EVALUATING CLASSIFICATION AND REGRESSION TREES WITH CLINICAL TRIAL DATA

by

MATHEW JON GREGOSKI

(Under the direction of Ted A. Baumgartner)

ABSTRACT

An examination of classification and regression tree (CART) models among African American (AA) adolescents at risk for Essential Hypertension (EH) was conducted. The purpose of Study I was to compare multiple CART model rule creation and cross-validation techniques with each other using intervention data and validate the results with hierarchical regression models. The analyses utilized a data sample obtained from a randomized clinical trial of 181 AA adolescents considered to be at risk for EH. CART models were created using the Gini, Entropy, Class Probability, and Two-ing selection methods combined with the fraction of random cases, and V-fold cross-validation techniques The CART models examined behavioral stress interventions and the influence of underlying anthropometric, psychosocial, and behavioral variable and their impact on resting systolic blood pressure (SBP), diastolic blood pressure (DBP), and heart rate (HR). The findings imply that CART models using the "Gini or Entropy" selection methods combined with V-fold cross-validations were the best methods for use in clinical trial research. The results of the CART models agreed with previous regression analyses and in some circumstances provided additional information not captured by the regression models.

Study II utilized the same 181 participants in Study I and examined the same baseline anthropometric, psychosocial, and behavioral characteristics and treatment group effects. In addition, baseline characteristics, changes in anthropometric, psychosocial, and behavioral characteristics that occurred during the intervention period were also examined for the purpose of determining what treatments and characteristic lead to improved cardiovascular function in the natural environment as measured by 24 hour ambulatory SBP, DBP, and HR. Based on study I CART models using "Gini" and "Entropy" selection methods with V-fold cross-validation were constructed for ambulatory SBP, DBP, and HR. Hierarchical regression models were created that included variables and values based on the rules obtained from the CART analyses. Across all regression models significant effects were found for the subgroups formed from CART outputs. The studies show CART models created with "Gini or Entropy" selection methods combined with V-fold cross-validation are a useful method for maximizing clinical trial success rates at the individual level.

INDEX WORDS: Adolescents, African American, Ambulatory Blood Pressure, Essential Hypertension, Classification and Regression Trees, Psychosocial Characteristics.

EVALUATING CLASSIFICATION AND REGRESSION TREES WITH CLINICAL TRIAL DATA

by

MATHEW JON GREGOSKI

B.A., Elon University, 2000

M.S., Augusta State University, 2005

A dissertation submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2009

EVALUATING CLASSIFICATION AND REGRESSION TREES WITH CLINICAL TRIAL DATA

by

MATHEW JON GREGOSKI

Major Professor: Ted A. Baumgartner

Committee: Rodney K. Dishman

Patrick J. O'Connor

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2009

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

A dramatic decline in the efficacy and efficiency of the former health care system led the National Institutes of Health into developing the 4P Medicine Policy (NIH, 2007). The former health care system could be described as disease-oriented, reactive, and sporadic, and in some cases interventions were utilized when they were least effective and most expensive. The 4P system has been described as personalized, predictive, preventive, and participatory. Some aims of the 4P system include: detect diseases earlier making them easier to and less expensive to treat effectively; place patients into groups that enable the selection of optimal therapy; reduce time, cost, and failure rate of clinical trials for new therapies; and shift the emphasis in medicine from reaction to prevention and from disease to wellness (Hood 2002). The approach to developing the 4P healthcare system requires an integration of expertise across multiple disciplines. Advancement of measurement and visualization technologies will assist in the transition to the 4P system, but new computational and mathematical tools are also important.

In trying to determine the important characteristics related to the 4P system current statistical models may not be the best methods to use. In many research fields, general linear models and ordinary least squares methods are widely advocated and used. Some examples of these include: ANOVA, ANCOVA, MANOVA, MANCOVA, and Linear Regression. Essentially, the statistical models all examine the same thing, which is the observed variance that is explained by group membership based on variables in the model compared to the random variance that exists which is not explained by any of the variable groups in the model. General linear models determine whether the difference between means of groups is larger than what might be expected by chance based on the amount of data, number of variables, and variance in the model. The ordinary least squares method commonly used to fit data in regression models also is dependent on explaining variance.

In order to determine the overall best set of predictors using the least squares method with linear regression models, beta terms are created by examining the explained covariance between a set of predictors and a dependent variable compared to the remaining variance not explained with the set of predictor variables. The general linear models and ordinary least squares methods are all dependent on group means. Results from these models may not generalize to the 4P system which emphasizes an individual personalized approach. Alternatively, it may be more feasible to determine what characteristics actually lead to individual improvement in the general population, and maximize the chance for improvement based on increasing these characteristics. The underlying problem is finding the best method for determining or learning what specific characteristics lead to successful improvement in terms of outcome variables at the individual level that also lead to success with new participants.

The previously mentioned statistical models utilize the general linear model or ordinary least squares approaches. An alternative approach is the use of minimax decision rules. Minimax decision rules have been used in many applications in order to minimize the maximum possible information loss. Information loss can be defined as the difference between the consequences of the best decision that could have been taken had the underlying circumstances been known and the decision that was in fact taken before they were known. When minimax rules are implemented to learn unknown structures of a dataset, it can be thought of as a way of maximizing the minimal gain. Once enough rules are created that define the structure of the dataset, participants can be then be matched by individual characteristics to the rules to increase the likelihood of an outcome. The approach matches well with the underlying purposes of the 4P system. In essence, the purpose of the 4P system is to maximize successful outcome using methods that are personalized, predictive, preventive, and participatory. Different variations of the minimax decision process exist. One minimax method that has been popularized is Classification and Regression Trees (CART) developed by Breiman, Friedman, Stone, & Olshen (1984). The CART technique does not rely on group means but uses rules learned from the model to create decision trees which graphically maps observations about a variable to conclusions based on the variable's predefined specified value. In other words, Decision tree designs are basically a graphic display of rules about the relationships between variables and values which are mapped to meet a specified outcome.

Figure 1.1 is a decision tree that was created to identify patients who are at risk for dying within 30 days. The patients had experienced a heart attack and survived for at least a 24 hour period past hospital admission. To identify what characteristics differentiated between heart attack patients who were "survivors" and those who experienced "early deaths" (dying within a 30 day period following the heart attack), Breiman et al. (1984) examined nineteen variables from the 215 heart attack patients. One hundred and seventy-eight patients were "survivors" and 37 experienced "early deaths". From the nineteen variables three were selected for use in the decision tree model. The three variables included: minimum systolic blood pressure, age, and whether the patient experiences sinus tachycardia (defined as being present if the sinus node heart rate ever exceeded 100 beats per minute during the first 24 hours after admission to the hospital).

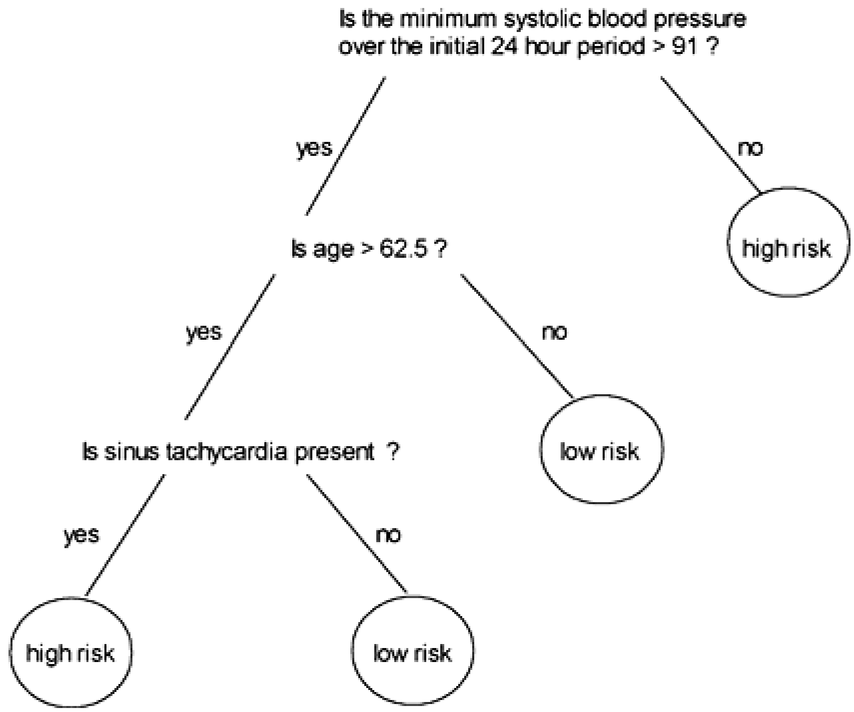

Figure 1.1: Simple Decision Tree Model for Assessing Risk for Early Death following a Heart Attack (Breiman et al. 1984)

In a CART model, rules are created on variables in the dataset depending on their likelihood to meet the specific criteria selected by the researcher (i.e. determining if the person at risk for early death following a heart attack). There are multiple methods that can be used to create rules on the variables and values but all are based on the way information is gained in the CART model. CART models are often created to increase rules with variables until no more rules are possible because all the participants have been classified (or maximum information is gained). In the final model one can easily determine treatment or risk just by following the rules in the model (i.e. in Figure 1.1 the presence of sinus tachycardia following a heart attack leads to high risk for an early death).

CART analyses have started to capture the attention of researchers in many disciplines and it appears its use is increasing. However, in many of the recently published studies CART methods were not adequately described. Primarily, CART has been used as a screening technique for simple dichotomous decisions such as the example provided in Figure 1.1 that displays decision rules to determine whether a patient is at high risk for an early death following a heart attack. No previous studies have implemented CART across different treatment groups in a standardized controlled trial with the attempt to create rules that classify individuals who will be successful in future studies based on a predefined clinically meaningful outcome.

The stance taken herein does not suggest that CART with its minimax decision rule based approach is the end all solution in clinical trial research and acknowledges that in many circumstances it is beneficial to use additional statistical models in conjunction with CART. However, it is plausible that in many circumstances it may be beneficial to use CART to help develop combinations of rules that lead to success on the individual level. Given the algorithms behind the different general linear models it is possible that CART may be better than general linear models for finding variables and values related to individual success rates that also generalize well to new individuals.

Purpose

The purpose of Study I is to compare multiple CART model rule creation and cross-validation techniques with each other using intervention data and to validate the results with previously conducted hierarchical regression models. In order to determine the best CART methods, all possible combinations of rule creating methods and cross-validations are examined. The results of Study I will help the researcher determine if these different methods and combinations lead to different interpretations of the data or have specific benefits in providing clearer and more concise explanations of the variable relationships. The CART models should help determine which additional statistical methods to use in follow up analyses. However, since the focus is on testing the different CART methods a specific criterion for the most accurate model needs to be set in advance. A previous set of hierarchical stepwise regressions will provide the reference and validation models for these analyses (Barns, Gregoski, Tingen, & Treiber, Under Review).

The purpose of Study II is to use determine the usefulness of CART methods in identifying the important baseline variables, and changes in those variables, related to individual success rates in improving ambulatory blood pressure and heart rate. In addition, a purpose is to determine how the variables identified in CART can be further used in hierarchical regression models. The variables and values from the CART model decision rules are used in hierarchical regression models which used continuous change scores rather than dichotomous scores.

CHAPTER 2

LITERATURE REVIEW

Few studies have examined the relationship of classification and regression trees (CART) models and statistical models used in clinical trials (Karels, Bryant, & Hik, 2004; Lemon, Roy, Clark, Friedmann, & Rakowski, 2003) and none have compared CART with hierarchical regression models. Because CART and hierarchical regression models have not been previously compared, much of the review will focus on why CART methods may outperform hierarchical regression models when the focus of the research is to maximize individual success rates. The personalized, predictive, preventive, and participatory medicine model is briefly discussed as its advocates may be particularly interested in CART methods. The data used in the study was obtained from a clinical trial examining African American adolescents at risk for essential hypertension. As a result other topics in the literature review focus on the mechanisms and psychosocial characteristics unique to the specific population being studied. The literature review is organized into four sections: (i) Personalized, predictive, preventive, and participatory medicine, (ii) essential hypertension among African American adolescents, (iii) psychosocial constructs, and (iv) CART analyses in making personalized predictive decisions.

Personalized, Predictive, Preventive, and Participatory Medicine

The aims of personalized, predictive, preventive and participatory medicine include: detecting diseases at earlier stages making them easier and more cost effective to treat; placing patients into the optimal treatment groups increasing the likelihood of effective therapy; reducing adverse reactions to prescribed drugs by detecting individual drug responses earlier; improving the selection of at-risk participants for new drug discovery; reducing time, cost, and failure rate of new therapies developed in clinical trials; and shifting the emphasis in medicine from reaction to prevention, and from disease to wellness (Hood & Galas, 2008). The Blue Ridge Academic Health Group emphasized the major consequence of focusing on disease treatment was the relative neglect of health promotion and disease prevention (BRAHG., 2003). Given the needed emphasis for health promotion and disease prevention, an argument can be made for the need to develop mathematical and computational methods for identifying the maximum number of behavioral and psychosocial characteristic related to successful intervention.

Essential Hypertension Among African American Adolescents

Globally, uncontrolled essential hypertension (EH) is the major cause of heart disease and stroke and the number one attributable risk factor for death (World Health Organization, 2007). African American (AA) individuals experience higher prevalence and earlier onset of EH compared to other ethnic groups in the United States (Burt, et al., 1995; Stamler, Stamler, & Neaton, 1993). In the past decade incidence of pediatric EH has been escalating (Muntner, He, Cutler, Wildman, & Whelton, 2004) with higher incidence reported among minority adolescents (Sorof, Lai, Turner, Poffenbarger, & Portman, 2004). Blood pressure (BP) ranking tracks from late childhood onward (Shear, Burke, Freedman, & Berenson, 1986) placing AA teens with high normal BP at particular risk for development of EH (Bao, Threefoot, Srinivasan, & Berenson, 1995). Non-pharmacological interventions in youth (e.g., physical activity, diet, electrolyte supplementation) have had mixed success rates and many researchers observe minimal to no impact upon BP in normotensive youth (Alpert, Murphy, & Treiber, 1994; Resnicow & Robinson, 1997). Given the increased incidence of EH in youth, new effective prevention programs need to be identified.

A number of behavioral stress-related factors have been identified as contributing to EH (e.g., aversive interpersonal interactions related to socioeconomic status associated inequality, racism, etc.; neighborhood and/or family dysfunction; ineffective anger management and coping skills) (Anderson, 1989; Clark, Anderson, Clark, & Williams, 1999; Player, King, Mainous, & Geesey, 2007; Williams, Yu, Jackson, & Anderson, 1997; Williams, et al., 2000). Behavioral stress induces increased retention of sodium (Harshfield, et al., 2007). A normal functioning renal sodium handling system results in a restoration of sodium balance following the cessation of stress via increased urinary release of sodium. African Americans, particularly those with EH, exhibit increased prevalence of sodium handling problems (Franco & Oparil, 2006; Nesbitt, 2004; Weinberger, 2006). Behavioral stress researchers examining the effects of acute laboratory stress have demonstrated a significant percentage of AAs retain sodium rather than exhibit the expected post-stress response of increased sodium excretion (Harshfield, Treiber, Davis, & Kapuku, 2002; Harshfield, Wilson, et al., 2002).

Hyperactivity of the sympathetic nervous system (SNS) has been implicated in the development of hypertension preceding cardiovascular complications (Reaven, Abbasi, & McLaughlin, 2004). For example, SNS activation that occurs during stress promotes sodium retention (Harshfield, Treiber, et al., 2002). Behavioral stress reduction practices may decrease neurohormonal activity, sodium appetite as indicated by decreased 24-hour urinary sodium excretion, and help control the hypothalamic-pituitary-adreno-cortical axis, and the renin-angiotensin-aldosterone system (Walton, Pugh, Gelderloos, & Macrae, 1995). Previous meditation studies involving adults have shown favorable results in, decreased 24-hour urinary sodium excretion, switching of sympathetic activity to parasympathetic activity with the end result creating favorable changes in cardiac-vagal function and improved metabolic functions related to inflammatory responses (Grossman, Niemann, Schmidt, & Walach, 2004). A few behavioral stress reduction interventions directed at BP control have been conducted with youth (Ewart, et al., 1987). Two controlled randomized clinical trials on AA adolescents with higher than normal BP found that after 2- and 4-month interventions, Transcendental Meditation® significantly lowered resting systolic blood pressure (SBP), resting diastolic blood pressure (DBP) and daytime ambulatory SBP/DBP compared to health education (HE) controls(Barnes, Treiber, & Davis, 2001; Barnes, Treiber, & Johnson, 2004). Three month breathing awareness meditation (BAM) programs involving AA teenagers with high normal SBP (i.e., pre-hypertensive) have shown favorable decreases in resting SBP and ambulatory SBP and decreased sodium excretion (Barnes, Pendergrast, Harshfield, & Treiber, 2008) when compared to (HE) control groups(Barnes, et al., 2001; Barnes, Treiber, et al., 2004). A study with normotensive middle school students using BAM reported greater decreases in resting SBP, and daytime ambulatory SBP and HR compared to a HE control group (Barnes, Davis, Murzynowski, & Treiber, 2004).

The need for anger and hostility management programs for youth within the school setting has been identified primarily from the perspective of reducing violence and conduct problems (Farrell & Meyer, 1997; Krajewski, Rybarik, Dosch, & Gilmore, 1996; Powell, et al., 1996). Findings have been mixed with some programs resulting in significant improvements (Ginsburg & Drake, 2002; Snyder, Kymissis, & Kessler, 1999) while others have revealed little or no change (Cirillo, et al., 1998; Orpinas, et al., 2000) in teacher or subject self-reported measures of behavior problems, fighting, etc. Life Skills Training (LST) has been well developed and methodically tested in schools as a substance abuse prevention program, particularly with inner city youth. LS has been found to facilitate development of important cognitive-behavioral skills for managing stress and anger, increase self-esteem, decrease anxiety and general stress and increase overall coping skills (Botvin, Baker, Renick, Filazzola, & Botvin, 1984; Botvin, Eng, & Williams, 1980; Botvin & Griffin, 2002).

Finding efficacious treatments that are personalized for at risk AA adolescents will require further investigation of underlying anthropometric, psychosocial, and behavioral characteristics. To date, a randomized clinical trial has not been conducted comparing several BP stress reduction interventions such as BAM, HE and LST among youth classified as pre-hypertensive due to high normal BP for their age, sex and height. In addition, previous studies have not considered how baseline values and changes in values obtained from personal anthropometric, psychosocial, and behavioral characteristic may influence the efficacy of stress reduction treatments on ambulatory BP and sodium handing among pre-hypertensive AA high school adolescents..

Psychosocial Constructs

Many measures represent constructs purported as influential to changes in hemodynamic function. While some of the measures are widely used, in many cases, few if any researchers have examined their adequacy among AA adolescents. Across different ethnic and racial backgrounds constructs may vary widely in their psychometric characteristics. As a result, brief descriptions are provided for each measure along with previous studies examining the constructs with adolescents and AA adolescents when available.

Adolescent Cook Medley Inventory. The twenty-three item Adolescent Cook Medley Inventory (ACMI) was developed from the full scale Cook-Medley inventory originally designed to test workgroup cohesion (Matthews, Gump, Block, & Allen, 1997). The current version of the ACMI uses a 4 point response scale compared to the original True or False format and the items have been altered to make them more appropriate for children. The possible scores range from 23 to 92. A previous studies has indicated good reliability with adolescent samples (Liehr, et al., 2000).

Family Environment Scale. The Family Environment Scale (FES) consists of twenty-seven items and was developed by Moos & Moos (1981) to provide information on family environment. The selected subscale items attempt to capture cohesion(coh), expressiveness(exp), and conflict(con). The subscales of the FES are purported to capture different aspects of the family environment deeming an internal consistency analyses inappropriate due to the heterogeneity of the measure. A sample of 1,067 families ranging from low socio-economic status (SES) to high SES had test-retest reliability across a 2 month interval ranging from .68 to .86. The scale has also effectively discriminated between families whose members have psychiatric problems and those who do not (Moos & Moos, 1981). Both parents and children completed the scale.

Spielberger Anger Expression Inventory. The Spielberger Anger Expression Inventory (SPAX) scales consists of twenty-four items developed to capture Anger Out (AO), the frequency of anger expressed, Anger In (AI), the frequency of anger experienced but not expressed, Anger Control (AC), the frequency an alternative to anger was utilized. Anger Expression (AE) is also sometimes computed by combining the AI and AO scores and subtracting out the AC. Johnson et al. (1987) reported the inventory to have good internal consistent reliability (alpha = .84) among a sample of 350 adolescent females of which 171 were black (Johnson, 1992).

Perceived Stress Survey. The brief version of the Perceived Stress Survey (PSS) contains four items and describes life stress in terms of feeling in control. The PSS measures perceptions of life overload, and lack of predictability and control. Items are coded numerically from 0 to 4 making a possible 0 to 16 range across the survey. No psychometric evidence for a population of adolescents was found. However, the PSS was designed for use with samples with at least a junior high education and therefore should be suitable for adolescents (Ng & Jeffery, 2003).

Everyday Discrimination Scale. The Everyday Discrimination Scale (EDS) consists of nine items answered on a six point Likert scale format. Items were designed to assess chronic, routine, and relatively minor experiences of unfair treatment. Perceived discrimination is measured from within the context of unfairness (i.e., being treated with less respect, courtesy, receiving poorer service, etc.) as opposed to the context of gender, race-ethnicity, or social class. Williams et al., (1997) found the scales to be internally consistent but indicated that black adults had significantly higher scores compared to whites (Williams, Yan, Jackson, & Anderson, 1997).

City Life Inventory. The City Life Inventory (CLI) consists of thirty-six items answered in a Likert scale format with a four point scale. The original thirty-six items were derived from Project Heart, a series of community based studies in Baltimore that investigated relationships between cardiovascular risk and emotional stress in urban adolescents (Fitzgerald, Brown, Sonnega, & Ewart, 2005). In a previous study Ewart and Suchday (2002) validated the eighteen items with objective indices of environmental quality using an adolescent sample (Ewart & Suchday, 2002).

Hollingshead Index. Socio-economic status was measured using the Hollingshead (HH) four factor indexes (Hollingshead, 1981) . The maximum score from the four factor index has been shown to have good psychometric properties in the United States with diverse populations (Cirino, et al., 2002).

Youth Risk Behaviors Survey. The Youth Risk Behaviors Survey (YBRS) was created to capture substance intake among adolescents. The scale assesses smoking behavior, use of alcohol, and illicit drugs (Kann, et al., 1998). In a study examing 1,679 students in grades 7 through 12 the scale was reported to reliably capture substance abuse (Brener, Collins, Kann, Warren, & Williams, 1995).

CART Analyses in Making Personalized Predictive Decisions

The creation of rule based classification through methods such as CART has been widely available for many years (Breiman, et al., 1984). However, many programs and methods used to created decision trees and decision rules differ from those in the original CART book. Many studies which have examined decision trees (Barriga, Hamman, Hoag, Marshall, & Shetterly, 1996; Podgorelec, Kokol, Stiglic, & Rozman, 2002), or use alternative methods and algorithms labeled as “machine learning” such as ID3 or C4.5 (Quinlan, 1993; Witten & Frank, 2005) but none of these studies examined the original CART program published in “Classification And Regression Trees” by Breiman et al. (1984). Although, their book offers thorough real world examples and precise derivation of the tree models using probability and set theory, it does not include instruction on assistance with software to implement the programs. At the time the book was written most of the software packages that are commonly used today were unavailable.

The major problem with CART analyses in its original development was inadequate computer resources. The creation of rules to use in large decision trees requires the use of large amounts of computer random access memory. CART remained in the literature but it is speculated that the use of difficult notation combined with the extensive programming needed to work around computer systems with less memory made methods difficult to comprehend for novice researchers. Researchers not familiar with set theory or computer programming would have difficulty deciphering programs and likely turn towards alternative methods.

Despite the limitations in resources for learning CART methods and the lack of commercial programs available to conduct the methods a handful of researchers have examined CART methods with the original programming code. In a study focusing on improving population screen for glucose tolerance, researchers found CART to be a useful tool producing adequate specificity and sensitivity in determining high risk subjects (Barriga, et al., 1996). In another study researchers found that CART methods were able to screen individuals who received flu vaccines, and results were validated by logistic regression models (Lemon, et al., 2003). In addition, an Ecology study provided evidence to support that CART performed as well as discriminant function analyses (DFA) in age classification of marmots with complete datasets, and that CART outperformed DFA when cases were missing data (Karels, et al., 2004).

Previous researchers have provided overviews of the benefits of decision trees and their use in medicine (Podgorelec, et al., 2002). In addition some have even provided a methodological review (Lemon, et al., 2003). However, previous researchers have not implemented CART models with the attempt to create decision rules related to individual success rates as an alternative to statistical techniques that use general linear modeling or ordinary least squares approaches such as hierarchical regression models. In one study researchers examined logistic regression and signal detection methods which are similar to CART on the ability to identify different subgroups at risk (Kiernan, Kraemer, Winkleby, King, & Taylor, 2001). The researchers found that logistic regression models identified individuals that were homogeneous in outcome but heterogeneous in risk predictors. However, signal detection methods identified individuals that were homogeneous in both outcome and risk predictors. These results are strong evidence for advocating the use of CART methods in order to create subgroups based on decision rules that identify the specific variables and values related to success.

References

Alpert, B. S., Murphy, J. K., & Treiber, F. A. (1994). Essential hypertension: Approaches to prevention in children. *Medicine and Exercise in Nutrition and Health, 3,* 296-307.

Anderson, N. B. (1989). Racial differences in stress-induced cardiovascular reactivity and hypertension: Current status and substantive issues. *Psychological Bulletin, 105*(1), 89-105.

Bao, W., Threefoot, S. A., Srinivasan, S. R., & Berenson, G. S. (1995). Essential hypertension predicted by tracking of elevated blood pressure from childhood to adulthood: The bogalusa heart study. *American Journal of Hypertension, 8*, 657-665.

Barnes, V. A., Davis, H. C., Murzynowski, J. B., & Treiber, F. A. (2004). Impact of meditation on resting and ambulatory blood pressure and heart rate in youth. *Psychosomatic Medicine, 66*(6), 909-914.

Barnes, V. A., Pendergrast, R. A., Harshfield, G. A., & Treiber, F. A. (2008). Impact of breathing awareness meditation on ambulatory blood pressure and sodium handling in prehypertensive African American adolescents. *Ethnicity and Disease, 18*(1), 1-5.

Barnes, V. A., Treiber, F. A., & Davis, H. (2001). Impact of transcendental meditation on cardiovascular function at rest and during acute stress in adolescents with high normal blood pressure. *Journal of Psychosomatic Research, 51*(4), 597-605.

Barnes, V. A., Treiber, F. A., & Johnson, M. H. (2004). Impact of transcendental meditation on ambulatory blood pressure in african-american adolescents. *American Journal of Hypertension, 17*(4), 366-369.

Barns, V. A., Gregoski, M. J., Tingen, M. S., & Treiber, F. A. (Under Review). Factors impacting improved hemodynamic function among african american adolescents engaged in behavioral stress reduction programs.

Barriga, K. J., Hamman, R. F., Hoag, S., Marshall, J. A., & Shetterly, S. M. (1996). Population screening for glucose intolerant subjects using decision tree analyses. *Diabetes Research in Clinical Practice, 34 Suppl*, S17-29.

Botvin, G. J., Baker, E., Renick, N. L., Filazzola, A. D., & Botvin, E. M. (1984). A cognitive-behavioral approach to substance abuse prevention. *Addictive Behaviors, 9*(2), 137-147.

Botvin, G. J., Eng, A., & Williams, C. L. (1980). Preventing the onset of cigarette smoking through life skills training. *Preventive Medicine, 9*(1), 135-143.

Botvin, G. J., & Griffin, K. W. (2002). Life skills training as a primary prevention approach for adolescent drug abuse and other problem behaviors. *International Journal of Emergency Mental Health, 4*(1), 41-47.

BRAHG. (2003). The Blue Ridge Academic Health Group. Reforming medical education: Urgent priority for the academic health center in the new century., 2008, from http://www.whsc.emory.edu/blueridge/_pdf/blue_ridge_report_7_2003may.pdf

Breiman, L., Friedman, J. H., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. Belmont, Calif.: Wadsworth International Group.

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*(6), 575-580.

Burt, V. L., Whelton, P., Rocella, E. J., Brown, C., Cutler, J. A., Higgins, M., et al. (1995). Prevalence of hypertension in the us adult population: Results from the third national health and nutrition examination survey, 1988-1991. *Hypertension, 25*, 305-313.

Cirillo, K. J., Pruitt, B. E., Colwell, B., Kingery, P. M., Hurley, R. S., & Ballard, D. (1998). School violence: Prevalence and intervention strategies for at-risk adolescents. *Adolescence, 33*(130), 319-330.

Cirino, P. T., Chin, C. E., Sevcik, R. A., Wolf, M., Lovett, M., & Morris, R. D. (2002). Measuring socioeconomic status: Reliability and preliminary validity for different approaches. *Assessment, 9*(2), 145-155.

Clark, R., Anderson, N. B., Clark, V. R., & Williams, D. R. (1999). Racism as a stressor for African Americans: A biopsychosocial model. *American Psychologist, 54*(10), 805-816.

Ewart, C. K., Harris, W. L., Iwata, M. M., Coates, T. J., Bullock, R., & Simon, B. (1987). Feasibility and effectiveness of school-based relaxation in lowering blood pressure. *Health Psychology, 6*(5), 399-416.

Ewart, C. K., & Suchday, S. (2002). Discovering how urban poverty and violence affect health: Development and validation of a neighborhood stress index. *Health Psychology, 21*(3), 254-262.

Farrell, A. D., & Meyer, A. L. (1997). The effectiveness of a school-based curriculum for reducing violence among urban sixth-grade students. *American Journal of Public Health, 87*(6), 979-984.

Fitzgerald, S. T., Brown, K. M., Sonnega, J. R., & Ewart, C. K. (2005). Early antecedents of adult work stress: Social-emotional competence and anger in adolescence. *Journal of Behavioral Medicine, 28*(3), 223-230.

Franco, V., & Oparil, S. (2006).Salt sensitivity: A determinant of blood pressure, cardiovascular disease and survial. *Journal of the American College of Nutrition, 25,* 247S-255S.

Ginsburg, G. S., & Drake, K. L. (2002). School-based treatment for anxious african-american adolescents: A controlled pilot study. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*(7), 768-775.

Grossman, P., Niemann, L., Schmidt, S., & Walach, H. (2004). Mindfulness-based stress reduction and health benefits: A meta-analysis. *Journal of Psychosomatic Research, 57*(1), 35-43.

Harshfield, G. A., Hanevold, C., Kapuku, G. K., Dong, Y., Castles, M. E., & Ludwig, D. A. (2007). The association of race and sex to the pressure natriuresis response to stress. *Ethnicity and Disease, 17*(3), 498-502.

Harshfield, G. A., Treiber, F. A., Davis, H., & Kapuku, G. K. (2002). Impaired stress-induced pressure natriuresis is related to left ventricular structure in blacks. *Hypertension, 39*(39), 844-847.

Harshfield, G. A., Wilson, M. E., Hanevold, C., Kapuku, G. K., Mackey, L., Gillis, D., et al. (2002). Impaired stress-induced pressure natriuresis increases cardiovascularload in african american youths. *American Journal of Hypertension, 15*(10 Pt 1), 903-906.

Hollingshead (1981). Four factor index of social status. Dept of Sociology, Yale University.

Hood, L. E., & Galas, D. J. (2008). P4 medicine: Personalized, predictive, preventive, participatory a change of view that changes everything, from http://www.cra.org/ccc/initiatives

Johnson, E. H. (1992). *Personality, elevated blood pressure and essential hypertension*. Washington DC: Taylor & Francis.

Kann, L., Kinchen, S. A., Williams, B. I., Ross, J. G., Lowry, R., Hill, C. V., et al. (1998). Youth risk behavior surveillance--united states, 1997. *MMWR CDC Surveillance Summaries, 47*(3), 1-89.

Karels, T. J., Bryant, A. A., & Hik, D. S. (2004). Comparison of discriminant function and classification tree analyses for age classification of marmots. *Oikos, 105*(3), 575-587.

Kiernan, M., Kraemer, H. C., Winkleby, M. A., King, A. C., & Taylor, C. B. (2001). Do logistic regression and signal detection identify different subgroups at risk? Implications for the design of tailored interventions. *Psychological Methods, 6*(1), 35-48.

Krajewski, S., Rybarik, M., Dosch, M., & Gilmore, G. (1996). Results of a curriculum intervention with seventh graders regarding violence in relationships. *Journal of Family Violence, 11*, 93-112.

Lemon, S. C., Roy, J., Clark, M. A., Friedmann, P. D., & Rakowski, W. (2003). Classification and regression tree analysis in public health: Methodological review and comparison with logistic regression. *Annals of Behavioral Medicine, 26*(3), 172-181.

Liehr, P., Meininger, J. C., Mueller, W. H., Chan, W., Frazier, L., & Reyes, L. R. (2000). Psychometric testing of the adolescent version of the cook-medley hostility scale. *Issues Comprehensive Pediatric Nursing, 23*(2), 103-116.

Matthews, K. A., Gump, B. B., Block, D. R., & Allen, M. T. (1997). Does background stress heighten or dampen children's cardiovascular responses to acute stress? *Psychosomatic Medicine, 59*(5), 488-496.

Moos, B. S., & Moos, R. H. (1981). *Family environment scale manual*. Palo Alto, CA: Consulting Psychologist Press.

Muntner, P., He, J., Cutler, J. A., Wildman, R. P., & Whelton, P. K. (2004). Trends in blood pressure among children and adolescents. *JAMA, 291*(17), 2107-2113.

Nesbitt, S. D. (2004). Hypertension in black patients: Special issues and considerations. *Current Cardioogyl Reports, 6*(6), 416-420.

Ng, D. M., & Jeffery, R. W. (2003). Relationships between perceived stress and health behaviors in a sample of working adults. *Health Psychology, 22*(6), 638-642.

NIH (2007). The promise of personalized medicine. *NIH Medline Plus, 4*(1), 2-4.

Orpinas, P., Kelder, S., Frankowski, R., Murray, N., Zhang, Q., & McAlister, A. (2000). Outcome evaluation of a multi-component violence-prevention program for middle schools: The students for peace project. *Health Education Research, 15*(1), 45-58.

Player, M. S., King, D. E., Mainous, A. G., 3rd, & Geesey, M. E. (2007). Psychosocial factors and progression from prehypertension to hypertension or coronary heart disease. *Annals of Family Medicine, 5*(5), 403-411.

Podgorelec, V., Kokol, P., Stiglic, B., & Rozman, I. (2002). Decision trees: An overview and their use in medicine. *Journal of Medical Systems, 26*(5), 445-463.

Powell, K. E., Dahlberg, L. L., Friday, J., Mercy, J. A., Thornton, T., & Crawford, S. (1996). Prevention of youth violence: Rationale and characteristics of 15 evaluation projects. *American Journal of Preventive Medicine, 12*(5 Suppl), 3-12.

Quinlan, J. R. (1993). *C4.5 : Programs for machine learning*. San Mateo, Calif.: Morgan Kaufmann Publishers.

Reaven, G., Abbasi, F., & McLaughlin, T. (2004). Obesity, insulin resistance, and cardiovascular disease. *Recent Progress in Hormone Research, 59*, 207-223.

Resnicow, K., & Robinson, T. N. (1997). School-based cardiovascular disease prevention studies: Review and synthesis. *Annals of Epidemiology, 7*(S7), S14-S31.

Shear, C. L., Burke, G. L., Freedman, D. S., & Berenson, G. S. (1986). Value of childhood blood pressure measurements and family history in predicting future blood pressure status: Results from 8 years of follow-up in the bogalusa heart study. *Pediatrics, 77*(6), 862-869.

Snyder, K. V., Kymissis, P., & Kessler, K. (1999). Anger management for adolescents: Efficacy of brief group therapy. *Journal of the American Academy for Child and Adolescent Psychiatry, 38*(11), 1409-1416.

Sorof, J. M., Lai, D., Turner, J., Poffenbarger, T., & Portman, R. J. (2004). Overweight, ethnicity, and the prevalence of hypertension in school-aged children. *Pediatrics, 113*(3 Pt 1), 475-482.

Stamler, J., Stamler, R., & Neaton, J. D. (1993). Blood pressure, systolic and diastolic, and cardiovascular risks. U.S. population data. *Archive of Internal Medicine, 153*(5), 598-615.

Walton, K. G., Pugh, N. D., Gelderloos, P., & Macrae, P. (1995). Stress reduction and preventing hypertension: Preliminary support for a psychoneuroendocrine mechanism. *Jorunal of Alternative and Complementary Medicine, 1*(3), 263-283.

Weinberger, M. H. (2006). Pathogenesis of salt sensitivity of blood pressure. *Current Hypertension Reports, 8*, 166-170.

Williams, D. R., Yan, Y., Jackson, J. S., & Anderson, N. B. (1997). Racial differences in physical and mental health: Socio-economic status, stress and discrimination. *Journal of Health Psychology, 2*(3), 335-351.

Williams, D. R., Yu, Y., Jackson, J. S., & Anderson, N. B. (1997). Racial differences in physical and mental health. *Journal of Health Psychology, 2*, 335-351.

Williams, J. E., Paton, C. C., Siegler, I. C., Eigenbrodt, M. L., Nieto, F. J., & Tyroler, H. A. (2000). Anger proneness predicts coronary heart disease risk: Prospective analysis from the atherosclerosis risk in communities (aric) study. *Circulation, 101*(17), 2034-2039.

Witten, I. H., & Frank, E. (2005). *Data mining: Practial machine learning tools and techniques* (Second ed.). San Francisco: Elsevier Inc.

World Health Organization (2007). *The world health report 2007 - A safer future: Global public health security in the 21st century*. Geneva, Switzerland: World Health Organization.

CHAPTER 3

PERSONALIZED PREDICITIVE DECISION METHODS: A COMPARISON OF CART AND HIERARCHICAL LINEAR REGRESSION APPROACHES[1]

[1] Gregoski, M.G., Baumgartner, T.A; University of Georgia. To be submitted to *Psychological Reports*

Abstract

Objectives: To compare multiple Classification and Regression Tree (CART) models created with different rule creation and cross-validation methods with hierarchical regression models using intervention data from a randomized control trial.

Design: CART models were conducted using different combinations of rule selection and cross validation methods. Model outcomes were compared to previous hierarchical regression models which used the same data. Data was obtained from a randomized clinical trial that investigated the impact of behavioral stress reduction techniques, and baseline anthropometric, psychosocial, and behavioral characteristics on essential hypertension. The participants consisted of 181 African American adolescents who had blood pressure values between the 75^{th} to 95^{th} percentiles based on age, height, and gender.

Results: The CART models conducted with the Gini and Entropy rule selection methods, and V-fold cross-validation were the best methods and revealed similar results when compared to hierarchical regression models. In addition, creating a hierarchical regression models based on CART results revealed that although some variables were missing data, they actually proportionally accounted for more variance.

Conclusions: CART models should be conducted using both the Gini and Entropy rule selection methods, and V-fold cross-validation. CART models provide additional diagnostic information such as finding curvilinear relationships and interactions that may be missed when only using traditional regression analyses.

INDEX WORDS: Adolescents, African American, Blood pressure, Essential hypertension, Classification and Regression Trees, Psychosocial characteristics.

Introduction

Across multiple disciplines emphasis has been placed on the need for personalized planned preventive treatment programs. Researchers have agreed that specific individually tailored treatment programs must be developed and implemented in order to reduce the current negative health trends inflicting many individuals in the general population (Heymann, Prentice, & Reinders, 2007). While research programs are often designed to address these demands for individual specificity, the statistical methods and analyses used in these research programs may not always be the best methods available to help determine what individual factors are related to whether a treatment program will be successful at the individual level.

Parametric statistics are used in many research programs; these statistics assumes data come from a type of probability distribution and makes inferences about the parameters of the distribution (Geisser & Johnson, 2006). Parametric statistics with continuous outcome variables are often preferred over non-parametric statistics because they are thought to be a more sensitive statistical method. However, parametric models with continuous outcomes may not always be the best choice when the overall goal is to develop personalized prevention strategies (Kraemer, et al., 1999). If two or more treatment groups have an equal amount of improvement in terms of the measured outcome the researcher must either determine if the treatments are equally effective across all participants, or try and find some other explanation within each of the groups to account for the differences between participants that received a treatment and improved from those who did not improve. Alternatively, it may be more feasible to ignore the sample distribution and dichotomize the measured outcome into success and unsuccessful cases and use a statistical model to determine what characteristics actually lead to an improvement in the participants. Based on the statistical model results, the chance for success could be maximized in future studies by basing decisions on the characteristics related to success. The underlying problem is finding the best method for determining or learning what specific characteristics lead to personalized improvement and also translate to real world application.

In research situations it is generally assumed that individual differences are randomly distributed and are equally accounted for through the research design. Unfortunately the variables that need to be distributed equally may not be measured and in many cases the researcher may never know whether the assumptions they attribute to the research design actually accounts for all of the relevant individual differences that exist. In addition, it is possible that a specific treatment does work at a clinically meaningful level but the criteria for statistical significance is not met, or that the researcher did not realize that variables which were not the focal point of the study were actually more related to the outcome than the treatments being investigated. In these research situations personal characteristics influence the results and are not accounted or used in an efficient manner by the statistical methods implemented. Many statistical models that are regularly used were not designed to capitalize on personal characteristics; rather their methods were designed to capitalize on group separation. As a result these models will not be efficient unless there are distinct subgroups within the sample. Researchers sometimes try to overcome the problem of not having distinct subgroups in the sample by creating their own subgroups, such as grouping individuals as above or below the median. The creation of subgroups is a subjective decision. As a result the method used to create subgroups in one study may not translate well to a sample in a different study.

As previously mentioned, parametric statistics are a classification of statistics where inferences are made from the results of the data and the decisions on which the inferences are made are based on rules of probability related to the parameters of the distribution (Geisser &

Johnson, 2006). These parametric methods are the ones most widely taught for making statistical inference (Cox, 2006). The use of parametric methods requires more assumptions than non-parametric methods. When the assumptions are correct, parametric methods can produce more accurate and precise estimates and are said to have more statistical power than alternative non-parametric techniques. However, if those assumptions are incorrect, parametric methods can be very misleading. Because parametric statistics have strict guidelines related to the probability rules of their sample distributions, in certain circumstances they may be limited in their effectiveness (Hoaglin, Mosteller, & Tukey, 2000).

Purpose

One purpose of the current study is to compare distribution free Classification and regression tree (CART) models which create decision rules based on dichotomous target (dependent) variables (Success or Failure) with parametric hierarchical regression models that have continuous dependent variables. The comparison of techniques will determine if CART models provide additional information beneficial to personalized predictive treatments.

To create decision rules on dichotomous target variables, CART models use iterative selection methods and separates variables and values based on their ability to maximize success rates on the dichotomous target variable. The selection method used by CART models can vary and to evaluate the effectiveness of CART decision rules on personalized predictive treatments the different methods that are available must be compared. In addition to different iterative methods used to determine the decision rules, different methods for cross-validation also exist. The method chosen for cross-validation can also influence the effectiveness of CART models and their capacity to assist in personalized predictive treatments. A second purpose of the study is to compare of these iterative selection and cross-validation methods.

Linear Regression Methods

When dependent variables are in a continuous form a linear regression approach is often implemented. The results of a linear regression analysis help establish that a set of independent variables are linearly related to a dependent variable and provides an index of the proportion of explained variance (R^2) in the dependent variable. When variables are linearly related to the dependent variable and they explain a large enough proportion of variance in that variable that exceeds what might occur by chance, they are said to be statistically significant. In addition the linear model can establish the relative predictive importance of the independent variables. Also, interaction terms can be added into the model as independent variables, which help determine whether unique effects due to the cross membership of subgroups exist (i.e. a person is male, and as has a college education). The examination of two R^2's determine if adding an independent variable or interaction term to the model helps explain a significant amount of additional variance. Hierarchical regression is a linear regression method that provides a measure of how much variance in the dependent variable can be explained by one or a set of additional independent variables, over and above what was explained by variables already included in the model.

In some circumstances linear regression methods are useful when attempting to measure group separation or differences. However, linear regression by definition follows the assumption that variables are linearly related, and because of this assumption several issues can develop from its use. From a logical perspective interaction terms always have less statistical power than main effects, because main effects must be entered into the regression model prior to examining the interaction effects. In studies with few variables this is usually not an issue. However, in many studies this may discourage the researcher from exploring what may be important interactions since using this approach and adding more variables into the model reduces the chance the main effects in the regression model will be significant. In addition, independent variables can be highly correlated causing the variance attributed to these variables to be inflated, and can result in incorrect decisions about the relationships between included independent variables and the dependent variable. The solution taken is often excluding one of the highly correlated variables, but this approach may also reduce the amount of variance obtained for a subgroup of cases in the sample.

Another major issue in using a regression approach is the possible influence of the order independent variables are entered into the regression model. While stepwise regression approaches exist and will mathematically attempt to determine the best order of variables, they are not recommended (Derksen & Keselman, 1992). A more accepted approach is creating all possible regression models, in which every possible ordering of predictor variables are examined. The use of all possible regression models can become a daunting task as the number of models that must be examined is proportional to the number of predictor variables; in large studies all possible models may be unreasonable. Possibly the biggest flaw to the regression approach is the existence of curvilinear relationships between an independent variable and the dependent variable. In these instances, the researcher has to somehow transform the independent variable so it meets the linear assumption of the model. Previously mentioned, the use of researcher created subgroups almost always involves some level of subjective decision making that may not always be accepted by others.

The hierarchical regression models described are a parametric approach that tries to fit a linear equation in order to explain variation in a dependent variable using a combination of independent variables. Rather than using equations based on a linear combination of independent variables, another method to analyze data is the use of decision rules. Decision rules do not depend on theoretical distributions, and as a result may be a better approach when making decisions about individual treatments. With decision rules, once the rules are created a researcher can examine individual characteristics and easily prescribe the most effective treatment plan just by following the rules in the model.

Decision Rules

In order to use a decision rule methods it is essential to first have a meaningful outcome often called a target variable that the rules tested will be based on. For example, many individuals are tested regularly for hypertension in hopes of determining its onset early so it can be controlled, reducing the risk for cardiovascular disease in the long term. The target variable is hypertension and could be dichotomized so that a person either has hypertension (BP ≥ 140/90 mmHg) and is assigned a score of one, or does not have hypertension and is assigned a score of zero. Based on individual characteristics or personal attributes it is possible to create sets of rules which differentiate between those who have hypertension and those who do not. A personal attribute can be anything measured about the person, and can be physical characteristics (i.e. gender, height, weight), or personal characteristics (e.g. family history, perceived stress). The purpose is to create good decision rules from the data which can be applied to new individuals in order to classify them correctly.

1R Rule

The simplest form of decision rule methods is called the 1R method (1R = 1-rule). This method helps to create a one level decision tree. The decision tree in Figure 3.1 is an example used to guide emergency medical teams as to whether a patient might be at risk for early death (i.e. dying within a 30 day period) following a heart attack. As shown in Figure 3.1, there are yes and no rules for each variable in the decision tree, and based on the answers at each variable, the medical team for the patient can quickly make an accurate decision on risk for early death. The 1R method for creating a decision tree is a simple and quick method that often helps create a single rule that is quite good for determining the influence or effect the single most important independent variable from a data set on a target variable.

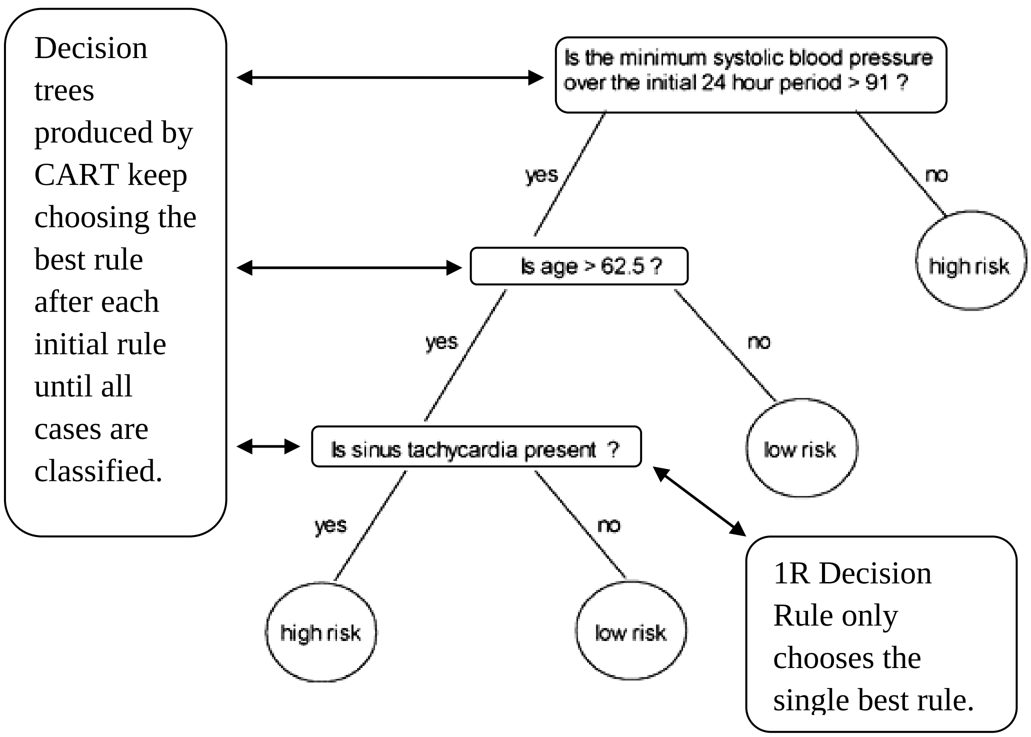

Figure 3.1: Simple Decision Tree Model for Early Death Risk for Heart Attack Patients (Breiman, Friedman, Stone, & Olshen, 1984) edited to display 1R rule and CART procedures

In order to implement the 1-R rule, the following steps are needed. First there must be a target variable to create the cart model from (i.e. early death, did the person die within 30 days; yes, or no). Then for each variable in the data set (i.e. lowest systolic blood pressure (SBP) rating in past 24 hours, age, and whether they had sinus tachycardia[2]) determine how often each level (i.e. how many had sinus tachycardia) appears in terms of the target variable (if they experience early death, or not). Once this is completed, find the most frequent level and make the rule assign that class to this variable-level. (i.e. for minimal SBP; exceeding 91 mmHg was the most frequent level that patients experienced a early death following heart attacked, for age; exceeding 62.5 was the most frequent level). The next step is to calculate the error rate (proportion of correct classifications based on the rule) for all of the rules and choose the rule with the smallest error rate (with heart attack, increased tachycardia, would be a better classifier than minimum SBP, or age). With continuous variables there are way too many individual values to classify (i.e. both minimum SBP, and age are examples). In these instances, the data should be made into dichotomous subgroups in which the most natural breaking point is chosen to base the rule on (i.e. if every person that experienced a heart attack and died early had age > 62.5, it would make sense to make this the natural break point). With the 1R approach, a single rule is chosen as the best classifier (for early death from heart attack, tachycardia would be chosen). The method is often useful but more than one rule is needed when attempting to determine what type of treatment is best (dietary, behavioral, pharmacologic), or what mechanism is leading to the elevated blood pressure (e.g., salt sensitivity, renin homeostasis, insulin resistance, genetics factors).

Classification and Regression Trees (CART)

The 1R approach is not the only methodology that is able to determine rule inference based on individual characteristics. More advanced method of creating classification rules have

[2] defined as being present if the sinus node heart rate ever exceeded 100 beats per minute during the first 24 hours after admission to the hospital (Breiman, et al., 1984).

been available for many years. The most widely cited method for creating Classification and Regression Trees (CART) was developed and popularized by Breiman et al. (1984). CART is a nonparametric model that was originally derived from signal detection theory (Breiman, et al., 1984). However, now many variations of CART are widely used (Witten & Frank, 2005). As shown in Figure 3.1, the CART technique output can be made easily accessible through a decision tree design where graphical maps display the rules about relationships between independent variable values and how they related to the target variable similar to the 1R method. However, CART models use a combination of rules displayed in a decision tree to provide a graphic display of the important variables and values related to the target variable. The previously 1R method used error rates to determine the best rule to use. CART models use the same approach but also take into account the use of multiple rules and how the order of variables can affect variable and their values for selection in the model.

Similar to 1R, CART models create rules by choosing the most important variables and then creating rules for the data using values from those variables to form subgroups that are based on the likelihood of success on the meaningful target variable. In the previous example (see figure 3.1) the target variable was an early death, (i.e. if a heart attack was experienced and the patient died within 30 days then target = 1, if the patient did not die within 30 days then target = 0). There are multiple ways to determine how the values and variables are chosen to form rules on the target variable, all of which are based on some form of information gained by the subgroups that are created when the rules are implemented. The process is sometimes referred to as inferring rudimentary rules (Witten & Frank, 2005). The rules are known as rudimentary because prior to analyzing the dataset, nothing is known about its underlying structure. In order to determine the structure, information is gained by creating rules and then determining how well the rules classify all the data in the dataset. CART models use this process, and information is gained from the accuracy of the rule that is determined by how well the rule classifies the entire sample in terms of the target variable. If every person in the sample that experience a heart attack and had their lowest SBP less than 91 mmHg also experienced an early death attack, the information gained is that for admitted heart attack patients who lowest SBP greater than 91 mmHg is equal to early death 100% of the time. However, we may have some heart attack patients who have a SBP greater than 91 who also experience an early death, and so another rule is needed for these individuals. CART models often create rules based on important variables and the values best related to a successful outcome until no more splits are possible (everyone is classified 100% correctly), the researcher then examines the rules and the subgroups they classify and removes the ones that do not account for many cases using a process called pruning the model.

Several advantages are present with the use of CART. For example, the researcher can set limits such as not allowing rules to be created in the model unless they contain a certain percentage of individuals. This procedure increases the likelihood of obtaining a more parsimonious model. In addition, if some participants are missing an important variable for which a rule is created, another variable called a surrogate variable can be used. Surrogate variables are other variables in a CART model which have similar characteristics to the missing variables in terms of how the rules create subgroups. Another advantage of CART may be the interpretability of its end result. The final product of a CART model is a decision making tree which is easy to read and graphically displays the optimal breaks among variables showing the simple rules that will best lead to the target variable. In addition, at the end of each branch of a CART model is what is called the node information. The node information includes the percentages of successes and failures and describes the number of cases per split as well as relevant class proportions.

CART began with very simple information gain methods and since its introduction the computer algorisms used to implement the information gain methods have been greatly improved. In the field of computer sciences where many of the advances in decision tree analyses have been developed CART is better known as a method of “Machine Learning” (Witten & Frank, 2005).   The field of machine learning focuses on the technical aspects of extracting useful and accurate patterns that can be generalized to make accurate predictions in the future. CART is a form of machine learning; much like linear regression is a form of parametric statistics.

The potential for CART has not been fully investigated in terms of predicting outcomes that are clinically meaningful. In addition, the specific methodologies utilized in CART models have not been compared with each other using clinical trial data to determine if there are differences in the final decision rules. However, CART has been compared to parametric techniques with discrete outcomes including logistic regression (Kiernan, Kraemer, Winkleby, King, & Taylor, 2001; Lemon, Roy, Clark, Friedmann, & Rakowski, 2003) and discriminant function analyses (Karels, Bryant, & Hik, 2004). In all of these studies CART performed as well or better than the parametric technique in reaching the correct decision. Ironically, there are no published studies comparing the final decision rules from CART to the results from a parametric model like hierarchical regression that uses a continuous variable as the dependent variable. In other words, no one has tested whether transforming a continuous variable into a dichotomous variable and implementing a rule induction approach adds any additional information beyond the original parametric model and whether this approach is useful for making predictive decisions.

CART Rule Selection Methods

Breiman et al. (1984) reviewed the rule selection methods available and determined that in many research studies similar decision trees are produced with different rule selection methods but on occasion different rule selection methods can produce slightly different trees. The rule selection procedures include: Gini, Entropy, Class Probability, Twoing, and Favoring Even Splits.

Gini index. The Gini index is a measure of dispersion examining the inequality between groups at each variable split. A Gini index of zero is indicative of an equal number of participants on both sides of the split (i.e. 50% success 50% failure), and an index of one is indicative of unequal group dispersion (i.e. one group has 1%, the other has 99%). Rules are formulated with the attempt to create the smallest Gini index coefficient. Breiman et al. (1984) favor the Gini approach.

Entropy. The entropy of a measure is also sometimes described as *gain*. Quinlan (1998) describes entropy by stating the information conveyed by any problem depends on its probability and can be measured in bits and defined as minus the logarithm to base 2 of that probability. A bit is another term for single piece of information that is related to the certainty of an outcome. With Entropy, information and uncertainty refer to the process that involves selecting one or more objects from a set of objects. If a coin was one sided and always displayed heads the randomness in information gained with Entropy would be $\log_2 1$ or 0 bits. In other words the information gained from the event happening has 100% certainty, no matter what the coin is always heads. With a fair coin the outcome is always heads or tails (50% certainty). With entropy this would mean for an outcome we are uncertain by $\log_2 2$, or 1 bit of information. When Entropy is used as a rule selection method, the goal is to find the variable with the lowest

Entropy value, as this is the most certain (least random) variable. Entropy doesn't assume that probabilities are always equal and trees created with the entropy method attempt to find variables that have high information gain but also highly distinguish between successes and failures.

Class Probability. The class probability tree creates rules similar to a tree created with the Gini method, but rules represent the highest probability of the successful events occurring rather than trying to maximize equal group dispersion across the rules. The displayed variables show the overall probability of variables success and failures based on the variables that have the highest proportion of successes to failures. Class probability trees are often larger than Gini trees, and their terminal nodes (place where no additional splits can occur) are less reliable because they represent chance occurrences and may only contain a few cases. However, the details of the data structure they reveal are often valuable.

Two-ing. With Two-ing instead of searching for a single variable and trying to evening distribute the number of successes and failures, the CART model searches for two variables that when combined account for more than 50% of the data . The model then searches for variables that will maximize the successes based on the original two variables chosen. Unlike Gini or Entropy which can easily produce 90/10 splits the two-ing procedure will tend to produce 50/50 splits and is likely to be the splitting criteria of choice when aiming to obtain more balanced splits.

Favoring Even Splits. Favoring even splits is not a rule selection method but sometimes is labeled as a boosting technique. A weighted average is multiplied to the selection criteria given variables in the model that create rules for subgroups that are more equal in size more importance so they are more likely to be selected.

CART Cross-validation Methods

One of the major advantages with CART is the ability to conduct cross-validation from within the same sample the analyses is conducted on using an iterative sub sampling approach. Cross-validation can be conducted on any statistical model, and is usually completed using some form of a hold-out sample. With a hold-out sample, the original sample is divided into two subsets, one for constructing the model and a second for testing the model. If a hold-out sample is not used validation can be tested using a separate independent sample. Regardless of the approach cross-validation is reasonable. However, with large data sets it often makes more sense to use a hold-out sample. Cross-validation can provide some useful insight regarding the sensitivity of results to small changes in the data. However, even with large sets of data there may only be a few scores on a given variable of interest. Breiman et al. (1984) provides a more detailed description of cross-validation techniques however a review of two popular methods with CART is warranted. These methods are: Fraction of Random Cases, and V-Fold Cross-validation.

Fraction of Random Cases. Selecting a fraction of random cases simply lets the researcher select a random percentage of the data to be split into a model construction set and a model testing set. There is no optimal fraction that is best for all situations and the use of this procedure improves as the data sets increase in size. Breiman et al. (1984) suggested a 2/3, 1/3 model construct/model test split. In other words they recommended the test fraction be set to .33, which would use 66% of the data to construct the model with and the remaining 33% would be used to test the constructed model on.

V-Fold Cross-validation. The v-fold cross-validation method is a good way to make the maximal use of sample data. The v-fold cross-validation method allows the CART tree to be constructed using all of the data but never uses the entire data set for construction during a single run. If a 10-fold cross-validation is conducted, the model is constructed using 90% of the data, and during the testing phase the constructed model is tested on the remaining 10% of the data. The use of 10-fold cross-validation requires running an additional 10 trees (i.e. 10-fold cross-validation) each of which is constructed on a different 90% of the dataset and then tested on a different 10% of the data set. The results from the cross-validation runs are combined and put in a table of synthesized test results. Reducing the number of v-fold below 10 is not recommended and Breiman et al. (1984) report that reliability of cross-validation results is reduced when the number of folds is less than 10. In addition, they report for classification trees there is very little benefit from going up to 20 folds.

CART Best Tree and Pruning Methods

In order to determine whether the current number and selection of variables for classifications in a decision tree fit the data reasonably well, a few aspects of a CART model need to be examined. First, every tree will have a specified selection method, as well as a specified cross-validation method. In addition, other parameters of the model may be specified. For example costs can be assigned to individual predictor variables with certain variables receiving a penalty if they are selected. Finally, other intricacies such as favoring even splits can be defined in advance. Once these model parameters are set, the analyses is conducted and cases are examined using the decision rules created from the model. Using the rules of the model one of four outcomes is possible. The model predicted the case is a success, and it is true the case is a success. It is labeled a *True Positive (TP).* In a similar fashion the model may predict the case is a failure and it is true the case is a failure. This case is labeled a *True Negative (TN).* Unfortunately the model is not always correct. If the rules from the model label a case as a success, but in truth it is a failure, the case is labeled a *False Positive (FP)*. The only other possible outcome is the model labels a case is a failure, but in truth it is a success. The case is labeled a *False Negative (FN)*.

The four outcomes: True Positive (TP), True Negative (TN), False Positives (FP), and False Negatives (FN) are usually displayed in a 2 (Actual Outcome) X 2 (Predicted Outcome) table and these tables are used to calculate rates. If a researcher is interested solely in just the positive cases they may calculate the true positive rate which is calculated by taking the TP and dividing it by the total number of positives (TP + FN). The false positive rate is calculated by taking the FP and dividing it by the total number of negatives (FP +TN). The overall success rate is the number of correct classifications (TP + TN) divided by the total number of classifications (TP + TN + FP + FN). Finally to get the error rate just subtract the overall success rate from one.

Gains Charts. In CART, the accuracy and costs of different decisions trees are not known before the CART models are built. In building the decision tree researchers will want to examine several sets of decision rules based on different combinations of variables and how well they predict success. For any CART analysis, multiple CART models are calculated to be compared. Suppose a sample consists of 250 cases, 30 predictor variables (including a treatment variable), and the target variable has a 50% success rate (125 successes, 125 fail). Also suppose a researcher built a CART model which identified a success rate for 83 individuals and uses 10 variables to define the 83 successful individuals. The CART model has a 66% effectiveness rate and cost 10 variables (including the treatment) to define the model. It may benefit the researcher to only administer the treatment to the 83 individuals who were identified by the 10 variables. However, this model is not the only model that can be developed by using the CART methodology. The same CART method may find that using 11 variables can identify a success rate for 110 cases. This model has an 88% effectiveness rate. Another CART model uses 7 different variables but it defines 90 cases correctly which is a 72% effectiveness rate. Thus, often there are tradeoffs between the cost of the model (number of variables used to split data) and the number of cases correctly identified.   For any given CART analysis it is beneficial to examine several possible models. The goal is to find a CART model that identifies a high percentage of positive instances, yet uses as few variables as possible. For any model the proportion of positive instances (cases labeled as success/over total cases) is computed. If this value is then divided by the success rate when using all variables, the result is the amount of gain. A gains chart visually displays different decision trees on the X-axis and the proportion of the sample size accounted for on the Y-axis. The goal is to choose a model with relatively few rules that yields a high gain. In ideal situations this type of model is located on the left side of the gains chart meaning a few rules had a high gain. A CART model using multiple rules that correctly classifies all the data will always be in the highest point on the right side of the gains chart, but will not be cost effective because it uses all of the variables. The final model selected should be a model that is cost effective enough that it lies above the 45° diagonal that bisects the gains chart.

Receiver Operator Characteristic Curves. To determine the adequacy of the final CART model, Receiver Operator Characteristic (ROC) curves are often used. ROC curves show the ratio of sensitivity and specificity of classifications based on the selected variables, the value chosen to split on those variables, and the current number of splits. The sensitivity refers to the proportion of cases that are truly a success and also test positive and is a measure of how good the CART model is at classifying success. Sensitivity provides a measure of the proportion of cases selected by the rules of the CART model relative to all cases that actually are successful. The specificity refers to the proportion of cases that fail and also test negative. In other words, the specificity is a measure of how good the CART model is at picking out cases that are not successful. Sensitivity and specificity are both measure of correct decision made by the CART model. In addition to sensitivity ROC curves display the False positive rate (1 – specificity). In a ROC plot, sensitivity is plotted on the Y-axis, and the false positive rate is plotted on the X-axis. The plotting of these two rates into an ROC curve provides information of several aspects of the CART model. First it shows the tradeoff between sensitivity and specificity (an increase in sensitivity will be accompanied by a decrease in specificity). Second the closer the curve is to the left side of the upper left-hand corner of the ROC plot the more accurate the CART model. The closer the ROC curve comes to the 45 degree diagonal of the ROC plot, the less accurate the CART model. Third, the slope of the tangent line to ROC curve represents the likelihood ratio (LR) for that value of the test. With CART, multiple trees are compared by examining the different slopes created by the different decision tree models. Finally, the area under the ROC curve represents a measure of test accuracy or how well the test separates the group being tested into those who succeed and fail based on the selected variables and rules.

In order to examine multiple trees with different numbers of splits and alternative splitting variables, ROC curves and relative error curves are created for each tree and these results are used to increase or decrease splits in order to find the optimal decision tree. The relative error curve is used to produce a relative cost profile and trace the relationship between classification errors and tree size. The relative error curve is scaled between zero and one, where a zero represents no error or perfect fit, and one represents the performance of random guessing. The ROC curve is also scaled between zero and one but is scaled differently. The ROC curve represents the probability of a correct decision based on the model parameters, so if the ROC is .50 the model predicts no better than random guessing. Values for the ROC closer to 1 are better. When deciding how many rules to use it is important to consider both the relative error of the model as well as the percentage of correct classifications based on the test model from the cross validation data. For example if an ROC curve revealed a value of .72, one would expect the model to accurately classify participants 72 percent of the time. The value from the ROC can be tested using a cross validation subset of the current sample or by using a separate independent sample and classifying the participants by implementing the rules of the CART analyses and then identifying the percentage of correct classifications. In addition, one can compute standard deviations and error bars around the ROC value for the chosen tree in order to provide an impression of the reliability of making a correct decision (Metz, 1978).

Methods

To determine whether CART models provide information useful for developing personalized strategies aimed at prevention, a comparison to current statistical methods is needed. It is also necessarily to compare different CART combinations of splitting selection criteria (gini, entropy, class probability, two-ing), and cross-validation methods (v-fold, fraction of random cases). The data utilized in the study came from a randomized clinical trial. The results from CART are compared to stepwise hierarchical regression models from the clinical trial study. The sample of 181 consisted of African American (AA) adolescents that are considered at risk for essential hypertension (EH) (Barnes, Gregoski, Tingen, & Treiber, Under Review). Since the focus of the study is to examine statistical methods; information not related to examination of the statistical methods from the stepwise hierarchical regression models is omitted but is available in the original manuscript.

Independent variables in the CART and regression models were taken from anthropometric measures, treatment group assignment, and questionnaires. Anthropometric measures were collected prior to intervention and all questionnaires were administered at baseline and following intervention using a personal computer that was specially programmed to prevent entry errors (e.g. errors of omission, out of range responses). The previous manuscript utilizing regression analyses found the scales to have adequate internal consistency in the current sample.

Independent Variables

Treatment Groups. As part of the Barnes et al. (Under Review) study participants received one of four randomly assigned intervention treatments that were conducted by health education teachers during their regular class periods. The sample consisted of 181 participants that were randomly assigned to one of the following four treatment groups: Breathing Awareness Meditation (BAM) 52 participants (19 males), Health Education (HE) 59 participants (26 males), LifeSkills Training (LST) 58 participants (27 males), and Combination (Combo): 12 participants (5 males). Attendance of treatment sessions was examined among treatment groups prior to any additional analyses. No between-groups differences in attendance were found (F [3, 177] = .587, p = .624).

Anthropometric Characteristics. The Anthropometric measures (i.e., body mass index [BMI], waist circumference) were collected according to standard procedures at pre-test and 3-month post-test (NCHS, 1988).

Adolescent Cook Medley Inventory (ACMI). Twenty-three items from an adapted version of the Cook-Medley (HO) by Matthews (1997) were used. The current version of the HO scale uses a 4 point response rather than the original True/False format. Items were altered to make them more appropriate for children. Possible scores ranged from 23 to 92. Previous studies have indicated good reliability with adolescent samples (Liehr, et al., 2000).

Family Environment Scale (FES). Twenty-seven items developed by Moos & Moos (1981) provide information on family environment and the selected subscale items attempt to capture cohesion, expressiveness, and conflict. The subscales of the FES are purported to measure different aspects of the family environment deeming an internal consistency analyses inappropriate due to the heterogeneity of the measure. A previous study of 1,067 families including participants who varied in socio-economic status found stability reliable of the FES subscales across a 2 month interval with test-retest coefficients ranging from .68 to .86. The FES has also effectively discriminated between families whose members have psychiatric problems and those who do not (Moos & Moos, 1981). Both parents and children completed this measure.

Spielberger Anger Expression Inventory. Twenty-four items developed to measure Anger Out, the frequency of anger expressed, Anger In, the frequency of anger experienced but not expressed, Anger Control, which is the frequency an alternative to anger was utilized. Anger Expression (AE) is also sometimes computed by combing the Anger In and Anger Out scores and subtracting out the Anger Control. Johnson et al. (1987) reported the scale to have good internal consistent (alpha = .84) among a sample of 350 adolescent females of which 171 were black (Johnson, 1992).

Perceived Stress Scale. The 4-item Perceived Stress Scale describes life stress in terms of feeling in control and is a measure of life overload, and lack of predictability and control. Items are coded numerically from 0 to 4 making possible a Perceived Stress Score of 0 to 16. No psychometric evidence for the Perceived Stress Score using a population similar to African American adolescents was found. However, the Perceived Stress Scale was designed for use with participants with at least a junior high education and therefore should be suitable for use with 9th grade adolescents (Ng & Jeffery, 2003).

Everyday Discrimination Scale. The nine items were answered using a six point Likert scale format. Items were designed to assess perceived chronic, routine, and relatively minor experiences of unfair treatment. Perceived discrimination is formed within the context of unfairness (i.e., being treated with less respect, courtesy, receiving poorer service, etc.) as opposed to the context of gender, race-ethnicity, or social class. Williams et al., (1997) found the scales to be internally consistent but indicated that black adults had significantly higher scores compared to whites (Williams, Yan, Jackson, & Anderson, 1997).

City Life Inventory. Thirty-six items were administered and responses were answered using a Likert scale format with four possible answers. The original 36 items were derived from Project Heart, a series of community based studies in Baltimore that investigated relationships between cardiovascular risk and emotional stress in urban adolescents (Ewart & Suchday, 2002).

Socioeconomic Status (SES). SES was measured using the Hollingshead four factor indexes (Hollingshead, 1981) and was completed by the parents. The four factor index has been shown to have good psychometric properties in the United States with diverse populations (Cirino, et al., 2002).

Health History Form. (Treiber., Musante, Baranowski, Strong, & Levy, 1991). Due to potential confounding of treatment effects by substance intake, subjects completed a brief questionnaire which assessed smoking behavior, use of alcohol, and illicit drugs.

Health Belief Scale. Individual item assessments with Likert formats (0-4) were created and given to determine the participant's beliefs about the efficacy of each of the four treatment programs.

Lifestyle Behavior Form. Participants indicated hours per week spent watching television and exercising. A single item measured if the participant smoked cigarettes. If they indicated yes they also recorded the number of cigarettes they smoked per week (Kann, et al., 1998).

Dependent Variables

Dependent variables were changes (Post-test – Pre-test) in SBP, Diastolic blood pressure (DBP) and heart rate (HR) during supine rest periods assessed on one day, taken before interventions began and again three months later. At each of these evaluations, resting hemodynamic data were recorded three times every two minutes following a 10-minute rest using established protocols (Dysart, Treiber, Pfleiger, Davis, & Strong, 1994; Treiber, et al., 1993) with Dinamap Vital Signs Monitors 1846SX (Wattigney, Webber, Lawrence, & Berenson, 1996). The Dinamap is a valid device (Whincup, Bruce, Cook, & Shaper, 1992) for use in pediatric research (Park & Menard, 1987). The averages of the last two readings collected during each evaluation were used in the statistical analyses.

Target Variables

Hierarchical regression models from the Barnes, et al. (Under Review) study had change scores (Post-test – Pre-test) created from continuous dependent measures for SBP, DBP, and HR. In order to develop dichotomous target variables for CART, the continuous change score variables (Post-test – Pre-test) for SBP, DBP, and HR were transformed. Any SBP or DBP change score which was at least -3mmHg or lower was coded as a success (1), and any change score for SBP or DBP greater than – 3mmHg was coded as a failure (0). Any HR change score which was at least -3 beats per minute (BPM) or lower was coded as a success (1), and HR change scores greater than -3 BPM were coded as a failure (0). The values for success and failure were chosen because they are indicative of a reduction from the 95^{th} to the 75^{th} percentile based on normative data for this adolescent group (WHO, 1999).

Statistical Analysis

CART analyses were constructed using all possible combinations of splitting selection criteria (gini, entropy, class probability, two-ing), and cross-validation methods (v-fold, fraction of random cases) with the CART 6.0 program developed by Salford Systems (San Diego California, USA; www.salfordsystems.com). The different model splitting and cross-validations combinations created a total of 16 different CART models for SBP, DBP, and HR totaling 48 models.

During CART model construction, decision rules were developed using the previously described independent variables. In the Barnes et al. (Under Review) regression study, treatment grouping variables were forced entered in the hierarchical regression analysis first. As a result, CART models were conducted with treatment groups forced entered so rules were first created on treatment groups. Once models were completed, the success rates classified by the CART model rules were examined with Fisher's exact test.

Results

The criteria for determining the best model for CART was choosing the model with the lowest cost (fewest variables) and highest ROC. In many of the models, the model with the lowest cost was also the model with the highest ROC. In some circumstances two models had extremely similar ROC values (less than .10). When ROC values were that similar the model with the lower cost was chosen. In one circumstance no model could be constructed and as a result this model does not have costs reported. For some of the probability models cost exceeded 1.0. Because these models were not efficient, no CART node information is reported. If the

ROC for any model was less than .50 (worse than a random guess), it is reported as < .50. The decision rules for all final CART models were examined using of Fisher exact test to determine if success and failure rates differed significantly. The results are organized such that regression results from the previous (Barnes, et al., Under Review) study are described first, followed by the subsequent CART models.

Systolic Blood Pressure

Barnes, et al. (Under Review) SBP Results. In the previous study, the final regression model examining changes in SBP revealed main effects for BAM ($p \leq .001$), LST ($p = .057$), and child-reported family cohesion ($p = .004$). Independent of treatment, participants that reported higher levels of family cohesion had greater reductions in systolic blood pressure. For each one-point increase in reported family cohesion, there was a reduction in systolic blood pressure of 1.34 mmHg. In addition, the analyses revealed two significant two-way interactions which were child-reported family cohesion with BAM ($p \leq .001$) and with LST ($p = .012$). The combined predictors accounted for 15.9% of the total variance ($F[5, 166] = 6.28$, $p \leq .01$). No other predictors were statistically significant (all p's>0.07).

CART Models for SBP (Current Study). The results for the SBP CART models are displayed in Table 3.1. The columns in the table show how the model was created and describe the rule splitting and cross-validation methods used and whether the model was boosted with favoring even splits. The relative cost and ROC values are shown for the final model that was selected. The number of nodes describes the number of rules selected by the model, and the Fisher exact test describes whether the subgroups formed from these rules were significant. Finally there is a column describing whether the rules from CART agreed with the previous regression results. In some models CART agreed with the regression results but created additional rules. When this occurred the last column describing model agreement also included the number of additional decision rules created by the CART models. The different model combinations and cross validation techniques for the most part agreed with the regression model in that the optimal tree had three important nodes or predictors. For the best model, CART first separated the HE treatment from the BAM and LST treatments. The next rule was on family cohesion and indicated that participants, who received BAM or LS and also had reported family cohesion greater than 6.50 on the FES cohesion scale, had higher success rates in terms of improving SBP of at least 3mmHg than other participants in the study. These results indicate the same relationship provided by the regression model used as the reference model. Some model combinations had a slightly better ROC value when four variables were selected for the final tree (i.e. models 2, 4, 6, 8, 12, 14, 16). In all of these models, the additional rule was created for the HE group, and the BMI variable was chosen for the one additional rule. Of the participants who received HE and were successful, the rate of success was more likely for the participants whose BMI was greater than 27.69. Across all model combinations ROC and costs were extremely similar and only varied by a few tenths of a point with the exception of class probability (models 9, 10, 11, 12) which was not surprising as the class probability method was expected to have both higher costs and lower ROC values because of the way subgroups are created with this method.

Table 3.1 Regression and CART results for Systolic Blood Pressure

Model	Split Method	Cross-Validation	Favor Even Splits	Cost	R^2 or ROC	# of Nodes| Predictors	Fisher's Exact test P<.05?	Do Splits Agree with Reference?
Regression	N/A	N/A	N/A	N/A	R^2|.16	3	Yes	N/A
CART Model 1	Gini	V-Fold	No	.66	ROC|.65	3	Yes	Yes
CART Model 2	Gini	Fract Rnd	No	.68	ROC|.69	4	Yes	Yes+1
CART Model 3	Gini	V-Fold	Yes	.66	ROC|.69	3	Yes	Yes
CART Model 4	Gini	Fract Rnd	Yes	.68	ROC|.69	4	Yes	Yes+1
CART Model 5	Entropy	V-Fold	No	.69	ROC|.63	3	Yes	Yes
CART Model 6	Entropy	Fract Rnd	No	.68	ROC|.69	4	Yes	Yes+1
CART Model 7	Entropy	V-Fold	Yes	.66	ROC|.65	3	Yes	Yes
CART Model 8	Entropy	Fract Rnd	Yes	.68	ROC|.69	4	Yes	Yes+1
CART Model 9	Probability	V-Fold	No	1.14	ROC|.55	N/A	Yes	Yes
CART Model 10	Probability	Fract Rnd	No	.94	ROC|.69	5	Yes	Yes+2
CART Model 11	Probability	V-Fold	Yes	1.09	ROC|.57	N/A	Yes	Yes
CART Model 12	Probability	Fract Rnd	Yes	.91	ROC|.69	4	Yes	Yes+1
CART Model 13	Two-ing	V-Fold	No	.66	ROC|.65	3	Yes	Yes
CART Model 14	Two-ing	Fract Rnd	No	.68	ROC|.69	4	Yes	Yes+1
CART Model 15	Two-ing	V-Fold	Yes	.66	ROC|.65	3	Yes	Yes
CART Model 16	Two-ing	Fract Rnd	Yes	.68	ROC|.69	4	Yes	Yes

Diastolic Blood Pressure

Barnes, et al. (Under Review) DBP results. The regression models from the Barnes, et al. (Under Review) study conducted on changes in resting DBP scores revealed no predictors that were statistically significant, and no model results were reported.

CART Models for DBP (Current Study). The regression model for DBP from the Barnes et al. (Under review) study was not significant, indicating there was not a clear linear relationship between treatments and/or variable combinations on changes in resting DBP. The results shown in Table 3.2 reveal that CART models were unable to create meaningful decision rules based on the relationships between treatment and/or variable combination and resting DBP. Across all of the model combinations, the cost exceeded .90 suggesting high costs for the variables used. In addition, all of the ROCs were minimally greater than the chance values of .50.

Table 3.2: Regression and CART results for Diastolic Blood Pressure

Model	Split Method	Cross-Validation	Favor Even Splits	Cost	R^2 or ROC	# of Nodes| Predictors	Fisher's Exact test P<.05?	Do Splits Agree with Reference?
Regression	N/A	N/A	N/A	N/A	R^2|N/A	N/A	N/A	N/A
CART Model 1	Gini	V-Fold	No	.93	ROC|.53	12	Yes	N/A
CART Model 2	Gini	Fract Rnd	No	.96	ROC|.56	4	Yes	N/A
CART Model 3	Gini	V-Fold	Yes	.94	ROC|.56	10	Yes	N/A
CART Model 4	Gini	Fract Rnd	Yes	1.05	ROC|.51	N/A	N/A	N/A
CART Model 5	Entropy	V-Fold	No	.90	ROC|.55	16	Yes	N/A
CART Model 6	Entropy	Fract Rnd	No	.96	ROC|.56	4	Yes	N/A
CART Model 7	Entropy	V-Fold	Yes	.97	ROC|.54	5	Yes	N/A
CART Model 8	Entropy	Fract Rnd	Yes	1.05	ROC|.51	N/A	N/A	N/A
CART Model 9	Probability	V-Fold	No	1.50	ROC|.53	N/A	N/A	N/A
CART Model 10	Probability	Fract Rnd	No	N/A	ROC<.50	N/A	N/A	N/A
CART Model 11	Probability	V-Fold	Yes	N/A	ROC<.50	N/A	Yes	N/A
CART Model 12	Probability	Fract Rnd	Yes	N/A	ROC<.50	N/A	N/A	N/A
CART Model 13	Two-ing	V-Fold	No	.96	ROC|.55	14	Yes	N/A
CART Model 14	Two-ing	Fract Rnd	No	.96	ROC|.56	4	Yes	N/A
CART Model 15	Two-ing	V-Fold	Yes	.95	ROC|.55	10	Yes	N/A
CART Model 16	Two-ing	Fract Rnd	Yes	1.04	ROC|.51	4	Yes	N/A

Heart Rate

Barnes, et al., (Under Review) HR results. The regression models from the previous study conducted on change in resting heart rate included the following variables: Child-reported family cohesion (p=.02) and City Life Exposure to Violence (p=.039). In addition, two significant two-way interactions were found. The model accounted for 13.9% of the total variance with (F[7, 164]=3.794, p=.001). One interaction involved the BAM treatment and parent-reported family cohesion (p=.002). The greatest reductions in resting HR were associated with higher levels of family cohesion if the participant received the BAM treatment. The second two-way interaction was observed between groups receiving the BAM treatment and HE benefits expectancy (p=.016). Heart rates improved significantly for those who received BAM and had low HE benefits expectancy. Child-reported family expression was omitted from the final model although it was statistically significant (*p*=.041), it was not significantly correlated with the dependent variable at the zero-order level (*p*>.16).

CART Models for HR (Current Study). The results for HR CART models are displayed in Table 3.3. As shown in the table none of the models were in agreement with the regression model from the Barnes et al. (Under Review) study. The reference model found main effects for child reported family cohesion (p = .02), and city life exposure to violence (p =.039). In addition the reference model two significant two-way interactions. The first interaction was between the BAM treatment and parent-reported family cohesion (p = .002) where receiving BAM and having high cohesion led to larger reductions in HR. The second interaction was between BAM and the expectancy of HE (p = .016). Participants that rated HE low and received BAM displayed larger reductions in HR. All CART models partitioned the BAM and HE groups together and set the LST group apart. Models 1 and 5 had the lowest cost and highest ROCs. Both models revealed the same splits which suggested the in order of importance, the top four variables chosen by the CART were: Socio-economic status as indicated by the Hollingshead Max score, Spielberger Anger Expression, Spielberger Anger Out, and lifestyle behavior hours of exercise reported per week.

The overall best CART models separated LST from BAM and HE. If participants received LST, success rates were best if they did not have overly high Anger Out, and if they reported at being at the midpoint or higher on how much they exercised each week compared to the rest of the participants in the study. If participants received BAM or HE, there was a curvilinear relationship for socio-economic status that seemed to interact with Anger Expression levels. Participants whose parents reported lower SES (less than 25^{th} percentile) had higher success rates; and of those participants success rates improved if they reported higher Anger Expression, than if they reported lower levels of Anger Expression. However, if participants reported lower levels of Anger Expression but their parents reported higher levels of socio-economic status success rates improved.

Table 3.3: Regression and CART results for Heart Rate

Model	Split Method	Cross-Validation	Favor Even Splits	Cost	R^2 or ROC	# of Nodes|Predictors	Fisher's Exact test P<.05?	Do Splits Agree with Reference?
Regression	N/A	N/A	N/A	N/A	R^2|14	4	Yes	No
CART Model 1	Gini	V-Fold	No	.70	ROC|.66	7	Yes	No
CART Model 2	Gini	Fract Rnd	No	.80	ROC|.63	6	Yes	No
CART Model 3	Gini	V-Fold	Yes	.94	ROC|.56	13	Yes	No
CART Model 4	Gini	Fract Rnd	Yes	.92	ROC|.60	3	Yes	No
CART Model 5	Entropy	V-Fold	No	.68	ROC|.66	7	Yes	No
CART Model 6	Entropy	Fract Rnd	No	.80	ROC|.63	6	Yes	No
CART Model 7	Entropy	V-Fold	Yes	.84	ROC|.58	4	Yes	No
CART Model 8	Entropy	Fract Rnd	Yes	.72	ROC|.62	12	Yes	No
CART Model 9	Probability	V-Fold	No	1.02	ROC|.62	6	Yes	No
CART Model 10	Probability	Fract Rnd	No	.94	ROC|.62	4	Yes	No
CART Model 11	Probability	V-Fold	Yes	1.08	ROC|.61	4	Yes	No
CART Model 12	Probability	Fract Rnd	Yes	1.16	ROC|.54	6	Yes	No
CART Model 13	Two-ing	V-Fold	No	.81	ROC|.65	5	Yes	No
CART Model 14	Two-ing	Fract Rnd	No	.80	ROC|.63	6	Yes	No
CART Model 15	Two-ing	V-Fold	Yes	.85	ROC|.59	4	Yes	No
CART Model 16	Two-ing	Fract Rnd	Yes	.66	ROC|.60	3	Yes	No

Discussion

The results for the CART models varied across the different target variables. The models created for SBP had the most consistent results, possibly because the reference model accounted for more variance compared to DBP and HR. The SBP models are displayed in Table 3.1, and with the exception of the class probability rule splitting method, all models were very similar in terms of model cost, and ROC values. When the v-fold technique was selected as the method for model cross validation, the models all produced 3 nodes and the decision rules matched the relationships found by the hierarchical regression approach used as the reference model. First, the CART models separated the HE treatment from BAM and LST treatments. Next, the models rules indicated if participant's received BAM or LST treatment their level of success depended on child reported family cohesion. The rule on child reported family cohesion separated participants that chose a score of at least 6.50 from those that had selected lower values suggesting that participants who received BAM or LST and also reported higher family cohesion would be the most likely to benefit in terms of resting SBP reduction. When fraction of random cases was chosen as the validation approach, the CART models were constructed using 2/3 of the data and validated by testing the models with remaining 1/3. During this approach the number of rules created increased across all splitting methods. When Gini, Entropy, or Twoing was the selection method, the CART analyses selected four rules. The rules were identical to what was found when v-fold cross validation was used, but an extra rule was made on the HE group. With this approach CART models created the rule, that if participants who received HE also had a higher initial BMI values, success rates were higher than HE participants with lower initial BMI values. BMI was not part of the original regression model, and from the results, seems to be an artifact of the fraction of random cases methods.

Based on the SBP results, the Gini and Entropy rule selection methods are both recommended and the V-fold is recommended as the cross-validation method. These combinations seem to be the best approaches for agreement with the Barnes et al. (Under Review) regression models used for validation. Class Probability had the highest overall model costs, and during v-fold cross-validation produced the worst ROC values. Regardless of cross-validation method, model costs were always much higher for Class Probability compared to the other methods. As a result the Class Probability method is not recommended as the method for selecting decision rules. In general when v-fold was used, favoring even splits had no effect on the rules chosen by the models but did slightly increase the ROC values. It appears the selection method accompanied by the v-fold cross-validation automatically favors even splits. A recommendation would be to run future CART methods selecting this option and if CART is unable to split the variables favoring even splits the researcher could try rerunning the model and not selecting this option. If the goal is to account for more information, the researcher seeks variables that affect all cases. If this can happen with a single variable where the variable does separated success and failure into even groups, the variable is extremely useful.

The regression model for change in DBP was not significant, but CART models were conducted anyway to determine if CART may produce meaningful outcomes that were not found by the initial regression models. The CART models did not find any meaningful relationships in regards to DBP improvement. All models had extremely high costs, and none of the models could produce an ROC that was much better than chance. The fact that CART did not produce any meaningful decision tree models should not be viewed as a weakness of the method. Similar to parametric statistics, CART methods are not able to find relationships that randomly exist.

The HR models produced by CART did not have the same relationships provided by the reference model created with the hierarchical linear regression approach. In trying to determine why the CART models agreed with the hierarchical regression for SBP but not for HR, a few points need to be addressed. First, the overall variance accounted for by the HR hierarchical regression model was lower when compared to the variance accounted for in the SBP hierarchical regression model. Second, the final model for HR hierarchical regression had more variables included in the final hierarchical regression model for SBP. With more variables and lower variance; the relationships revealed in this model were not as easy to distinguish. In an attempt to determine why the HR models from CART did not agree with the original reference model the variables chosen by CART were entered into a hierarchical regression model using the same methods as the previous study (Barnes, et al., Under Review). First, CART selected Socio-economic status (SES) measured by the Hollinghead scale into its model, and the results between SES and a successful improvement in HR appeared to be curvilinear. For 64 of the cases the Hollingshead variable was missing. It also appears this variable was missing proportionately more for the BAM treatment than for the LST and HE groups. When a regression model was conducted that included the variables selected from the CART method (LST, SES, anger expression, and exercise) the regression model was significant ($F[4,112] = 2.95$, $p = .023$; $R^2 = .095$). The variance accounted for in the model was less than the original regression model, 9.5% compared to 13.9%. However, proportionally the new model accounted for more variance when controlling for the fact the regression model created from the best overall CART model had fewer cases.

In addition to calculating a regression model for changes in HR using the variables recommended by the results from CART, the frequency distributions for the continuous HR change score variable (Post-test – Pre-test) used in the previous regression model and dichotomous target HR variable used in the CART models were examined by group for both BAM and LST. Overall, the continuous change score HR variable was lower for LST -.66 +/- 7.16 compared to BAM .55 +/- 8.07. In addition, the LST dichotomized target variable had a higher percentage of success 40% compared to the BAM group 36.7%. The values for both groups were examined for distribution violations, and both were normally distributed. The 25^{th,} 50th, and 75th percentiles for LS were -5.44, -.375, and 5.46 compared to -5.25, 1.325, and 5.31 for BAM. A final CART model was conducted using only the BAM group with the Gini splitting method and v-fold cross validation. The CART model had high costs, and low ROC values suggesting little information could be gained from only examining the BAM group. The only variable selected in the model was efficacy for the HE, the same variable found to have an interaction effect for BAM in final hierarchical regression model.

Overall, CART performed as well as the Barnes et al. (Under Review) hierarchical regression models. In addition the CART procedure found that for the BAM and HE treatments, SES had a curvilinear relationship in terms of HR success. Also, the CART model revealed for improvement in HR, the LST treatment is best, unless the participant has a high amount of anger out. As the LST treatment was designed as a method of cognitively reducing stress levels, it is logical that participants receiving LST would have the best success rates, but having a high amount of anger out trait would reduce the potential for success. For HR, CART methods revealed that individuals who reported some exercise also had better success than participants who reported little to no exercise. It is possible that exercise help participants reduce stress during everyday life providing some increase chance for improvement. Another interesting finding is the additional rules that were created during the CART model construction for improvement in SBP when using fraction of random cases for cross-validation. The models suggested that HE was not as successful as BAM and LST, but that participants who received HE had better success if they had higher BMI values than participants who received HE but had lower BMI values. In this finding CART revealed the importance of baseline difference that other statistical methods may not adequately take into account. The HE treatment was basically a control group where researchers distributed materials on healthy diet and exercise. While the treatment is not as efficacious as BAM or LST, it is basically free and CART while it did not lead for a successful reduction for everyone, some participants did benefit from its use.

The data used in this study did not have overly strong underlying relationships but it is an accurate representation of results obtained from a clinical trial setting. As found with most statistical techniques, both stronger effect sizes and more data increase the likelihood of finding consistent relationships and significant results. The differences between the SBP and HR results support this general notion. CART provides a way of deciding which variables are important in terms of success rates, and also have practical implications such as considering treatments for personalized predictive medicine.

Regression models will determine whether a variable or treatment, or interaction of both is linearly related to an outcome. It does not however, account for the number of individual cases that change. For example, out of 30 participants 25 may have little to no change, but five may have huge decreases on a variable of interest. As long as the distribution assumptions are not violated, parametric approaches such as linear regression will not discriminate between five out of 30 cases that have large changes, or 15 cases out of 30 cases have moderate changes. When the goal is risk reduction, large reductions may not be needed, and the researcher may want to reach as many individuals as possible with a moderate change. It is this aspect that researchers who are focusing on situations such as predictive personalized medicine decision may appreciate the most with procedures such as CART. The target value for success is determined a-priori and with this method, the CART model can be created to capture all individuals who have a change that is clinically meaningful. The process allows researchers to determine what is needed to reach a clinically important change, allowing the model to select the variables that reach the most people, not just the ones that demonstrate the biggest changes. The differences between the CART model output and the regression model output demonstrate this point nicely.

Finally, model significance is an issue that is constantly evaluated. If model significance is a necessity, X^2 statistics, or Fisher's exact test can easily be computed from the information provided in CART outputs. As shown in the tables 3.1- 3.3, the methods used for CART model splitting in combination with the cross-validation methods, only tend to choose splits that are statistical significant in terms of proportions. In addition, when CART results are submitted to create follow-up parametric statistics, the results from these parametric models also tend to be significant. For the epidemiologist or practicing clinician, CART automatically produces decision tree outputs that can easily be displayed to explain a clinically meaningful outcome or an important epidemiological trend. The easy to interpret graphical result in concert with clinically useful statistics such as intent to treat analyses, allow CART to be a strong candidate for making decisions across many fields. Based on its ease of interpretation and versatility, CART models make their results are capable of aiding decisions made at any level, ranging from local clinical trials to policy changes at the national level.

In examining the best CART decision rule selection method and cross-validation combination, only the Gini, and Entropy methods combined with V-fold cross validation for SBP revealed the same relationships as the reference model. However, those methods also created a significant regression model when implemented into a subsequent hierarchical regression for the HR data. Although additional testing is needed, the results of the study provide evidence that CART methods using Gini, and Entropy rule selection methods combined with V-fold cross validation, provide additional information that may be missed when only using regression approaches.

References

Barnes, V. A., Gregoski, M. J., Tingen, M. S., & Treiber, F. A. (Under Review). Factors impacting improved hemodynamic function among African American adolescents engaged in behavioral stress reduction programs. *International Journal of Behavioral Medicine*.

Breiman, L., Friedman, J. H., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. Belmont, Calif.: Wadsworth International Group.

Cirino, P. T., Chin, C. E., Sevcik, R. A., Wolf, M., Lovett, M., & Morris, R. D. (2002). Measuring socioeconomic status: Reliability and preliminary validity for different approaches. *Assessment*, *9*(2), 145-155.

Cox, D. R. (2006). *Principles of statistical inference* (illustrated ed.). Cambridge, N.Y.: Cambridge University Press.

Derksen, S., & Keselman, H. J. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology, 45*, 265-282.

Dysart, J. M., Treiber, F. A., Pfleiger, K., Davis, H., & Strong, W. B. (1994). Ethnic differences in the myocardial and vascular reactivity to stress in normotensive girls. *American Journal of Hypertension*, *7*, 15-22.

Ewart, C. K., & Suchday, S. (2002). Discovering how urban poverty and violence affect health: Development and validation of a neighborhood stress index. *Health Psychology*, *21*(3), 254-262.

Geisser, S., & Johnson, W. M. (2006). *Modes of parametric statistical inference*. Hoboken, N.J.: John Wiley & Sons.

Heymann, D. L., Prentice, T., & Reinders, L. T. (2007). *The world health report 2007- A safer future : Global public health security in the 21st century*. [Geneva]: World Health Organization.

Hoaglin, D. C., Mosteller, F., & Tukey, J. (2000). *Understanding robust and exploratory data analysis*. New York: Wiley-Interscience.

Hollingshead (1981). Four factor index of social status. Dept of Sociology, Yale University.

Johnson, E. H. (1992). *Personality, elevated blood pressure and essential hypertension*. Washington DC: Taylor & Francis.

Kann, L., Kinchen, S. A., Williams, B. I., Ross, J. G., Lowry, R., Hill, C. V., et al. (1998). Youth risk behavior surveillance--United States, 1997. *MMWR CDC Surveillance Summary, 47*(3), 1-89.

Karels, T. J., Bryant, A. A., & Hik, D. S. (2004). Comparison of discriminant function and classification tree analyses for age classification of marmots. *Oikos, 105*(3), 575-587.

Kiernan, M., Kraemer, H. C., Winkleby, M. A., King, A. C., & Taylor, C. B. (2001). Do logistic regression and signal detection identify different subgroups at risk? Implications for the design of tailored interventions. *Psychological Methods, 6*(1), 35-48.

Kraemer, H. C., Offord, D. R., Jensen, P. S., Kazdin, A. E., Kessler, R. C., & Kupfer, D. J. (1999). Measuring the potency of risk factors for clinical or policy significance. *Psychological Methods, 4*, 257-271.

Lemon, S. C., Roy, J., Clark, M. A., Friedmann, P. D., & Rakowski, W. (2003). Classification and regression tree analysis in public health: Methodological review and comparison with logistic regression. *Annals of Behavioral Medicine, 26*(3), 172-181.

Liehr, P., Meininger, J. C., Mueller, W. H., Chan, W., Frazier, L., & Reyes, L. R. (2000). Psychometric testing of the adolescent version of the Cook-Medley hostility scale. *Issues Comprehensive Pediatric Nursing*, *23*(2), 103-116.

Matthews, K. A., Gump, B. B., Block, D. R., & Allen, M. T. (1997). Does background stress heighten or dampen children's cardiovascular responses to acute stress? *Psychosomatic Medicine, 59*(5), 488-496.

Metz, C. E. (1978). Basic principles of roc analysis. *Seminar Nuclear Medicine, 8*(4), 283-298.

Moos, B. S., & Moos, R. H. (1981). *Family environment scale manual*. Palo Alto, CA: Consulting Psychologist Press.

NCHS. (1988). Nhanes III anthropometric procedures (video). Washington DC: US Dept of Health and Human Services, US Government Printing Office.

Ng, D. M., & Jeffery, R. W. (2003). Relationships between perceived stress and health behaviors in a sample of working adults. *Health Psychology, 22*(6), 638-642.

Park, M. K., & Menard, S. M. (1987). Accuracy of blood pressure measurement by the dinamap monitor in infants and children. *Pediatrics, 79*, 907-914.

Treiber, F. A., McCaffrey, F., Musante, L., Rhodes, T., Davis, H., Strong, W. B., et al. (1993). Ethnicity, family history of hypertension and patterns of hemodynamic reactivity in boys. *Psychosomatic Medicine, 55*, 70-77.

Treiber., F. A., Musante, L., Baranowski, T., Strong, W. B., & Levy, M. (1991). *Validation of a family health history of cardiovascular dieseases questionnare.* Paper presented at the 12th annual Meeting of the Society of Behavioral Medicine.

Wattigney, W. A., Webber, L. S., Lawrence, M. D., & Berenson, G. S. (1996). Utility of an automatic instrument for blood pressure measurement in children: The Bogalusa heart study. *American Journal of Hypertension, 9*(3), 256-262.

Whincup, P. H., Bruce, N. G., Cook, D. G., & Shaper, A. G. (1992). The dinamap 1846sx automated blood pressure recorder: Comparison with the hawksley random zero sphygmomanometer under field conditions. *Journal of Epidemiology in Community Health, 46*(2), 164-169.

WHO (1999). Predictive preventive medicine for hypertensive cardiovascular disease. Proceedings of Satellite Symposium to WHO Forum '97. Kyoto, Japan, October 1997. *Clinical and Experimental Pharmacology and Physiology, 26*(7), 527-580.

Williams, D. R., Yan, Y., Jackson, J. S., & Anderson, N. B. (1997). Racial differences in physical and mental health: Socio-economic status, stress and discrimination. *Journal of Health Psychology, 2*(3), 335-351.

Witten, I. H., & Frank, E. (2005). *Data mining: Practial machine learning tools and techniques* (Second ed.). San Francisco: Elsevier Inc.

CHAPTER 4

THE IMPACT OF BEHAVIORAL STRESS REDUCTION PROGRAMS AND INTERPERSONAL CHANGE UPON AMBULATORY HEMODYNAMIC FUNCTION AMONG AFRICAN AMERICAN ADOLESCENTS[1]

[1] Gregoski, M.G., Baumgartner, T.A; University of Georgia. To be submitted to *The Journal of Psychosomatic Research*

Abstract

*Objectives***:** To utilize CART models to define the combination of treatment groups, baseline interpersonal characteristics, and changes in interpersonal characteristics that lead to improvement in ambulatory hemodynamic function with adolescent clinical trial data.

Design: CART models were conducted on dichotomous target (dependent variables) that represented if participants had a clinically meaningful improvement in ambulatory hemodynamic function measured by: systolic blood pressure (SBP), diastolic blood pressure (DBP), and heart rate (HR) recorded across hourly periods throughout the day. Predictors in the CART models included treatment groups; baseline anthropometric, psychosocial, and behavioral characteristics; and changes in anthropometric, psychosocial, and behavioral characteristics. Separate CART models where created for SBP, DBP, and HR across periods of: 7am to 3pm, 3pm to 10 pm, 12pm to 7am, and 24 hours. The predictor variables and values found as important to success rates of clinically meaningful change by CART models were then entered into hierarchical regression models that used the continuous change scores as the dependent variable.

Results: CART models produced success rates for all ambulatory measures that ranged from 25% to 44%. All hierarchical regression models created with the variables and values chosen by the CART models were significant. The variance accounted for from these models ranged from 28% to 55%.

Conclusions: The use of CART models prior to hierarchical regression allows a researcher to determine the most important changes in variables and values related to clinically meaningful improvements.

Keywords: Ambulatory Blood Pressure, Adolescents, African American, Behavioral Stress Reduction, Classification and Regression Trees, Hierarchical Regression

Introduction

Uncontrolled essential hypertension (EH) is the major cause of heart disease and stroke and globally is the number one attributable risk factor for death (Heymann, Prentice, & Reinders, 2007). Researchers examining the recent National Health and Nutrition Survey (NHAINES III) have agreed that EH rates among youth are rising and that the global obesity epidemic is at least partially responsible for the shift in blood pressure with elevated systolic blood pressure (SBP) and diastolic blood pressure (DBP) levels attenuating once body mass index (BMI) is statistically controlled (Muntner, He, Cutler, Wildman, & Whelton, 2004). In reports on the examination of NHAINES III researchers have only discussed the relationship between increased BMI and hypertension, but results from several independent clinical trials have shown that environmental stress factors also significantly contribute to the development of EH (Anderson, 1989; Clark, Anderson, Clark, & Williams, 1999; Williams, Yu, Jackson, & Anderson, 1997). Regardless of its contributing factors, elevated blood pressure among youth is particularly relevant because blood pressure levels track from childhood into adulthood (Shear, Burke, Freedman, & Berenson, 1986). In addition, pediatric autopsy researchers have shown increased atherosclerosis at higher blood pressure levels in youth (Homma, et al., 2001; Tracy, et al., 1995). In order to reduce long term vascular damage, the early detection and management of essential hypertension is necessary.

Although obesity has been reported to be partially related to recent increases in blood pressure found among youth, it does not completely explain racial discrepancies reported in both epidemiological and clinical studies. Incidence of pediatric EH is much higher among African American (AA) adolescents when compared to other ethnic groups in the United States (Nesbitt & Victor, 2004). In addition, non-pharmacological interventions aimed at EH reduction (i.e.

physical activity, dietary education) have been met with mixed success, and many interventions resulted in little to no impact on blood pressure improvement among normotensive youth (Alpert, Murphy, & Treiber, 1994; Resnicow & Robinson, 1997). It has been well established that AA adolescents have more exposure to daily stressful events based on racism (i.e. social inequality, aversive neighborhood characteristics) than any other minority groups (Williams & Williams-Morris, 2000) and exposure to racial discrimination is positively related to elevated levels of blood pressure (Anderson, 1989; Armstead, Lawler, Gorden, Cross, & Gibbons, 1989). It is probable the higher incidence rate of increased blood pressure among AAs than other groups is at least partially due to high daily stress exposure. Unfortunately, there is lack of effective easily implemented EH interventions for at risk AA pediatric populations.

Physiological researchers have shown that behavioral stress leads to increased sympathetic nervous system activity, which in turn leads to increases in blood pressure and sodium excretion (Harshfield, et al., 2007; Harshfield, Treiber, Davis, & Kapuku, 2002; Reaven, Abbasi, & McLaughlin, 2004). In most individuals, the renal sodium handling system results in a restoration of sodium balance following the cessation of stress through the increased urinary release of sodium. However, in previous behavioral stress studies examining the effects of acute laboratory stressors AA's compared to European Americans tend to exhibit significantly higher blood pressure increases, and have a greater likelihood to retain sodium rather than exhibit the typical post-stress response of increased sodium excretion (Harshfield, Treiber, et al., 2002; Harshfield, Wilson, et al., 2002).

The growing evidence that psychosocial stress contributes to the development of EH has resulted in numerous behavioral stress reduction programs among adults with EH (Anderson, Liu, & Kryscio, 2008; Dickinson, et al., 2008; Rainforth, et al., 2007). In addition a few researchers examining adolescents have shown that meditation programs resulted in favorable hemodynamic change with reductions in resting and 24-hour ambulatory blood pressure as well as decreased 24-hour urinary sodium (Barnes, Davis, Murzynowski, & Treiber, 2004; Barnes, Pendergrast, Harshfield, & Treiber, 2008; Barnes, Treiber, & Johnson, 2004). In a study comparing breathing awareness mediation (BAM) with health education (HE), participants who received BAM elicited greater decreases in resting SBP, ambulatory SBP, and decreased sodium excretion (Barnes, et al., 2008).

It is purported that AA's have high psychosocial stress and the increases in blood pressure due to stress exposure may be reduced with effective stress-related coping strategies. Life Skills Training (LST) has been found to facilitate the development of important cognitive-behavioral skills for managing stress and anger, increase self-esteem, lower anxiety and general stress, and increase overall coping skills (Botvin & Griffin, 2002). Two recent clinical trial studies have been conducted in which the efficacy of BAM , LST, and HE among youth classified as pre-hypertensive due to elevated normal blood pressure for age, sex, and height have been examined. The first study examined the impact of BAM, LST, HE or a combination of BAM and LST, in concert with underlying anthropometric, and psychosocial characteristics on the outcomes of resting SBP and DBP, and heart rate (HR). The study found positive benefits for participants who received BAM or LST. In addition the participants who received these treatments had even greater improvement if they also reported high family cohesion (Barnes, Gregoski, Tingen, & Treiber, Under Review). In the second study (Gregoski, Barnes, Tingen, Harshfield, & Treiber, Under Review) researchers examined the effect of the same treatments on ambulatory SBP, DBP, and HR across daytime, nighttime, and 24 hour periods, as well as the impact of the interventions on sodium handling. Overall, participants who received BAM had the most improvement in nighttime, and 24 hour blood pressure, as well as lower sodium intake indicating less stress responses following the intervention. No studies have examined the treatment effects of BAM and LST, in concert with the influence of baseline and changes in anthropometrics and psychosocial characteristics to determine what characteristics lead to improvement in ambulatory SBP, DBP, HR, and 24 hour sodium excretions.

The purpose of the study is to investigate the BAM, LST and HE treatment effects found in the previous research (Barnes, et al., Under Review) and to determine whether background anthropometric or environmental factors or changes in these factors are related to ambulatory hemodynamic improvement. In addition, the purpose of study is to investigate what changes within each treatment intervention occurred and whether these changes or baseline underlying psychological characteristics are in concert with each other in order to determine what changes are needed to reach a clinically meaningful improvement.

Previous adolescent researchers have provided rationale for the influence of underlying psychosocial relationships and reported significant relationships between variables related to hostility, and family functioning with and changes in resting blood pressure (Barnes, et al., Under Review; Clark & Armstead, 2000; Yan, et al., 2003). However, these researchers have not investigated change of psychosocial characteristics due to treatment effects, or examined the influence of underlying psychosocial relationships with ambulatory blood pressure. When examining adolescent blood pressure, several advantages have been reported for the use of ambulatory blood pressure measures over resting blood pressure measures including: higher correlations with risk factors for organ damage, and superior sensitivity (Urbina, et al., 2008).

Researchers have often used regression approaches to examine how psychosocial characteristics affect changes in blood pressure (Barnes, et al., Under Review). Another approach is to examine changes in blood pressure using classification and regression trees (CART). With CART approaches a meaningful change in blood pressure is set in advance. CART models then create rules using the variables and values in the dataset in order to determine what characteristics lead to success in terms of the meaningful change in blood pressure. In a previous study with similar data, Gregoski & Baumgarter (Under Review) compared different methods for creating Classification and Regression Tree (CART) models and validated them with hierarchical linear regression models. In the study the researchers determine the best methods for CART model construction and recommend using Gini and Entropy methods combined with V-fold cross validation. In addition, the researchers advocate CART methods because of its ability to find homogeneous predictors. The results of the previous study demonstrated the CART approach allows the detection of variables that affect many participants at a pre-determined clinically meaningful level rather than finding interactions that have large differences but only affect a select few individuals which may occur when using hierarchical linear regression models.

The process used by CART becomes less cost effective as more variables are selected into the model (Gregoski & Baumgartner, Under Review). As a result, attempts have to be made to find the one or two variables for each treatment that provided the best success rated based on the selected criteria (i.e. meaningful blood pressure change). In addition, CART modeling can be used to determine which treatments differ in terms of variables that lead to higher success rates. Based on the CART results, dummy coded interaction variables can be created and used in hierarchical regression models. In order to create the dummy coded variables for regression models, group separation has to be first examined. Because CART models are built in terms of success rates treatment groups that are the most similar in terms of their success rates are always placed together on one side of the model; the remaining treatment groups are placed on the other side of the model. Once treatments are examined CART models creates rules for each set of treatment groups using the independent variables. The CART model rules attempt to find the characteristics that are homogeneous in terms of identify individuals that were successful on clinically meaningful levels previously described. The final product is a regression model that includes the predictors related to a clinically meaningful change but is also maximized in that it affects a high proportion of participants.

Methods

CART models were created to determine whether underlying psychological characteristics or changes in psychological characteristics (Post-test – Pre-test) influence individual levels of clinically defined success rates for ambulatory SBP, DBP, and HR. Variables identified by CART models as important to success rates for ambulatory SBP, DBP, and HR were examined with hierarchical regression analysis to assess whether the variables important to individual success rates were significant when examined with hierarchical regression models that used continuous change scores (Post-test – Pre-test). In order to determine what characteristics lead to improvement over the 24 hour period, 24 hour models were constructed first. However, daytime periods (7am -3pm), after school periods (3pm – 10pm) and nighttime periods (12pm – 7am) were also examined to determine if the same characteristics were selected by CART models across different periods of the 24 hour period day.

Subjects

Consort information on subject acquisition and attrition for the data utilized in the current study has been previously reported (Gregoski, et al., Under Review). The sample of 181 participants consisted of 52 (19 males) receiving the BAM treatment, 58 (27 males) receiving the LST treatment, 59 (26 males) receiving the HE treatment and 12 (5 males) receiving a combination of BAM and LST (Combo) treatments. All research testing personnel were blind to treatment assignments.

Anthropometric Procedures

Height was measured by stadiometer and weight by Detecto CN20 scale (Cardinal Scale Manufacturing Co., Webb City, MO). Seated resting SBP, DBP and HR measures were recorded using Dinamap 1846SX monitors (Critikon, Inc. Tampa, FL) for 10 minutes (minutes 5, 7 & 9 averaged). The first measurement each day was discarded and the other 2 measurements were averaged.

Interventions

The treatment interventions were conducted by six high school teachers during regular class periods. Teachers participated in training programs for the intervention they were randomly assigned to teach. Each teacher was certified as being competent to teach by the training program instructors. In addition, qualitative assessments of the teachers' implementations of the programs were conducted on a weekly basis. Reports using Likert scale ratings (0-4 scale) with four as the best score demonstrated that instructors were perceived as competent with mean of the ratings: 3.34±0.26 for thoroughness; 3.28±0.32 for class attentiveness; and 3.31±0.27 for enthusiasm. In addition treatment intervention attendance did not differ among treatment groups $F[3, 177] = .587$, $p = .624$.

Breathing Awareness Meditation (BAM). BAM is exercise one of the Mindfulness-Based Stress Reduction Program (Kabat-Zinn & Hanh, 1990). Practice involves focusing upon diaphragm movement, sustaining attention on the breathing process and passively observing thoughts. The individual sits upright in a comfortable position with eyes closed while breathing in a slow, deep, and relaxed manner. Ten-minute sessions were prescribed at school and home each week day. On weekends, subjects were instructed to practice 10 minute sessions twice daily. Compliance for BAM practice at home was 86.6±7.4 percent determined by self-report. In-school attendance for participants receiving the BAM intervention averaged 81% of total sessions.

Health Education (HE). Health education lessons were provided weekly and consisted of 50-minute sessions on cardiovascular health-related lifestyle behaviors based upon National Heart, Lung and Blood Institute guidelines for youth. The lessons included brochures, handouts, videotapes, discussions and recommendations for increasing physical activity (e.g., walking, sports, etc.), and establishing and maintaining prudent diet (e.g., reducing fat and sodium intake). No other stress reduction techniques were administered to those receiving HE treatment. In-school attendance for participants receiving HE intervention averaged 80% of total sessions.

LifeSkills Training (LST). LST program lessons lasted 50-minutes and were provided weekly. Lessons involved group discussions, passive and active modeling, behavioral rehearsal, feedback, reinforcement and behavioral homework assignments. The components selected for LifeSkills lessons included: problem-solving skills, reflective listening, conflict resolution, and anger management to enhance social skills, assertiveness, and personal and social competence (Botvin, Baker, Renick, Filazzola, & Botvin, 1984). Relaxation or stress reduction techniques were not administered to the LST group. In-school attendance for members receiving the LST intervention averaged 85% of total sessions.

Combination of BAM and LST (Combo). Participants assigned to the Combo intervention received weekly 50-minute sessions identical to the description provided for LST. BAM was practiced at school the other 4 days each week and at home each weekday night and twice on the weekend days. The Combo group was discontinued after the first cohort due to reports from instructors suggesting the intervention required too much of the regular class schedule. In-school attendance for participants in the Combo group averaged 82% of total sessions.

Ambulatory Measures

Before and following the intervention, ambulatory SBP, DBP, and HR were recorded for 24 hours. Measurements were recorded every 30 minutes during school, every 20 minutes during self-reported after school waking hours, and every 30 minutes during self-reported sleep hours using Spacelabs 90207 monitors (SpaceLabs, Inc., Issaquah, WA). Ambulatory BP monitoring has been shown to more precisely measure changes in BP in normal daily activities than resting blood pressure measures (Urbina, et al., 2008) and is particularly useful in research with children and adolescents (Graves & Althaf, 2006). Researchers from previous studies have found the instrument to be valid (O'Brien, Mee, & O'Malley, 1991) and values accepted for ambulatory readings were based on previously established criteria (Harshfield, Wilson, et al., 2002). Hourly averages were obtained by averaging all readings for each clock hour across the following time periods: day-time at school (7 a.m. to 3 p.m.), after-school (3 p.m. to 10 p.m.), night-time (12 to 7 a.m.), and 24 hours.

Sodium Excretion

Overnight urine samples were collected for examination of urinary sodium excretion rate (UNaV; mEq/hour) at the same time ambulatory measures were recorded. Participants were provided urine collection containers along with written and verbal instructions for collection and were instructed to note urine collection times at bedtime and upon morning awakening. Out of the 181 subjects, 40 either were non-compliant with urine collection, or failed to provide adequate urine volume during one of the sampling periods (< 80 ml for seven hours). Seven of the remaining 141 subjects had a low amount of creatinine clearance (< 2.94 mg/kg for seven hours). As a result these subjects were excluded from the sodium excretion analysis because they did not meet the criteria for adequate urine collection. No manipulation was made across groups for sodium intake, and all subjects maintained their normal free-living diet.

Independent Variables

Anthropometrics (i.e., body mass index [BMI], waist circumference) were measured according to standard procedures at pre-test and three month post-test (National Center for Health Statistics, 1988). Participants also completed surveys of lifestyle behaviors, (i.e., self-reported physical activity, sedentary behavior, TV viewing, smoking), stress-related coping styles and background environmental factors including Spielberger Anger Expression Scale (Spielberger, et al., 1985); Adolescent Cook Medley Hostility Inventory (ACMI) (Liehr, et al., 2000); City Life Events (Ewart & Suchday, 2002); Family Environment Subscales (FES) of Cohesion, Expression & Conflict (Moos & Moos, 1986); and Everyday and Perceived Discrimination Scales (Kessler, Mickelson, & Williams, 1999). The reading levels of the various surveys were determined to be below the grade levels of the participants and had been previously validated for use with adolescents (15-17 years). When available, parents also completed the family environment scale. Also, At pre-test, parents completed the Hollingshead four factor index of social status (Hollingshead, 1975). Finally the participants were surveyed on how efficacious they thought each intervention treatment was. The measures have been previously described as having good reliability and validity with AA adolescents (Barnes, et al., Under Review).

Statistical Procedures

Change scores (Post-test – Pre-test) were computed for all ambulatory measures, and independent variables with the exception of Hollingshead index and perceived efficacy of intervention treatments that were only measured at pre-test. In addition target scores which were dichotomized were created for all ambulatory variables except sodium handling. If a participant's change score (Post-test – Pre-test) improved by at least 3 (mmHg for ambulatory BP; beats per minute for HR) his target score was coded as a success (1), otherwise he was coded as a failure (0). The value of 3mmHg and 3 beats per minute were chosen as clinically meaningful levels because they exceed the expected normal tracking value of increases for BP and (decreases for HR) across any two consecutive years of age regardless of the participants, percentile rank or gender (Rabbia, et al., 2002; Urbina, et al., 2008). Not enough information currently exists on sodium excretion among healthy adolescents to formulate a clinically meaningful level of change; as a result it was not analyzed further in this study.

The procedures used for CART were chosen based on results of previous research examining resting hemodynamic function (Gregoski & Baumgartner, Under Review). CART models were computed for all ambulatory measures using the Gini and Entropy rule creation methods, combined with V-fold cross-validation set to 10 iterations. Final CART variables were chosen based on the best relative cost (lowest) and highest ROC values. The results of the CART analyses were further submitted to a hierarchical linear regression models by creating dichotomous variables based on the variables and values found to have the best success rates with CART. When creating the dichotomous variables, participants were given a one if they displayed the variables and values selected by the CART decision rules that indicated success otherwise they were given a zero. In building regression models, all treatment grouping variables were entered in the first step. In the second step any variables that were part of an interaction with the variables and values selected by the CART analyses were entered to control for initial main effects or baseline differences. In the third and final step the dummy coded dichotomous variables created from the CART output were entered.

Results

Ambulatory SBP

Variables that impacted ambulatory SBP success rates across a 24 hour period are displayed in Table 4.1. When examined in intervals throughout the 24 period, the same variables were not always selected by the CART models. Hollingshead scores were not available for all participants because in about 1/3 of the sample the parent was not present either during the pretest or posttest period and another relative accompanied the participant. As a result only 124 participants had Hollingshead data. Even though, there were only 124 Hollingshead scores, it was more important in terms of success rates than some variables that had all 181 participants in some of the CART models. Intervention groups did not differ significantly in response rate for the Hollingshead measure, therefore when CART selected it as an important variable, it was kept in the model and follow-up regression analyses only analyzed cases that had both pre and post Hollingshead scores.

Table 4.1: Variables and Values Selected by CART for Ambulatory Systolic Blood Pressure

Dependent Variable	Variables and Values selected by CART Models	Total Success Rate		R^2 from Regression
24 hour Systolic BP	LST = HH>23.50 & ΔSp_AO ≤ -.450	46/123	37%	.28
	BAM&HE = ΔBMI ≤ -.52 & LSTefficacy < 2.83			
12am to 7am Systolic BP	LST = FES_Expression > 8	70/179	39%	.41
	BAM&HE = BAMefficacy >.50			
7am to 3pm Systolic BP	LST= No other predictors	68/186	37%	.36
	BAM&HE = ΔACMI ≥ 8.5 & ΔBMI > -.52			
3pm to 10pm Systolic BP	LST= ΔSp_AI ≥ .50	80/188	43%	.32
	BAM&HE = Sp_Ax ≤ 12.50			

Δ : preceding variable names represents change scores, otherwise scores are from Pre-test values only. BAM = Breathing Awareness Meditation Treatment. LST = LifeSkills Treatment. HE = Health Educatation Treatment. BP = Blood Pressure. ACMI: Adolescent Cook Medley Inventory. FES_Expression: Family Environment Scale Expression. BMI: Body Mass Index. HH: Hollinghead Max score. Sp_AO: Spielberger Anger-Out. Sp_AI:Spielberger Anger-In. Sp_AX: Spielberger Anger-Expression.

During 24-hour SBP, the most important contributors to success as indicated by an improvement of 3mmHg or more where change in BMI, change in Spielberger Anger Out, Hollingshead Max score, and treatment intervention. The CART method identified that during the 24 hour period the LST intervention overall had a lower success rate 29% or (16/55), than BAM and HS with 46% or (31/68). Out of the 29% (16/55) participants who received LST and improved by at least 3mmHg, Hollinghead Max scores, and reported changes in Spielberger Anger Out were the next most important indicators in terms of success rates. Seventy-five percent or (12/16) of the participants with an improvement of 3mmHg that received the LST treatment had parents that reported a score of greater than 23.50 on the Hollinghead Max scale, and the participant reported changes (Post-test – Pre-test) in Spielberger Anger Out less than or equal to -.450. From the 46% or (31/68) of participants improved by 3mmHg that received BAM or HE , changes in BMI and initial ratings on the efficacy of the LST were the next best indicators of success rates. Of the successful cases for BAM and HE, 20% or (6/31) had a change in BMI less than or equal to -.52 and also reported a lower efficacy of the LST treatment at baseline with scores less than 2.83. Overall, the variable combinations used to create subgroups based on the CART model results were present in 38% (18/47) of the participants that improved at the clinically meaningful target level of 3mmHg. To determine the impact of these interactions on 24 hour systolic BP as a continuous score, a hierarchical regression was computed with the ambulatory SBP change scores (Post-test – Pre-test) as a continuous dependent variable. Interaction effects from CART were entered after both treatment effect and main effects for potential moderators were entered into the first step of the model. Important variables in CART were entered into the model through the use of dummy coded variables. The final regression model for 24-hour SBP was significant with $F(7,115) = 6.45$, $p \leq .01$, $R^2=.28$.

During the nighttime hours between 12am to 7am, CART procedures revealed different variables as contributing to success for SBP as defined by an improvement of 3mmHg. Overall, the BAM and HE treatments had slightly higher rates of success 42% or (48/114), than the LST treatment 34% or (22/65). CART revealed that between these hours if participants received BAM or HE, and they also had reported they believed the BAM treatment would be at least somewhat effective (greater than .50 on scale measuring BAM efficacy) the treatment was highly successful. Because all successful participants exceeded this value, success rates did not change for this variable 42% or (48/114). For participants that received the LST treatment, higher levels of family expression (scores greater than 8) was the next important variable in terms of success rates. Twenty-three percent or (5/22) of the participants who received the LST treatment and successfully improved by at least 3mmHg also reported pre-intervention family expression scores exceeding eight. The hierarchical regression model examining ambulatory SBP with change scores (Post-test – Pre-test) during the nighttime hours of 12am to 7am was significant with $F(6, 178) = 19.54$, $p \leq .01$, $R^2=.41$.

During the daytime hours between 7am and 3pm CART procedures identified, treatment group, changes in the Adolescent Cook Medley Inventory (ACMI), and changes in BMI as the most important variables related to success as defined by an improvement of at least 3mmHg. The overall, success rates for participants receiving BAM and HE treatments was 41% or (49/119). The overall success rates for participants who received the LST treatment was 28% or (19/67). For participants that received the LST treatment no other variables were identified by CART as important in terms of success. For BAM and HS, changes in ACMI, and BMI were selected as important variables in terms of success rates. Twenty-seven percent or (13/49) of the successful participants had change scores on the ACMI scores less than or equal to 8.5 and indicated a change in BMI greater than -.52. The hierarchical regression model examining ambulatory SBP with change scores (Post-test – Pre-test) during 7am – 3pm was significant such that $F(5, 178) = 19.70$ $p \leq .01$, $R^2=.36$.

During the afternoon and early evening daytime hours between 3pm and 10 pm CART procedures identified Spielberger Anger Expression, change in Spielberger Anger In, and treatment group as the most important contributors to success in terms of an improvement of at least 3mmHg. The overall success rate for participants who received BAM and HE treatments was 44% (53/121). The next important variable selected by CART in terms of success rates for

BAM and HE treatment groups was baseline Spielberger Anger Expression scores. Of the 53 successful cases 64% (34/53) reported baseline anger-expression scores at pretest to be less than or equal to 12.50. The total rate of success for participants receiving the LST treatment was 40% (27/67). Reported changes on Spielberger Anger On scores were the next best indicator of success for this group. Of the 40% of cases that received the LST treatment and had a successful improvement of at least 3mmHg, 74% of those cases or (20/27) also reported changes on anger-in scores less than or equal to .50. The hierarchical regression model examining the ambulatory SBP change scores (Post-test – Pre-test) for 3pm to 10pm was significant with $F(6, 157) = 13.91$ $p \leq .01$, R^2=.32.

Ambulatory DBP

Results for all time periods for ambulatory diastolic blood pressure are displayed in Table 4.2. Across the total 24 hour period CART procedures identified BMI at pre-test, amount of reported television hours watched per week, treatment group, change in Spielberger Anger Out, and change in BMI as the most important contributors to success in terms of improving by at least 3mmHg. Overall, BAM and HE treatments had a 30% or (36/122) success rate for 24-hour DBP and participants that received the LST treatment had an 18% or (12/68) success rate overall. For participants receiving the LST treatment, the next most important variable related to success was the reported hours of television watched per week. All LST participants 100% or (12/12) that improved by at least 3mmHg reported they watched less than 34.5 hours of television per week. If participants received BAM or HE, success was related to initial BMI, change in anger-out, and change in BMI. Ninety-four percent or (34/36) of the successful participants in BAM or HE had an initial BMI greater than 18.73, a change in Spielberger Anger Out less than or equal to 6.50 and a change in BMI less than or equal to .38. The regression model examining the ambulatory DBP change scores (Post-test – Pre-test) for the 24 hour period was significant with $F(7, 193) = 32.62$, $p \leq .01$, $R^2 = .53$.

Table 4.2: Variables and Values Selected by CART for Ambulatory Diastolic Blood Pressure

Dependent Variable	Variables and Values selected by CART Models	Total Success Rate		R^2 from Regression
24 hour Diastolic BP	LST = Hours per week of TV > 34.5	48/190	25%	.53
	BAM&HE = BMI>18.73, ΔSP_AO ≤ 6.50 & ΔBMI ≤ .38			
12am to 7am Diastolic BP	LST = ΔPSS ≤ -.63	39/114	34%	.31
	BAM&HE = HH > 49, or HH ≤ 49 + ΔCL_FinStrain< -.50			
7am to 3pm Diastolic BP	BAM = HH ≤ 33.50 & SP_AO > 15.50	44/110	40%	.40
	LST&HE = BAMefficacy ≤ 3.17 & ΔACMI > 5.50			
3pm to 10pm Diastolic BP	LST = Δ CL_ExpVio ≤ -1.50	67/188	36%	.47
	BAM&HE = WaistAvg > 25.25			

Δ : preceding variable names represents change scores, otherwise scores are from Pre-test values only. BAM = Breathing Awareness Meditation treatment. LST = LifeSkills Treatment.
HE = Health Educatation treatment. BP = Blood Pressure.
ACMI: Adolescent Cook Medley Inventory. BMI: Body Mass Index.
PSS: Perceived Stress Survey. CL_Financial Strain: City-Life Financial strain.
CL_ExpVio: City Life Exposure to Violence. HH: Hollingshead Max score.
SP_AO: Spielberger Anger Out.

The examination of ambulatory DBP during the nighttime hours of 12am to 7am with CART revealed SES, change in exercise behavior, change in reported City Life Financial Strain, and change in Perceived Stress Survey variables as important contributors to success defined by an improvement of at least 3mmHg. For nighttime hours CART procedures selected SES as an important variable and some of the ambulatory measures did not meet the criteria for adequate measures so the overall number of cases was reduced to 114. Out of the 114 cases, LST had an overall lower rate of success 25% or (13/52) than participants receiving BAM and HE with an overall success rate of 42% or (26/62). If participants received the LST treatment, the next best indicator of success was change in perceived stress. Forty-six percent or (6/13) participants who received LST and improved by at least 3mmHg also reported a change in perceived stress less than or equal to -.63. If the participants received the BAM or HE success rates were better if parents reported higher Hollingshead scores (approximately 50th percentile or better), or lower Hollingshead scores (approximately less than 50th percentile) if the participant reported improvement in their perceived financial strain indicated by the City Life Financial Strain index. Sixty-two percent or (16/26) participants either had either a Hollingshead greater than 49, or a score less than or equal to 49 accompanied with reported change in City Life Financial Strain less than -.50. The regression model examining change scores (Post-test – Pre-test) on ambulatory DBP during the hours of 12am to 7am was significant with $F(7,113) = 6.86$, $p \leq .01$, $R^2=.31$.

The CART models conducted on daytime ambulatory measures for the daytime hours of 7am to 3pm identified the Hollingshead score, Spielberger Anger Out, change in ACMI, efficacy towards the mediation treatment, and treatment group as the most important variables in terms of success, defined by an improvement of at least 3mmHg. Overall success rates were higher for the BAM treatment with 41% or (13/32) of the participants being successful. For participants that received LST and HE overall success rates were 35% or (31/88) Out of the 13 successful participants that received BAM, 69% or (9/13) had parents that reported a Hollingshead score of less than or equal to 33.50 and they also reported an initial Spielberger Anger Out score greater than 15.50. For participants who received HE or LST, 80% or (28/35) of the participants who improved by at least 3mmHg rated the BAM treatment as having low efficacy less than or equal to 3.17 and also reported change scores on the ACMI to be greater than -5.50. The regression model examining change scores (Post-test – Pre-test) for ambulatory DBP during the hours of 7am to 3pm was significant with $F(8,111) = 9.44$, $p \leq .01$, $R^2=.40$.

CART models created ambulatory DBP for the afternoon and evening daytime hours between 3pm and 10 pm revealed that participants who received BAM or HE had higher success rates 40% or (48/121) than participants who received LST 28% or (19/67) where success was defined as an improvement of at least 3mmHg. One hundred percent or (48/48) of the participants who received BAM or HE treatment and were successful reported a waist average at pretest greater than 25.25. For participants receiving the LST treatment the next important variable in terms of success was change in reported City Life Exposure to Violence. Sixty-three percent or (12/19) participants who were successful had a City Life Exposure to Violence change score less than or equal to -1.50. The regression model examining ambulatory DBP change scores (Post-test – Pre-test) for the hours of 3pm to 10pm was significant with $F(6, 185) = 26.48$, $p \leq .01$, $R^2=.47$.

Ambulatory HR

Across the 24 hour period for ambulatory HR results seemed to be somewhat more consistent than the results for SBP and DBP. Results for HR are displayed in Table 4.3. For 24-hour ambulatory HR, CART models identified exercise reported at pre-test, Adolescent Cook Medley Inventory (ACMI) pre-test scores, and treatment group as the most important variables related to success with success defined as an improvement of at least 3 beats per minute. Overall participants receiving the BAM treatment had a higher rate of success 42% (27/64) compared to the LST and HE treatments 32% (40/126). For participants who received BAM, the next best indicator of success was the ACMI. For those who were successful, 93% (25/27) reported an

ACMI score at pretest greater than 51.50. If participants received LST or HE, the next best indicator of success was reported exercise. Out of the total successes for LST and HE, 28% (11/40) reported exercise at pretest to be less than 2.5 hours per week. The regression model examining the ambulatory HR change scores (Post-test – Pre-test) for the 24 hour period was significant with $F(6,183) = 13.41$ $p \leq .01$, $R^2=.31$.

Table 4.3: Variables and Values Selected by CART for Ambulatory Heart Rate

Dependent Variable	Variables and Values selected by CART Models	Total Success Rate		R^2 from Regression
24 hour Heart-rate	BAM = ACMI > 51.50 LST&HE = Exercise (pre-test) < 2.5 hours per week	67/190	35%	.31
12am to 7am Heart-rate	BAM = No other predictors LST&HE = Δ exercise ≤ -7 hours per week	65/179	36%	.55
7am to 3pm Heart-rate	LST = ΔSP_AC > .50 BAM&HE = No other predictors	67/189	35%	.47
3pm to 10 Heart-rate	HE = Exercise per week > 11.50 hours per week BAM&LST = ΔTV hours watched per week ≤-13.50 & CL_ExpVio < 20.50	79/181	44%	.53

Δ : preceding variable names represents change scores, otherwise scores are from Pre-test values only. BAM = Breathing Awareness Meditation Treatment. LST = LifeSkills Treatment.
HE = Health Educatation Treatment. BP = Blood Pressure.
ACMI: Adolescent Cook Medley Inventory. SP_AC: Spielberger Anger Control.
CL_ExpVio: City Life Exposure to Violence.

The CART models created for ambulatory HR during the nighttime hours of 12am to 7am selected change in reported exercise behavior and treatment group as the most important variables in terms of success as defined by an improvement of at least 3 BPM. Participants who received the BAM treatment had a higher success rate 46% (28/61), than participants who received HE and LST 31% (37/118). If participants received the BAM treatment no other important variables were found. For participants who received the HE or the LST treatment, change in reported exercise behavior was the next most important variable in terms of success. Of the 37 successful cases, 35% (13/37) reported a change in exercise behavior less than or equal to -7 hours per week. The regression model examining ambulatory heart rate over from the hours of 12am to 7am using change scores (Post-test – Pre-test) was significant with $F(5, 171) = 42.29$ $p \leq .01$, R^2=.55.

The CART models created for ambulatory HR during the daytime hours of 7am to 3pm selected change in Spielberger Anger Control and treatment group as the most important variables related to success as defined by an improvement of at least 3BPM. If participants received the HE or BAM treatment the overall success rate was 39% (47/122) and no other variables were selected as being important to success. If participants received the LST treatment the overall success rate was 30% or (20/67) and the next important variable in terms of success was Spielberger Anger Control. Seventy percent or (14/20) of the successful LST participants reported Spielberger Anger Control greater than .50. The regression model examining ambulatory HR for the hours of 7am to 3pm using change scores was significant with $F(5, 178) = 10.21 \leq .01$, R^2=.47.

The CART models created for ambulatory HR during the daytime afterschool and evening hours of 3pm – 10pm identified exercise at pretest, change in the amount of hours watching television, treatment group, and the City Life Exposure to Violence reported at pretest as the most important variables related to success as defined by an improvement of at least 3BPM. Overall if participants received HE they had a higher rate of success 47% (27/57), than if they received BAM or LST 42% (52/124). Seventy-eight percent or (21/27) participants that received HE and were successful reported exercise pretest scores less than or equal to 11.50. If participants received BAM or LST, success was related to the change in amount of television watched and change in City Life Exposure to Violence. Fifty-six percent or (29/52) of the successful participants reported a change in the amount of hours of television watched per week to be less than or equal to -13.50, or the change in television hours watched per week to be greater than -13.50 but the exposure to violence at pretest to was less than 20.50. The regression model examining ambulatory HR for the hours of 3pm to 10pm using change scores (Post-test – Pre-test) was significant with $F(7, 176) = 28.81 \leq .01$, $R^2=.53$.

Discussion

Ambulatory SBP

The variables important to clinically meaningful success as defined by an improvement of at least 3mmHg for ambulatory SBP were similar to findings reported in previous research. In addition there was some consistency in variables selected across the different time periods throughout the day which adds some validity to their use. For 24 hour SBP, successful participants that received the LST treatment also reported Hollingshead Max scores exceeding 23.50 and changes in Spielberger Anger-Out to be less than or equal to -.45. In general across many interventions, higher levels of SES usually coincide with higher levels of success or improvement. In this study a value of 22 was the lowest quartile for the Hollingshead Max score, so the rationale that participants with scores higher than 23.50 would have more success is reasonable. The Spielberger Anger-Out score indicates the frequency of anger that is expressed. A change in Anger-Out of less than or equal to -.45 may seem like a small change, change scores were created by subtracting pre-test scores from post-test scores (post-test – pre-test), so having a lower change value means improvement from pre-test to post-test with negative scores meaning more of a reduction. The Spielberger Anger-Out score was only chosen as an important variable for those who received the LST treatment. It is plausible that participants, who improved on this variable, did so as a result of receiving the LST treatment. For participants that received BAM and HE, the variables related to success also seems to follow previous research. Participants that received BAM or HE, were more likely to be successful if they had a change in BMI less than or equal to -.52, and if they had rated their beliefs about the efficacy of LifeSkills to be less than 2.83.

The reduction in BMI supports what previous researchers have found. In lowering SBP, a score less than 2.83 on the efficacy towards LST basically means the participants were among those who rated LST the highest. The 75th percentile for LST efficacy was 2.33, so those who scored above 2.83 and received BAM or HE may have been less successful because they did not get the treatment they really wanted. The total rate of success for the 24 hour SBP was 37% or (46/123). The regression model that included the suggested CART variables and values and examined the 24 hour systolic BP change scores (Post-test – Pre-test) was significant and accounted for about 28% of the variance in the model. This suggests that the variables and values chosen by CART were influential even when examined with change scores.

For ambulatory SBP measured between the nighttime hours of 12am and 7am, participants that received the LST treatment and also had a score of greater than 8 on the Expression component of the Family Environment Scale had higher success rates. A value of 7 was at the 75th percentile of the Family Environment Scale Expression scores in the sample. It is plausible that participants with high family expression were also discussing with their family what they were learning during the LST treatment and this added to the reinforcement of its use. Participants who received BAM or HE, and were successful had also rated the efficacy of BAM to be greater than .50. The efficacy measure has a range of 0 to 4, with the lowest 25% of the participants responding 1.0. Participants scoring lower than .50 may have felt the treatment was so ineffective that they did not take it seriously or comply with practicing at home. The total success rate for SBP change measured from 12am to 7pm was 39% or (70/179). The regression model with the variables and values chosen by the CART process examining the change scores (Post-test – Pre-test) was significant and accounted for 41% of the variance in the model. It appears that for total success based on the clinically meaningful changes, important variables chosen by CART during the nighttime period had less variability compared to the important variables selected for the 24 hour period or the time period during the school day.

The time period from 7am to 3pm is the period when participants where in school for most of the time. During this period, if participants received the LST treatment no other variables were found to be related to success. For participants that received BAM or HE, the same change level for BMI that was found for 24 hour SBP was important to success, however during this time period the selected value related to success included participants who had BMI changes greater than -.52 and this was only the case if their reported change in the Adolescent Cook Medley Inventory (ACMI) was greater than or equal to 8.5. The ACMI has been purported to measure both cynicism and hostility. An ACMI value greater than or equal to 8.5 indicates a score that became severely worse (above the 75th percentile), hence very few participants reached this level. The basic interpretation of this time period was that BAM and HE appeared to have protective benefits, as long as participants did not overly reduce their BMI or greatly increase their cynicism/hostility. Overall, the total success rate for systolic BP change during the hours from 7am to 3pm was 37% or (68/186). The regression model examining the change scores

(Post-test – Pre-test) using the variables and values from CART was significant and the overall amount of variance accounted for by the regression model was 36%.

The hours from 3pm to 10pm are during the nighttime period when participants were most likely to be at home and be awake. During this period, participants who received the LST treatment and reported changes in the Spielberger Anger-In scale greater than or equal to .50 had higher success rates. The Anger-In scale represents the amount of anger experienced but not expressed. It is likely that participants who received the LST treatment would have at least a slight increase on the Anger-In scale as the LST treatment lessons should help them redirect some of their anger inward rather than express it in a fashion that would have negative consequences. For participants who received BAM or HE, success rates were higher for those who had a Spielberger Anger-Expression score less than or equal to 12.50. Anger-Expression represents the amount of anger held in (experienced but not expressed), plus the amount of anger expressed outward minus the amount of anger control (ability to redirect ones anger). The mean score for Anger-Expression was 13.73 and the median was 13.00. It seems rationale that participants who did not have high anger expression at pre-test may be more susceptible to engage in the materials provided for BAM and HE. The total success rate for clinical change from 3pm to 10pm was 43% or 80/188. The regression model examining the CART selection variables and values on systolic BP change scores (Post-test – Pre-test) was significant, and accounted for 32% of the variance.

Ambulatory DBP

The overall results for DBP tended to support previous research, and variables measuring the same attribute were often chosen and in the directions the one might expect to be more related to success. For 24 hour DBP, successful participants that received the LST treatment also reported their television watching to be less than 34.5 hours per week. While 34.5 hours of television per week seems excessive, the average television watched in the sample was 49.51 hours per week with the 50th percentile score at 46 hours per week. It is plausible that participants who were watching less television were taking more time to participate in other behaviors which may be related to improvement in DBP such as physical activity. For participants who received BAM or HE, improvement was related to an initial BMI exceeding 18.73, a reduction in the amount of Spielberger Anger-Out less than or equal to 6.50, and a change in BMI less than or equal to .38. The 25th percentile for BMI in the sample was 21.18, so a value exceeding 18.73 included most participants. For changes in the Spielberger Anger-Out, the 75th percentile was a value of two which also included most of the participants. For changes in BMI, the upper 75th percentile had a value of .18. In other words participants just needed a slightly elevated BMI at baseline, no big increase in their Spielberger Anger-Out, and no big increase in their BMI, in order to have positive results. The outcome of this model suggests, the treatment alone was really the driving force behind the improvement. Overall the participants having variables and values selected by CART had an overall success rate of 25% or (48/190). The regression model examining the same variables and values with change scores (Post-test – Pre-test) for DBP over the 24 hour period was significant and approximately 53% of the variance was accounted for by the model.

For the period measuring from 12am to 7am, stress and financial measures were related to success rates. For participants who received the LST treatment, success was related to changes in the perceived stress survey less than or equal to -.63. The 25th percentile for changes in the perceived stress survey was -.50. Previous studies support the linkage between stress and diastolic BP, so a reduction of stress should be related to higher success rates. For participants, who received BAM or HE, reduction was related to socio-economic status. Participants receiving either of these treatments had more success if their parent reported a Hollingshead Max score exceeding 49, or if their parents reported a score less than or equal to 49 but the participant reported a reduction less than -.50 in their perceived financial strain on the City Life Financial Strain index. The 75th percentile for the Hollingshead Max score was 45 in the sample and the perceived City Life Financial Strain had a mean change score of .19 with a score of -1 in the top 25th percentile. In other words, participants who receive BAM or HE were most likely to be successful if their parents had reported a higher level of SES or if their parents reported a lower SES yet they perceived their financial status as improving. Overall, the success rate for the 12am to 7am period was 34% or (39/114). The regression model examining this period was significant and accounted for 31% of the variance when examined with continuous change scores (Post-test – Pre-test). The R^2 for this analysis was slightly lower, possibly because the sample size was reduced due to fewer available Hollingshead Max scores.

During the daytime hours of 7am to 3pm when participants were in school, participants who received the BAM treatment were more likely to have successful DBP change if their parents reported a Hollingshead Max score less than or equal to 33.50 and the participant reported an initial Spielberger Anger Out score to be less than 15.50. The mean score for the Hollingshead Max was 33.69 for the entire sample and the mean Spielberger Anger Out score was 16.75. So, if participants were near average values for these measures, they were likely to be successful in DBP change. For participants who received LST or HE, success was related to initial ratings on the efficacy of the BAM treatment of less than or equal to 3.17 and a change in the Adolescent Cook Medley Inventory (ACMI) of less than 5.50. The 75th percentile was a score of 2.33 for BAM efficacy and a score of 4 for the ACMI change scores. As long as participants were not among the top ratings for BAM efficacy and they didn't greatly increase their ACMI, success rates were higher. The overall success rate for the daytime hours of 7am to 3pm was 40% or (44/110) and the overall variance accounted for by the regression model for daytime DBP change scores (Post-test – Pre-test) was 40%.

For afternoon and nighttime DBP during the hours of 3pm to 10pm participants that received the LST treatment had higher success rates if they also reported changes in the City Life Exposure to Violence less than or equal to -1.50. The mean change score was - .45, and the 25th percentile was a score of -3. A portion of the LST treatment is supposed to teach the participants to seek alternative behaviors to violence; hence an improvement in this score may reflect the understanding and utilization of the LST treatment lessons. For participants who received BAM or HE, success rates were higher if the participants had a waist average exceeding 25.25 inches. The 25th percentile for this measure was 28 inches, so values exceeding 25.25 included almost all participants. The total success rate for all participants was 36% or (67/188). The regression model examining the DBP change scores (Post-test – Pre-test) using the CART variables and values was significant, and accounted for 47% of the variance in the model.

Ambulatory Heart rate

For the ambulatory HR measures, variables related to physical activity, and television watching were important to success rates across most of the time intervals throughout the day. During the 24 hour measure of ambulatory HR, for participants who received the BAM treatment, pretest scores on the Adolescent Cook Medley Inventory (ACMI) were found to be important to success. Participants receiving BAM and indicating ACMI scores less than 51.50

were more likely to have success than participants exceeding this value. The mean score for this scale was 57.50 and a score of 51.50 was indicative of the 25^{th} percentile. Given the relationship with the ACMI and hostility/cynicism, it seem rationale that participants lower on this measure would have better chances towards success. For participants who received the LST and HE treatments success rates were related to initial hours of exercise reported at pre-test with scores of less than 2.5 hours a week having higher success rates. The number of participants that exercised less than 2.5 hours per week was low with only 28% (11/40) of the successful participants indicating a score this low. However, for these participants it is plausible they only had room for improvement on this variable causing it to be related to success. Overall, the success across all groups during this time period was 35% or (67/190). The regression model examining the CART variables and values on the 24 hour HR change scores (Post-test – Pre-test) was significant and the variance accounted for by the model was 53%.

When examining the nighttime period from 12am to 7pm, no other variables were related to success rates for participants who received the BAM treatment. For participants who received LST and HE treatments, success rates were related to change in exercise behavior with success rates being higher if the participants change in exercise was less than or equal to -7 hours per week. Given that for the 24 hour period success rates for participants receiving LST or HE was related to initial low exercise values it is likely these participants were also the ones that increased their exercise levels. The total success rate across all treatments was 36% or (65/179). The regression model examining the CART variables and values from 12am to 7pm change scores (Post-test – Pre-test) for this period was significant and the variance accounted for by the model was 55%.

For daytime ambulatory HR measures from 7am to 3pm, no other variables besides treatment group were related to success rates for participants that received the BAM and HE treatments. For participants that received the LST treatment, a change in Spielberger Anger Control score exceeding .50 was related to success. The mean change for Spielberger Anger Control was -.32, and the 50^{th} percentile was a change score of 0. It is reasonable that the LST lessons were responsible for the change in Spielberger Anger Control, and that better anger control helped improve daily resting heart rate values. For the total sample, success rates were 35% or (67/189). The regression model examining the CART variables and values from 7am to 3pm for ambulatory HR change scores (Post-test – Pre-test) accounted for 47% of the variance.

For ambulatory HR measures from 3pm to 10pm, if participants received the HE treatment success rates were related to exercise reported at baseline, with participants who reported exceeding 11.5 hours per week having higher success rates. The mean reported value of exercise at pre-test across all groups was 11.23 hours. It is likely those who already were exercising more, were engaging in it because they found it more enjoyable. If this were the case it is also likely these participants would adhere to other provided health recommendations. For participants that received BAM and LST, success rates were related to changes in television watched per week less than or equal to 13.50, and a City Life Exposure to Violence scores less than or equal to 20.50. If participants watched less television, they would have more time to practice other healthier alternatives, and may have also been more likely to engage in BAM or LST practice at home. In addition, if participants perceived their City Life Exposure to Violence to be reduced between pre-test and post-test periods, it would be expected to coincide with a reduction in ambulatory HR. Out of the total sample the success rate was 44% or (79/181). The regression model examining the CART variables and values with HR change scores (Post-test – Pre-test) during 3pm to 10pm was significant and the model accounted for 53% of the variance.

Summary for All CART Models

These CART techniques used in this study probably will not work well for all research scenarios. It is important to remember that homogeneous characteristics with treatment groups have to be present in order for CART to extract them. In other words, CART cannot find clinically meaningful subgroups or variables that help determine success rates if they are not present. In addition if the overall success rate for a treatment is low, there will be nothing to build the CART model with. It is important to have a data set that includes a mixture of participants, some who succeed in terms of the clinically meaningful outcome, and some who fail. In addition, it is important to remember the follow-up regression models were built in terms of clinically meaningful success levels. The success rates from CART models will not always be similar to the variance accounted for in the final regression models and this is shown in Tables 4.1-4.3. When interpreting these values even though they are were calculated using the change scores (Post-test – Pre-test) they really are only accounting for the variance of successful participants and should be interpreted in that fashion. For example, when examining ambulatory BP during the 24 hour period, there were a total of 37% or (46/123) successful cases across all three treatment groups. The $R^2 = .28$ for the hierarchical regression model created using the variables and their respective cut points from the CART models. While it is not incorrect to say these variables accounted for 28% of the variance in the total regression model, it would be more correct to say, out of the 37% of successful cases, the regression model was able to explain 28% of the total variance.

In ideal situations a researcher has a treatment that yields high success rates, and a few homogeneous variables are important for explaining those high success rates. Unfortunately, obtaining these types of results will not always occur. In general the chance for obtaining these type of results is increased by either a large effect size, or having a larger sample. In terms of clinically meaningful success levels, it is more important to have a larger sample with a small to moderate effect size for all participants opposed to a smaller sample or a large sample with a large effect size that only affects a few individuals. The importance of this effect was shown with the Hollingshead Max score and some of the models for DBP. Even with a much smaller sample size, the information gained by keeping the Hollingshead superseded other variables that had more participants in the dataset. If regression were the only method being used for analyses, it is likely the Hollingshead Max score would have been omitted in the analyses due to the reduced overall power, and this information would have been lost.

CART models are based on techniques which use rules selection methods and cross-validation methods to determine the best sets of variables and values related to success rate. As a result variables and values chosen in the model may not always be clinically useful. As shown in the CART model examining DBP from 3pm to 10pm, for participants that received the BAM and HE treatments, a waist average exceeding 25.25 inches was chosen by CART as the next important variable in terms of success rates. Almost all of the participants met this criterion, so an argument could be made on whether it is usefulness. However, no other variables were chosen. Because almost all the participants met this criterion, an alternative interpretation to make for these success rates is that for the BAM and HE treatments, no other variables contributed to clinically meaningful changes for the DBP from 3pm to 10pm.

Some variables and values were chosen by CART for both SBP and DBP and they were in the same direction at the same time period which adds strength to the acceptability of the CART methods. For SBP ambulatory measures taken from the 7am to 3pm time period that often occurs during school time, success rates for the BAM and HE treatments were related to a reduction in ACMI scores. During the same time period a reduction in ACMI scores was related to success for DBP among the LST and HE treatments. The finding supports the argument decreasing scores on the ACMI is a driving force in reaching clinically meaningful changes. Unfortunately why this occurs is not well established.

Conclusions

The purpose of this study was to determine what treatment groups, individual characteristics, and changes in individual characteristics lead to a clinically meaningful level of success in ambulatory SBP, DBP, and HR using a combination of CART and regression approaches. The methods used investigated not only baseline individual characteristics but also examined how individual changes that occurred during the treatment period affected the rate of success. Overall, the combination of methods worked extremely well. CART models indicated variables and values related to success rates, and the variables selected by CART models worked well when examined with changes scores in hierarchical regression methods. Although, the methods worked well for model building in the present, future studies using the rules created by the CART model still need to be conducted. A replication of the present findings would add validity to support these methods.

References

Alpert, B. S., Murphy, J. K., & Treiber, F. A. (1994). Essential Hypertension: Approaches to prevention in children. *Medicine and Exercise in Nutrition and Health, 3,* 296-307.

Anderson, J. W., Liu, C., & Kryscio, R. J. (2008). Blood pressure response to transcendental meditation: A meta-analysis. *American Journal of Hypertension, 21*(3), 310-316.

Anderson, N. B. (1989). Racial differences in stress-induced cardiovascular reactivity and hypertension: Current status and substantive issues. *Psychological Bulletin, 105*(1), 89-105.

Armstead, C. A., Lawler, K. A., Gorden, G., Cross, J., & Gibbons, J. (1989). Relationship of racial stressors to blood pressure responses and anger expression in black college students. *Health Psychology, 8*(5), 541-556.

Barnes, V. A., Davis, H. C., Murzynowski, J. B., & Treiber, F. A. (2004). Impact of meditation on resting and ambulatory blood pressure and heart rate in youth. *Psychosomatic Medicine, 66*(6), 909-914.

Barnes, V. A., Gregoski, M. J., Tingen, M. S., & Treiber, F. A. (Under Review). Factors impacting improved hemodynamic function among African American adolescents engaged in behavioral stress reduction programs. *International Journal of Behavioral Medicine*.

Barnes, V. A., Pendergrast, R. A., Harshfield, G. A., & Treiber, F. A. (2008). Impact of breathing awareness meditation on ambulatory blood pressure and sodium handling in prehypertensive African American adolescents. *Ethnicity and Disease, 18*(1), 1-5.

Barnes, V. A., Treiber, F. A., & Johnson, M. H. (2004). Impact of transcendental meditation on ambulatory blood pressure in African American adolescents. *Americal Journal of Hypertenssion, 17*(4), 366-369.

Botvin, G. J., Baker, E., Renick, N. L., Filazzola, A. D., & Botvin, E. M. (1984). A cognitive-behavioral approach to substance abuse prevention. *Addictive Behaviors, 9*(2), 137-147.

Botvin, G. J., & Griffin, K. W. (2002). Life skills training as a primary prevention approach for adolescent drug abuse and other problem behaviors. *International Journal of Emergency Mental Health, 4*(1), 41-47.

Clark, R., Anderson, N. B., Clark, V. R., & Williams, D. R. (1999). Racism as a stressor for African Americans: A biopsychosocial model. *American Psychologist, 54*(10), 805-816.

Clark, R., & Armstead, C. (2000). Family conflict predicts blood pressure changes in African-American adolescents: A preliminary examination. *Journal of Adolescence, 23*(3), 355-358.

Dickinson, H. O., Campbell, F., Beyer, F. R., Nicolson, D. J., Cook, J. V., Ford, G. A., et al. (2008). Relaxation therapies for the management of primary hypertension in adults: a Cochrane review. *Journal of Human Hypertension*.

Ewart, C. K., & Suchday, S. (2002). Discovering how urban poverty and violence affect health: development and validation of a Neighborhood Stress Index. *Health Psychology, 21*(3), 254-262.

Graves, J. W., & Althaf, M. M. (2006). Utility of ambulatory blood pressure monitoring in children and adolescents. *Pediatric Nephrology, 21*(11), 1640-1645.

Gregoski, M., Davis, H., Barnes, V., Tingen, M., Dong, Y., & Treiber, F. (2009). *Interactive Effects of the ADRB2 T-47C polymorphism and stress reduction interventions upon cardiovascular reactivity to stress among African American (AA) adolescents*. Paper presented at the American Psychosomatic Society.

Gregoski, M. J., Barnes, V. A., Tingen, M. S., Harshfield, G. A., & Treiber, F. A. (Under Review). Impact of meditation and behavioral life skills interventions upon ambulatory hemodynamic function and sodium handling among African American adolescents. *Journal of Psychosomatic Research*.

Gregoski, M. J., & Baumgartner, T. A. (Under Review). Personalized Predictive Decision Methods: A Comparison of CART and Hierarchical Regression Approaches. *Psychological Methods*.

Harshfield, G. A., Hanevold, C., Kapuku, G. K., Dong, Y., Castles, M. E., & Ludwig, D. A. (2007). The association of race and sex to the pressure natriuresis response to stress. *Ethnicity and Disease, 17*(3), 498-502.

Harshfield, G. A., Treiber, F. A., Davis, H., & Kapuku, G. K. (2002). Impaired stress-induced pressure natriuresis is related to left ventricular structure in blacks. *Hypertension, 39*(39), 844-847.

Harshfield, G. A., Wilson, M. E., Hanevold, C., Kapuku, G. K., Mackey, L., Gillis, D., et al. (2002). Impaired stress-induced pressure natriuresis increases cardiovascularload in African American youths. *American Journal of Hypertension, 15*(10 Pt 1), 903-906.

Heymann, D. L., Prentice, T., & Reinders, L. T. (2007). *The world health report 2007 - A safer future : Global public health security in the 21st century*. [Geneva]: World Health Organization.

Hollingshead, A. B. (1975). Four Factor Index of Social Status. Yale University.

Homma, S., Ishii, T., Malcom, G. T., Zieske, A. W., Strong, J. P., Tsugane, S., et al. (2001). Histopathological modifications of early atherosclerotic lesions by risk factors--findings in PDAY subjects. *Atherosclerosis, 156*(2), 389-399.

Kabat-Zinn, J., & Hanh, T. N. (1990). *Full catastrophe living: Using the wisdom of your body and mind to face stress, pain and illness. The program of the stress reduction clinic at the University of Massachusetts medical center.* New York: Delta.

Kessler, R. C., Mickelson, K. D., & Williams, D. R. (1999). The prevalence, distribution, and mental health correlates of perceived discrimination in the United States. *Journal of Health and Social Behaviors, 40*(3), 208-230.

Liehr, P., Meininger, J. C., Mueller, W. H., Chan, W., Frazier, L., & Reyes, L. R. (2000). Psychometric testing of the adolescent version of the Cook-Medley hostility scale. *Issues in Comprehensive Pediatric Nursing, 23*(2), 103-116.

Moos, R. H., & Moos, B. H. (1986). *Family Environment Scale Manual*. Palo Alto, CA: Consulting Psychologists Press.

Muntner, P., He, J., Cutler, J. A., Wildman, R. P., & Whelton, P. K. (2004). Trends in blood pressure among children and adolescents. *Journal of the American Medical Association, 291*(17), 2107-2113.

National Center for Health Statistics (1988). NHANES III Anthropometric procedures (video),. Washington DC: United States Department of Health and Human Services, United States Government Printing Office.

Nesbitt, S., & Victor, R. G. (2004). Pathogenesis of hypertension in African Americans. *Congestive Heart Failure, 10*(1), 24-29.

O'Brien, E., Mee, F., & O'Malley, K. (1991). Accuracy of the Spacelabs 90207 determined by the British Hypertension Society protocol. *American Journal of Hypertension, 9*(6), 573-574.

Rabbia, F., Grosso, T., Cat Genova, G., Conterno, A., De Vito, B., Mulatero, P., et al. (2002). Assessing resting heart rate in adolescents: determinants and correlates. *Journal of Human Hypertension, 16*(5), 327-332.

Rainforth, M. V., Schneider, R. H., Nidich, S. I., Gaylord-King, C., Salerno, J. W., & Anderson, J. W. (2007). Stress reduction programs in patients with elevated blood pressure: A systematic review and meta-analysis. *Current Hypertension Reports, 9*(6), 520-528.

Reaven, G., Abbasi, F., & McLaughlin, T. (2004). Obesity, insulin resistance, and cardiovascular disease. *Recent Progress in Hormone Research, 59*, 207-223.

Resnicow, K., & Robinson, T. N. (1997). School-based cardiovascular disease prevention studies: review and synthesis. *Annals of Epidemiology, 7*(S7), S14-S31.

Shear, C. L., Burke, G. L., Freedman, D. S., & Berenson, G. S. (1986). Value of childhood blood pressure measurements and family history in predicting future blood pressure status: results from 8 years of follow-up in the Bogalusa Heart Study. *Pediatrics, 77*(6), 862-869.

Spielberger, C. D., Johnson, E. H., Russell, S. F., Crane, R. J., Jacobs, G. A., & Worden, T. J. (1985). The experience and expression of anger: Construction and validation of an anger expression scale. In M. A. Chesney & R. H. Rosenman (Eds.), *Anger and Hostility in Cardiovascular and Behavioral Disorders* (pp. 5-30). New York: Hemisphere/McGraw-Hill.

Tracy, R. E., Newman, W. P., 3rd, Wattigney, W. A., Srinivasan, S. R., Strong, J. P., & Berenson, G. S. (1995). Histologic features of atherosclerosis and hypertension from autopsies of young individuals in a defined geographic population: the Bogalusa Heart Study. *Atherosclerosis, 116*(2), 163-179.

Urbina, E., Alpert, B., Flynn, J., Hayman, L., Harshfield, G. A., Jacobson, M., et al. (2008). Ambulatory blood pressure monitoring in children and adolescents: Recommendations for standard assessment: A scientific statement from the American Heart Association Atherosclerosis, Hypertension, and Obesity in Youth Committee of the Council on Cardiovascular Disease in the Young and the Council for High Blood Pressure Research. *Hypertension, 52*(3), 433-451.

Williams, D. R., & Williams-Morris, R. (2000). Racism and mental health: The African American experience. *Ethnicity and Health, 5*(3-4), 243-268.

Williams, D. R., Yu, Y., Jackson, J. S., & Anderson, N. B. (1997). Racial differences in physical and mental health. *Journal of Health Psychology, 2*, 335-351.

Yan, L. L., Liu, K., Matthews, K. A., Daviglus, M. L., Ferguson, T. F., & Kiefe, C. I. (2003). Psychosocial factors and risk of hypertension: The Coronary Artery Risk Development in Young Adults (CARDIA) study. *Journal of the Americal Medical Association, 290*(16), 2138-2148.

CHAPTER 5

SUMMARY

Using a sample of adolescents currently at risk for essential hypertension the researcher examined the use of different CART techniques and compared them with hierarchical regression models on resting and ambulatory systolic and diastolic blood pressure, and heart rate. In Study I, CART models were conducted on resting systolic, and diastolic blood pressure, and heart rate using different combinations of rule selection (Gini, Entropy, Class Probability, and Two-ing) and cross-validation (Fraction of random cases, V-fold) methods. In addition, decision rules formed from the CART models using dichotomous dependent variables were compared to previous hierarchical regression models using continuous change scores for the dependent variable. The results of Study I showed that the Gini and Entropy rule selection methods combined with V-fold cross-validation agreed with the results of the hierarchical regression models. In addition, CART method examining heart rate revealed that although some variables were missing data, they actually proportionally accounted for more variance when included in regression models. Finally for the decision rules created by the CART models revealed a curvilinear relationship that would be hard to detect when only using hierarchical regression methods.

In Study II CART models were conducted on dichotomous target (dependent) variables that represented whether participants had a clinically meaningful improvement in ambulatory blood pressure and heart rate measures. CART models included three treatment groups designed for behavioral stress reduction; baseline anthropometric, psychosocial, and behavioral characteristics; and changes in anthropometric, psychosocial, and behavioral characteristics as predictor variables. Dichotomous target variables were created from continuous change scores (Post-test – Pre-test) for ambulatory systolic blood pressure (SBP), diastolic blood pressure (DBP), and heart rate (HR). Participants were assigned a target score of one if they met the criteria for a clinically meaningful improvement (improved 3 mmHG for SBP, DBP; 3 beats per minute for HR), otherwise they were assigned a target score of zero. The predictor variables and values that were found as important predictors by the CART models were further submitted to hierarchical regression models that used the continuous change scores as the dependent variable. CART models produced success rates for ambulatory SBP, DBP, and HR outcomes that ranged from 25% to 44%. All hierarchical regression models created with the variables and values chosen by the CART models were significant. The variance accounted for from these models ranged from 28% to 55%.

CONCLUSION

Specifically, the previous studies were conducted to determine: the best combinations of rule selection and cross-validation methods for CART models, if these models would agree with previous hierarchical regression models, and whether CART models could create decision rules using the variables and values related to improving ambulatory systolic and diastolic blood pressure, and heart rate.

Study I

CART models should be conducted using both the Gini and Entropy rule selection methods, combined with V-fold cross-validation. CART models effectively agreed with previous hierarchical regression models for systolic blood pressure and diastolic blood pressure.

Finally, CART added useful information which was either not found or not reported in previous hierarchical regression analyses.

Study II

CART models allow the researcher to determine the most important changes in variables and values related to clinically meaningful improvements. In addition, the study showed the decision rules created with CART models, also produced significant results when implemented into hierarchical regression models.

Studies I and II

The differences between regression models and CART models in these studies are shown in how the models work and what they produced. Hierarchical regression models found effects (post-test – pre-test change scores) based on group separation. It did not matter if the effect occurred for a high proportion of participants, or if it only occurred for a few participants but with great magnitude. Either way, the regression model showed significance. CART models found effects (a clinically meaningful change) only if the effect occurred for a high proportion of participants. While the findings of the current two studies are promising, validation with future studies is needed.