction such that as group sample size, *n*, and *p* variable intercorrelation, **R**, increases for a *p* x 1 vector of effects, the proportion of vectors fitting the rule increases. However, as the number of continuous variables, *p*, increases, the proportion of vectors fitting the rule decreases. This may be an issue of power; as the number of *p* variables increases, multivariate power generally decreases (Stevens, 2002).

Table 18. Proportions of Elements Fitting the Rule for the *p* = 5 Vector in Table 14 Population Effects: Two Small, One Medium, Two Large Effects, Respectively

n	*R*	.00 - .20	.21 - .40	.41 - .60	.61 - .80	.81 - .99
50		.3098	.3270	.2794	.1978	.0470
		.3010	.3334	.2860	.1964	.0452
		.8294	**.8458**	**.8314**	.7452	.4154
		.9896	**.9768**	**.9906**	**.9866**	**.8960**
		.9904	**.9766**	**.9902**	**.9880**	**.8980**
		.2206	*.2546*	*.1956*	*.0906*	*0*
		.2162	*.2506*	*.2000*	*.0940*	*0*
		.7626	*.7980*	*.7500*	*.5602*	*.0012*
		.9872	***.9762***	***.9868***	***.9650***	*.0774*
		.9884	***.9766***	***.9860***	***.9676***	*.0758*
100		.2096	.2676	.2348	.2110	.1086
		.2140	.2668	.2452	.2104	.1186
		.9168	**.9424**	**.9290**	**.9182**	**.8294**
		1.000	**1.000**	**.9998**	**.9998**	**.9994**
		1.000	**1.000**	**.9998**	**.9998**	**.9998**
		.1214	*.1912*	*.1454*	*.1186*	*.0302*
		.1200	*.1876*	*.1480*	*.1218*	*.0326*
		.8482	***.9096***	***.8776***	***.8466***	*.6020*
		.9998	***1.000***	***.9996***	***.9996***	***.9956***
		1.000	***1.000***	***.9996***	***.9996***	***.9960***
150		.1634	.2440	.1882	.2060	.0038
		.1608	.2430	.1980	.2032	.0028
		.9574	**.9776**	**.9628**	**.9752**	.4410
		1.000	**1.000**	**1.000**	**1.000**	**.9916**
		1.000	**1.000**	**1.000**	**1.000**	**.9926**
		.0810	*.1584*	*.1040*	*.1268*	*0*
		.0740	*.1562*	*.1130*	*.1222*	*0*
		.8970	***.9560***	***.9232***	***.9430***	*.0006*
		1.000	***1.000***	***1.000***	***1.000***	*.1784*
		1.000	***1.000***	***1.000***	***1.000***	*.1862*
200		.1392	.1836	.1470	.1056	.0160
		.1338	.1922	.1494	.1042	.0166
		.9792	**.9864**	**.9808**	**.9678**	**.8188**
		1.000	**1.000**	**1.000**	**1.000**	**1.000**
		1.000	**1.000**	**1.000**	**1.000**	**1.000**
		.0624	*.1032*	*.0680*	*.0388*	*.0002*
		.0612	*.1036*	*.0740*	*.0344*	*.0002*
		.9424	***.9678***	***.9520***	***.9100***	*.3130*
		1.000	***1.000***	***1.000***	***1.000***	***.9956***
		1.000	***1.000***	***1.000***	***1.000***	***.9964***

Note: S_T are in Roman font, and S_W are italicized. Proportions $\geq$ .8 are in bold.

For vectors with mixed medium and large effects, the *n* x **R** interaction was also evident, but not for the highest level of intercorrelation, **R** = .81 - .99. For these mixed vectors, the fit of the rule dropped for certain group sample sizes *n* when intercorrelation was highest. Furthermore, the drop was more pronounced for $\mathbf{s_W}$ than $\mathbf{s_T}$. Proportions fitting the rule for mixed vectors containing small effects appeared less stable at high intercorrelation, with drops evident for certain *n* cells with intercorrelation **R** = .61 - .99 (e.g., Table 12). However, as previously noted, proportions of vectors fitting the rule were low where mixed vectors included small effects. Thus, the rule of thumb works better where *p* variable intercorrelation is low to moderate. If a researcher reduces collinearity in the continuous variable set in order to achieve a more parsimonious model for conducting MANOVA (Stevens, 2002), then the researcher might still confidently apply the rule of thumb to a *post hoc* DDA, provided that anticipated effects are medium and/or large.

References

Cohen, J. (1988). *Statistical power analysis for the social sciences* (2nd. ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.

Cooley, W. W., & Lohnes, P. R. (1971). *Multivariate data analysis*. New York: John Wiley & Sons.

Dalgleish, L. I. (1994). Discriminant analysis: Statistical inference using the jackknife and bootstrap procedures. *Psychological Bulletin*, *116*, 498-508.

Huberty, C. (1994). *Applied discriminant analysis*. New York: Wiley.

Huberty, C. (1975). The stability of three indices of relative variable contribution in discriminant analysis. *Journal of Experimental Education*, *44*, 59-64.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). Fort Worth, TX: Harcourt Brace College Publishers.

SAS Institute (1999). *SAS/IML user's guide: Version 8*. Cary, NC: Author.

Schneider, M. K. (2002). *A Monte Carlo investigation of the Type I error and power associated with descriptive discriminant analysis as a MANOVA post hoc procedure*. Unpublished doctoral dissertation. University of Northern Colorado.

Schneider, M. K. (2004). Comparison of the usefulness of within-group and total-group structure coefficients for identifying variable importance in descriptive discriminant analysis following a significant MANOVA: Examination of the two-group case. *Multiple Linear Regression Viewpoints*, *30*, 8-18.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum.

Tatsuoka, M. M. (1988a). Multivariate analysis of variance. In J. R. Nesselroade and R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed.). New York: Plenium Press.

Tatsuoka, M. M. (1988b). *Multivariate analysis: Techniques for educational and psychological research* (2nd ed.). New York: Macmillan Publishing.

Send correspondence to: Mercedes K. Schneider, Ph.D.
Ball State University
Email: mschneider@bsu.edu

Regression-Discontinuity with Nonparametric Bootstrap

Randall E. Schumacker
University of North Texas

Robert E. Mount
Dallas Independent School District

The regression-discontinuity design (RD) is a powerful methodological alternative to the quasi-experimental design when conducting evaluations. RD designs involve testing post-test mean treatment differences between the experimental and comparison group regression lines at the centered cutoff point for statistical significance. This study simulated a RD treatment effect of 7 points in simulated normal and non-normal data distributions. The bootstrap technique was then used to estimate stability of estimates. Evaluation data oftentimes is non-normal, so understanding whether this impacts the RD design analysis is important. The bias (difference between the observed treatment effect and bootstrap estimate) and the confidence intervals are reported. Bootstrap estimates are useful in understanding whether the treatment effect is stable and the amount of estimation error present in RD given underlying normal and non-normal distributions. Results indicated that estimates of RD treatment effects are not severely impacted by non-normal, positive skewed, distributions. Consequently, robust estimation methods and/or data transformations such as probit are most likely sufficient to provide accurate stable estimates of treatment effects when concerned about meeting assumptions in regression analyses.

The regression-discontinuity (RD) design is different from a quasi-experimental design in that the assignment status is determined on the basis of a cutoff score on the pre-test measure and the pre-test measure can be different than the post-test measure (Cappelleri & Trochim, 2000). The statistical analysis of data in an RD design involves testing post-test mean differences between the group regression lines at the centered cutoff point for statistical significance, i.e., treatment effect. The RD design is therefore very useful when researching programs, procedures, or treatments given on the basis of need or merit. The basic regression-discontinuity equation can be expressed as: $Y_{Post} = b_0 + b_1 X_{Pre} + b_2 Z + e$, where Y_{Post} = post-test measure, Z = assignment status, X_{Pre} = pre-test measure, e = residual error, and the b's are estimated sample regression weights. The regression weight for Z, the treatment effect variable, indicates the amount of gain or loss in the post-test assessment measure, i.e., a positive sign indicates gain, while a negative sign indicates loss. This study will explore RD with nonparametric bootstrap under normal and non-normal distributions. The purpose of this study will be to determine if the application of bootstrap is helpful in obtaining more stable estimates of treatment effect under non-normal data conditions, especially since RD is used in program evaluation where non-normal data is commonly encountered. The comparison of RD treatment effects using normal and non-normal data under simulated conditions will provide an understanding of how results may be affected.

Development of the regression-discontinuity design began in 1958 with a problem faced by Donald T. Campbell and his colleagues (Trochim, 1984). They were trying to assess differences between National Merit Scholarship Program participants and non-participants. In this situation, experimental design randomization was not possible, rather students above a cutoff point on the exam would receive a scholarship and those below the cutoff point would not. The regression discontinuity approach was developed because educational researchers quite often encountered this type of evaluation design. The researchers' dilemma was compounded by the fact that around the cutoff point, some students were awarded scholarships and some were not, depending on variables chosen for the analysis. Inconsistent assignment to groups based on the pre-test measure, around the cutoff point, was later named "fuzzy" group assignment. Specific techniques were developed for use in the situations where assignments of subjects with borderline pre-test scores were no longer made solely on the basis of the pretest score.

The regression-discontinuity design (RD) is a powerful methodological alternative to the quasi-experimental design when conducting evaluations. Thistlethwaite and Campbell (1960) proposed the RD approach to avoid problems inherent in ex-post facto designs that required matching of subjects. Bottenberg and Ward (1963) described the RD design as a special type of regression analysis involving two mutually exclusive groups that didn't necessarily require random assignment of subjects to an experimental and control group. The basic RD design requires a pre-test measure, a post-test measure, and an assignment status, i.e., received treatment or did not receive treatment. RD analyzes the treatment difference between the regression lines of the two groups at the cutoff score that was used to assign group membership (Trochim, Internet article). If the regression lines differ, then there is a *discontinuity*, the magnitude of which can be statistically interpreted. Regression-discontinuity designs do not randomly assign subjects nor control for extraneous variables. Unlike quasi-experimental designs, the regression-discontinuity design allows for some control over group assignment through use of a cut-off score.

Following its initial conceptualization, regression-discontinuity was not widely used until 1974 with the passage of Public Law 93-380, which attempted to standardize evaluation of the programs that various school districts were implementing to make use of Title I funds. Three evaluation models that school districts could use to evaluate various compensatory programs were created. RD was one of the models (Trochim 1984, McNeil 1984). Unfortunately, many school districts did not adopt the RD design, possibly because it was a more complex statistical analysis technique (Trochim & Davis, Internet article). In the decades that followed, the regression-discontinuity design has remained an underutilized technique in educational evaluation.

One of many problems educational researchers face when using regression-discontinuity is that data must meet all assumptions in regression analysis. Real-world data often has the potential to violate the assumptions of normally distributed data, equal variance, normally distributed residuals, and linear relationship between the predictor and outcome variables. Regression-discontinuity analysis in circumstances where such violations exist may produce biased estimates of the treatment effect. The bootstrap technique will be undertaken in order to assess the impact on RD of non-normal data. Not everyone is in agreement about using bootstrap techniques in least-squares regression to resolve problems related to these violations of assumptions. For example, Venables and Ripley (2001) view bootstrap as having little use in least-squares regression because if residual errors are close to being normally distributed, the standard theory applies. If not, robust regression estimation methods are available (p. 175) (Schumacker, Monahan, & Mount, 2002). A data transformation approach using a profile likelihood function is also suggested (p. 182). We have found that non-normal data can be transformed using probit to yield a more normal distribution prior to statistical analyses.

Method and Procedures

Nonparametric bootstrap is a resampling procedure with replacement (Searle, Internet article). In brief, bootstrap involves using the sample data to construct a theoretical pseudo-population, composed of repeated random samplings from the original data set. Each additional sample can be equal to the number in the original sample. If the original sample included a full range of the values that exist in the actual population, then each additional sample can be thought of as representative of the true population distribution. When the desired statistic is computed for the original sample and for all additional bootstrap samples, a reasonable idea of the population distribution and the error distribution of the test statistic may be obtained. In this study, the non-parametric bootstrap technique in S-PLUS was utilized (S-PLUS, 2005).

Data Sources

The normal and non-normal data distributions for the study were simulated using two S-PLUS script programs written by the author and run in S-PLUS (S-PLUS, 2005). The normal distribution used the function, *rnorm*, and the non-normal distribution used the function, *rexp* (Johnson & Kotz, 1970).

The normal distribution true score was based on a sample size of 500, mean of 14, and standard deviation of 1. Random residual error was added to the true score based on a sample size of 500, mean of 0, and standard deviation of 1. The pre-test (X) mean value of 14 was used to determine group assignment (z), i.e., z = 1 if x <= 14, else z = 0. The z = 1 denoted the experimental group and Z = 0 the comparison group. A 7-point post-test gain was added to values for members in the group assigned z = 1, i.e., a 7-point treatment effect was introduced (post-test – pre-test = 7 point gain). These results were centered prior to RD regression analysis (XC = X – 14) (Cohen, Cohen, West & Aiken, 2003). The resulting display of RD data for the experimental and comparison group are in Figure 1.

The non-normal distribution true score was based on a sample size of 500 and a rate of 2 (inverse of mean to produce skew) that yielded a skewed distribution with mean of 14 and a positively skewed long right tailed distribution. Residual error was similarly added to the true score, but used a rate of 4. The pre-test (X) mean value of 14 was once again used to determine group assignment (z), i.e., z = 1 if x <= 14, else z = 0. A 7-point post-test gain was therefore added to values for members in the group assigned z = 1, i.e., a 7-point treatment effect was introduced (post-test – pre-test = 7 point gain). These results were also centered prior to RD regression analysis (XC = X – 14). The *Moment* formula for skewness and kurtosis should be used rather than the *Fisher* formula when resampling using bootstrap or jackknife procedures (S-PLUS, 2005). Figure 2 displays of RD data for the experimental and comparison groups.

The pre-test mean was 14, the post-test mean was 21, and therefore a known treatment effect of 7.0 was specified in both the normal and non-normal distributions. The non-normal distribution however was created to be positively skewed (Figure 2). For both types of distributions, the resulting treatment effect mean bootstrap estimate was compared to the pseudo population treatment effect mean with confidence intervals based on 500 bootstrap samples with a bootstrap sample size of 500. A comparison of the normal and non-normal treatment means as well as a comparison of each to the known treatment effect was conducted using an independent t-test at the .05 level of significance.

Results

In this study, the experimental and comparison group regression lines were compared at a centered cutoff point for differences in treatment effect. The cutoff point was set to maximize the magnitude of the “discontinuity” between groups observed at the cutoff point. Schumacker (1992) has pointed out the importance of carefully considering the cutoff score in actual RD designs and discussed methods for locating the most useful cutoff score. Trochim (1984) had earlier suggested situations involving multiple comparison groups in which it might be helpful to use more than one cutoff score. Multiple comparison groups and cut-off scores however were not employed in this study. The comparative results for the normal versus non-normal distributions using simulated data are presented next.

The mean, standard deviation, median, and skewness values for the pre-test scores in the normal and non-normal data distributions are in Table 1. The normal distribution indicates the same mean and median values with skewness close to zero, as expected. The non-normal data indicates a median value that is lower than the mean and a skewness value indicating of a positively skewed distribution. Both distribution types were simulated to have the same pre-test mean. In the RD design, the pre-test scores are used to determine group assignment, i.e., experimental (treatment) versus comparison (non-treatment) groups. If the pre-test mean was equal to or less than 14, a person was assigned to the treatment group, else assigned to the non-treatment group. Results of this group assignment are in Table 2. In the normal distribution, we would expect the same number in each group (50/50); however, some sampling error is present and expected. In a non-normal distribution, we would expect a larger number in the treatment group (60/40) due to the positively skewed distribution (The median value is less than the mean value in a positively skewed distribution). Centering at the cut-off score was accomplished by taking the pretest score minus 14 (XC = X – 14). The expected mean for this cut-off value in both distributions is zero (normal distribution mean = 0.009; non-normal distribution mean = 0.001).

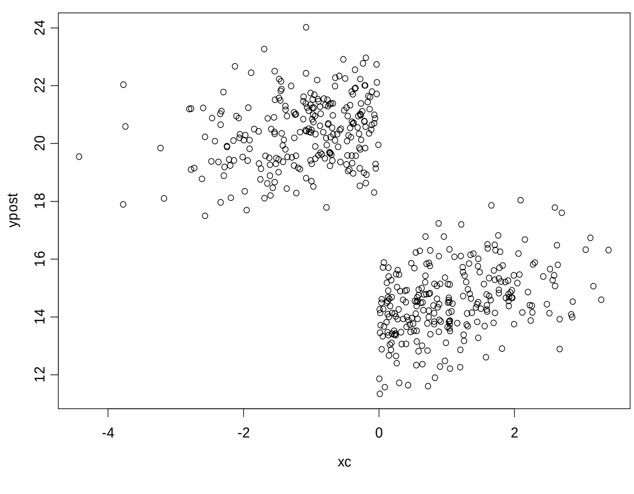


Figure 1. RD experimental and control group data at cut-off for normal distribution

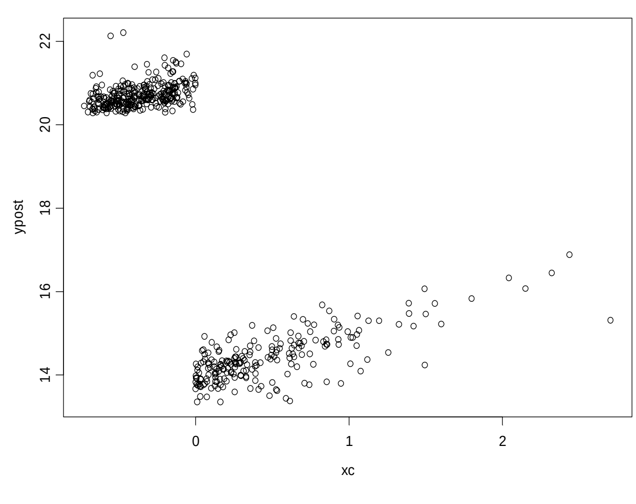


Figure 2. RD experimental and control group data at cut-off for non-normal distribution

Table 1. Descriptive Statistics

Distribution	Mean	SD	Median	Skewness
Normal	14	1.37	14	-.08
Non-Normal	14	0.54	13.86	1.42

Table 2. Group Assignment using Pre-test score

Distribution	Group	*N*	Percent
Normal	Non-treatment (0)	256	0.51
	Treatment (1)	244	0.49
Non-Normal	Non-treatment (0)	211	0.42
	Treatment (1)	289	0.58

Table 3. RD Regression analyses (n=500)

Distribution	Intercept	Slope	Treatment
Normal	14.01	.46	6.89
Non-Normal	14.00	.86	7.02

Table 4. RD Bootstrap Estimates

Distribution	Observed	Bootstrap	Bias	SE
Normal	7.167	7.167	0	.002
Non-Normal	7.294	7.294	0	.001

Table 5. Independent t-test between distribution means and known treatment mean

Distribution	Mean	Difference	*SE*	*t*
Normal	6.89	-.13	.11	-1.18
Non-Normal	7.02			
Normal	6.89	-.11	.17	-.65
Population	7.00			
Non-Normal	7.02	.02	.05	.40
Population	7.00			

* critical $t = 1.96$, $p<.05$

The RD regression analyses for simulated data from both the normal and non-normal distributions are in Table 3. The RD regression equation is:

$$\hat{Y}_{Post} = b_0 + b_1 X_{Pre} + b_2 Z$$

where a post test score (Y) is predicted using pretest score (X) and group assignment (Z), experimental versus comparison. The regression weights refer to b_o = intercept, b_1 = slope, and b_2 = treatment effect (positive value is gain; negative value is loss). An intercept of 14 is expected for both types of distributions, however, the treatment effects (b_2) are expected to differ if skewness affects the RD design analysis. The treatment effects were similar (normal = 6.895; non-normal = 7.02), with any differences due to sampling error.

A non-parametric bootstrap was applied to the 500 simulated data set results for both the normal and non-normal distributions. Results are presented in Table 4. The observed mean treatment effect departed only slightly from the known specified treatment effect of 7.0. The difference in the normal distribution can be attributed to sampling error. The difference in the non-normal distribution can be contributed to sampling error and skewness. Bootstrap for both the normal and non-normal distributions reproduced the observed mean values, thus no difference in the expected outcome, i.e., bias = 0. Consequently, very little standard error was present in the bootstrap estimates. The 5% and 95% confidence intervals for the non-normal distribution results were: (7.292; 7.296) or 7.294 +/- .002. The 5% and 95% confidence intervals for the normal distribution were: (7.163; 7.171) or 7.167 +/- .004. The 5% and 95% confidence intervals around the bootstrap estimate contain 2 standard errors (SE).

An independent *t*-test was used to test whether the observed means were different from the known treatment mean and between themselves. These results are in Table 5. Results indicated that the treatment means were not statistically significantly different nor were each different from the known specified population treatment mean.

Conclusions

Results indicated similar treatment effects whether normal or non-normal distributions were used with a centered cut-off score. Given a know treatment effect of 7.0 with random sampling error, bootstrap estimates were similar to observed estimates from pseudo-populations (bootstrap samples). No bias was reported and the findings indicated stable estimates in the presence of non-normality. The treatment effect estimated in RD using normal and non-normal simulated data distributions indicated that the RD design is not severely impacted by skewed data distributions commonly found in program evaluation. Robust estimation methods and/or data transformations such as probit are most likely sufficient to provide accurate stable estimates of treatment effects when concerned about meeting assumptions in regression analyses.

Regression discontinuity is appropriate for evaluation data and should be used more often in lieu of not meeting assumptions in quasi-experimental designs, especially in analysis of covariance. RD analyses can explore treatment effect differences at different cutoff points, use different pre-test measures than post-test measures, does not require matching of subjects, and can use multiple comparison groups with different cut-off scores. Educational researchers should therefore make increased use of the regression-discontinuity technique for program evaluation.

References

Bottenberg, R.A. and Ward, J.H. (1963). *Applied Multiple Linear Regression*. Lackland Air Force Base, San Antonio, TX: Aerospace Medical Division, No. AD413128.

Cappelleri, J.C. and Trochim, W. (2000). Cutoff Designs. In Chow, Shein-Chung (Ed.) *Encyclopedia of Biopharmaceutical Statistics*, 149-156. Marcel Dekker: New York, NY.

Chambers, J. M., Mallows, C. L. and Stuck, B. W. (1976). A Method for Simulating Stable Random Variables. *Journal of the American Statistical Association*, *71*, 340-344.

Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Lawrence Erlbaum, Mahwah: NJ.

Johnson, N. L. and Kotz, S. (1970). *Continuous Univariate Distributions, vol. 1 and 2*. Houghton-Mifflin, Boston.

McNeil, K. (April, 1984). *Random Thoughts on Why the Regression Discontinuity Design Is Not Widely Used*. Paper presented at the American Educational Research Association Annual Meeting. New Orleans, LA.

Searle, W. *Bootstrap Confidence Intervals*. Internet article located at: http://www.pitt.edu/~wpilib/statfaq/bootfaq.html

Schumacker, Randall E. (April, 1992). *Factors Affecting Regression-Discontinuity*. Paper presented at the American Educational Research Association Annual Meeting. San Francisco, CA.

Schumacker, R.E., Monahan, M.P., & Mount, R.E. (2002). A Comparison of OLS and Robust Regression using S-PLUS. *Multiple Linear Regresssion Viewpoints*, *28*(*2*), 10 - 13.

S-PLUS (2005). *S-PLUS 6 User's Guide for Windows*. Insightful, Inc., Seattle, WA.

Thistlethwaite, D.L. & Cambell, D.T. (1960). Regression-discontinuity analysis: An alternative to the ex-post facto experiment. *Educational Psychology*, *51*(*6*), 309-317.

Trochim, William M. K. (1984). *Research Design for Program Evaluation, the Regression Discontinuity Approach*. Sage Publications: Beverly Hills, CA.

Trochim, William and Davis, Sarita. *Computer Simulations for Research Design*. Internet article located at: http://www.socialresearchmethods.net/simul/simul.htm

Trochim, W. *The Regression-Discontinuity Design*. Internet article located at: http://www.socialresearchmethods.net/kb/quasird.htm

Venables, W.N. & Ripley, B.D. (2001). *Modern Applied Statistics with S-PLUS* (3rd Edition). Springer Verlag: Statistics and Computing Series, New York: NY.

Send correspondence to: Randall E. Schumacker, Ph.D.
University of North Texas
Email: RSchumacker@unt.edu

An Evaluation of a Family Preservation Juvenile Justice Program with a Cox Regression Model

Don D. Allen
Columbia Consulting, Inc.

John W. Fraas
Ashland University

This article illustrates the use of Cox regression to analyze recidivism among felony offending juveniles who were assigned to one of two criminal youth programs. The program that employed an intensive home-based family preservation model was identified as the experimental group. The other program, which did not utilize an intensive home-based family preservation model, was labeled the control group. The study used a quasi-experimental design and Cox regression analysis to compare recidivism outcomes of juvenile offenders treated in the Partners Program with a control group of juvenile offenders who were not given the treatment. The Cox regression analysis revealed that for the juveniles treated in the Partners Program their length of time to recidivism was longer and risk of being re-arrested was lower than for the juveniles in the control group, adjusting for the various covariates.

One alternative to the traditional juvenile justice program is an intensive home-based program strategy utilizing multisystemic treatment. This program had shown significant improvement in reducing recidivism and improving the lives of juveniles who committed serious crime. In a study released in 1992, Henggeler, Melton, and Smith found that a multisystemic treatment strategy (MST) used in South Carolina reduced the rates of criminal activity initially and at the 2.4-year follow-up point (Henggeler, Melton, & Smith, 1992; Henggeler, Melton, Smith, Schoenwald, & Hanley, 1993).

Based on the MST principles and the MST model as published by Henggeler and his associates (Henggeler, 1994), a program labeled The Partners Program was designed. This program, which is a home-based family preservation program, was implemented in January 1995 as a pilot program for the Richland County Juvenile Court in Mansfield, Ohio. The Partners Program was perceived to be successful in reducing juvenile recidivism but had not been evaluated empirically through the use of (a) a quasi-experimental research design and (b) the analysis of the data with a Cox regression model.

The Partners Program

The Partners Program offers the opportunity for juveniles adjudicated delinquent for a felony offense and committed to the Department of Youth Services (DYS) to remain with their families in their homes. The Partners Program provides a community-based intervention program at a cost that is substantially lower than the cost incurred when juveniles are sent to the Department of Youth Services' detention facility (Allen, 1996).

Program Eligibility and Interventions

A juvenile residing with his/her family in Richland County, Ohio having committed a felony and subsequently admitted to or been found guilty of the offense, is adjudicated delinquent. If the offense is great enough to warrant confinement in one of the Department of Juveniles Services' correctional facilities, the Richland County Juvenile Court judge makes a determination in cooperation with the court staff to either send the juvenile to the Department of Youth Services detention facility or offer the juvenile and his/her family the opportunity of going into the Partners Program.

Upon entry into the Partners Program, the juvenile is released from the Richland County juvenile jail to the custody of his/her parent(s) or guardian(s). During the release, the family and delinquent juvenile meet informally with the Partners Program supervisor, the juvenile court director, and the direct service provider that will be personally overseeing treatment and intervention.

Immediately, the direct service provider arranges a meeting with the immediate family and the delinquent juvenile. Rules, expectations, and general structure of the Partners Program are explained. This is individually tailored to the juvenile and family in question. Both the juvenile and the family are involved in the creation of this plan. Direct contacts with the family and the delinquent juvenile are done at the family's home and in the family's neighborhood. Assessment is ongoing, interactive, and designed for continued growth and skill development. The basic principles and tenets of MST as published by Henggeler, Schoenwald, Borduin, Rowland, and Cunningham (1998) are intrinsic to the formation and implementation of the Partners Program.

Prior to the Partners Program, a juvenile committing a serious crime or a sufficient pattern of significant delinquent behavior would be sent to the Department of Youth Service Correctional facility for a specified time determined by the Richland County Juvenile Court Judge. The Partners Program allows juveniles the opportunity to remain in the community in their homes with their families and receive the intensive home-based treatment necessary to improve social and personal skills that may reduce the likelihood of engaging in felony crime.

Once the initial phase of contact, rapport, and basic implementation of the intensive home-based program are established, community, extended family, peer support, and other systems are brought in and a comprehensive list of intervention needs over eight major life areas are evaluated. The areas of life most commonly reviewed are: (a) spiritual, (b) health, (c) family, (d) social, (e) school, (f) employment, (g) financial, (h) hobbies and recreation, and (i) legal (court expectations). The strengths of the family and the delinquent juvenile are established, all the identified problems in the major life areas are recognized, and plans are implemented to remedy these problems. Individuals brought in from the community including extended and immediate family are called upon to oversee progress with the targeted problem areas. The direct service provider then oversees, coordinates, and remains in contact with all the individuals working to improve the situation. This allows the direct service provider an opportunity to work with all the people connected with the delinquent juvenile as a means to further assess and evaluate how best to help. Further, those individuals in the delinquent juvenile’s life that are deleterious to the overall success of the Partners Program and the juvenile can also be addressed.

This process continues to occur over several months depending upon the needs of the delinquent juvenile and his/her family. Once sufficient success has occurred and the staff feels comfortable with the level of skill attained by the delinquent juvenile and his/her family, the juvenile then graduates from the Partners Program. The juvenile is then overseen through the probation department (Aftercare) and tangentially by the Partner’s staff and his/her original direct service provider.

Research Methodology

This study utilized a quasi-experimental program evaluation design with non-random non-equivalent groups (Cook & Campbell, 1979; Royse, 1995). This design utilized two non-equivalent groups that consisted of the juveniles treated in the Partners Program (experimental group) and juveniles treated in the traditional DYS system (control group).

Rossi and Freeman (1993) address some of the challenges of an impact assessment/program evaluation study. One challenge is in assessing gross outcomes versus a net outcome. Gross outcomes, which encompass net outcomes, consist of all changes observed as an outcome measure. With regard to Partners Program evaluation, the gross outcome was whether or not the intervention produced a reduction in felony recidivism between the control and experimental groups. Net outcomes were "those results that can be reasonably attributed to the intervention free and clear of the effects of any other causes that may be at work" (Rossi & Freeman, 1993, p. 221). When reviewing the results of this study, one should keep in mind the difficulty encountered in separating the net effect from the gross effect when a quasi-experimental program evaluation design with non-random non-equivalent control groups is used.

Sample Selection

The experimental and control groups consisted of 130 juveniles who were adjudicated delinquent for a first to fourth degree felony. The control group consisted of youth entering the study from January of 1993 through December of 1994. The experimental group consisted of youth entering the study from May 31, 1996 though June 30, 1998. Ohio Revised Code (ORC) numbers were used to ensure consistency of felony degree between the offenses of juveniles in the experimental and control groups. As stated earlier, the juvenile court judge would have referred juveniles from the control group for the Partners Program had it been available at the time the control group juveniles were adjudicated. Juveniles, who were evaluated to be too dangerous for the Partners Program, would have been screened out at this phase by the juvenile court judge and not considered as part of the control group population.

The control group consisted of juveniles who were adjudicated delinquent for a first to fourth degree felony. They were subsequently committed by the Richland County Juvenile Court to the State of Ohio's Department of Youth Services (DYS) correctional facility. To determine what facility best suited the delinquent juveniles, they began by serving a 30-day evaluation period at the Circleville detention site. They were then sent to a detention facility that was best suited to their needs. Upon serving their sentence at the recommended Department of Youth Services' corrections facility, they were released back to the custody of the Richland County Juvenile Court and subsequently returned to their family. At this time, they were entered into the study. The juveniles from Richland County, Ohio who participated in the traditional juvenile court program from January 1, 1993, through December 31, 1994 comprised the control group, which consisted of 45 felony offending juveniles. One youth under age 13 did go to (DYS) detention.

In the case of the juveniles in the experimental group, the director of the juvenile court identified a juvenile that ordinarily would be sent to the DYS. A referral was made to the Partners Program staff. The supervisor of the Partners Program examined the fit between each juvenile's identified problems and the resources within the court, the family, and the community that would enable the Partners Program to intervene successfully. If the supervisor believed there were sufficient resources and all parties agreed to the terms and rules of the Partners Program (Partners), the juvenile was accepted into the program. The director conferred with the judge, the judge ordered the referral, and a Partners Program staff member was then assigned to the case. There has never been a juvenile or family accepted into Partners who refused to join the program. Any youth over the age of 13 accepted to the program and included in this study would have been incarcerated at the DYS. The five youth in the sample under age 13, would have received county based probation and/or detention in the county juvenile facility. Juveniles in the experimental group were eligible for entry into the study at the point they were released to the Partners Program. All juveniles who participated in the Partners Program of Richland County, Ohio, from May 31, 1996, through July 31, 1998, comprised an experimental group of 85 juveniles who committed felony offenses.

Three factors added to the strength of the selection process and this program evaluation research. First, the four delinquency professionals who conducted subject selection were at the Richland County Juvenile Court for the entire time of the study. Second, Judge Ronald Spon, who was not involved in subject selection but consistently, presided over the Richland County Juvenile Court during the entire length of the study, remains incumbent. His involvement, noted later, added more consistency to the adjudication and incarceration process. Third, in prior research on home-based family preservation programs not all of the juveniles included were actually removed from the home. All juveniles in this study regardless of whether they were in the control and experimental groups over age 13 would have been removed from their homes through their incarceration periods.

Cox Regression and the Dependent Variable

Since the number of days until recidivism occurred for each juvenile was recorded for the data set used in this study, a Cox regression model was used to evaluate the differences between the recidivism rates of the experimental and control groups. An analysis of a Cox regression model is a form of survival analysis that allows the researcher to use various factors to model the length of time it has taken for an event to occur (e.g., re-arrest) even when some of the participants have not experienced the event (i.e., the censored cases).

Cox regression model rather than logistic regression model or a multiple linear regression model was the analytical approach used for two reasons. First, valuable information, specifically the length of time until a felony re-arrest, could be utilized. Such information would not be used if logistic regression had been used to analyze a dichotomized variable in which a juvenile would be simply be classified as being re-arrested or not re-arrested. Second, all cases are selected, not just the ones who where re-arrested (i.e., non-censored cases). It would be possible for a researcher to use only non-censored cases or assign the censored cases the maximum time observed in the study. Such data could be analyzed with a multiple linear regression model with the dependent variable consisting of time to the event. If only the censored cases were analyzed or the censored cases were assigned the maximum time observed, however, the true survival period would be underestimated (Adams, 1996).

To understand what serves as the dependent variable in a Cox regression model, three concepts need to be understood: (a) survival probability, (b) survivor function, and (c) hazard rate. The survival probability is the probability that a juvenile will not be re-arrested until a given point in time. The survivor function depicts the relationship between estimated survival probabilities over time. When graphed, the survival function for this study shows the proportion of juveniles not re-arrested by a specified point in time. According to Blossfeld and Mayer (1989) the hazard rate is the instantaneous rate of change in the survivor function. The hazard rate for this study indicates the instantaneous rate at which juveniles are re-arrested.

In a Cox regression model the dependent variable is the hazard rate. In order to allow the SPSS® computer software to generate the hazard function, the length of time between the day the participants were released from prison (for members of the control group) or the day the participants began the Partners Program (for members of the experimental group) and the day they committed another felony must be entered as a variable in the data set. It is important to note that the entry point for each control group participant was the day the participant was released from DYS detention, while the entry point for each experimental participant was day the participant entered the Partners Program. Participants who were not re-arrested for felony within 850 days were assigned a value of 850.

The number of censored cases was 73, which was 56% of the 130 juveniles. A total of 17 of the 45 juveniles (37.8%) in the control were not re-arrested (censored cases); while 56 of the 85 juveniles (65.9%) in the experimental groups were not re-arrested (censored cases). The median number of days until re-arrested for the 28 non-censored cases in the control group was 241, while the number of days for the 29 non-censored cases in the experimental group was 400.

Cox Regression and the Independent Variable

Based on empirical findings of other researchers (Henggeler, Schoenwald, Borduin, Rowland, & Cunningham, 1998; Allen, 2004), seven independent variables were used as covariates in the Cox regression model. The seven covariates and the group membership variable were defined as follows:

1. Ages of the participants at the time of commitment (X_1) -- Age was defined as an interval measure for any youth entering the study before the age of 18. The zero age point is the point of study entry. Age was portioned into years and months. The year is shown as 13, 14, 15, etc. The months were added to the year as a decimal divisible by 12. So, the data for a person who is 15 years and six months old is quantified as 15.5 (i.e., 15 and 6/12).
2. Ages of the participants at the time they entered the study (X_2) -- The data for this variable were recorded as described in the previous variable.
3. Gender (X_3) -- Gender was defined as a discrete/binomial measure dummy coded as one for male and zero for female.
4. Race (X_4) -- Race was defined and coded as a discrete/binomial measure with one for white and zero for black, with white as stipulated in the guidelines established by the Department of Youth Services in the State of Ohio.
5. Frequency of prior probation and/or misdemeanors (X_5) -- Probation/misdemeanor violations were defined under the Ohio Revised Code and monitored by the Richland County, Ohio court record. Misdemeanor and probation violations were quantified as the frequency of occurrence.
6. Frequency of prior felony convictions (X_6) -- Prior felony violations were defined under the Ohio Revised Code and monitored by the Richland County, Ohio court record. Prior felonies were quantified as the frequency of felonies.
7. Loss due to parental inaccessibility at or prior to the time the juveniles entered the study (X_7) -- Loss due to a parental inaccessibility (the parent is for all intents and purposes unavailable or inaccessible to the child, that is, in prison, lives out of the state, etc.), was coded as dummy coded, one for presence of the event, zero for absence of the event.
8. Group membership (X_8) -- This variable identified whether a juvenile was exposed to the Partner Program (experimental group) or not exposed to the Partner Program (control group). The control and experimental groups were assigned values of zero and one, respectively.

The mean and standard deviation values for seven of the independent variables are listed in Table 1 (X_8, the group variable was excluded). The differences between the mean values of the control and experimental groups were not statistically significant for the following four variables: (a) the ages of the participants at the time of commitment, X_1; (b) gender, X_3; (c) frequency of prior probation and/or misdemeanors, X_5; and (d) frequency of prior felony convictions, X_6.

Table 1. *Descriptive Statistics for the Independent Variables*

Variables	Control Group Mean[a]	SD	Experimental Group Mean	SD	Total Mean	SD
X_1 (Commitment Age)	16.10	1.03	15.73	1.36	15.85	1.26
X_2 (Age at Entry)[b]	16.72	1.05	15.92	1.34	16.20	1.30
X_3 (Gender)	0.91	0.29	0.87	0.34	0.88	0.32
X_4 (Race)[b]	0.49	0.51	0.67	0.47	0.61	0.49
X_5 (Prior Prob./Misd.)	6.07	4.85	4.94	5.05	5.33	4.99
X_6 (Prior Felony Con.)	2.22	1.48	1.84	1.31	1.97	1.38
X_7 (Loss of Access)[b]	0.42	0.50	0.68	0.47	0.59	0.49

[a] The means for the Gender, Race, and Loss of Access variables are the proportions of juveniles in who were male, white, and experienced loss due to parental inaccessibility, respectively.
[b] Differences between the means of the control and experimental groups were significant at the .05 level.

Statistically significant differences existed between the control and experimental groups with respect to three of the independent variables. With respect to the mean age of the participants at the time they entered the study, which was variable X_2, the mean was higher for the control group ($\bar{X}_C$=16.73) and the experimental group ($\bar{X}_E$=15.92). An analysis of variable X_4, which indicated whether each juvenile was white or non-white, revealed that the proportion of white juveniles in the control group (.49) was less than the proportion in the experimental group (.67). And the analysis of variable X_7, which noted whether a juvenile had experienced parental loss due to parental inaccessibility, indicated that the proportion in the control group (.42) was lower than the proportion in the experimental group (.68). Additional group comparison data can be found in Allen (2004).

In Cox regression analysis independent variables are identified as either time-invariant variables or time-varying covariates. As the names imply, time-invariant variables are independent variables that do not change over time, and time-varying variables do change over time. As noted by Adams (1996):

> Some time-varying covariates such as address and income may change relatively quickly; others, such as the level of education may change more slowly. This distinction is important because some time-varying covariates such as age and education may be treated as time-invariant covariates for practical purposes. (p. 274).

It should be noted that the independent variables used in this study were treated as time-invariant variables for the Cox regression analysis.

Cox Regression and the Proportional Hazards Assumption

As noted by Cox (1972), a Cox regression model is a semiparametric regression model. The model is based on the assumption that the groups defined by the covariates have the same underlying hazard function. Adams (1996) noted that:

> The various parameters for each [covariate] group act to shift the hazard function up or down. Because the Cox regression model assumes that the hazard functions are proportional to one another, it is necessary to check the covariates to determine whether they meet this proportional hazard (PH) assumption. (p. 274)

One method used to assess whether PH assumption is met is to plot the log of the negative log transformation of the survival function. If the PH assumption is met, the curves for the covariate groups of interest should not differ from proportionality in a substantial way.

To check the PH assumption we plotted the log of the negative log transformation of the survival function with each variable to be checked acting as the stratification variable. In this procedure we created categorical variables for the continuous variables. The various plots appeared to be proportional. Thus, we assumed the PH assumption was met.

Cox Regression Analysis Results

The results of the Cox regression analysis are listed in Figure 1 and Table 1. The survival function estimates are depicted in Figure 1 for each group (i.e., the experimental group and the control group). Figure 1 displays the estimated differences between the experimental group and the control group, holding the other variables constant.

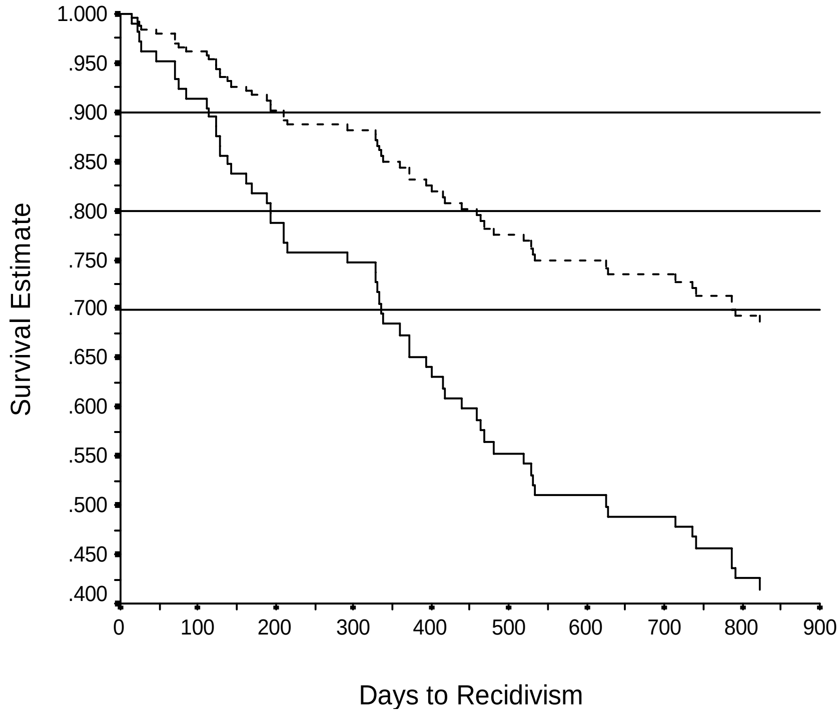


Figure 1. Survival Functions for the Experimental and Control Groups.
[a] The dashed line represents the survival function of the experimental group, while the solid line represents the survival function of the control group.
[b] Reference lines are placed at the .90, .80, and .70 levels.

Figure 1 also contains three horizontal reference lines to indicate the 90%, 80% and 70% survival points. A review of these survival points revealed the following:

1. The percent of juveniles in the experimental group who had not been re-arrested decreased to the 90% level at approximately day 200. For the control group, this 90% level had been reached in approximately 110 days.
2. The percent of juveniles in the experimental group who had not been re-arrested decreased to the 80% level at approximately day 450. For the control group, this 80% level had been reached in approximately 195 days.
3. The percent of juveniles in the experimental group who had not been re-arrested decreased to the 70% level at approximately day 790. For the control group, this 70% level had been reached in approximately 330 days.

Thus, as depicted in the two survival curve estimates, the members of the experimental group (juveniles in the Partners Program) reflected prolonged time to re-arrest when compared with members of the control group (juveniles in the Partners Program).

Table 2 contains further results of the Cox regression analysis. The coefficient for the variable representing the group membership (X_8), which was -0.843, was statistically significant at the established alpha level of .05 (Wald statistic = 6.72, p = .01). This value of -0.843 indicates that being a member of the experimental group reduces the log of the hazard (the hazard of committing another crime) by 0.843, controlling for the other variables in the model. This value can be better understood by interpreting its antilog value, that is, $\exp(\beta_i)$. The antilog of the coefficient for the group membership variable was 0.43. This value, which is referred to as a risk ratio or effect, indicates the risk of experimental group members being re-arrested was 43% of the risk of control group members, holding constant the other variables.

Table 2. *Cox Regression Analysis Results* [a]

Variables	Coefficients	SE	Wald	*p-value*
X_1 (Commitment Age)	0.087	0.283	0.09	0.76
X_2 (Age at Entry)	-0.267	0.288	0.86	0.35
X_3 (Gender)	1.209	0.731	2.73	0.10
X_4 (Race)	-0.522	0.294	3.15	0.08
X_5 (Prior Prob./Misd.)	0.062	0.029	4.49	0.03
X_6 (Prior Felony Con.	0.136	0.089	2.35	0.13
X_7 (Loss of Access)	-0.348	0.283	1.50	0.22
X_8 (Groups)	-0.843	0.325	6.72	0.01

[a] The χ^2 value for the change in the -2 times the log likelihood value when the eight independent variables were added to the analysis was 35.00 ($p < .001$).

It is of interest to note that one other predicted variable, which was a behavioral variable, was statistically significant. The frequency of prior probation and/or misdemeanors (X_5) had a coefficient value of 0.062 (Wald statistic = 4.49, $p = .03$). The antilog of this coefficient 1.064 indicates that an increase of one prior conviction increases the log of the hazard (the hazard of committing another crime) by 6.4%, holding constant the other variables.

Summary and Implications

The analysis of the Cox regression model indicated that compared to the juveniles receiving traditional DYS intervention (control group), the juveniles in the Partners Program (experimental group): (a) reflected prolonged time to re-arrest and (b) lower risks of being re-arrested. Of the other predictor variables entered into the Cox regression model, only prior misdemeanors/probation violations was significant. Thus behavioral and not temporal measures were related to juvenile delinquency.

This study has practical implications for juvenile court administrators who are interested in reducing felony re-arrest rates or substantially prolonging the days until a youth does get re-arrested. The intervention and supervision strategies utilized in the Partners Program appear to create a greater involvement in the lives of the delinquent juveniles and their families. Although such interventions might cause a higher incidence of misdemeanor and probation violation occurrence (see Allen, 2004), it appears to improve the life skills of the youth and those surrounding them, resulting in a reduction in felony re-arrest and subsequent removal from the community. While remaining in the community, these juveniles have the opportunity to learn and grow from more suitable role models (including Partners Program direct service providers) than if in (DYS) detention.

Prior study research shows that each probation or misdemeanor offense increases the likelihood of a felony occurrence by greater than six percent. As noted here, prior misdemeanors/probation violations were the only significant predictor of felony recidivism. Based on all these findings, juvenile court administrators may want to pay greater attention early on to those youths who are repeat misdemeanor and probation violation offenders. Offering greater structure and supervision modeled after the Partners Program intervention strategy may reduce the occurrence of felony offenses in the future.

It should be noted that the home-based family preservation model, generally, has been criticized because of study design. Detractors argue children included in the study may or may not have been placed out of the home thereby creating a sample population that was not truly at risk of placement. With the exception of the six children under age 13 referenced earlier, this critique does not apply to this research as all the youth included in the Partners Program study experimental group would have been incarcerated as those in the control group were, thereby adding important information to the research literature. This research supports prior research on the efficacy of family preservation strategies.

It is important to keep in mind that this study used a non-randomized quasi-experimental design which prohibits one from assuming causation making it difficult to generalize to other populations. The above findings suggest the Partners Program family preservation model should place even more emphasis on intervention with delinquent juveniles at the earliest sign of frequent misdemeanor/probation violation occurrence, thus increasing the likelihood of successful intervention and decreasing the likelihood of delinquent behavior in the future. Further, because the Richland County, Ohio's Partner Program model successfully allows youth to avoid incarceration into DYS detention, allowing them to remain in the care of their families and community, the family preservation programs modeled after the Richland County, Ohio Partners Program warrant further study.

Although the exact cost savings of the Partners Program intervention is beyond the scope of this study, it is still important to note the economic ramifications of such findings. Since the Partners Program is community based and able to operate at a lower cost than the Department of Youth Services detention facilities, (Allen, 1996), the program is able to save Ohio tax dollars. When a juvenile is able to enter the adult community with skills that allow that juvenile to avoid a criminal career, cost savings is substantial (Snyder & Sickmund, 1999). Cost analysis of the Richland County Juvenile Court's Partners Program merits further study.

References

Adams, G. J. (1996). Using a Cox regression model to examine voluntary teacher turnover. *Journal of Experimental Education*, *64*(2), 267-285.

Allen, D. (1996, June). *The Partners Program*. Paper presented at the meeting of the Department of Youth Service's Reclaim Ohio Conference, Columbus, OH.

Allen, D. (2004). *The Juvenile Court of Richland County, Ohio's Partner Program: A Long Term Program Evaluation Study*. Dissertation Abstracts International 65(07). (UMI No. 3138326)

Blossfeld, H. & Mayer, K. (1989). *Event history analysis: Statistical theory and application in the social sciences.* Hillsdale, NJ: Erlbaum.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Cook, T., & Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. New York: Rand McNally College Publishing Company.

Henggeler, S. W., et al. (1994). *Treatment manual for family preservation using multisystemic therapy*. Charleston, South Carolina: South Carolina Health and Human Services Finance Commission.

Henggeler, S., Schoenwald, S., Borduin, C., Rowland, M., Cunningham, P. (1998). *Multisystemic treatment of antisocial behavior in children and adolescents*. New York: Guilford Press.

Henggeler, S., Melton, G., Smith, L., Schoenwald, S. & Hanley, J. (1993). Family preservation using multisystemic treatment: Long term follow-up to a clinical trial with serious juvenile offenders. *Journal of Child and Family Studies*, *2*, 283-293.

Henggeler, S., Melton, G., & Smith, L. (1992). Family preservation using multisystemic therapy: An effective alternative to incarcerating serious juvenile offenders. *Journal of Consulting and Clinical Psychology*, *60*, 953-961.

Rossi, P., H., & Freeman, H., E., (1993). *Evaluation: A systemic approach* (5th ed.). London: Sage.

Royse, D. (1995). *Research methods in social work* (2nd ed.). Chicago: Nelson-Hall Inc.

Sampson, R., and Laub, J., (1994). Urban poverty and the family context of delinquency: A new look at structure and process in a classic study, *Child Development*, *65*, 523-540.

Snyder, H. & Sickmund, M. (1999). *Juvenile offenders and victims:1999 national report* (NCJ178257). Rockville, Maryland: Office of Juvenile Justice and Delinquency Prevention.

Sondheimer, D., Schoenwald, S., & Rowland, M. (1994). Alternatives to the hospitalization of youth with a serious emotional disturbance. *Journal of Clinical Child Psychology*, *23* (suppl.), 7-12.

Send correspondence to: John W. Fraas, Ph.D.
Ashland University
Email: jfraas@ashland.edu

Youth Risk Behavior Survey: A Breakdown of Adolescent Risk Behaviors

Shannon Williams
University of Northern Colorado

Barrett Dunn
Boulder County Health Department

Daniel J. Mundfrom
University of Northern Colorado

Cindy Kronauge
University of Northern Colorado

The purpose of this study was to examine the relationship between students with high risk behavior activity and the behaviors in which they participate. The Youth Risk Behavior Survey (YRBS) was developed to monitor significant adolescent risk behaviors. The YRBS was administered to a sample of 9th through 12th grade students in two school districts in northern Colorado. A logistic regression model was used to compare students with high risk behavior to their peers with low risk behavior activity. The aim of the study was to assist in the identification of significant risk behaviors that may expand the corresponding knowledge base by which new community and school-based programs could be developed to address these risk behaviors.

Adolescence is a very impressionable time. Beliefs and values are formed, parental influence diminishes, and peer pressure increases (Gunbaum, Basen-Engquist, 1993). As a result of this suggestible time, many health related behaviors develop. These behaviors include those that affect health during adolescence and later in life. In the United States, homicide, suicide, motor vehicle crashes, and other unintentional injuries account for almost three-quarters of all deaths among youth 10-24 years of age (Grunbaum, Kann, Kinchen, Ross, Gowda, Collins, & Kolbe, 2000). For adults greater than 24 years, nearly two-thirds of deaths result from cardiovascular disease and cancer often from causes initiated during adolescence (Grunbaum et al., 2000). Adolescents may or may not be aware of the repercussions of their behaviors. Jessor (1998) describes risk behaviors as "risk factors for personally or socially or developmentally undesirable outcomes." Subsequently, it is important to detect, monitor, and ultimately, intervene in these behaviors.

In response to this need, in 1988, the Centers for Disease Control and Prevention developed the Youth Risk Behavior Surveillance System (YRBSS). The YRBSS was developed with three main goals: 1) to focus on specific behaviors among youth that cause important health problems; 2) determine whether those behaviors increase, decrease, or remain the same over time; and 3) provide comparable data among national, state, and local samples of youth (Brener, Collins, Kann, Warren, & Williams, 1995). The YRBSS identifies and measures behavior and activity in the following categories: (1) tobacco use; (2) alcohol and other drug use; (3) unhealthy dietary behaviors; (4) physical activity; (5) sexual behaviors that contribute to unintended pregnancy and sexually transmitted diseases; and (6) behaviors that contribute to unintentional and intentional injuries (Brener et al., 1995). By examining these categories, the YRBSS is able to determine the prevalence of health risk behaviors, compare this data on the national, state, and local levels, and assess whether risk behaviors increase or decrease over time (CDC, 2003). Every two years, the YRBSS conducts the Youth Risk Behavior Survey (YRBS). By utilizing the results of this survey, dominant risk behaviors are detected and school-based educational and prevention programs are developed.

Past analyses of the YRBS have focused on entire group analyses along with the comparison of subgroups within the student population. Analyses comparing male and female students have shown males and females differ significantly on risk behaviors (Grunbaum et al. 2000). Comparisons have also been made between racial groups, for example, Grunbaum et al. (2000) reported differences in harassment, drug use, and suicide, among other behaviors that have been explored across White, Black, and Hispanic subgroups. Also, Garofalo, Wolf, Kessel, Palfrey, & DuRant (1998) found associations between sexual orientation and risk behaviors. Specifically, they found gay, lesbian, and bisexual youth were more likely to be threatened, have suicidal ideation and attempts, and be engaged in substance use than were students with other sexual orientations.

The present study was designed to compare students with low risk behavior to their peers with high risk behavior. More specifically, differences in prevalence and likelihood of variables associated with gender, race, sexuality, alcohol and drug use, sexual activity, weight loss, fighting, and suicide were examined.

Method

The data used in this study were collected by a county health department in the western United States in collaboration with two local school districts. A two-stage cluster sample design was utilized. From within the school districts, alternative and private schools were excluded from the sample. All other schools were selected and within those schools, classrooms were randomly chosen to participate. Every student enrolled in the classes selected was eligible for inclusion in the study. An informational packet regarding the survey with an option to exclude their student was sent to the parents of all students selected to participate. Students were granted anonymity and voluntary participation.

Upon completion, the health department received 1,957 completed surveys. To reduce sample selection bias, a weighting factor was applied. The weighting factor caused the distribution of the sample to match the distribution of males and females by school grade. This allowed inferences to be generalizable to all 9-12th grade students at participating schools.

All analyses were conducted using SUDAAN (Research Triangle Institute, Research Triangle Park, NC). As a result of using a clustered sampling method, there is a lack of independence in the error terms when aggregate-level data are used. SUDAAN accounts for the clustered sample design of the YRBS and the lack of independence in the error terms (LeClere, Soobader, 2000). SUDAAN produces robust variance estimates which account for the intra-cluster correlation, weighting, and without-replacement sampling used in the YRBS.

Variables for Study

A list of health risk behaviors common to adolescents was compiled. These were: (1) alcohol consumption and driving; (2) carrying a weapon; (3) physical fighting; (4) attempted suicide; (5) current tobacco use; (6) binge drinking; (7) current marijuana use; (8) current cocaine use; (9) current inhalant use; (10) sexual activity; (11) failure to use a condom; (12) alcohol and drug use at last intercourse; and (13) using an unhealthy weight loss method. Each risk behavior was weighted equally and combined into a score. This score consisted of a count of affirmative responses to questions about these risk behaviors. Individual scores ranged from 0 (no risk behaviors present) to 13 (all thirteen risk behaviors reported). Based on the number of risk behaviors exhibited, each student was classified as having a high level of risk behavior (4 or more risk behaviors) or a low level of risk behavior (fewer than 4 risk behaviors).

A major purpose of this study was to identify significant risk behaviors and to expand the knowledge base by which new community and school-based programs could be developed. To facilitate this, the thirteen risk behaviors were grouped according to content and consolidated into a new variable or content area. Merging similar behaviors will allow school districts and communities to focus on a topic area rather than a specific action. The thirteen risk behaviors were consolidated as follows. Using an unhealthy weight loss method consisted of using at least one of the following methods: fasting, unprescribed diet pills, or a laxative or vomiting. Binge drinking and driving or riding with someone while under the influence of alcohol were combined into a measure of alcohol behavior. Weapon carrying and fighting were combined together and finally, marijuana, inhalant, and cocaine use were combined into one common drug use category. All other risk behaviors were unchanged. The resulting content areas included in the study were: (1) use of an unhealthy weight loss method; (2) alcohol related behaviors; (3) behaviors involving fighting and weapons (4) current drug use; (5) current tobacco use; (6) attempted suicide; and (7) current sexual activity.

A cross-tabs analysis was used to identify the prevalence of each of the seven risk behaviors identified. A logistic regression was then used to examine the seven major content areas and frequently studied demographics including sexual identity, gender, and race. This procedure helped to determine which behaviors are more common among high-risk behavior students. Odds ratios (OR) and associated 95% confidence limits (95% CI) are presented for each content area and demographic.

Results

1,957 completed surveys were returned. Table 1 provides a description of the study population according to the seven areas of study and two demographics, sexuality and race. Gender was not included because the logistic regression model found no significant differences between males and females. All other variables were significant and remained in the model. For each of the seven areas of interest, the proportion of students in each group who participate in each risk is higher for the group classified as high-risk. Not all students responded to each question. As a result, each sample size may not be equivalent.

For the low-risk group, only two risk behaviors were prevalent: alcohol related behaviors and behaviors involving fighting and weapons. For the low-risk group 23.4% of students participated in alcohol related behavior and 18.9% of students carried a weapon or have been in a physical fight. These numbers are compared to 91.5% and 64.9%, respectively for the high-risk group of students. For the low-risk behavior group, all other risk behaviors were relatively low.

A logistic regression model yielded large odds ratios. These odds ratios are shown in Table 1. The largest odds ratios are for students who exhibit alcohol related behavior and those who are sexually active. These students are 292.89 and 232.58 times more likely to be a part of the high-risk behavior group. Not quite as large, but of definite importance are students who fight or carry a weapon. Those students are 139.76 times more likely to be in the high-risk group. As demonstrated by all the large odds ratios, students who exhibit any of the behaviors in the seven risk areas are much more likely to be classified into the high-risk group.

Table 1. Distribution of risk behaviors for high and low levels of risk behavior.

Risk Behavior	High-level of risk behavior		Low-level of risk behavior			
	N	%	*N*	%	OR	(95% CI)
Unhealthy weight loss method						
Yes	147	28.3	100	7.4	98.53	(48.37, 200.73)
No	370	71.7	1298	92.6	1.00	
Alcohol related behaviors						
Yes	473	91.5	333	23.8	292.89	(150.22, 571.04)
No	39	8.5	1058	76.2	1.00	
Fighting and weapons						
Yes	327	64.9	269	18.9	139.76	(80.88, 241.53)
No	188	35.1	1121	81.1	1.00	
Current drug use						
Yes	375	72.8	134	9.1	44.16	(27.13, 71.89)
No	141	27.2	1270	90.1	1.00	
Current tobacco use						
Yes	422	81.3	180	13.0	43.61	(26.16, 72.70)
No	99	18.7	1219	87.0	1.00	
Attempted Suicide						
Yes	173	36	126	8.9	32.99	(18.59, 58.55)
No	343	64	1267	91.1	1.00	
Currently sexually active						
Yes	347	68.4	157	10.8	232.58	(135.01, 400.64)
No	160	31.6	1244	89.2	1.00	
Sexuality						
Not Heterosexual	59	12.5	45	3.4	6.71	(2.53, 17.78)
Heterosexual	446	87.5	1338	96.6	1.00	
Race						
White	367	69.5	1094	77.1	3.76	(1.86, 7.60)
Hispanic	106	21.7	192	14.1	3.99	(1.75, 9.13)
Asian	17	3.1	40	3.4	2.39	*(0.64, 9.00)
Other	28	5.7	75	5.4	1.00	

*not statistically significant at $\alpha = .05$

Conclusions

There were three major findings from this study. First, the majority of students exhibit a low level of risk behavior. Of the students who responded, 74.8% or 1410 reported fewer than four risk behaviors and 25.2% or 521 reported four or more risk behaviors. This shows that although the risk behaviors used in the study are important, the majority of students do not participate in most of them. Thus, education promoting positive and safe decision making from parents, teachers, the school district, and the community should be maintained and promoted.

Second, for those students who do exhibit a high level of risk behavior, risk behaviors were consistent across each of the seven areas of interest. High-risk behavior students consistently acted out in their use of unhealthy weight loss methods, alcohol, fighting and weapon carrying, drugs and tobacco, involvement in sexual activity, and attempted suicide. Likelihood ratios showed that sexual activity, alcohol use, and fighting and weapon carrying are of highest concern. The majority of students with a high level of risk behavior exhibit these behaviors. Although students who participate in any of the seven risk behavior areas are more likely to be classified in the high level risk behavior group, these three areas, sexual activity, alcohol related behaviors, and fighting and weapon carrying, are much more prominent. As a result, programming, although touching on each of the seven risk behavior areas, should focus on these three particular areas. These three areas are the most abused behaviors and should be emphasized.

Third, in contrast to prior studies, this study found no significant difference between males and females across level of risk behavior activity. Both groups were equally likely to exhibit a high level of risk behavior. However, as in past studies, this study did find differences amongst sexuality and race. Students who identify themselves as something other than heterosexual were significantly more likely to be in the high risk behavior group. Also, there is a difference in the likelihood of being in the high risk group across Whites and Hispanics. There was no significance difference among Asians and all others.

The results of this study have implications for the content of prevention and education programs. Such programs need to address multiple-risk behaviors among all adolescents. Adolescents that exhibit risk behaviors do not do so in an exclusive manner. Most students classified into the high-risk group exhibited multiple risk behaviors. Thus, education and prevention needs to incorporate a multiple-risk behavior approach. Programs need to focus on why some risk behaviors may inherently lead to others. Students exhibiting a high level of risk behavior, although much more likely to carry a weapon or fight, use alcohol, and be sexually active, were much more likely to exhibit behaviors present in the other risk behavior categories as well. This high occurrence of all risk behaviors shows the need for prevention and education programs to incorporate a multiple-risk behaviors philosophy.

The results of this study emphasize the need for young adult health risk behaviors to be monitored. These behaviors need to be monitored for prevalence, comparison, and trend. With this knowledge, new programs designed to intervene, educate, and deter risk behaviors can be developed that utilize school, parent, and community resources. In conjunction, new methods of detection need to be found that do not focus solely on easily observed behaviors. By monitoring changes over time, school districts can evaluate the progress and effectiveness of these new programs, which are ultimately developed to reduce unhealthy risk behaviors. Monitoring behaviors over time will enable effective programs to flourish, less effective programs to be modified and some programs to even be eliminated. Ultimately, monitoring these programs and their impact on student behavior will result in effective school programming which will involve not only students, but teachers, parents, and other community members and resources.

References

Brener, N.D., Collins, J.L., Kann, L., Warren, C.W., & Williams, B.I. (1995). Reliability of the Youth Risk Behavior Survey Questionnaire. *American Journal of Epidemiology*, *141*, 575-580.

Center for Disease Control and Prevention (2003). 2003 *National School-Based Youth Risk Behavior Survey* Data Documentation Manual. Available online at http://www.cdc.gov/nccdphp/dash/yrbs/data/index.htm

Garofalo, R., Wolf, RC., Kessel, S., Palfrey, SJ., & DuRant, RH. (1998). The association between health risk behaviors and sexual orientation among a school-based sample of adolescents. *Pediatrics*, *101*, 895-902.

Grunbaum, J., & Basen-Engquist, K. (1993). Comparison of health risk behaviors between students in a regular high school and students in an alternative high school. *Journal of School Health*, *63*, 421-426.

Grunbaum, J., Kann, L., Kinchen, S.A, Ross, J.G, Gowda, V.R, Collins, J.L., & Kolbe, L.J. (2000). Youth Risk Behavior Surveillance National Alternative High School Youth Risk Behavior Survey, United States, 1998. *Journal of School Health*, *70*, 5.

Jessor, Richard. (Ed.). (1998). *New perspectives on adolescence risk behavior*. New York: Cambridge Press.

LeClere, F. B., Soobader, M. (2000). The effects of income inequality on the health of selected US demographic groups. *American Journal of Public Health*, *90*, 1892-1897.

Send correspondence to: Daniel J. Mundfrom
University of Northern Colorado
Email: daniel.mundfrom@unco.edu

Multiple Linear Regression Viewpoints
Information for Contributors

Multiple Linear Regression Viewpoints (*MLRV*) is a publication sponsored by the American Educational Research Association's Special Interest Group on Multiple Linear Regression: The General Linear Model (MLR/GLM SIG). It is published twice a year to facilitate communication among professionals who focus their research on the theory, application, or teaching of multiple linear regression models and/or the general linear model. Manuscripts submitted to *MLRV* should conform to the language, style, and format of the Publication Manual of the American Psychological Association (5^{th} ed., 2001). Three copies (two blind) of a doubled spaced manuscript (including equations, footnotes, quotes, and references) of approximately 25 pages in length, a 100 word abstract, and an IBM formatted diskette with the manuscript formatted in WordPerfect or Word should be submitted to one of the editors listed below.

Mathematical and Greek symbols should be clear and concise. All figures and diagrams must be photocopy-ready for publication. Manuscripts will be anonymously peer reviewed by two editorial board members. Author identifying information should appear on the title page of only one submitted manuscript. The review process will take approximately 3 months. A letter acknowledging receipt of the manuscript will be sent to the first author, and upon review, a letter indicating the peer review decision will be sent to the first author. Potential authors are encouraged to contact the editors to discuss ideas for contributions or determine if their manuscript is suitable for publication in *MLRV*.

EDITOR INFORMATION

Randall E. Schumacker, Editor *MLRV*
Educational Research, Chair
P.O. Box 311335
Mathews Hall 304
University of North Texas
Denton, Texas 76203-1335
(940) 565-3962 Office
(940) 565-2185 Fax
rschumacker@unt.edu

T. Mark Beasley, Associate Editor
Department of Biostatistics
School of Public Health
309E Ryals Public Health Bldg.
University of Alabama, Birmingham
Birmingham, AL 35294
(205) 975-4957 (voice)
(205) 975-2540 (fax)
mbeasley@uab.edu

ORDER INFORMATION

Jeffrey B. Hecht, MLR/GLM SIG Executive Secretary
Department of Educational Technology, Research & Assessment
Northern Illinois University
DeKalb, IL 60115-2854
jbhecht@niu.edu

Check out our website at: http://www.coe.unt.edu/mlrv/

POSTMASTER: Send address changes to:
Jeffrey B. Hecht, MLR/GLM SIG Executive Secretary
Department of Educational Technology, Research & Assessment
Northern Illinois University
DeKalb, IL   60115-2854

Multiple Linear Regression Viewpoints (ISSN 0195-7171) is published by the
AERA Special Interest Group on Multiple Linear Regression: General Linear Model
through the **University of Alabama at Birmingham** and the **Dallas Independent School District**.